Achieving price stability in an uncertain world

Assessing uncertainty in DSGE models

Chamon Wieles

TU Delft

Achieving price stability in an uncertain world

Assessing uncertainty in DSGE models

Master thesis submitted to Delft University of Technology
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering & Policy Analysis**

Faculty of Technology, Policy and Management

by

Chamon Wieles

Student number: 4448200

To be defended in public on November 20^{th}, 2023

Graduation committee

Chair:	Prof. dr. ir. J.H. (Jan) Kwakkel
First supervisor:	Dr. ir. W.L. (Willem) Auping
Second supervisor:	Dr. S.T.H. (Servaas) Storm
External supervisor:	Dr. W.A. (Jan Willem) van den End

TU Delft

Preface

This thesis marks not only the completion of my master's program in Engineering and Policy at Delft University of Technology, but also the end of my life as a student. I started off this part of my life with a bachelor's degree in Civil Engineering and later started the International Bachelor's in Economics and Business Economics and a Bachelor's in Law. Doing the Master's in Engineering and Policy Analysis was the perfect next step to apply my technical and analytical skills in a social context. When it was time to graduate, I wanted to use what I learned in the master's program in an economical context. This project provided an ideal opportunity to do so. With this thesis, I have explored the use of a methodology that is not yet used in macroeconomics. I invite others to continue to develop the best way to use exploratory modeling and analysis in their policy-making process.

Completing this research would not have been possible without the insightful feedback, endless patience, and contagious positivity of my graduation committee. I would like to thank prof. dr. ir. J.H. Kwakkel for allowing me to take on this project, taking on the responsibility of chair in my graduation committee, and providing a safe environment for me to grow. I am grateful to dr. Ir. W.L. Auping for being my first supervisor and for meeting with me on a weekly basis. These meetings have been of great importance in writing this thesis and in my personal development. The critical view of dr. S.T.H. Storm has helped me take this research to a new level by providing interesting insights and ideas, for which I am extremely grateful. Finally, I would like to thank dr. W.A. van den End, for his patience, hospitality, and active engagement. Due to his curiosity, this research was initiated.

In addition to my graduation committee, I would like to thank my friends and family for their comfort, advice, and understanding. The distance between us and the time since we last saw each other may be long, but the love and support they have given me is overwhelming.

Lastly, I am proud of myself for the steps I have taken in the last few months, not only academically but in all aspects of life. The journey of life is a multi-objective optimization problem, and achieving the best performance on all objectives at the same time is simply impossible. During this thesis, I have gained more knowledge about myself, developed my skills, and gained a better idea of what excites and motivates me. In a world filled with uncertainty, I hope to shed some light and support to achieve stability.

Chamon Wieles
Delft, November 2023

Summary

The central bank plays an important role in achieving price stability in a country. High levels of inflation, or deflation, are undesirable and cause economic instability. It is up to the central bank to set the interest rate in such a way that inflation stays within an acceptable range. Especially after the economy experiences a shock, it is expected to mitigate the economic consequences of this shock. This requires them to adjust the interest rate, which is one of its main courses of action to achieve price stability. To find out which interest rate is optimal, one of the models it uses is dynamic stochastic general equilibrium (DSGE) models. DSGE models simulate the behavior of consumers, firms, and the interest rate setting behavior of the central bank. The economy is simplified to two variables: inflation and the output gap. Inflation refers to the change in price over time, and the output gap is the difference between the actual output of an economy and the total output that an economy is assumed to achieve, also called the potential output. More specifically, the model calculates the deviations in inflation and output gap from a steady state with a constant inflation rate and constant economic growth. The main economic features that are included in our model are the degree of price stickiness in an economy, the future expectations of the economy, the substitution behavior between goods of consumers, and how the economy will behave under certain interest rates. Price stickiness is determined on the basis of the share of firms that can adjust their prices at each time step and how much future values are discounted. Lastly, the shock that is applied can be modified in size and persistence. On the basis of experience, assumptions are made about these features in the economy and about the characteristics, especially the persistence, of a shock.

When setting the interest rate for a specific quarter, the central bank takes into account three components: the interest rate in the previous period, current inflation, and current output gap. The central bank can decide the extent of the influence of these three components. DSGE models help to decide its best approach. The assumptions it makes about the economy and the shock might not be completely true, which means that it has to take into account parametric uncertainty and shock uncertainty. Although this uncertainty is in place and the exact state of the economy might not align with their expectations, it has to design policies that will still achieve the desired goals. This research has evaluated how these uncertainties influence the results of DSGE models and designed a policy that takes these uncertainties into account. This is done through Exploratory Modeling and Analysis (EMA), a methodology that is not yet used in (macro)economics. The basic idea behind EMA is to run a large number of experiments with changing combinations of parameter values and finding out how these combinations affect the outcomes of the model.

The experiments showed different behavioral characteristics of the economic variables, mainly in the magnitude of the deviation and whether the deviations oscillated over time. Larger deviations indicate that the shock has a stronger effect on the economy. More extreme deviations indicate economic instability, since the state of the economy varies greatly over time. Both are unwanted by the central bank as it pursues economic stability.

To find out which parameters are likely to cause these behavior characteristics, a behavior-based scenario discovery using time series clustering is conducted using the output gap and inflation time series. By clustering cases that have similar characteristics, a statistical method (patient rule induction method) is used to identify which ranges of which parameters are most likely to have caused these distinctive features. Based on the relative quality of the clustering using the Silhouette Index as an internal validation measure and a visual inspection of how distinctive the behavioral characteristics are, a number of clusters are chosen. Of the computed clusters, only clusters with more than 500 experiments are included to ensure that we get significant results. Subsequently, we find which combinations of parameters cause these behavioral characteristics. Important to note is that this means that the results do not show the effect of changing one variable, ceteris paribus.

The most influential parameters for the behavior of the economy over time in response to a positive demand shock are the share of firms that adjust their prices at a certain time step, how conservative the central bank is in its interest rate setting behavior related to the previous interest rate, and the persistence of the shock. The most extreme behaviors in inflation and the output gap (large deviations

and most extreme oscillation) go hand in hand and are likely to occur when a large share of firms can adjust their prices, while the central bank is conservative in its behavior, after a brief shock. This indicates that the economy might be overcompensating even though the shock is over, and the central bank is lagging both in response and after the initial shock as well when the economy is moving back to a steady state. In addition to this, the share of firms that can adjust their prices influences whether there are large deviations in either output gap or inflation. If this share is small, meaning that prices are relatively sticky, the economy experiences less inflation, but the actual economy is outperforming its potential economy. In practice, this could mean that there is low unemployment and that there is more demand in an economy than supply, resulting in shortages. If this share is large, prices will increase (higher inflation), causing demand and supply to balance out (smaller output gap).

After a negative supply shock was applied, the output gap initially deviated from the steady state both positively and negatively. Over time, all output gap deviations turned negative before moving back to steady state. Whereas the different groups after a demand shock showed distinct behavioral characteristics, our results are more indicative of the magnitude of the deviation for the supply shock. Compared to the analysis of the experiments after a positive demand shock, fewer behavioral distinctions were found after a negative supply shock.

Second, a many-objective optimization approach is used to find what kind of behavior the central bank should be ideal most optimal to control inflation after a supply shock. Economic stability is translated into objectives suitable for this model, namely minimizing the maximum deviation and minimizing the period-to-period change for both output gap and inflation, as well as additional objectives related that incorporate the acceptable range of inflation. Based on these objectives, Pareto optimal policies are found: policies in which the performance on one objective cannot be improved without reducing the performance on at least one other objective. Policies that have the best performance on at least one objective are then subjected to a wide range of combinations of the uncertain parameters. This allows us to assess the robustness of each of the policies and check whether the policy will achieve its desirable result irrespective of the exact state of the economy regarding its price stickiness, etc. For this assessment, robustness metrics are defined. The robustness metrics that are used measure how well a specific policy is performing compared to other policies, as well as the share of economic states under which it will achieve certain set thresholds.

A policy in which the current interest rate is minimally influenced by the previous interest rate and responses heavily to the current inflation minimizes inflation and is best for inflation control. This policy does result in larger deviations from in output gap than other policies that might be less responsive and more conservative. The fact that the central bank is inflation-focused can be seen in the objectives and robustness metrics, which, unavoidably, affects the outcomes of the policy design.

In general, we see that EMA is able to provide additional information when analyzing DSGE models and contribute to the design of robust policies by the central bank. EMA provides additional ways to analyze policy problems compared to conventional methods by evaluating and comparing the performance of different policies. Instead of solely using a loss function to assess the performance, it assesses the robustness on different objectives, which represent the interests of the central bank, and provides ways to show the trade-offs the central bank faces. In addition, it provides novel ways to understand the dynamics of DSGE models and shows which input parameters are influential on the behavior of the model. Importantly, we have to acknowledge that these parameters are exogenous and that it is thus not within the power of the central bank to adjust these parameters. The quality of predictions, however, can be improved upon by understanding which parameters might need extra attention, as these are critical in the expected behavior of the economy. However, the results of the open exploration of the supply shock are limited, indicating a need for an adjustment in the applied method for this case, such as another clustering algorithm, or a change in methodology, such as basing the scenario discovery on another characteristic of the outcomes. Before the central bank can incorporate EMA in their standard analyzes, additional research is required. We considered a theoretical policy problem with the simple DSGE model. Additional research could include real-world data, work with an extended DSGE model, or the central bank could have more specific interests in more practical policy problems. This method did not compensate for general critics on the use of DSGE models Collaborative research between economists and EMA modelers is encouraged to optimally combine the functionalities of DSGE models and the possibilities EMA provides. In future research, specific attention should be paid to ensure the validity of DSGE models when EMA is applied.

Contents

1
Introduction

Declined private consumption due to the pandemic, unstable oil prices and a price war between two oil exporters, and a disrupted European energy market due to the Russian invasion of Ukraine have severely affected the European economy (Baumeister, 2023; Muggenthaler et al., 2021). These immense external shocks resulted in an inflation surge in 2021-2022, reaching a peak in October 2022 with an annual inflation rate of 10.6% in the euro area (Lane, 2023). In this area, inflation, the change in price levels over time, is targeted at 2% annually (De Nederlandsche Bank, 2021). The increase in energy prices caused prices to increase throughout the economy, which consequently affected the purchasing power of consumers (N. Hansen et al., 2023). To compensate for this loss in purchasing power, the employees asked for an increase in their wages. Consequently, this increased the costs of firms and contributed to an upward pressure of prices. Theoretically, this could result in a wage-price spiral, where wages and prices cause each other to increase in a vicious circle (The Economist, 2023). This is one of the ways that high inflation could cause economic instability.

To counteract the high inflation rates, the European Central Bank, responsible for stabilizing prices in the economy, has adjusted the interest rate multiple times. Increasing the interest rate increases the cost of borrowing, dampening demand in an economy, and consequently decreasing inflation rates (Matheson, 2019). On the other hand, deflation, negative inflation, is also undesired. When deflation occurs, the cost of outstanding debt increases, and consumers are likely to increase their savings, which causes firms' profits to decrease. In turn, unemployment might increase because consumers are likely to increase their savings even more. This phenomenon is called a deflationary spiral (Den Haan et al., 2018). By lowering the interest rate, the central bank discourages saving and stimulates spending in an attempt to raise the inflation rate. However, it is common belief that the central bank is limited in its actions once the interest rate has reached the zero lower bound (Buiter, 2003; Krugman & Wells, 2021). In 2014, the European Central Bank implemented a negative interest rate, showing that there is still room for central banks to take action even if the interest rate is low (Altavilla et al., 2022).

To help their policy decisions, central banks use models to simulate the economy, such as dynamic stochastic general equilibrium (DSGE) models (Tovar, 2009). These models explain and predict the behavior of economic variables over time and are used for forecasting, story telling, and policy experiments (Del Negro & Schorfheide, 2013). Starting in a steady state, the model simulates the effect of a shock on the economy, represented by several economic variables, over time by forcing a deviation from the steady state. DSGE models include economic features, such as the degree of price stickiness, the interest setting behavior of the central bank, and the influence of the economic variables on each other. For each quarter, the economic variables are calculated. Simple versions of DSGE models include the output gap, inflation, and the interest rate as economic variables. Over time, many extensions of the DSGE models have been developed, including different economic characteristics and additional macroeconomic variables such as investment, employment, and real wages (Smets & Wouters, 2003).

The central bank has to make assumptions about the economic state and the behavior of economic agents, translated into the input values for the parameters in the model. This can be done using standard values based on knowledge and experience, or estimating values based on historical data (Fernández-Villaverde & Guerrón-Quintana, 2021). Usually, an initial estimate is updated using Bayesian methods based on observations from the real world (Ferroni et al., 2015). It is common for policy makers to take the economic model with certainty, considering fixed values of the parameters and taking the median or mean of a posterior distribution (Górajski et al., 2023; Taylor & Williams, 2010). Another uncertainty the central bank has to deal with is the size and persistence of the shock.

When uncertainty is discussed in relation to DSGE models, this uncertainty is usually only considered in one or a few features of the model. Xiao et al. (2018), for example, incorporated predefined uncertainties, such as technology uncertainty and fiscal policy uncertainty, as a set of discrete values for each uncertainty that are representative of their different expectations. To deal with parameter uncertainty, a Bayesian approach to econometrics, which can be applied in a limited information setting, and Markov-Chain Monte Carlo analysis have been applied Boivin and Giannoni, 2006; Christiano et al., 2010. However, new ways to assess this uncertainty are also researched, such as using info-gap theory, which is done by Ben-Haim and Van den End, 2019.

Some papers have applied a sensitivity analysis to DSGE models, such as Čapek et al., 2009, and Ratto, 2008. Ratto, 2008 has applied two classes of global sensitivity analysis methods: Monte Carlo filtering techniques and high-dimensional model representation. Whereas in the Monte Carlo filtering a multi-parameter Monte Carlo simulation is performed, in the high-dimensional model representation a generic function is decomposed into a finite number of functional terms of increasing dimensionality (Ratto, 2008). These are found to be useful methods in analyzing the behavior of DSGE models and are also included in the software most used to solve DSGE models, namely DYNARE (Ratto et al., 2016).

Although this uncertainty is present, policy makers must design policies that achieve a desired goal. This is called robust policies. Maier et al. (2016) states that robustness can be thought of as a measure of the insensitivity of the performance of a given strategy to future conditions. This implies that, even though the future might turn out differently than expected, the policy will still achieve the desired result. Several articles investigate the creation of robust policies using DSGE, such as (Górajski et al., 2023). Robustness of policies is quantified on the basis of their average or maximum welfare loss for different assumed states of the world. On the basis of real data, they have estimated distributions in which the parametric uncertainties can lie.

To contribute to the growing desire for robust central bank policies, this research will use exploratory modeling and analysis (EMA) to assess uncertainty in DSGE models. We will apply methods to assess uncertainty and create robust policies that are, to the best of our knowledge, not yet applied in relation to DSGE models. Unlike commonly used methods, this research considers continuous ranges (instead of discrete values) for parametric uncertainties and deviates from the conventional way of evaluating the performance and robustness of different policies.

EMA is commonly used for policy problems in which deep uncertainty plays an important role. Kwakkel, Walker, et al. (2016) states that 'deep uncertainty means that the various parties to a decision do not know or cannot agree on the system and its boundaries; the outcomes of interest and their relative importance; the prior probability distribution for uncertain inputs to the system (Lempert, 2003; Walker et al., 2013); or decisions are made over time in dynamic interaction with the system and cannot be considered independently (Haasnoot et al., 2013; Hallegatte et al., 2012).' This research will focus on parametric uncertainty and shock uncertainty.

The basic idea of EMA is to run the model of the policy problem a large number of times with different input values that represent the whole range of uncertainties, and analyze the range of outcomes of the models. Approaches rooted in these methods have emerged more and more in recent years (Bartholomew & Kwakkel, 2020). Two approaches can be distinguished in EMA: open exploration and directed search. In our open exploration, we run the model for a large number of different combinations of uncertainties and, optionally, policies and analyze outcomes of the models to understand the effect specific uncertainties have on the outcomes of our model (Kwakkel & Jaxa-Rozen, 2016). Directed search allows us to search for policies, sets of actions that the policy maker can take that will best achieve the objectives of the policy maker and, optionally, other stakeholders in the policy problem (Kwakkel, 2017).

EMA requires us to look at the DSGE models differently than usual. Contrary to other papers, we will not make use of a welfare-loss function, but we will define specific objectives and robustness metrics that conform to the interests of the central bank. Commonly used loss functions are comparable to the the canonical loss function in Carceller and van den End (2023), in which they sum the square of the deviations in inflation from the the steady state inflation rate and the square of the output gap multiplied by a certain relative weight. EMA does not make use of a loss function, but considers and compares the performance of different policies on a certain objective with the help of robustness metrics. In addition to this, EMA is able to provide us with substantially different insights. While many papers decide on the input parameters and analyze the outcomes, EMA provides us with the opportunity to analyze the input parameters based on a certain outcome. EMA could provide economists with new tools to assess parametric uncertainty and create robust policies. We will find out whether our methods are suitable for DSGE models and identify steps that must be taken before a central bank could incorporate EMA into its daily practice.

The application of EMA to a new type of model contributes to the world of EMA by exploring its usability on DSGE models and its added benefits in the world of macroeconomics. Because one of the methods we use, behavior-based scenario discovery using time series clustering, is a relatively novel method, we may find pitfalls and ways to improve this methodology.

We will perform our analysis on a simple New Keynesian DSGE model, as described by Galí (2015). This model consists of three equations that compute the output gap, inflation, and the interest rate over time. The time series of the output gap and inflation are the outcomes of the model in which we are mainly interested. The output gap is the difference between the actual output of the economy and the total output that the economy is assumed to potentially have (Jahan & Mahmud, 2013). A positive output gap means that the economy is outperforming its potential and usually implies low unemployment and higher demand for goods than supply. A negative output gap means that the economy is underperforming, resulting in higher levels of unemployment and greater supply than demand. The interest rate is determined by three components: the interest rate of the previous period, the current inflation rate, and the current output gap. The (relative) influence of these three components on the interest rate is decided upon by the central bank and represents the levers in the model. The economic features that are included in our model are the expectations of the future economic state, the price stickiness in an economy, and the response of consumers to current inflation and interest rates. The parameters that describe these features are our parametric uncertainties. Lastly, the main responsibility of the central bank to provide economic stability is translated into objectives. These are to minimize the deviation in inflation from the steady state inflation rate, to minimize the output gap over time, and to minimize the period-to-period change in these variables. In addition, a desired inflation range is included, representing its inflation control behavior.

Because this is the first attempt to apply EMA to DSGE models, this research is exploratory of nature. This is reflected in our main research question, which is the following:

What can be learned by applying exploratory modeling and analysis methods to dynamic stochastic general equilibrium models and how can the central bank implement this when deciding the interest rate?

This question will be answered by applying EMA to specific theoretical policy problems for which the central bank uses DSGE models. We will consider the additional insight that EMA can provide to the policy maker, the quality of the policy advice we can give to deal with the problems, and the limitations of and recommendations for future research on the applicability of EMA on DSGE models.

The first theoretical policy problem will be focused on open exploration, increasing our understanding of the dynamics of the model. By doing this, we gain insight into what kinds of dynamics can occur given a variety of uncertainties (Kwakkel & Pruyt, 2013). More specifically, we are interested in what parameters cause certain behaviors of the economy when a shock has occurred. This is formalized in our first subquestion:

SQ1. What influence do uncertain parameters and the shock uncertainty have on the behavior of inflation and the output gap over time after a supply shock and after a demand shock?

A behavior-based scenario discovery using time series clustering is performed, as described by Steinmann et al. (2020), to find out which parameters (and ranges of parameters) cause certain behavioral characteristics in the output gap and inflation time series. Scenario discovery helps us identify commonalities in scenarios that cause similar results (Bryant & Lempert, 2010). By clustering time series based on their behavior, we can identify certain (unique) characteristics of these clusters. Different characteristics are the magnitude of the output gap and inflation and how these variables move towards the steady state over time (think of (lack of) oscillation or a fast/slow adjustment to the steady state). Scenarios per cluster are analyzed for common parameter ranges (Kwakkel & Jaxa-Rozen, 2016). This allows us to understand how different economic states might respond to the demand or supply shock. On the basis of this understanding, the policy maker is aware of the economic features that might need to be studied more extensively. In addition to this, it might help the central bank understand which assumptions they should revisit if future economic behavior turns out differently than expected.

In the second part of our analysis, we will study how the central bank should set the interest rate to robustly maintain economic stability after a supply shock has occurred. Our second subquestion is a representative question for the directed search part of EMA:

SQ2: What are the optimal values for the smoothing parameter for the interest rate and the policy response to inflation that the central bank should follow to robustly achieve stability in the economy after a supply shock?

The policy levers for the central bank are the smoothing parameter for the interest rate and the policy response to inflation, meaning these are the ones they control and can be changed in order to robustly achieve its objectives. A many-objective optimization approach will find a set of Pareto optimal policies, policies of which the performance on one objective cannot be improved without hurting the performance on at least one other objective. This is done by searching through the range of values for the policy levers in the baseline scenario.

In the existence of a trade-off in the objectives, we will have at least two policies that have the most desirable performance on at least one objective. An important notion in the world of robust decision making is that a policy should be evaluated on how it performs in different states of the world and not on the performance in only one state (Homaei & Hamdy, 2020). The policy that has the most desirable performance on a specific objective in one scenario (like the baseline scenario)[1] is not necessarily considered the best policy. It may be possible that this policy underperforms in many of the other scenarios and turns out not to be the most robust policy.

With the help of predefined robustness metrics, we can assess the robustness of (a selection of) the different policies. The results of this assessment and the performance of the objectives are interpreted and related to the interests of the central bank. On the basis of this, a policy advice can be written.

More information about our DSGE model is presented in Chapter 2. This chapter is concerned with some basic information about DSGE models and the information needed to properly apply EMA on this model, such as the definition of our objectives. In addition, we will discuss the applied shocks and the uncertainty surrounding these shocks. The next chapter, Chapter 3, considers the methodology of open exploration and directed search with which we will answer the research (sub)questions. In this chapter, additional information is presented on the basics of EMA and an outline of the experimental setup is presented, including the ranges and distributions we assign to the uncertainties. The thesis will then go on to the results of subquestion 1 in Chapter 4, followed by the results of subquestion 2 in Chapter 5. Before answering our research question in Chapter 7, we first critically discuss our findings, the limitations of this research, and present further recommendations in Chapter 6.

[1]The values for the uncertainty which are generally used by the central bank is called the baseline scenario.

2
Model explanation

First, we will discuss the model and look at the parameters and show how we can connect the model to the exploratory modeling and analysis methods, which will be explained in the next chapter. Note that we have not developed this model and we take this model for granted. The assumptions we made and the conclusions we draw are thus solely based on the theoretical world of DSGE models and are exclusively applicable to this research. Therefore, conclusions, statements, and recommendations should not be taken out of context.

2.1. Introduction to the DSGE model

Macro economists want to make informed and efficient policy decisions, which means that they have to have some knowledge about the behavior of the economy and its response to its policy decision. As they cannot experiment with the real economy and compare results before deciding on a definitive policy, so they have to confide themselves into models, simplifications of reality. One of the models they use are dynamic stochastic general equilibrium (DSGE) models. More specifically, the subset of DSGE models our model falls under is called New Keynesian DSGE models. The difference between the different subsets is caused by the assumptions made in the derivation of the model equations (Slanicay, 2014). For the rest of this thesis, we will refer to the model as the DSGE model.

Our DSGE model consists of three equations that represent the behavior of households, firms, and the behavior of the central bank in setting interest rates. Each equation computes one of the variables: the output gap $\tilde{y}$, inflation π, and the interest rate i. It is calibrated around a steady state with 2% annual inflation rate and, after deviating from this steady state due to a shock, it will return to this equilibrium. As the causal loop diagram in Figure 2.1 shows, the variables are connected through a balancing loop. This visualization provides an initial impression of the dynamics of the model. Leaving out constant parameters, it shows the variables that change over time (output gap ($\tilde{y}_t$), inflation (π_t, the interest rate (i_t) and the expected values for the output gap (E$[\tilde{y}_t + 1]$ and inflation (E$[\pi_{t+1}]$) and the relations they have with respect to each other represented by the arrows. The '+' ('-') next to an arrow indicates that a change in one variable causes the affected variable to change in the same (opposite) direction. A chain reaction causing a variable to change because of his own initial change is called a loop and can be either reinforcing (both changes in the same direction) or balancing (the initial change caused a change in the opposite direction). We can observe that the interest rate has a negative relation on the output gap which positively affects the interest rate resulting in our first balancing loop. The output gap also positively influences the inflation rate which, in turn, positively affects the interest rate, resulting in a second balancing loop. This indicates the balancing behavior of the model and makes it not surprising that our model works its way back to a steady state over time. Inflation and the output gap and their respective expectations for the next period reinforce themselves, which is a well-known concept in economic theory.

2.2. Model equations

The model we will follow is based on the derivations described in Galí (2015). This model is based on three equations: the IS curve (Equation 2.2) representing the consumers in the economy, the Phillips

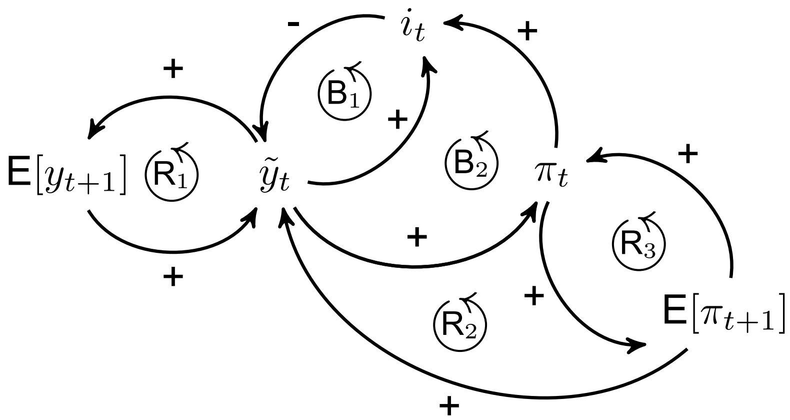


Figure 2.1: Causal loop diagram showing how the variables $\tilde{y}_t$, π_t, and i_t affect each other in our DSGE model. The arrows represent whether they affect each other positively (+) or negatively (-). Changes in a variable cause chain reactions affecting other variables (in)directly. A path through which a variable is influenced by itself is called a loop. A change in this variable can either be reinforcing itself by creating a change in the same direction as the initial change, or balancing if the loop causes a new change that contradicts the initial change. Two balancing loops (B_x) and three reinforcing loops (R_x) are identified.

curve (Equation 2.3) representing the firms in the economy, and the Taylor rule (Equation 2.4) which is the equation for the interest rate, which is set by the central bank. For every time step (quarters), the values of these three variables are calculated. Our agents are forward-looking, meaning that the expectations of the future economic state influence today's economy. The midpoint between the current value and the next value of the variables (the output gap and the output gap) is taken as the expected future value of the variables. The output gap and inflation, computed by the IS curve and the Phillips curve, respectively, are defined as deviations from the steady state. The exact steady state value of the output gap is unknown. In other words, they should be interpreted as a percentage change from the steady state equilibrium level of, respectively, output and the price level. An important assumption to keep in mind is that the change in values of the variables is small, allowing log-linearization when deriving the model equations. This is described in Equation 2.1.

$$\begin{aligned} x_t &\equiv \log X_t - \log X \\ &= \log \frac{X_t}{X} \\ &= \log(1 + \%\text{change}) \\ &\approx \%\text{change} \end{aligned} \tag{2.1}$$

In case the deviations are too large, this assumption will cause an exaggeration of the deviation, making the model invalid.

In this section, the formulas for the equations are presented, accompanied by a brief elaboration. More information is included in Appendix A.

IS curve (demand curve, or Euler equation)

It is assumed that households optimize their utility, which is a function of their consumption and the hours they work. For each period, the household budget, consisting of the hours worked multiplied by an hourly wage, the savings from the previous period, and a lump sum component of income, is divided into consumption and savings. Future utility is discounted with a discount factor. Log-linearizing the optimized consumption functions around a steady state with a constant rate of inflation and consumption growth gives us a level of consumption that is based on the future level of consumption, the interest rate, and the expected inflation. Subsequently, it is assumed that the output is identical to household consumption in an economy, ending with Equation 2.2.

$$\begin{aligned}
\widetilde{y}_t &= E\left(\tilde{y}_{t+1}\right) - \frac{1}{\sigma}\left(i_t - E\left(\pi_{t+1}\right) - r_t^n\right) - \left(\rho_d - 1\right) d \\
\text{in which } \widetilde{y}_t &= \text{output gap at time } t \\
E(y_{t+1}) &= \text{expected value of output at time } t+1 \\
\sigma &= \text{inverse of marginal elasticity of substitution} \\
i_t &= \text{interest rate at time } t \\
E(\pi_{t+1}) &= \text{expected value of inflation at time } t+1 \\
r_t^n &= \text{real natural rate of interest at time } t \\
\rho_d &= \text{autoregressive term of demand shock} \\
d &= \text{demand shock}
\end{aligned} \tag{2.2}$$

The output gap is the difference between the actual output (GDP) and the potential output of an economy, the total output the economy is assumed to be able to have. A positive (negative) output means that the actual output is larger (smaller) than the potential output. From Equation 2.2, we see that the expected future state of the economy positively influences the output gap, creating a reinforcing effect. The interest rate negatively impacts the output gap: a higher (lower) interest rate pushes (lifts) the actual output down (up). The degree in which the interest rate affects the deviation in the output gap is influenced by how easy consumers can switch between products; if it is easier to switch between products (1/σ is large), the interest rate (and the expected future value of inflation) has a greater effect on the output gap. Lastly, the output gap can experience a demand shock (both positive and negative) that can persist over multiple time steps.

Phillips curve (supply curve)

Firms in an economy want to maximize their profits, given by the prices times consumption minus the wages times hours worked. The relationship between consumption and hours worked is given by a production function. Not all firms are allowed to change their prices each period, causing prices to be sticky. Inflation, the change in price level over time, is then log-linearized around a steady state with a constant inflation rate, which we assume to be 2%. We end up with the Phillips curve, as presented in Equation 2.3.

$$\begin{aligned}
\pi_t &= \beta\, E\left(\pi_{t+1}\right) + \kappa\, \gamma\, \widetilde{y}_t + \mu \\
\text{in which } \pi_t &= \text{inflation at time } t \\
\beta &= \text{discount factor} \\
E(\pi_{t+1}) &= \text{expected value of inflation at time } t+1 \\
\kappa &= \text{degree of price stickiness} \\
\gamma &= \text{time demand parameter} \\
\widetilde{y}_t &= \text{output gap at time } t \\
\mu &= \text{supply shock}
\end{aligned} \tag{2.3}$$

Positive (negative) inflation means that prices increase (decrease) compared to the previous quarter. In our analysis, the quarterly inflation rate will be adjusted to the annual inflation rate. The expected inflation positively affects the current inflation, creating a strengthening effect. A supply shock to the economy is modeled through the Phillips curve by a forced change in inflation. The current output gap, adjusted for the price stickiness and the degree to which the economy responds to a change in interest rate, positively impacts inflation. Price stickiness (κ) is determined by the percentage of firms that can change their prices in a certain period (a higher percentage results in lower price stickiness) and the discount factor that determines how future prices are currently valued (if the discount factor is higher, future prices are currently valued higher, resulting in higher price stickiness).

$$\begin{aligned}
\kappa &= (1-\omega)(1-\omega)\beta/\omega \\
\text{in which } \omega &= \text{Calvo parameter} \\
\beta &= \text{discount factor}
\end{aligned}$$

The response of the economy to a change in the interest rate (γ) is determined by the response of the agents to a change in their real wage rate, the ease with which the agents switch products and the share of impatient agents in an economy.

$$\gamma = \frac{\chi\,(\chi - z) + \sigma}{1 - z}$$

in which $\chi =$ inverse Frisch elasticity of labor supply
$z =$ targeted share of impatient agents
$\sigma =$ inverse of marginal elasticity of substitution

Taylor rule (central bank reaction function)
In response to the current economic situation, the central bank adjusts the interest rate. Based on the interest rate in the previous period, the current deviation from the steady state inflation rate, and the current output gap, the interest rate is calculated each time step through the Taylor rule (Equation 2.4). If inflation is higher than the steady state and if the output gap is positive, the interest rate increases. As we have seen from the dynamics inside our model, as shown by the Causal Loop Diagram in Figure 2.1, this should push down the deviation in the inflation rate and the output gap.

$$i_t = \rho_r\, i_{t-1} + (1 - \rho_r)(\varphi_\pi\, \pi_t + \varphi_y\, \widetilde{y}_t) \tag{2.4}$$

in which $i_t =$ interest rate at time t
$\rho_r =$ smoothing parameter policy interest rate
$\varphi_\pi =$ policy response to inflation
$\pi_t =$ inflation rate at time t
$\varphi_y =$ policy response to output gap
$\widetilde{y}_t =$ output gap at time t

In the Taylor rule, the weight of each of the three components (previous interest rate, current deviation from steady state inflation rate, and the output gap) is determined. First, the smoothing parameter of the policy interest rate determines how much current influence can deviate from the previous policy interest rate. If the smoothing parameter in the policy interest rate was 1, the interest rate would not change with time. If this parameter were 0, the current interest rate would be fully composed of its response to current economic variables and would not take into account the previous interest rate. The share of the interest rate that is not defined by the previous rate is computed by its response to current economic variables. The weights of the deviation from the steady state inflation rate and the output gap are decided upon by the policy response to, respectively, the inflation rate and the output gap. This policy response indicates the magnitude of the response: A positive deviation from the steady state of 1 percentage point leads to an increase in the current interest rate of 1 percentage point if this policy response is 1. However, this policy response can also be smaller than or larger than 1, indicating the magnitude of the central bank response.

2.2.1. Shocks

The exogenous shocks in these equations are the demand shock (d) and the supply shock (μ). These are initially applied in the first quarter, but they can persist over time. These shocks are defined as follows:

$$d = \rho_d d_{t-1} + \varepsilon_d \tag{2.5}$$

in which $\rho_d =$ autoregressive term of the demand shock
$\varepsilon_d =$ white noise error term

$$\mu = \rho_\mu \mu_{t-1} + \varepsilon_\mu \tag{2.6}$$

in which $\rho_\mu =$ autoregressive term of the supply shock
$\varepsilon_\mu =$ white noise error term

The shocks we apply are a positive demand shock and a negative supply shock. A positive demand shock simulates an increase in spending caused by, for example, stimulus checks from the government to boost the economy. On the other hand, the restrictions imposed on the economy in the COVID-19 pandemic that cause consumers to stay home cause a decrease in spending and would be modeled as a negative demand shock (DNB, 2020).

The consequences of the Russian invasion in Ukraine could be modeled as a negative supply shock. This disrupted the European energy market and caused delays in production. Finding new retrievable natural resources that could be seen as a positive supply shock.

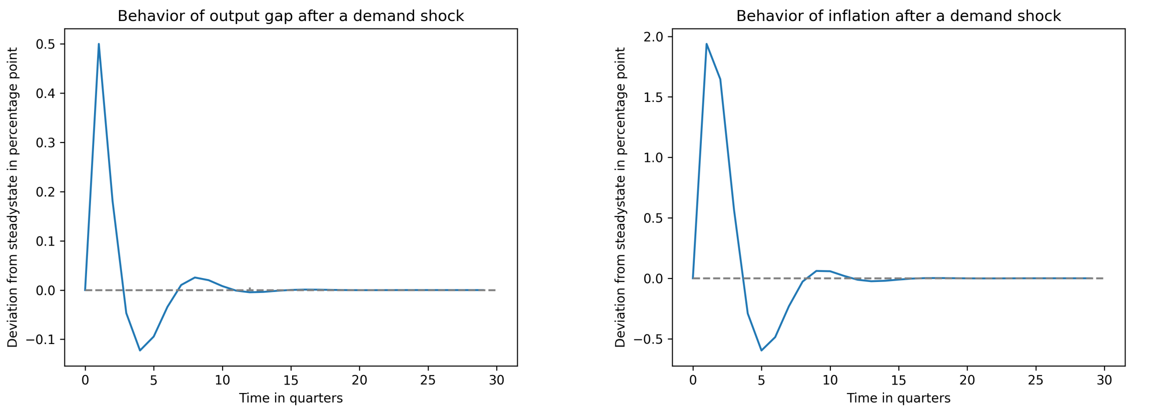
As we have seen in Equation 2.2 and Equation 2.3, shocks are simulated as a forced deviation from the steady state for the output gap or inflation. The output gap and inflation are defined as percentage changes from the steady state level of output gap and inflation, under the assumption that deviations are small. The shocks are defined in a similar way: A positive demand (supply) shock of 0.01 corresponds to a 1% increase in demand (supply).

In Appendix B, we show that the effect of a larger shock size results in a response that has a larger deviation with a similar relative change. A shock size of 0.1, 10 times larger than a shock of 0.01, resulted in a 10-times larger deviation from the steady state. Although this is largely the case, the ratio for the maximum or minimum values deviate slightly from the ratio of 10 for the supply shock. The maximum and minimum values are heavily influenced by the existence of outliers. In addition, in these cases, having assumed small deviations, the validity of the model should be questioned. Because we have no indication that, in addition to the linear multiplication of the magnitude of the deviation, other behavioral characteristics change when the shock size is changed, only one shock size is analyzed for each shock. For the model to be valid, the shock size has to be small. Therefore, a positive demand shock and a negative supply shock of 0.01 are chosen.

2.2.2. Behavior of the model

In the first quarter, a shock is applied to the economy. This can be either a positive demand shock, affecting the output gap, or a negative supply shock, affecting inflation. These two shocks have different effects on the economy.

Figure 2.2 shows the behavior of the economy after a demand shock. The increase in demand results temporarily in a larger actual economy than the potential economy, thus the output gap deviates in a positive way from the steady state. The inflation increases due to different reasons, with the most intuitive reason that the increase in demand leads to upward pressure on the prices. Another reason is the expected inflation rise due to this demand, resulting in an appeal to increase wages and, consequently, higher prices. As inflation increases, the output gap decreases as a result: An increase in prices results in a reduction in demand, which diminishes the potential economy.

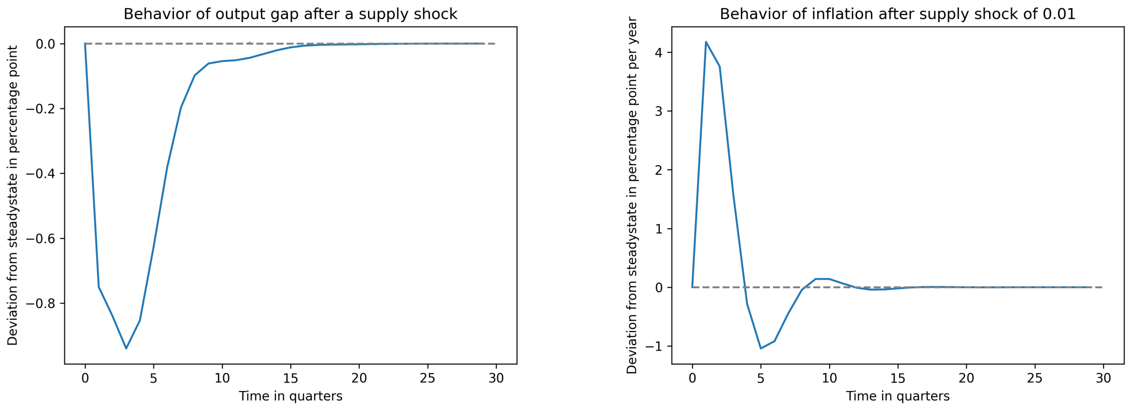


(a) The output gap deviates in a positive way from the steady state and oscillates before moving back to its steady state.

(b) The inflation deviates in a positive way from the steady state and oscillates before moving back to its steady state.

Figure 2.2: Behavior of the output gap and inflation after a demand shock. They show similar responses to a demand shock.

The behavior of the output gap and inflation is shown after a supply shock, such as an increase in oil price, is shown in Figure 2.3. Here, the output gap and inflation show opposite reactions. Whereas inflation increases, the output gap has a negative response to a supply shock. A cost-push shock results in higher costs for the production of goods, resulting in an initial positive effect on inflation. As a response to higher costs, demand in the economy decreases, and thus the actual economy drops compared to the potential economy. This effect might decrease inflation, causing inflation to oscillate before moving back to the steady state.



(a) The output gap deviates in a negative way from the steady state and slowly moves back its steady state.

(b) The inflation deviates in a positive way from the steady state and oscillates before moving back to its steady state.

Figure 2.3: Behavior of the output gap and inflation after a supply shock. They have opposite responses to a supply shock.

In our analyses, a large number of experiments with changing values for the parameters will be conducted. This will influence the behavior of the economy. The plot of all of these experiments will show differences in the magnitude of the deviation and its path to the steady state.

2.3. Interest of the central bank

The central bank, the only stakeholder in this model, uses its interest rate setting power to ensure economic stability. Within the scope of DSGE models, this means that the effect of a shock to the economy, represented by the output gap and inflation, is minimized. This means that the output gap and inflation should be close to the steady state over the entire time period and the period-to-period changes should be small. Besides this, the central bank, being inflation-focused, wants inflation to be back to normal 3 years (12 quarters) after the shock has occurred. The ideal annual inflation rate in the economy is 2%. In addition to this, deflation is undesirable due to the economic consequences of deflation (Krugman & Wells, 2021). No additional thresholds are included for the output gap.

Lastly, the central bank must be considered trustworthy. One of the implications of this need is that they want to have an accurate prediction of the consequences of their policy decisions and that their imposed policy has the same outcome no matter the exact state of the world. This is something that is typically assessed for in the robustness metrics and will not be translated into an objective.

In concrete terms, the central bank has the following interests:

1. Small output gap deviations;
2. Small inflation deviations;
3. Small period-to-period changes in output gap;
4. Small period-to-period changes in inflation;
5. Annual inflation rate after 12 quarters around 2 %;
6. No (or minimal) deflation in the economy;
7. Similar performance of the policy irrespective of the state of the world.

These interests should be reflected in the model objectives and robustness metrics for our directed search, which will be explained further in Section 3.2. Finding optimal policies will be done by optimizing our objectives, and the robustness of these optimal policies is assessed using the robustness metrics.

2.3.1. Objectives in our policy search

The objectives of our policy search are specific outcomes of the model we want to minimize or maximize. These objectives should align with the interests of the central bank. Therefore, eight objectives are defined as presented in Table 2.1. This table shows the objectives, the range in which it can lie, whether it should be minimized or maximized, and which interest it serves. Additional information on the objectives is presented in the Appendix C.

Table 2.1: Translation of the different interests into objectives that are either maximized or minimized.

Objectives			Interests					
Objective	Range	Maximize/ Minimize	Magnitude of deviations: Output gap	Magnitude of deviations: Inflation	Period-to-period change: Output gap	Period-to-period change: Inflation	Inflation after 12 quarters	Deflation
Maximum absolute output gap	Continuous [0,+∞)	Minimize	X					
Maximum absolute inflation	Continuous [0,+∞)	Minimize		X				
Standard deviation of output gap over time	Continuous [0,+∞)	Minimize			X			
Largest absolute change in inflation between 2 subsequent periods after Q2	Continuous [0,+∞)	Minimize				X		
Inflation rate at Q12	Continuous [0,+∞)	Minimize					X	
Quarters within inflation domain to Q12	Discrete [0,12]	Maximize		X		X	X	X
Quarters in which low inflation or deflation occurs to Q12	Discrete [0,12]	Minimize						X

2.4. Connecting DSGE to EMA

To connect our model to EMA, an XLRM-framework is designed, in which we distinguish the exogenous uncertainties (X), policy levers (L), the relations (R), and the measures (M) in the XLRM framework. This can be seen in Figure 2.4

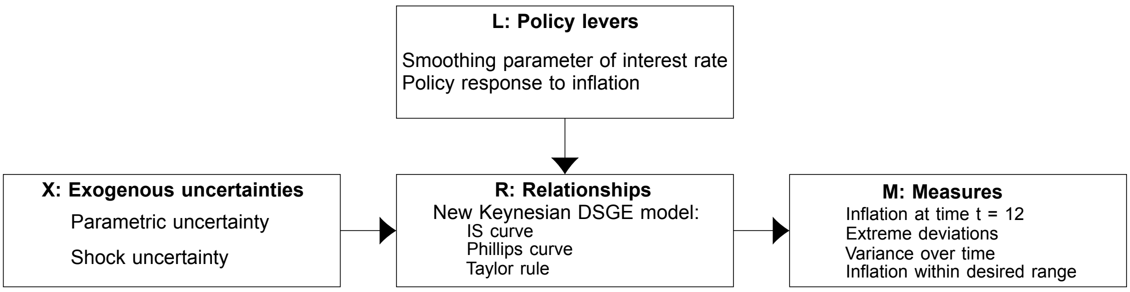


Figure 2.4: XLRM framework of the problem. X represents the exogenous uncertainties, L the policy levers, R the internal relations, and M the measures (or objectives) of the model.

All parameters, excluding policy levers, are considered uncertainties. According to the central bank's desires, only the interest rate smoothing parameter and the policy response to inflation are taken as policy levers.

3

Exploratory modeling and analysis

Before we go into the specifics of the methodology of this research, we have a general introduction to exploratory modeling and analysis (EMA). EMA is applied by policy makers to (1.) get a better understanding of the effect of uncertainties on a policy problem and (2.) design policies that will achieve its desired effect regardless of the exact state of the world (Bankes, 1993; Lempert, 2019). Methods to gain a better understanding of (the effect of the uncertainties on) the behavior of the model are part of the open exploration side of EMA. The other subset of EMA-based methods is directed search and is used for robust decision-making.

EMA requires us to dissect the model and the policy problem into different components: uncertainties, scenarios, levers, policies, outcomes, objectives, and robustness metrics. The uncertainties are the uncertain model parameters, which we assign a range of values to and a distribution. In our DSGE model, the uncertainties include the parametric uncertainties and the shock uncertainties. The different uncertainties together form the uncertainty space, as shown in Figure 3.1. A scenario is a point in this uncertainty space: a combination of specified values for the different uncertainties. The (policy) levers are parameters or components of the model that the decision maker can influence, to which we assign a range and distribution. These levers form a decision space (comparable to the uncertainty space), and a policy is a point in this decision space. In our case, the central bank controls how the interest rate is set, and the different policies thus refer to the overall interest rate setting behavior of the central bank in the model. Running the model for one scenario and one policy is referred to as an experiment, and for each experiment, a set of outcomes is obtained. When designing a policy, it is important to identify the interests of the stakeholders. These are translated into objectives, outcomes of the model we want to minimize or maximize, and robustness metrics, the criteria on which we assess the robustness. Like the levers and uncertainties, the outcomes and objectives form their own spaces: the outcome space and the objective space. In Appendix D additional information is found about this basic structure of EMA.

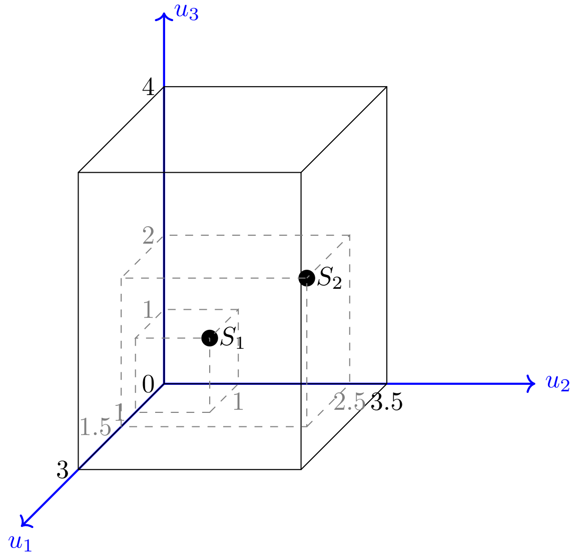


Figure 3.1: Visualization of the uncertainty space created by three theoretical continuous uncertainties u_1 (range (0,3)), u_2 (range (0,3.5)), and u_3 (range (0,4)). Two possible scenarios are shown: $S_1(u_1 = u_2 = u_3 = 1)$, and $S_2(u_1 = 1.5; u_2 = 2.5; u_3 = 2$.

It is inefficient, if not impossible, to compute the model for the whole uncertainty and/or decision space. To this end, we have to pick scenarios and policies at random, which is done through a Latin Hypercube sampler. Latin Hypercube first divides the ranges of each parameter into separate segments, thus dividing the uncertainty and decision space into separate spaces, cells, with all having the same probability of being chosen (McKay et al., 2000). It then picks a random value in a randomly chosen cell. Figure 3.2 shows a theoretical uncertainty space created by three uncertainties and divided into 27 cells. Although using Latin hypercube sampling helps us cover the whole uncertainty space, we are still required to use a large number of experiments to make sure we get a full understanding of the whole uncertainty space. The downside is the increase in computational time to conduct the experiments and subsequent analyses. By conducting 15000 experiments, we balance out covering the whole uncertainty space and the computational time.

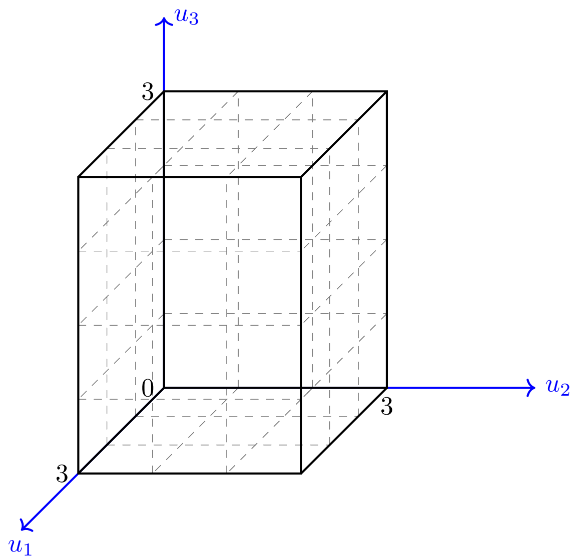


Figure 3.2: Uncertainty space in which the ranges from the uncertainties are divided into equal segments. Three uncertainties with each 3 segments results in 27 cells. Whereas Latin Hypercube sampling picks a cell at random and creates a gives a scenario from this cell, Full Factorial sampling returns a scenario of all cells.

As stated above, two different approaches exist in EMA: open exploration and directed search. In open exploration, we conduct computational experiments with different configurations that cover uncertainty space (Kwakkel & Pruyt, 2015). In practice, this means that we assign a range and distribution to the uncertain parameters, aligned with our expectations of the true value of these parameters, and analyze the outcome space following these uncertainties. Consequently, we see how the outcomes behave based on the different combinations of the parameter and lever values. This is the approach we will take to answer our first sub-question. In directed search, we search the space in a directed manner using some type of optimization approach, finding policies that satisfy our objectives (Kwakkel & Pruyt, 2015). To this end, we translate the interests of the central bank into objectives. The found policies, or a selection thereof, is then evaluated on its performance and robustness. The directed search is used for our second sub-question.

3.1. Open Exploration: Scenario discovery

Open exploration is, as the name suggests, exploratory of nature. By doing so, we want to improve our understanding of the effect of uncertainties on the outcome space. One of the ways to achieve this is through scenario discovery. In scenario discovery, scenarios are defined as a set of future states of the world that illuminate vulnerabilities of a proposed policy (Dalal et al., 2013). Halim et al. (2016) describes scenario discovery as 'a series of computational experiments to explore the consequences of various unresolved uncertainties' (Walker et al., 2013) that are 'analyzed through statistical machine learning to identify combinations of these uncertain developments that produce characteristic results.' In other words, with our scenario discovery, common input space properties are identified across ensembles of exploratory model runs (Steinmann et al., 2020).

Figure 3.3 shows how 5 different scenarios (2 uncertainties) from the uncertainty space (left, dotted) result in different outcomes (2 outcomes) lying in the outcome space (right, dashed), in the absence of multiple policies. In practice, many more experiments are done.

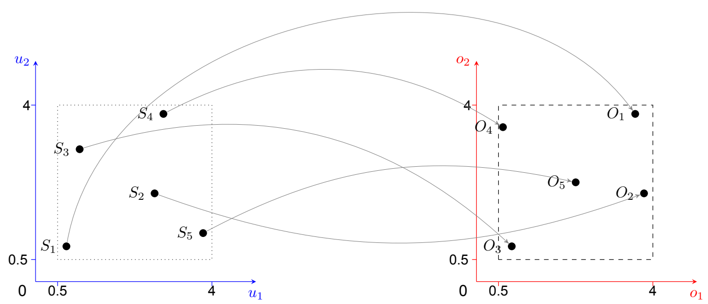


Figure 3.3: Representation of 5 experiments with random scenarios in the uncertainty space (dotted, left) and their corresponding outcomes in the outcome space (dashed, right) of a model with two uncertainties and two outcomes. No policies (or 1 policy) is included. As it is only a theoretical case, the same ranges are given to both uncertainties and both outcomes but this does not have to be the case.

From the outcome space, we take a characteristic we want to explore and compare the experiments that share these characteristics. To illustrate, in Figure 3.4 the dots in the uncertainty space (right), in the absence of policies, represent experiments and in the dots in the outcome space their corresponding outcomes. The area of the outcome space we are interested in is shown by the red dashed rectangle. In this area, as represented by the larger yellow dots, we can see that 11 experiments resulted in o_1 being greater than 3, which are visualized as larger blue dots in the uncertainty space. Analyzing these experiments, we can identify regions in the uncertainty space that are of interest (Bryant & Lempert, 2010; Kwakkel et al., 2013). In this way, we could find that the scenarios resulting in our outcomes of interest are located in the blue dashed box in the uncertainty space on the right. This gives us the indication that the case in which $u_1 \in (3, 4)$ and $u_2 \in (2, 4)$ are likely to result in a value of more than 3 for o_1. However, in the blue dashed area, we can also observe black dots, indicating that not all scenarios in this area will result in $o_1 > 3$. We will elaborate on this part of the analysis later when we discuss the Patient Rule Induction Method in Section 3.1.4.

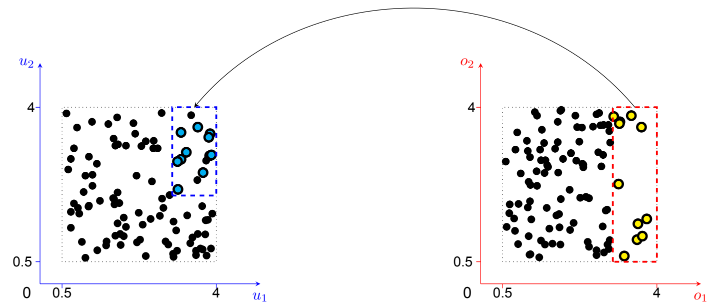


Figure 3.4: Uncertainty space and outcome space of a model with 2 uncertainties, 2 outcomes and no policy levers. 100 scenarios (experiments) (dots in the uncertainty space) and their corresponding results (dots) in the outcome space are visualized. Our region of interest in the outcome space (red, dashed rectangle) consists of the results (yellow-filled larger dots) of 11 experiments (blue-filled dots). These experiments lie in the blue, dashed area, but not all experiments in the blue, dashed area resulted in $o_1 > 3$.

As illustrated, a certain character of the outcomes forms the basis of our scenario discovery. The characteristic on which the selection is made is case-specific. In this research, the time series for inflation and the output gap show different types of behavior for different experiments, such as oscillation, the size of the deviations, and how long large deviations persisted over time. To find out what causes these different behaviors, the time series behavior is taken as the basis for our scenario discovery.

3.1.1. Behavioral based scenario discovery using time series clustering

In our scenario discovery, we do not consider any policies yet. The levers are considered as uncertainties in this step, so we can consider the effect of the policy levers on the outcome without having it set beforehand.

In this thesis, scenario discovery will be done in a relatively novel way: behavioral-based scenario discovery using time series clustering, as described by Steinmann et al. (2020). In other words, we will cluster the time series based on their behavior (such as the size of the deviation from the steady state or its path back to the steady state). A visualization of how this might look is given in Figure 3.5. In this figure, 10 time series are shown that are grouped into 3 clusters (red, blue, yellow). In this example, there are clear behavioral differences between the three groups. Note that this is a hypothetical case which will be used as an example in the following paragraphs.

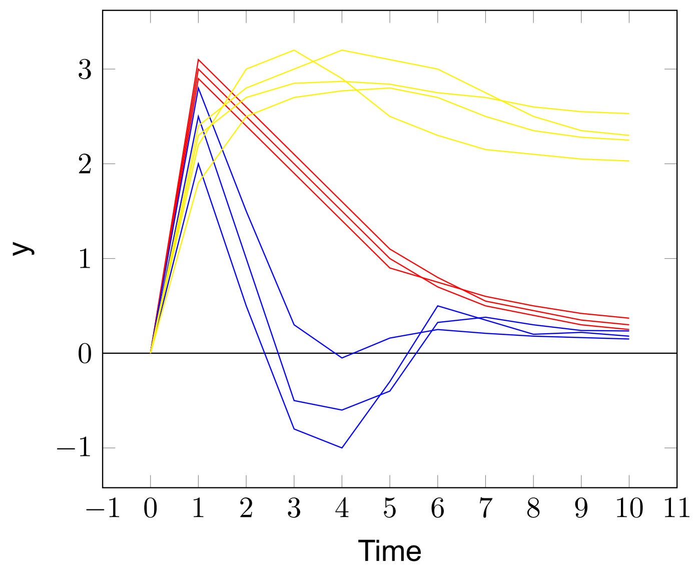


Figure 3.5: Visualization of 10 time series grouped into 3 clusters (red, blue, yellow). Time series showing comparable behavior are clustered into the same cluster.

The parameter values are taken from the parameter space that captures the aleatory and epistemic uncertainties of the model Hoffman and Hammonds (1994). We decided not to set the policies beforehand, so the levers are in this stage considered uncertainties. In Appendix D, the difference in approach is seen. Limited by the computer memory at a later stage of the analysis, the scenario discovery is done using 2000 scenarios (and thus 2000 experiments).

As described by Steinmann et al. (2020), the behavioral-based scenario discovery using time series clustering consists of three steps:

1. A number of simulation experiments are performed, randomly sampling from the input parameter space.
2. time series clustering is applied to find common macrolevel behaviors in the ensemble of output time series.
3. The patient rule induction method (PRIM) is performed for each time series cluster.

In the next paragraphs, the similarity metric that is used for clustering and the index that is used for internal validation will be discussed, before delving into detail on PRIM.

If no proper distinctions are found in the different clusters, we will adjust the dataset by removing the smallest clusters. Under the assumption that the experiments from the largest clusters are unlikely to be clustered in these smallest clusters as these are most likely to have a large distance, this has the same result as adding clusters. To decrease computation time, we have removed these experiments instead of simply adding numbers of clusters.

3.1.2. Similarity metric: Complexity-Invariant Distance

The similarity metric that is built into the EMA workbench is the Complexity-Invariant Distance. In this subsection, we will solely focus on explaining this algorithm and checking whether or not this algorithm is suitable for our case. [to be added]

To cluster the time series, the distance between each time series is calculated, corrected with a correction factor. The uncorrected distance between two time series, also called the Euclidean distance (ED), is shown in Figure 3.6. The ED between the black (solid) time series and blue (dashed) is shown in red (lined) in this figure. This value is calculated using Equation 3.1. Calculating the distance between time series Q and C with length n, the distance between $q(i)$ and $c(i)$ is calculated (the value for Q and C respectively at time step i).

$$ED(Q,C) \equiv \sqrt{\sum_{i=1}^{n}(q_i - c_i)^2} \quad (3.1)$$

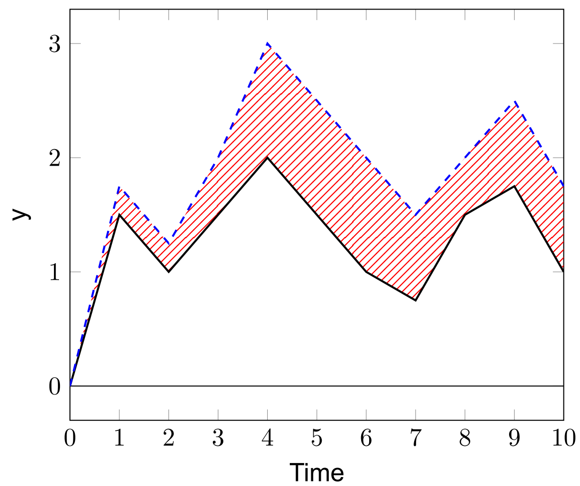


Figure 3.6: Euclidean distance (red lined area) between the black (solid) and blue (dashed) hypothetical time series.

The ED is corrected for the complexity of the time series, which is where the name 'complexity-invariant distance' comes from. Complexity refers to the 'peaks and valleys' in a time series: a straight line is less complex than a time series that goes up and down. The formula to compute this correction factor is given in Equation 3.2.

$$
\begin{aligned}
CID(Q,C) &= ED(Q,C) \times CF(Q,C) \\
CF(Q,C) &= \frac{\max(CE(Q), CE(C))}{\min(CE(Q), CE(C))} \\
CE(Q) &= \sqrt{\sum_{i=1}^{n-1}(q_i - q_{i+1})^2}
\end{aligned}
\tag{3.2}
$$

in which $CE(T)$: complexity estimate of a time series T

The complexity of each time series is calculated using the distance between two time steps, see Figure 3.7 For a horizontal line, the distance between two time steps is $1\times$ unit of the x-axis. In a diagonal line, this distance is larger: $\sqrt{(q_{i,y} - q_{i+1,y})^2 + (q_{i,x} - q_{i+1,x})^2}$. A more complex time series with more peaks and valleys results in a higher complexity estimate.

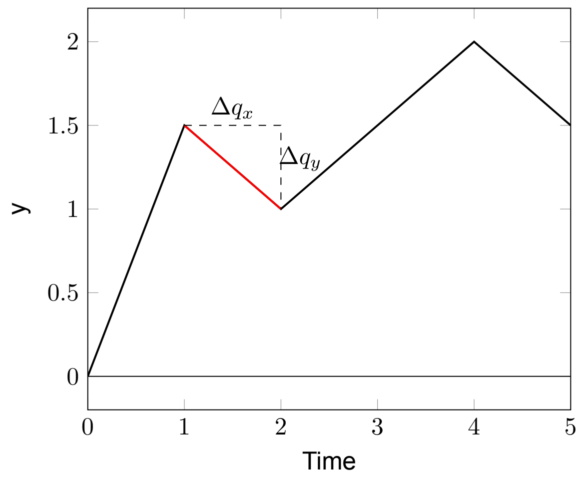


Figure 3.7: Distance between time steps 2 and 3 for time series Q.

The ED between two series is then corrected with the ratio of the complexity estimates of two time series. If the complexity is the same, the CID is equal to the ED. However, if these are unequal, the correction factor is larger than 1 (we divide the larger CE divide the smaller CE), resulting in a larger CID.

Whereas nearest neighbor classification, which does not correct for complexity and simply considers the ED, is usually a good classification method, it might cluster complex time series in simpler clusters (Batista et al., 2014). A case is shown in Figure 3.8. The black (solid) and blue (dashed) time series show similar behavior. However, because the distance between the simple (red, dotted) and the black line is smaller, the nearest neighbor classification will cluster these two together. In our case, as we focus on the behavior of the time series, we prefer the black and blue line to be clustered together. As the correction estimate of the blue and black line are more similar than the correction estimate of the black and red line, after the correction for complexity, the CID between the black and blue should be smaller than the CID between the black and red time series.

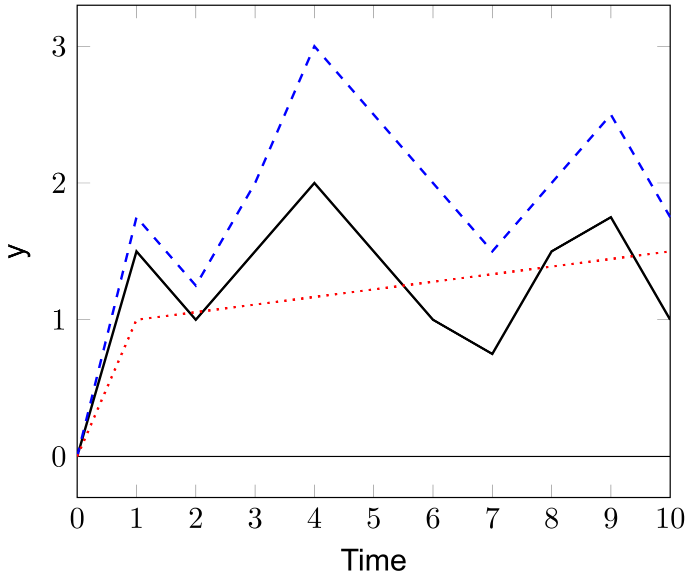


Figure 3.8: Two complex time series (black(solid) and blue (dashed)) and a simple (red, dotted) time series. Taking the nearest neighbor classification, the black and red time series might be clustered together, whereas the two complex time series show similar behavior.

3.1.3. Internal validation: Silhouette index

To quantify how well the clustering algorithm has worked, we use an internal validation index. There are several indices that we can use, with varying conditions and varying cases in which they do or do not work properly. The recommended approach by Arbelaitz et al. (2013) is to take the silhouette index (SI), due to its low computation time and simplicity (Dudek, 2020). However, caution should be taken when using the SI without checking its suitability to the clustering data of interest. For example, according to Jain et al. (2022) the SI is suitable for overlapping clusters (which is the case for our data), but less applicable when we have outliers (which these data do not have). Our data do not show any indication that the SI is not applicable for our case.

To calculate the SI, we have to know the tightness of the clusters (the distance between a time series and the other time series in the clusters, see Figure 3.9) and the separation (the distance between the time series of interest and the time series in other clusters, see Figure 3.10) (Rousseeuw, 1987). Knowing the tightness and the separation for each time series, we subtract the tightness from the separation and divide by the maximum of the two; see Equation 3.3. This will result in an SI of maximum 1 (good fit) and minimum -1 (bad fit). Let us consider the two extremes.

- **Good fit:**
  - Tightness is small: Distance to other time series in same cluster is small.
  - Separation is big: Distance to time series in other clusters is large.
  - Numerator $\approx$ separation.
  - Denominator $=$ separation (separation $>>$ tightness)
  - SI $= 1$
- **Bad fit:**
  - Tightness is big: Distance to other time series in same cluster is large.
  - Separation is small: Distance to time series in other clusters is small.
  - Numerator $\approx -$tightness.
  - Denominator $=$ tightness (tightness $>>$ separation)
  - SI = -1

$$SI = \frac{1}{n}\sum_{i=1}^{n}\frac{b(i)-a(i)}{max\{a(i);b(i)\}}$$

in which $n =$ number of time series

$b(i) =$ average distance between time series i and the time series in other clusters

$a(i) =$ average distance between time series i and the other time series in the same cluster (3.3)

One can imagine that 3 is the ideal number for our theoretical data, as shown in the figures in this section.

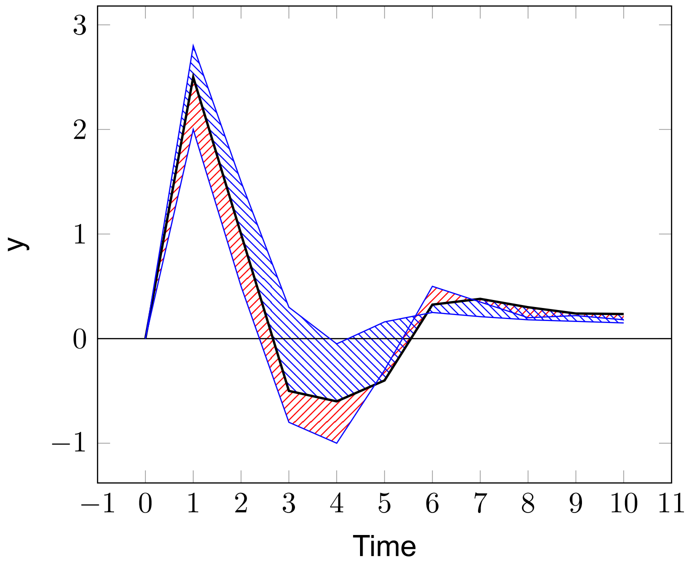


Figure 3.9: Calculation of distance between our time series of interest (black line) and the other time series in the cluster (blue time series). Distance between the lines is given by the blue and red area to the two other time series in the cluster.

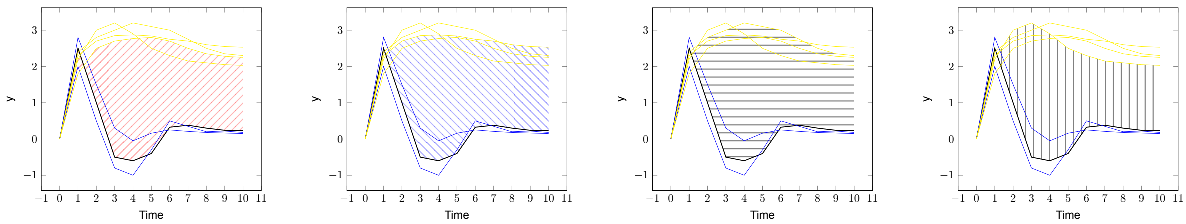
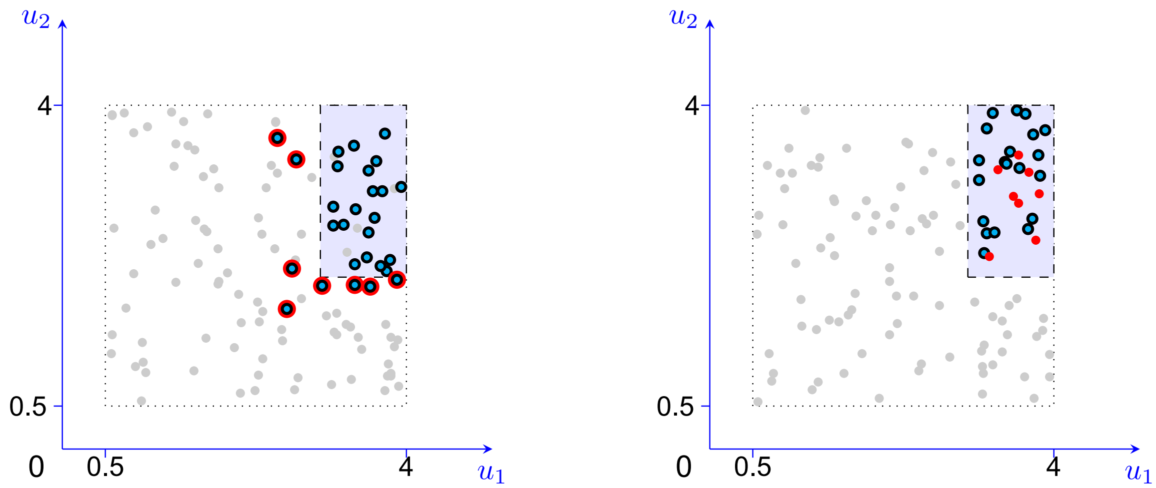


Figure 3.10: Distance between time series of interest and the time series in the yellow clusters.

Ideally, we would use the SI on the clustering of our final data. However, because computing the SI is computationally heavy as it has to calculate the distance between each and every experiment time series, we will use data with fewer experiments, namely 2000 experiments. This will be done for five experiment runs, thus creating 5 $\times$ 2000 experiments. Based on the five experiment runs we will see whether we can find an ideal number of clusters or whether we can find a relation between the number of experiments per cluster and the SI's.

3.1.4. Patient Rule Induction Method (PRIM)

The patient rule induction method (PRIM), as described by Friedman and Fisher (1999), aims to find combinations of values for uncertain input variables that result in similar characteristic values for outcome variables (Kwakkel, Haasnoot, et al., 2016). Now that the experiments are categorized, we search for commonalities among the experiments per category. Specifically, PRIM seeks a set of subspaces of the uncertainty space within which the value of a single output variable is considerably different from its average value throughout the domain (Kwakkel & Jaxa-Rozen, 2016). These subspaces give us a range for the parameters and a quasi-p-value that identifies whether the restriction imposed by the box is of statistical significance. Important to note is that this method considers the parameters to be uncorrelated. We consider three characteristics of these subspaces: coverage, density, and number of restrictions. As visualized in Figure 3.11a, coverage looks at the fraction of all experiments of interest that fall within the box, while density (see Figure 3.11b) is the fraction of experiments within the box that are of interest (Bryant & Lempert, 2010). The number of restrictions refers to the number of uncertain parameters that are used to describe the box. The algorithm presents us with a set of boxes that increase in the degree to which it is restricted. We can imagine that, as a subspace becomes more specific, it might lose some experiments of interest, resulting in a loss in coverage. However, we will also lose experiments that we are not interested in, increasing the density.



(a) The experiments of interest that are marked by a red circle are not included in our hyperrectangle (dashed, blue area). The share of experiments of interest that are included in the hyperrectangle is called coverage.

(b) The hyperrectangle includes both experiments of interest (larger dots, blue center), and experiments we are not interested in (red dots). The share of experiments of interest of all experiments in the hyperrectangle is referred to as the density.

Figure 3.11: Visualization of coverage and density of the PRIM hyperrectangles.

Usually, the last box with the largest density and the largest number of restrictions is the one we are interested in. However, other boxes that can be of interest are those with the largest density with fewer restrictions. Taking into account these boxes of interest, we make an informed decision on which box represents the scenario the best. This can be based on the range and the quasi-p-value of the parameters that describe the scenario the best for each box.

After finding the commonalities, we can discuss what might have caused the specific behavioral characteristics of the time series over time. Because these analyzes are done for the output gap and inflation separately, we can compare the results and compare how their behaviors are influenced by the uncertainties.

3.2. Directed Search

In the next step of our analysis, we introduce policies, combinations of policy levers. The same scenario can, under different policies, result in different outcomes. In Figure 3.12, the uncertainty space of two uncertainties (u_1 and u_2) result in three different outcome space (outcomes o_1 and o_2) due to three different policies (P_1, P_2, and P_3. For example, Policy 1 creates an outcome space that spans over a, relatively, large range for objective 1 and has, relatively, high values for objective 2. Note that, whereas we know the distributions of the uncertainties (in our case we assume a uniform distribution), we do not know the distributions of the outcomes in the different outcome spaces. This distribution of outcomes can differ per policy. The outcome spaces of these policies are the basis for determining the robustness of the policies, which is further explained in Section 3.2.2.

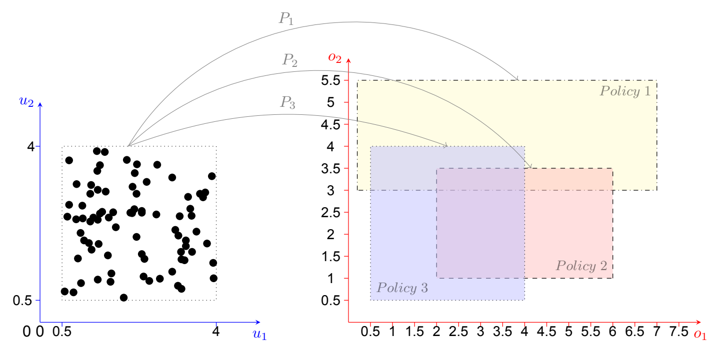


Figure 3.12: Three outcome spaces caused by the same uncertainty space under different policies (P_1, P_2, and P_3).

Before we can assess the robustness of different policies, these policies have to be designed. We do this based on the objectives of our model, which are defined in Section 2.3.1. Like we have seen for the uncertainty space and decision space, an objective space can be created using the objectives. We search for combinations[1] of policy levers that optimize our objectives. More specifically, we search for policies that are Pareto optimal: performance on one of the objectives cannot be improved without reducing performance on (at least) one other objective. Note that this is not Pareto optimality in the economic sense of the word but from a mathematical perspective. The plot of the outcomes of these Pareto optimal policies in our objective space is the Pareto front. This Pareto front shows the trade-offs of the system: as stated before, each point on the Pareto front represents outcomes from which cannot be deviated to improve the performance of one outcome without reducing the performance of at least one other outcome. It is important to note that we do not know the exact Pareto front, but we approximate it through our optimization methods. However, in this thesis, unless stated differently, we refer to these approximations as Pareto optimal solutions.

Finding these policies is a multi-objective optimization problem. Statnikov and Matusov (2012) describes the multiobjective optimization problem as the problem of finding “a vector of decision variables that satisfies the constraints and optimizes a vector function whose elements represent the objective functions. These functions form a mathematical description of performance criteria, which are usually in conflict with each other. Therefore, the term ’optimize’ means finding a solution that would give the values of all objective functions acceptable to the decision maker’ (Coello, 2007). In other words, we find the optimal values for the policy levers that satisfy a set constraints (optional) that optimize the outcomes (either minimizing or maximizing them). We do so through a multi-objective evolutionary algorithm, more specifically $\epsilon-$NSGAII as is further elaborated upon in (Kollat & Reed, 2007). The effectiveness of the optimization can be checked through performance metrics. An elaboration of these metrics and the results of these metrics can be found in the Appendix E. We find that the optimization convergences.

3.2.1. Orientation of optimal policies

The policies that are found by the algorithm are compared on their performance for the different objectives. As there are trade-offs in the model, different policies will have the most desirable performance on at least one objective.

For this thesis, the policies for further analysis are selected if they have the most desirable performance on at least 1 objective. It is possible that a policy is performing the best on multiple policies. Because we have two objectives that are discrete, it is likely that multiple policies perform the best on these objectives. Based on the number of policies that perform the best on this objective, we either make a decision to take both, and/ or check if some of these policies are performing also the best on other objectives.

It is possible to make other decisions, by considering the 2 or more best policies on an objective, or checking for a policy that is, for example, performing in the top 10% on all objectives.

[1] If the model has a trade-off in its objectives, it is impossible to find one policy that optimizes both objectives.

3.2.2. Robustness

It is important to assess policies on their robustness, and not on the outcome of a single case; policies that are optimal in some scenarios may underperform in other scenarios. To evaluate this robustness, we use the taxonomy of robustness frameworks as presented by Herman et al. (2015) This taxonomy consists of four components, which will be explained using our case and focused on the methods we use. Other methods for the different components are possible and can be found in Kwakkel (2017), but will not be considered in this research.

1. **Generation of policy options**: The generation of policies is done through a many-objective search. The model will present optimal policies, which essentially are optimal configurations of the policy levers (smoothing rate of interest rate and the policy response to inflation),
2. **Generation of states of the world**: Using the previously policy levers, we are interested in how these policies will perform in different states of the world, which can be predefined using results of the open exploration, or which can be sampled randomly.
3. **Vulnerability analysis**: Having created the policies and the states of the world, we can conduct a vulnerability analysis.
4. **Robustness evaluation**: To compare the different policies, we use different robustness metrics. These metrics have to be specified per case. We can distinguish three types of robustness metrics (Kwakkel, Eker, et al., 2016).

3.2.3. Robustness metrics

Three different types of robustness metrics can be considered: regret metrics, satisficing metrics, and descriptive metrics. Based on the case, the best (set of) robustness metrics has to be decided upon. A robustness metrics considers how well a policy is acting on a specific objective.

A regret metric measures on how well the policy performs compared to best performing policy for each scenario.[2] We then take the 10th (maximization of results) or 90th (minimization of results) percentile of these deviations from the best performance, to compensate for outliers (Herman et al., 2015; Kasprzyk et al., 2013). In Figure 3.13, we use the regret metric to compare the performance on objective 2, which we want to minimize. Policy 2 has the most desired performance for objective 2 for Scenario 1, meaning that $R_{1,2}$ (Regret for scenario 1, Policy 2) is 0. For Policy 1 and Policy 3 the 'regret' is, respectively, $R_{1,1}$ and $R_{1,3}$.

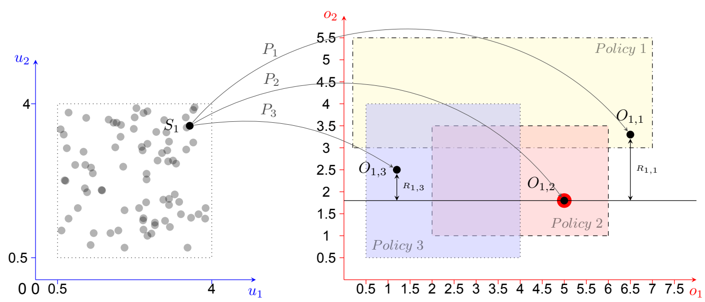


Figure 3.13: For scenario 1 (S_1), Policy 2 performs has the best performance on objective o_2. The distance between the most desired performance and the performance of Policy 1 and Policy 3 are given by, respectively, $R_{1,1}$ and $R_{1,3}$. $R_{1,2}$ equals 0.

In other scenarios, other policies might perform better on objective 2. For each scenario, the distance to the best performance is taken, as shown in Equation 3.4.

$$
\begin{aligned}
\textbf{Minimization}: r_i &= \frac{f_{i,90} - f_{i,optimal}}{f_{i,optimal}} \\
\textbf{Maximization}: r_i &= \frac{f_{i,10} - f_{i,optimal}}{f_{i,optimal}} \\
&\text{in which } i = \text{measure of interest}
\end{aligned}
\tag{3.4}
$$

[2]It is also possible to compare the policy to the baseline policy. In this case, the regret metric measures the distance between the performance of the policy of interest and the performance of the baseline policy.

We will use the regret metric for to assess the performance of the policies on the standard deviation of output gap, variance of inflation between $t = 0$ and $t = 12$, (absolute) maximum value for inflation and for output gap. For the objectives that consider certain behavior/events in each period (like the periods in which deflation occurs) this metric might be less insightful, because these are discrete values in the range [0,12]. It can still provide insights, which it will still be evaluated, but a satisficing metric is considered to be a better metric.

A satisficing metric measures how many experiments have a desired performance. In Figure 3.14, we show how the three policies perform for a maximum threshold of 3 for o_3.

We have two clear thresholds in our goals: no periods in which deflation occurs, and the number of periods inflation is not between 0% and 2% yearly. We can also add the level of inflation at time $t = 12$ which we want to be between 0% and 2% yearly. Using baseline/standard behavior of the model or the results of the Open Exploration, we can, for example, decide on a threshold for the maximum value or variance. For now, the performance on these objectives are measured by a regret metrics.

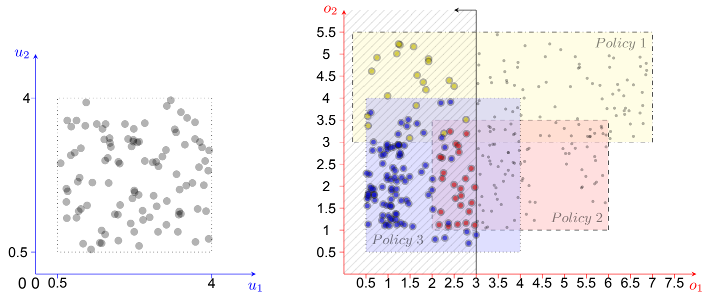


Figure 3.14: We want to have a policy that performs such that $o_1 < 3$. Policy 3 (blue outcomes), show the most experiments that fulfill this criterion.

Lastly, we can use descriptive measures to get an insight on the likelihood of the performance of the policy. As figu However, because we are unsure about the distribution of the input parameters (we have assumed a uniform distribution), we should be cautious using this metric. Only if we have more insights on the distribution of the input parameters becomes it insightful to

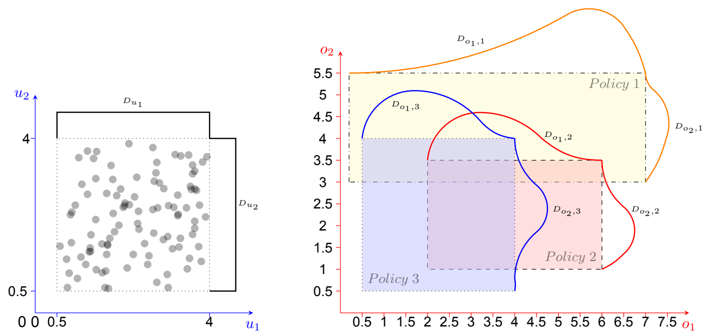


Figure 3.15: Distribution of performance of different objectives for Policy 1, 2, and 3 given by the $D_{objective,policy}$. Whereas the most ideal performances for 0_1 (minmized) might occur under Policy 1, it is more likely that Policy 1 will have high values for o_1, given its distribution $D_{o_1,1}$.

We will consider the descriptive metrics as this is in line with the desire of the central bank to have an accurate prediction, but we should be aware of the limitation.

3.3. Experimental set-up

In order to do conduct this research, we use Python. Conventionally, DSGE models are solved using the Dynare+-package in Matlab. One of the applications of this package is that the parameters of the DSGE model can be estimated using real-world data. However, the workbench we will use for EMA requires us to solve the DSGE models in Python. Ideally, to allow for future research that includes more functionalities of the DSGE, a package is found in Python that is similar to the Dynare+-package in Matlab. In addition to this, this Python package has to be suitable for working with the ema-workbench and, ideally, is up to date and documented. This led us to the pydsge package developed by Gregor Boehl, as described by Boehl (2022).[3]

Furthermore, we use the open source EMA workbench from https://github.com/quaquel/EMAworkbench (Kwakkel, 2017). The complete code and analyses can be found in https://github.com/cwieles/pydsge-ema_workbench. The simulations are run for 30 time steps [0,29] and the shocks are simulated to occur at time step t=1. Latin Hypercube is used for the randomization.

Table 3.1: Experimental set-up of analyses done in this research.

Experiments [seed]	Shocks	Uncertainties	Scenarios	Levers	Policies	Outcomes
5 x 2000 [0, 3, 17, 23, 31]	Demand (0.01)	Uncertainties Levers	5 x 2000	-	-	Output gap Inflation
15000 [0]	Demand (0.01)	Uncertainties Levers	15000	-	-	Output gap Inflation
2 x 15000 [0, 3]	Demand (0.1)	Uncertainties Levers	15000	-	-	Output gap Inflation
5 x 2000 [0, 3, 17, 23, 31]	Supply (0.01)	Uncertainties Levers	5 x 2000	-	-	Output gap Inflation
15000 [0]	Supply (0.01)	Uncertainties Levers	15000	-	-	Output gap Inflation
2 x 15000 [0, 3]	Supply (0.1)	Uncertainties Levers	15000	-	-	Output gap Inflation
5 x nfe=100000* [0, 1, 2, 3, 4]	Supply (0.01)	Uncertainties	-	Levers	-	Optimal policies, performance metrics, performance on objectives
15000 [0]	Supply shock (0.01)	Uncertainties	15000	Levers	Selected policies	Output gap Inflation Interest rate

Uncertainties: discount factor, inverse marginal elasticity of substitution, Calvo parameter, persistence of demand shock, persistence of supply shock, inverse Frisch elasticity of labor supply, share of impatient agents.
Levers: policy response to inflation, smoothing parameter of the interest rate.
The randomization of scenarios is done using a Latin hypercube.
* This experiment is an optimization function.

The ranges of the uncertainties we have included in the open exploration part of the thesis are presented in Table 3.2. This is also the range on which we have conducted our multi-objective optimization in the directed search. The ranges are based on the expert knowledge and, broadly speaking, are uniform distributions with the mean being the standard used value in non-estimated DSGE models.

Table 3.2: Model uncertainties for open exploration

Parameter	Distribution	Lower limit	Upper limit
Discount factor	Uniform	0.95	1
Inverse of marginal elasticity of substitution	Uniform	1	2.5
Calvo parameter	Uniform	0.6	0.95
Autoregressive term of demand shock	Uniform	0.5	0.95
Autoregressive term of supply shock	Uniform	0.5	0.95
Inverse Frisch elasticity of labor supply	Uniform	0.6	1.5
Targeted share of impatient agents	Uniform	0.2	0.5
Policy response to inflation	Uniform	1	2.5
Smoothing parameter policy interest rate	Uniform	0.5	1

On the basis of open exploration after the negative supply shock, an additional set of uncertainty ranges is identified. This is done to avoid extreme deviations from the steady state, thereby ensuring the validity of the model. This additional set is presented in Table 3.3 and is used for the evaluation of the robustness in the directed search.

[3]For more information visit the website of the researcher (gregorboehl.com), the documentation, and the Github repository.

Table 3.3: Model uncertainties for evaluation of policies in directed search

Parameter	Distribution	Lower limit	Upper limit
Discount factor	Uniform	0.95	1
Inverse of marginal elasticity of substitution	Uniform	1	2.5
Calvo parameter	Uniform	0.6	0.81
Autoregressive term of demand shock	Uniform	0.5	0.95
Autoregressive term of supply shock	Uniform	0.5	0.83
Inverse Frisch elasticity of labor supply	Uniform	0.6	1.5
Targeted share of impatient agents	Uniform	0.2	0.5

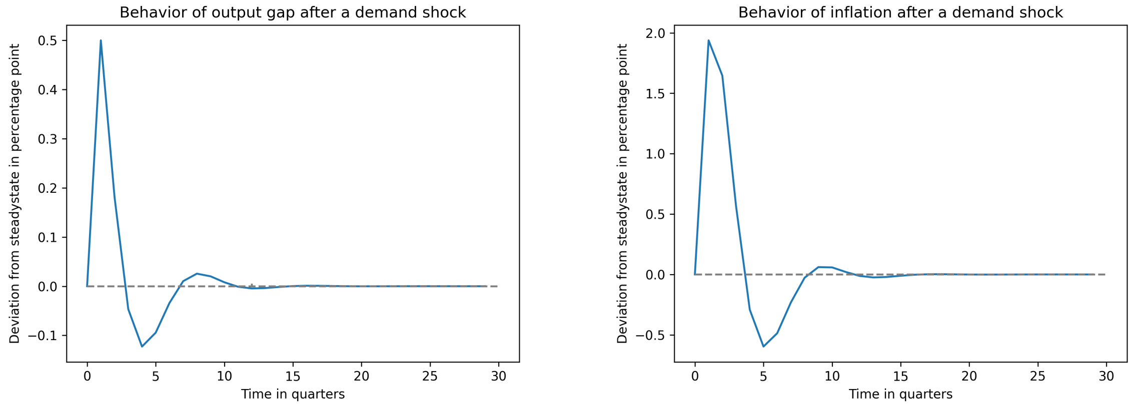
4

Results Open Exploration

In this chapter, we will show the results of the open exploration of our DSGE model after a demand and after a supply shock.

4.1. Demand shock

Figure 4.1, the same figure as we have seen in our model chapter, the standard behavior due to a demand shock is shown. We see similar behaviors for inflation and the output gap: an initial positive deviation, followed by an oscillation going back to the steady state.



(a) The output gap deviates in a positive way from the steady state and oscillates before moving back to its steady state.

(b) The inflation deviates in a positive way from the steady state and oscillates before moving back to its steady state.

Figure 4.1: Behavior of the output gap and inflation after a demand shock. They show similar responses to a demand shock.

The effect of changing the parameters slightly in 15000 experiments is shown in Figure 4.2. Compared to the standard behavior, we can see several differences in the behavior of the output gap. The peak deviation in output gap was 0.5 percentage point with normal parameters, but with changing parameters it can peak to more than 0.8 percentage point. Whereas the deviation is close to 0 percentage point at time t = 15 for the standard behavior, with changing parameters it might still be around 0.1 percentage point away from the steady state (both positive or negative) and around 0.05 percentage point from the steady state at time t = 30. In the standard behavior, we see a small oscillation from the peak positive deviation of around 0.5 percentage point from the steady state to a negative deviation of -0.1 percentage point from the steady state, back to a positive deviation of around 0.04 percentage point before going back to the steady state. We see multiple types of behavior in our experiments: oscillation like the standard behavior with different amplitudes (size of the deviation from the steady state), experiments going up once and moving back to the steady state without any negative deviations, and one oscillation and moving towards the steady state from after its dip without extra oscillations around

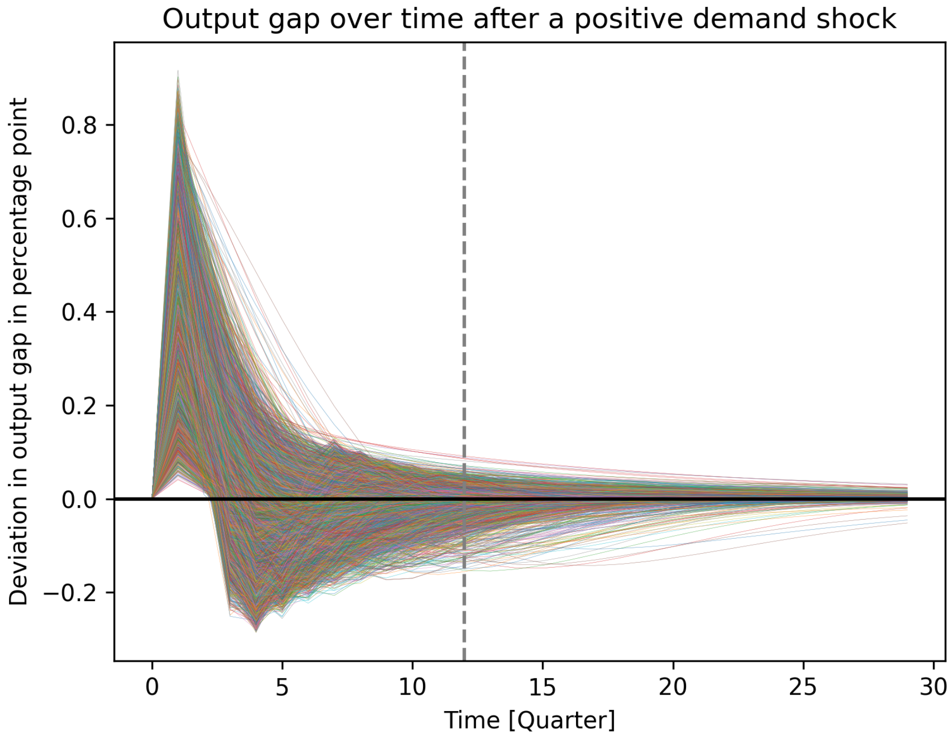


Figure 4.2: Output gap time series for a demand shock for 15000 experiments with a vertical line at time t = 12.

the steady state. The experiments also differ in the speed of the behavior: some experiments show a small lag in their behavior.

Figure 4.3 shows 15000 time series of inflation after a demand shock with slightly changing parameters. Compared to the standard behavior, we see similar differences as can be seen in the output gap time series. The peak can go up to 6 percentage point per year, compared to almost 2 for the normal behavior. Whereas inflation is near 0 percentage point at time t = 12 in our standard calibration, it has a value of around -0.5 and 2 percentage point in our experiments. At around t = 30, inflation is up to 1 percentage point per year in some of our experiments, whereas this is 0 for the standard behavior In the standard behavior, inflation oscillates from the peak positive deviation of around 2 percentage point per year from the steady state to a negative deviation of -0.5 percentage point from the steady state, back to a positive deviation of around 0.1 percentage point before going back to the steady state. Like the output gap time series, we observe differences in oscillation and in the speed with which the behavior occurs. Oscillation like the standard behavior are observed with different amplitudes (size of the deviation from the steady state). Experiments with a delayed peak and then moving back to the steady state without any negative deviation can be seen. Lastly, we see time series that show one oscillation (one peak and one dip) and then move back to the steady state while the deviation stays negative.

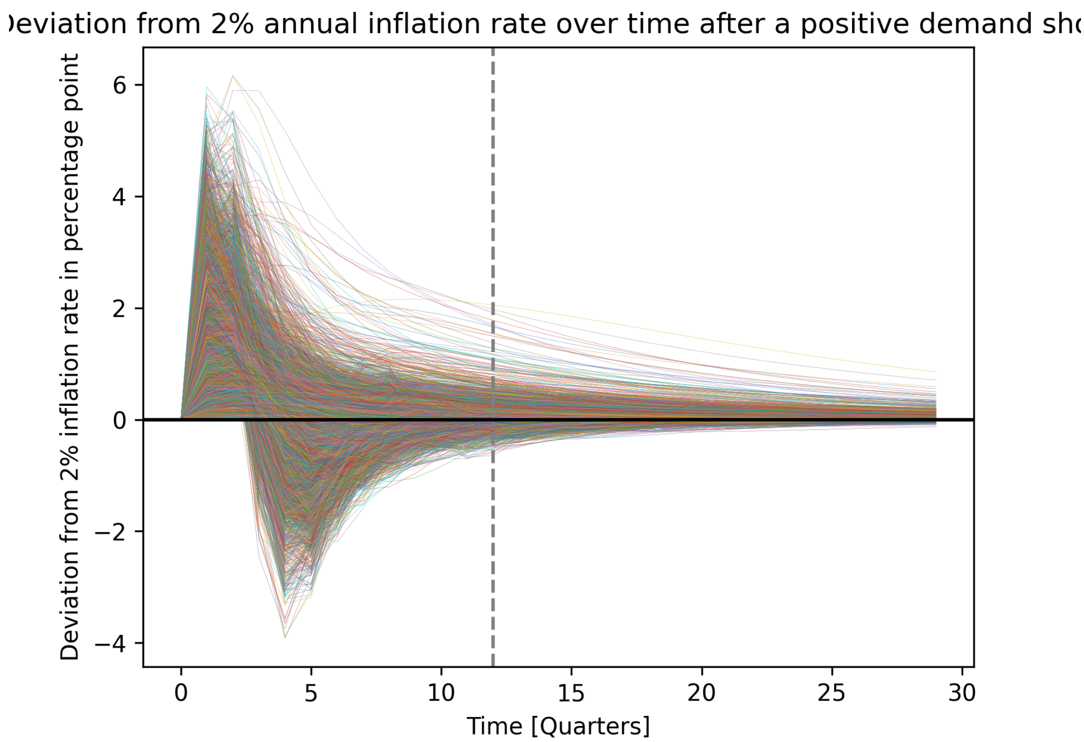


Figure 4.3: Inflation time series for a demand shock for 15000 experiments with vertical line at time t = 12.

Because we see that the time series have different behaviors, behavior based scenario discovery using time series clustering should give us more insights in what determines what type of behavior.

4.1.1. Behavior based scenario discovery using time series clustering

The first step in our behavior based scenario discovery using time series clustering, is creating the clusters. To do this, we have to decide the number of clusters we want to create. This is done by comparing the Silhouette Index for different numbers of clusters for five different sets of 2000 experiments. The results of this analysis, which can be found in Appendix F, did not provide us with a single number of clusters that is optimal. Rather, a pattern is observed which indicates an improvement in the quality of the clustering whenever the addition of a cluster causes the biggest cluster to split. Therefore, we will select the number of clusters based on the change in the size of the biggest cluster. Consequently, we might end up with multiple numbers of clusters of interest. As can be seen in Table 4.1, the number of clusters in which we are interested is 2 (2 groups of interest), 6 (3 clusters of interest), 7 (4 clusters of interest), and 9 (6 clusters of interest). The clusters of interest are those with more than 500 experiments in order to have significant results in our further analyses.

Output gap clustering

Table 4.1: Number of experiments per cluster for output gap time series clustering for the experiment run with 15000 experiments

Number of clusters	Cluster number												Total
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	11911	3089											15000
3	11704	3089	207										15000
4	11704	3080	207	9									15000
5	11704	3080	196	11	9								15000
6	6480	5224	3080	196	11	9							15000
7	5224	3886	3080	2594	196	11	9						15000
8	5224	3886	2594	2227	853	196	11	9					15000
9	3931	3886	2594	2227	1296	853	196	11	9				15000
10	3931	3464	2594	2227	1293	853	422	196	11	9			15000
11	3931	3464	2594	2227	853	802	491	422	196	11	9		15000
12	3931	3464	2594	2227	853	802	491	422	149	47	11	9	15000

Based on a visual inspection of the behavioral characteristics of the clusters of interest for the selected numbers of clusters, we continue with the results of the 9 clusters (6 clusters of interest). As can be seen in Appendix F, fewer clusters of interest resulted in less distinguishable behavioral characteristics per cluster.

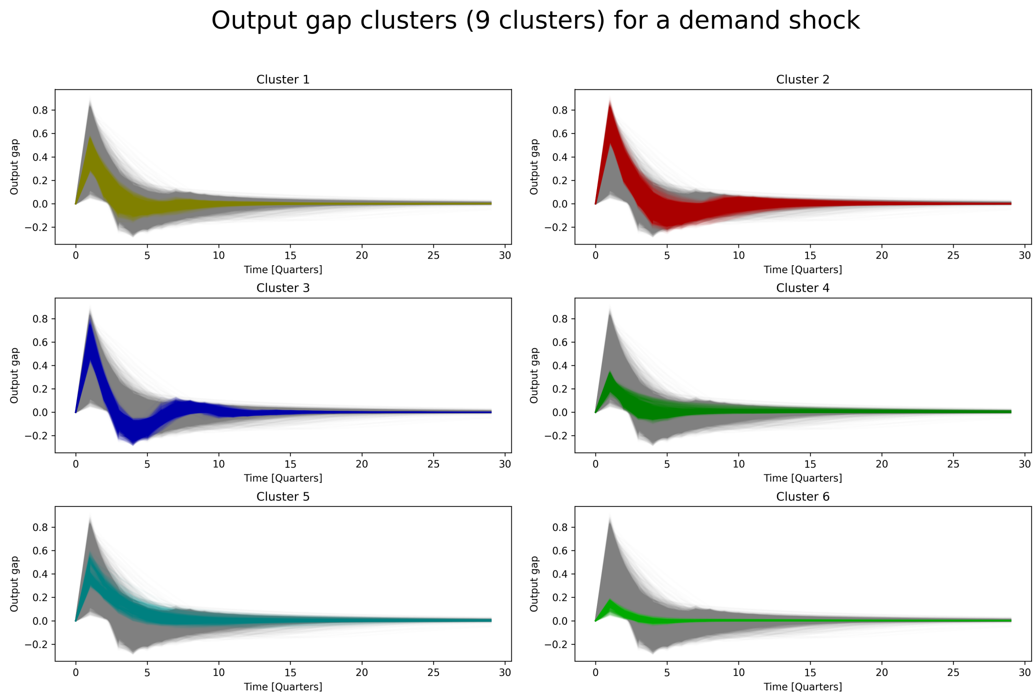


Figure 4.4: Visualization of the 6 clusters of interest when the output gap time series is clustered into 9 clusters.

The six clusters of interest when we cluster into nine clusters are shown in Figure 4.4. In this clustering approach, we see clear differences in behavior, which are summarized in Table 4.2. The experiments in cluster 3 are the experiments that have the clearest oscillation: They have the highest peak, immediately move to a dip and oscillate while going back to the steady state. Experiments in cluster 1 have a medium-sized deviation from the steady state and oscillates down to a negative deviation before moving back to the steady state. Experiments in cluster 4 show similar behavior, but have a smaller deviation. Cluster 2 includes the experiments that have the highest deviation from the peak, take a few time steps before reaching its dip, and move back to the steady state with or without indicating a positive deviation. Compared to the other experiments, the oscillating behavior occurs in later time steps, which we will refer to as 'lag'. Experiments in cluster 5 have a similar lag, a medium peak, and then move to the steady state without showing any oscillation and stay positive (almost) for the whole time series. The last cluster, cluster 6, shows very small deviations from the steady state and stay below 0.2 percentage point deviation from the steady state.

Table 4.2: Behavioral characteristics (magnitude of deviations, oscillation, lag) of the 6 clusters of interest based on the output gap time series after a demand shock.

Cluster	Magnitude of deviations	Oscillation	Lag
1	Medium	Small	No
2	Large	Medium	Yes
3	Large	Largest	No
4	Small - Medium	Mixed	No
5	Medium	None	Yes
6	Small	Small to none	No

In these different clusters, we apply the patient rule induction method (PRIM) to check for comparable parameter ranges that resulted in similar behavioral characteristics. Appendix F includes an in-depth description of how this is done per cluster. The final results of PRIM are shown in Table 4.3. Note that not all clusters are explained by the same parameters. Each hyperrectangle is uniquely formed with parameters that best describe this input space, and we have not forced a regression with these specific uncertainties to describe the input space that caused comparable behaviors.

Table 4.3: Output gap final for each cluster

Parameter	Range	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Smoothing parameter	[0.5 - 1]	[0.50 - 0.73***]	[0.63*** - 1]	[0.81*** - 1]	[0.5 - 0.75***]	[0.5 - 0.89***]	[0.5 - 0.75***]
Calvo parameter	[0.6 - 0.95]	[0.74***-0.87***]	[0.79*** - 0.95]	[0.60 - 0.80***]	[0.60 - 0.74***]	[0.85*** - 0.95]	[0.6 - 0.81***]
Autoregressive factor of demand shock	[0.5 - 0.95]	[0.58*** - 0.77***]	[0.50 - 0.73***]	[0.5 - 0.80***]	[0.65*** - 0.83***]	[0.77*** - 0.90***]	[0.84*** - 0.95]
Policy reaction to interest	[1 - 2.5]	[1.3*** - 2.5]			[1.6*** - 2.5]		[1.5*** - 2.5]
Coverage		0.168	0.548	0.486	0.283	0.505	0.639
Density		0.828	0.815	0.850	0.731	0.667	0.721

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Combining Tables 4.2 and Tables 4.3, we get an idea of how combinations of parameter values influence the behavior of the output gap. Important to note is that we do not have a situation where the value of one parameter has changed, ceteris paribus, but that the behavioral characteristics of the clusters may be caused by the combination of values of the parameters.

The four important parameters are the smoothing parameter, the Calvo parameter, the autoregressive factor of the demand shock, and the policy reaction to interest. First, let us briefly recap the different parameters and their role in the DSGE model. The smoothing parameter, found in the Taylor rule, indicates how much the current interest rate deviates from the previous interest rate. A higher smoothing parameter indicates smaller changes compared to the previous period. The Calvo parameter refers to the share of firms that are unable to adjust their prices in the current time step. A larger Calvo parameter indicates a higher degree of price stickiness. The autoregressive factor of the demand shock represents the persistence of a demand shock. Lastly, the policy reaction to inflation indicates how much the current inflation rate influences the current interest rate.

The most extreme oscillations were found for experiments in cluster 3. Cluster 3 has a high smoothing rate and low persistence of the demand shock. Because of the low persistence, the shock dies relatively quickly, resulting in a narrow peak. A large share of firms can adjust their prices at each time step (low Calvo parameter), while the central bank is conservative with its response (large influence of the previous interest rate). This indicates that a very responsive economy and a conservative interest rate setting approach might result in extreme oscillations, and thus economic instability, in the economy after a short shock.

Between clusters 1 and 2, the main difference is the lag in cluster 2. Table 4.3 shows that they differ mainly in the ranges of the smoothing parameter. The smoothing parameter is higher for cluster

2, showing that the effect of a higher smoothing parameter, given the ranges for the other parameters, results in a larger peak and slightly more and lagged oscillating behavior. The other parameters indicate a low persistence of the demand shock and a medium to high price stickiness. The central bank is limited in the speed with which they change their interest rate and thus cannot react quickly to the demand shock, resulting in a larger effect of the demand shock on the output gap. Cluster 4 shows parameter ranges similar to those of cluster 1, except for the Calvo parameter, which is lower for cluster 4. The range for the autoregressive factor of the demand shock overlaps mostly, but is somewhat higher for cluster 4. Cluster 4 experiments showed similar behavior to cluster 1 experiments, but with smaller deviations. This indicates that, together with a responsive approach of the central bank, more firms being able to adjust their prices might result in a lower peak. Firms can adjust their prices more easily, and then a high peak is prevented. A higher persistence of the shock might help decrease the oscillations of the economy.

Cluster 5 has a somewhat high persistence meaning and a high share of firms that are unable to change their prices at every time step. The output gap shows hardly any oscillations. As the persistence is low, the exogenous shock slowly dies out, meaning that the economy has time to adjust every time step. The economy is not overcompensating as the degree of price stickiness is high, and the central bank is not extremely conservative, but not necessarily very responsive in its behavior. This combination has prevented large oscillations and caused the output gap to slowly move towards the steady state.

In the experiments in cluster 6, the economy and central bank can respond quickly to the shock, as the Calvo parameter and the smoothing parameter are both low. The firms can adjust their prices quickly, and the central bank can adjust their interest rate without having to take the previously set interest rate into consideration. In addition to this, the shock has high persistence, meaning that the shock is slowly decreasing. We see that the responsive character of the economy and the slow change in external forces on the economy cause the output gap to stick close to the steady state.

With these observations in mind, we have the following indications of the causes of the specific behavior of the output gap due to a demand shock. A high persistence of the demand shock gives the economy a larger buffer to adjust at every time step, resulting in a straightforward path towards the steady state instead of showing oscillations. An economy with a large share of firms being able to adjust their prices and a responsive central bank is able to minimize the economic consequences of the shock. However, a large share of firms that are able to adjust when the central bank is acting conservatively may cause the large economic instabilities if a short demand shock occurs.

Inflation clustering

Table 4.4 describes the number of experiments per cluster when clustering the 15000 experiments into 2 - 12 clusters. A visualization of the clusters when the experiments are divided into two is found in Appendix F. The 3 clusters of interest when we cluster the experiments into 5 show more distinguishable behavior, which is why we will continue with these clusters.

Table 4.4: Number of experiments per cluster for inflation time series clustering for the experiment run with 15000 experiments

Number of clusters	**Cluster number**												**Total**
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	11366	3634											15000
3	11364	3634	2										15000
4	11364	3618	16	2									15000
5	7527	3837	3618	16	2								15000
6	7527	3837	3241	377	16	2							15000
7	7527	3837	3241	377	9	7	2						15000
8	7527	3837	3135	377	106	9	7	2					15000
9	7527	3837	3135	377	106	7	7	2	2				15000
10	7275	3837	3135	377	252	106	7	7	2	2			15000
11	7275	3837	3135	377	252	106	85	7	7	2	2		15000
12	7275	3837	3135	282	252	106	85	10	7	7	2	2	15000

The three clusters of interest when clustering into five clusters are shown in Figure 4.5. We see that cluster 1 has a medium deviation from the steady state and includes behavior that both oscillates and moves to the steady state after the peak without showing negative deviations from the steady state. Cluster 2 shows hardly any deviation from the steady state and remains relatively flat. Cluster 3 includes the experiments that show the most extreme behavior with the highest peaks and dips and most oscillations before moving to the steady state.

Table 4.5: Behavioral characteristics (magnitude of deviations, oscillation) of the 3 clusters of interest from the inflation time series after a demand shock.

Cluster	**Magnitude of deviations**	**Oscillation**
1	Medium	Mixed
2	Small	None
3	Large	Yes

Table 4.6 shows the results of PRIM, and presents the boxes that are considered to describe these experiments the best.

Table 4.6: Boxes describing the experiments in the 3 clusters of interest for the inflation time series clustering.

Parameter	**Range**	**Cluster 1**	**Cluster 2**	**Cluster 3**
Calvo parameter	[0.6 - 0.95]	[0.74*** - 0.83***]	[0.90*** - 0.95]	[0.60 - 0.71***]
Smoothing parameter	[0.5 - 1]	[0.50 - 0.86***]	[0.50 - 0.95***]	[0.78*** - 1.0]
Autoregressive factor of demand shock	[0.5 - 0.95]	[0.5 - 0.80***]		[0.50 - 0.89***]
Policy reaction to interest	[1 - 2.5]	[1.3*** - 2.5]		[1 - 2.0***]
Inverse marginal elasticity of substitution	[1 - 2.5]	[1.2*** - 2.5]		[1.5*** - 2.5]
Coverage		0.177	0.517	0.230
Density		0.948	0.982	0.994

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

We see the same parameters that are used to describe the output gap time series clustering. The experiments in cluster 2 have a high Calvo parameter, indicating that the inflation rate is small if the

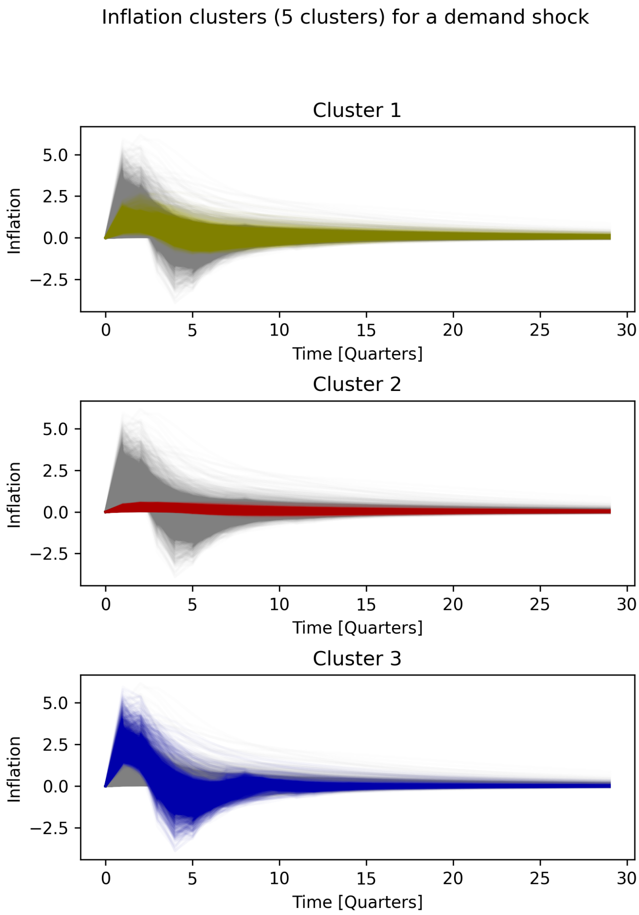


Figure 4.5: Visualization of the 3 clusters of interest when the inflation time series of a demand shock is clustered into 5 clusters.

economy is unable to adjust their prices. An economy experiences the largest deviations and most extreme oscillating behavior (cluster 3) if the economy is able to adjust its prices easily, while the central bank is relatively conservative with changing the interest rate from period to period.

Cross examination of output gap clusters and inflation clusters

In Figure 4.6 we see the behavior of the output gap (left) and inflation (right) time series for the inflation clusters. We see that inflation clusters 1 and 2 have overlapping output gap time series. The extreme behavior of inflation cluster 3 can also be seen in the output gap time series.

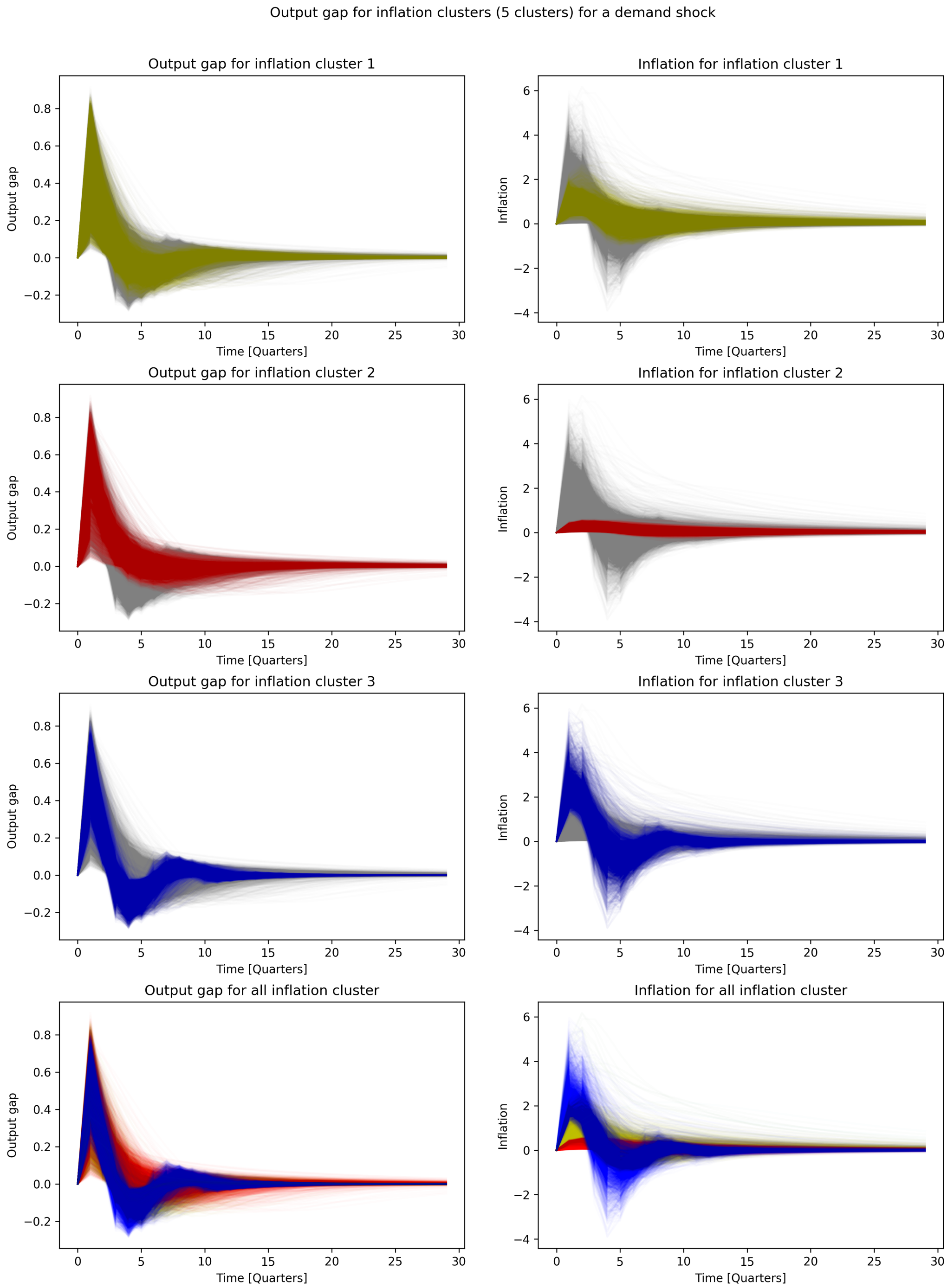


Figure 4.6: Output gap time series and inflation time series when the inflation clusters are applied for a positive demand shock.

In Figure 4.7, we see the behavior of the inflation time series for the output gap clusters. Clusters 1 and 2 for the output gap show similar behaviors in the inflation time series. The experiments in output gap cluster 3, characterized by having the most extreme behavior, also cause the most extreme oscillation and deviation in inflation. Clusters 4 and 6, experiments resulting in small deviations and hardly any oscillation show similar behavior in the inflation time series. Striking is the behavior of output gap cluster 5, which shows a medium peak in output gap but stays relatively close to the steady state for inflation, both without oscillating. The experiments in cluster 5 have a relatively high price stickiness, which, as we saw for inflation cluster 2, might cause inflation to stay close to the steady state.

The description of cluster 3 for the output gap clustering and cluster 3 for the inflation clustering is similar in the fact that the firms cannot adjust their prices easily, but the central bank can. Especially because the persistence of the demand shock is not very large, the economy might overcompensate (prices are quickly adjusted even though the shock is over), while the central bank, due to the restriction of the high smoothing parameter, is limited in adjusting their interest rate.

The effect of price stickiness is also seen when we compare inflation and the output gap for output gap clusters 5 and 6. These two represent a situation where the central bank is relatively free to move away from the previous interest rate and a relatively high persistence of the demand shock. However, a higher price stickiness (output gap cluster 5) minimizes inflation and causes the actual economy to perform above the potential economy (larger output gaps). A lower price stickiness (output gap cluster 6) results in firms adjusting their prices easily (higher inflation) and, by that, making sure the actual economy stays close to the potential economy a lot.

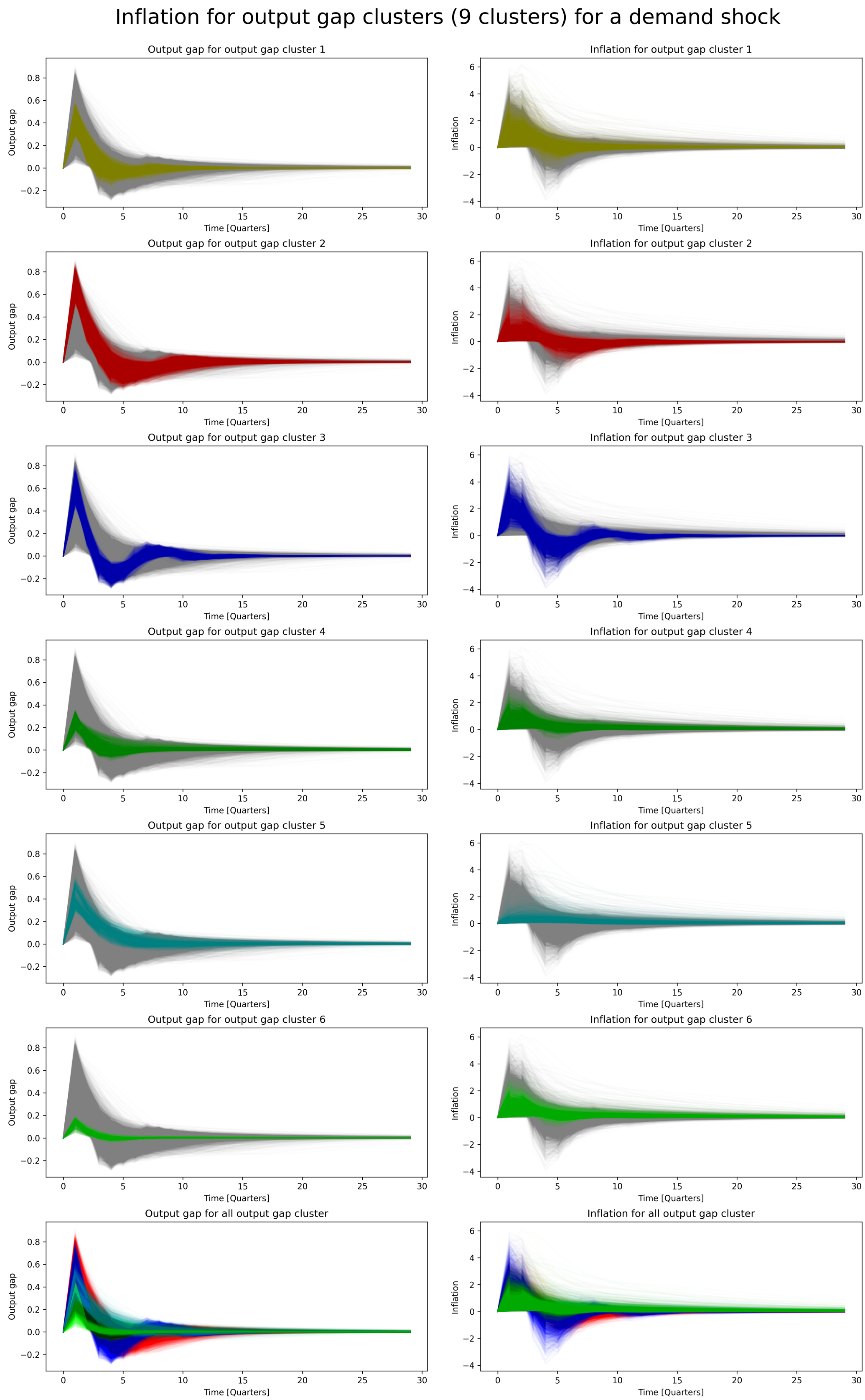
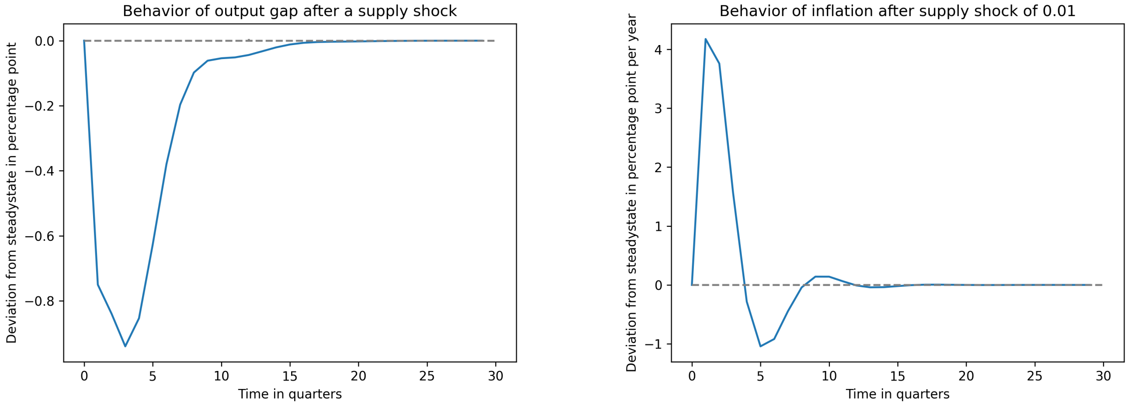


Figure 4.7: Output gap time series and inflation time series when the inflation clusters are applied for a demand shock.

4.2. Supply shock

The standard behavior of the output gap and inflation after a supply shock are presented in Figure 4.8. The output gap turns negative initially and moves from there to the steady state. Inflation shows an initial positive movement, followed by an oscillation down moving back to the steady state.



(a) The output gap deviates in a negative way from the steady state and slowly moves back its steady state.

(b) The inflation deviates in a positive way from the steady state and oscillates before moving back to its steady state.

Figure 4.8: Behavior of the output gap and inflation after a supply shock. They have opposite responses to a supply shock.

The output gap time series for our 15000 experiments are shown in Figure 4.9. We see that there are a lot of changes compared to the standard behavior of the supply shock. Whereas most experiments end up approaching the steady state from below (a negative deviation), the initial reaction to the shock differs a lot. The behavior we would expect is a negative shock and then moving to the steady state. Many of the experiments shown in Figure 4.9, however, show an initial positive deviation.

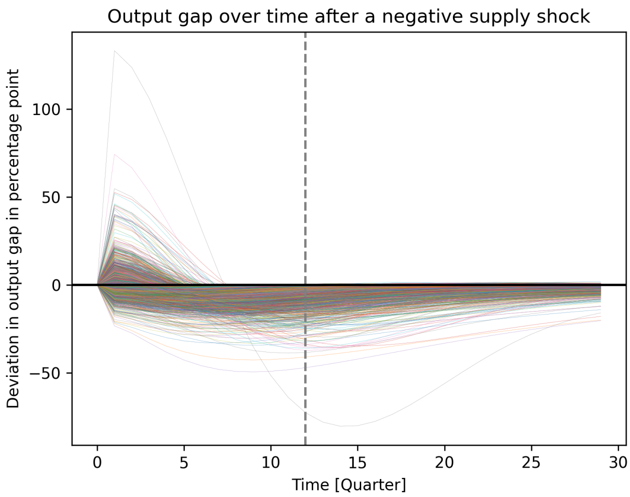


Figure 4.9: Output gap time series of 15000 experiments after a supply shock.

Figure 4.10 show the behavior of the inflation time series for our experiments. Like the output gap, the inflation time series show different behavior than the expected behavior. Many experiments hardly show any negative deviation and the size of the positive deviation is more than 40× larger.

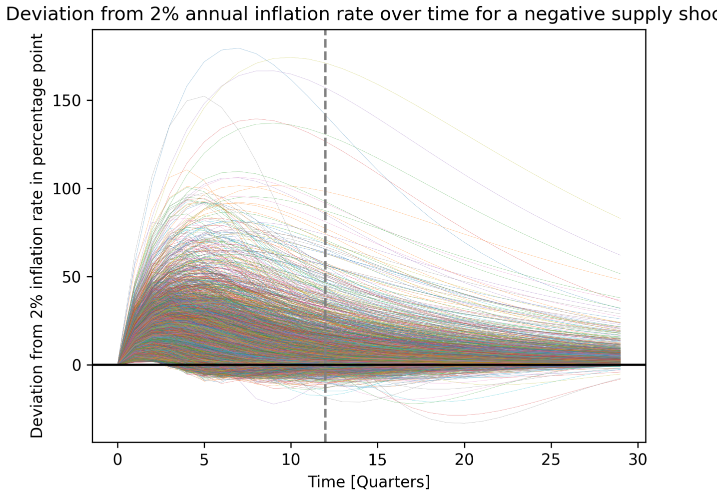


Figure 4.10: Output gap time series of 15000 experiments after a supply shock.

4.2.1. Behavioral based scenario discovery using time series clustering

When we cluster with different numbers of clusters, we can see that we stick with one big cluster, as shown in Table 4.7 for the output gap clustering and in Table 4.8 for the inflation clustering. When extra clusters are added, only the smaller clusters are divided into multiple clusters. In Appendix F, a visualisation is included of the 2 clusters of the output gap and inflation when we cluster into 2 clusters. Additionally, Appendix F, includes analyses done on an adjusted set of experiments in which the experiments causing the most extreme behavior are dropped. However, this has not improved the results significantly.

Table 4.7: Number of experiments per cluster for output gap time series clustering for the experiment run with 15000 experiments

Number of clusters	**Cluster number**												**Total**
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	14999	1											15000
3	14542	457	1										15000
4	14542	428	29	1									15000
5	14542	428	27	2	1								15000
6	14542	428	26	2	1	1							15000
7	14542	392	36	26	2	1	1						15000
8	14542	233	159	36	26	2	1	1					15000
9	14542	233	159	34	26	2	2	1	1				15000
10	14542	233	159	34	25	2	2	1	1	1			15000
11	11668	2874	233	159	34	25	2	2	1	1	1		15000
12	11668	2874	233	159	34	17	8	2	2	1	1	1	15000

Table 4.8: Number of experiments per cluster for inflation time series clustering for the experiment run with 15000 experiments

Number of clusters	Cluster number												Total
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	14889	111											15000
3	14889	106	5										15000
4	14889	105	5	1									15000
5	14889	83	22	5	1								15000
6	14889	83	22	3	2	1							15000
7	14889	76	22	7	3	2	1						15000
8	12443	2446	76	22	7	3	2	1					15000
9	12443	2446	76	22	7	2	2	1	1				15000
10	12443	2446	76	21	7	2	2	1	1	1			15000
11	12443	2446	76	13	8	7	2	2	1	1	1		15000
12	12443	2446	49	27	13	8	7	2	2	1	1	1	15000

Output gap clustering

Figure 4.11 shows the 2 clusters of interest when output gap time series are clustered into 11 clusters. We see that cluster 1, the largest cluster, has relatively small deviations from the steady state. Cluster 2 overlaps these small deviations, but also include larger deviations. This means that the clusters only provide us insights with which ranges of parameters the deviations are more likely to be within a small range from the steady state. Experiments in cluster 2 do not necessarily behave differently from the experiments in cluster 1.

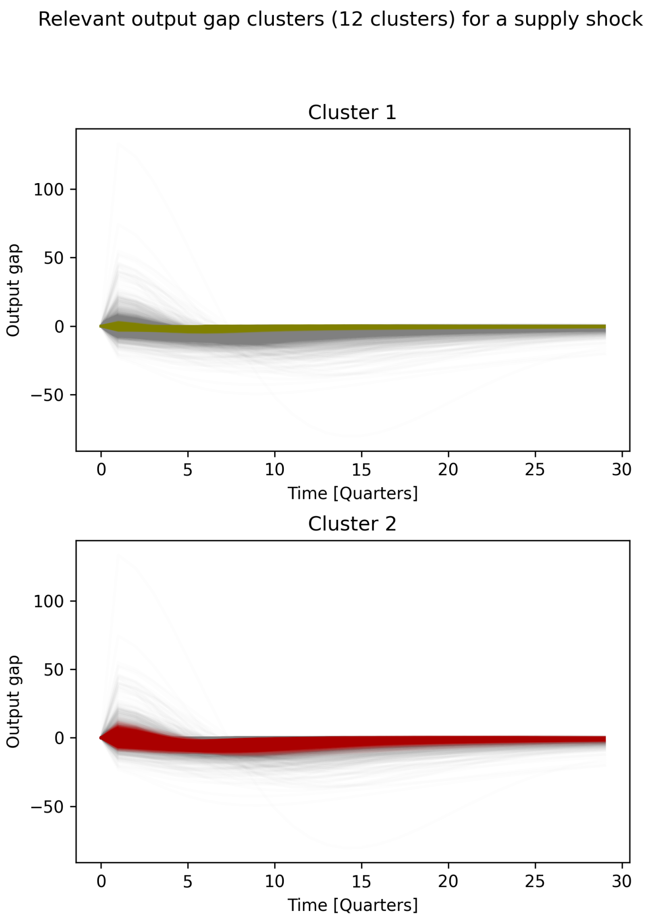


Figure 4.11: The 2 clusters of interest when output gap time series are clustered into 11 cluster.

If we take the two clusters of interest when the output gap time series are clustered into 11 clusters, we get the descriptions of the important boxes as stated in Table 4.9. With the exception of the Calvo parameter are the parameter ranges only exclude the extreme high values of the range we have given to the uncertainties. This limits the conclusions we can draw from these boxes.

What we can learn is that the deviations in output gap are likely to be limited if firms are able to adjust their prices quickly (low Calvo parameter).

Table 4.9: Output gap final for each cluster

Parameter	**Range**	**Cluster 1**	**Cluster 2**
Calvo parameter	[0.6 - 0.95]	[0.60 - 0.77***]	[0.84*** - 0.95]
Smoothing parameter	[0.5 - 1]	[0.50 - 0.95***]	[0.68*** - 0.93]
Autoregressive factor of supply shock	[0.5 - 0.95]	[0.5 - 0.89***]	
inverse Frisch elasticity of labor supply	[0.6 - 1.5]	[0.65 - 1.5*]	
Coverage		0.454	0.627
Density		1.00	0.655

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Inflation clustering

When we group the inflation time series into eight clusters, we end up with two clusters of interest, as shown in Figure 4.12. Contrary to the output gap time series, here we can distinguish the behavior of the two clusters more distinctly. The experiments in cluster 2 result in a more positive deviation from the steady state than in cluster 1 and take more time to move back to the steady state.

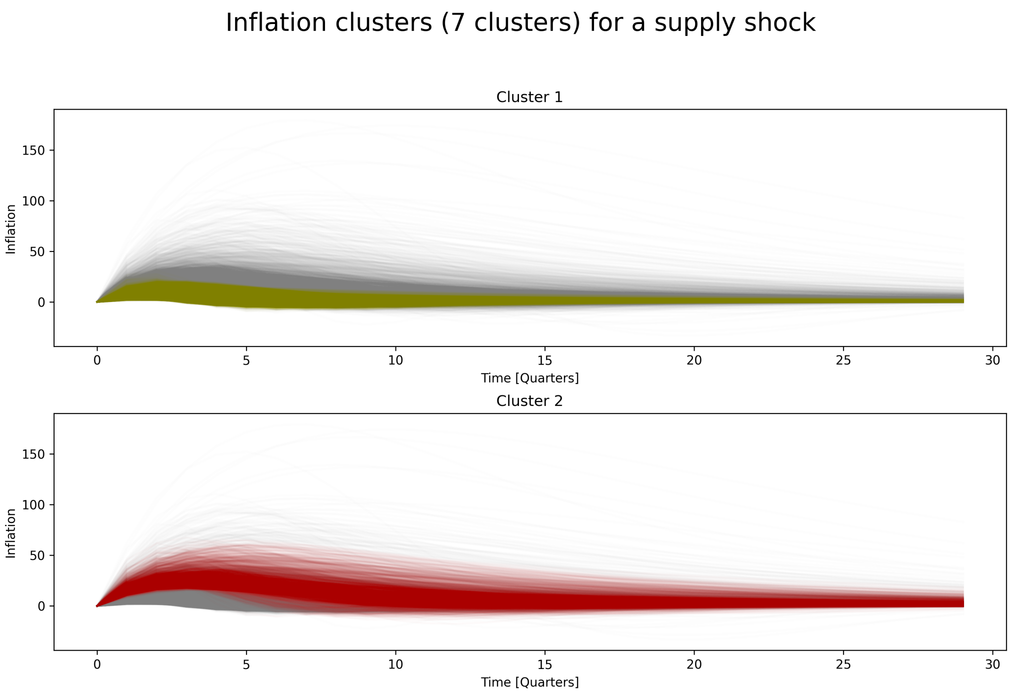


Figure 4.12: The 2 clusters of interest when inflation time series are clustered into 8 cluster.

The boxes that best describe the two inflation clusters are shown in Table 4.10. Like the output gap analysis showed, several of the parameter ranges are largely similar to the ranges we have given to the uncertainties, limiting the lessons we can draw from those ranges. The main lessons can be learned from the Calvo parameter and the persistence of the supply shock.

In cluster 1, a larger share of firms are able to adjust their prices at every time step than for the experiments in cluster 2. This would indicate that an economy that is able to adjust its prices easily could prevent large economic consequences due to a demand shock. However, it is important to note that a shock is defined as a forced deviation from the steady state in the inflation rate. Cluster 1 has a smaller persistence, meaning that the forced deviation over time is taken away faster than for those in cluster 2. Due to the lack of clusters it is not possible to qualify the exact contribution of the Calvo parameter and the persistence of the supply shock to the difference in deviation from the steady state between cluster 1 and cluster 2.

Table 4.10: Inflation final for each cluster

Parameter	Range	Cluster 1	Cluster 2
Calvo parameter	[0.6 - 0.95]	[0.6 - 0.81***]	[0.88*** - 0.95]
Autoregressive factor of supply shock	[0.5 - 0.95]	[0.5 - 0.83***]	[0.78*** - 0.93*]
Policy reaction to interest	[1 - 2.5]	[1.2*** - 2.5]	
Smoothing parameter	[0.5 - 1]		[0.5 - 0.98]
Discount factor	[0.95 - 1]		[0.95 - 0.99***]
Coverage		0.454	0.299
Density		1	0.953

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Cross examination of output gap clusters and inflation clusters

For both inflation and output gap time series clustering, we find that there are two main clusters. The PRIM results in Tables 4.9 and 4.10 show that there are different parameters of importance in creating the two different clusters. Figure 4.13 presents the time series of the output gap (left) and the deviation in inflation (right) for the different output gap clusters. Figure 4.14 shows the time series of the two economic variables for the two inflation clusters. The two figures in the bottom for both figures only show the time series in the clusters. This causes the scales of the y-axis to deviate from the other two figures.

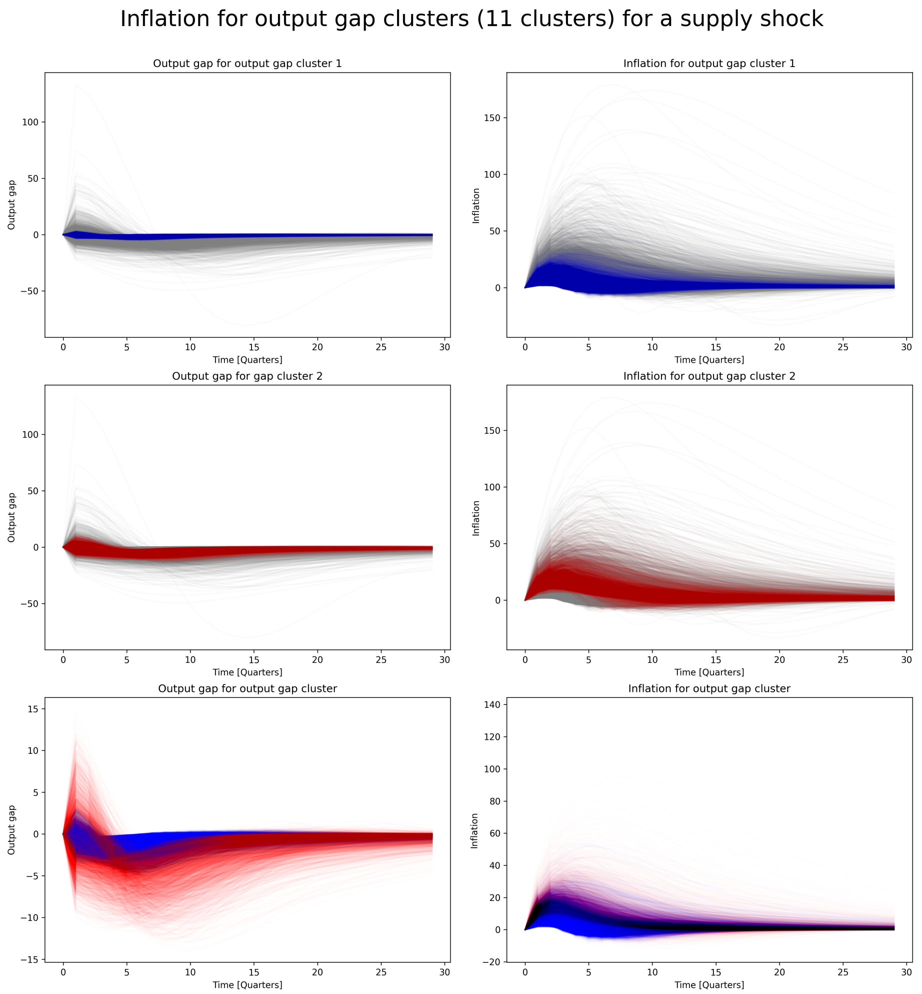


Figure 4.13: Output gap time series and inflation time series when the output gap clusters are applied for a negative supply shock. Note that the last figure does not include time series that are not included in either cluster.

We see that the two clusters are largely similar, but the inflation time series with a larger magnitude from inflation cluster 2 are not included in either of the two output gap clusters. This is clearly seen in the bottom right figure for both Figures 4.13 and 4.14. Here, for the inflation clusters, several time series with a deviation in inflation rate of up to 60 percentage points are included, whereas these are not included in the output gap clusters. The difference in scale does show that more extreme inflation time series are included in the output gap cluster, but these are not overly represented. Because the Calvo parameter ranges of both clusters are somewhat similar, it is not surprising to see comparable results between the two clusters.

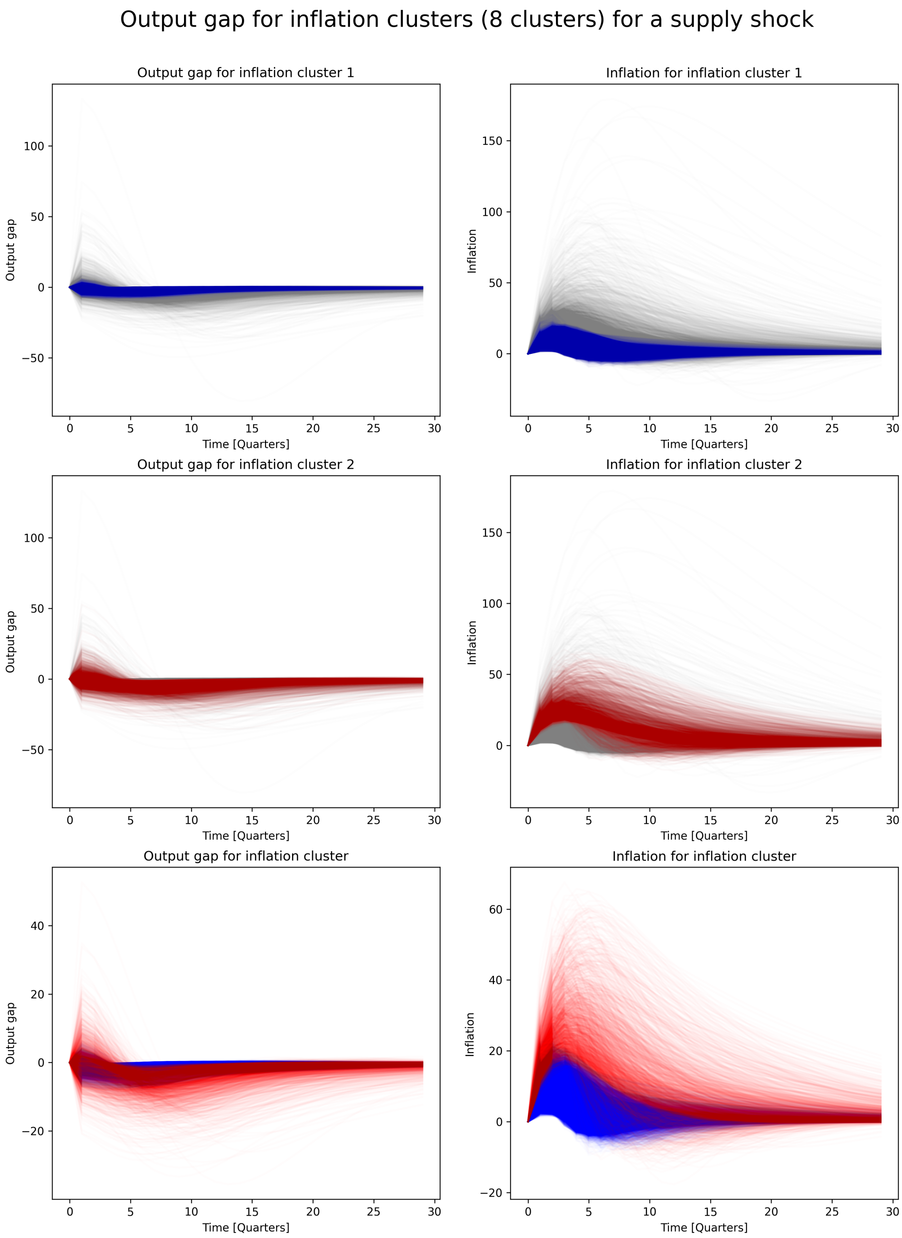


Figure 4.14: Output gap time series and inflation time series when the inflation clusters are applied for a negative supply shock. Note that the last figure does not include time series that are not included in either cluster.

5
Results - Directed Search

In this chapter, we discuss the results of our directed search. We have searched for Pareto optimal policies after a supply shock. This search is only conducted in the baseline scenario. First, we will visualize all policies and then compare the robustness of a set of chosen policies. The objectives, as described in Section 2.3.1, are the following:

- Minimizing inflation at time t = 12 (pi_12)
- Minimizing maximum inflation rates (pi_max)
- Minimizing quarters in which deflation occurs (deflation_quarters)
- Maximizing the quarters in which inflation is between 2% and 4% (quarters_inside_inflation_domain)
- Minimizing the period-to-period change in inflation after t=2 (largest_change_inflation)
- Minimizing maximum output gap (y_max)
- Minimizing the standard deviation of output gap over time (y_std)

For objectives with a continuous range, a regret metric is used. For objectives with a discrete range or when a threshold is included, we use a satisficing metric. Although the descriptive metric is presented, we should be aware that the distributions of the uncertainties are not validated. Therefore, our descriptive metrics are only valid if the uncertainties follow the uniform distribution we have assigned to them.

5.1. Pareto optimal policies

The performance of the different policies is visualized in Figure 5.1. Each objective has a separate vertical line that represents the (adjusted) range of all outcomes of all policies. This range is adjusted if we are certain of a minimum or maximum value, such as quarters in which deflation occurs and quarters in which inflation is within a certain domain (both have a range of [0,12]). For an objective that is minimized (maximized), the lower (upper) boundary of this range is on the bottom, and the top value is the upper (lower) boundary of this region. The lines connecting these vertical lines are separate policies, crossing the vertical objective lines at the value of the outcome this policy for this objective.

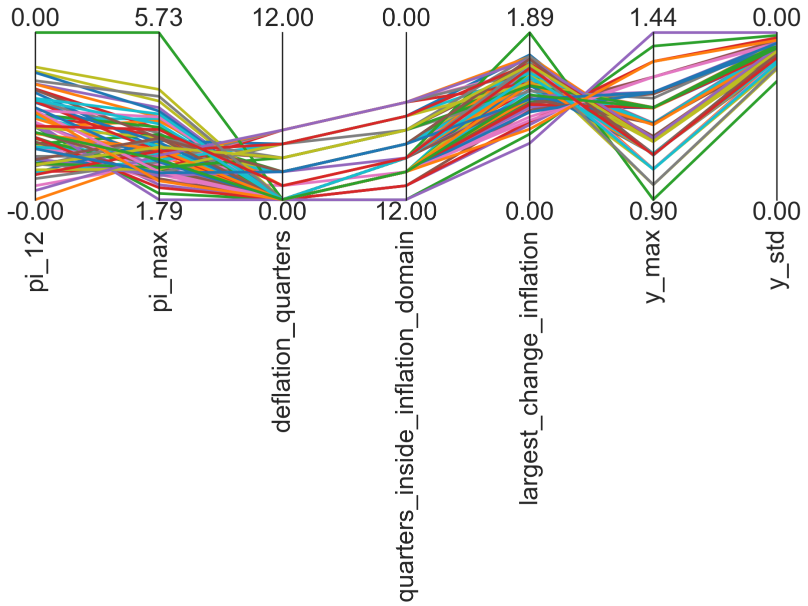
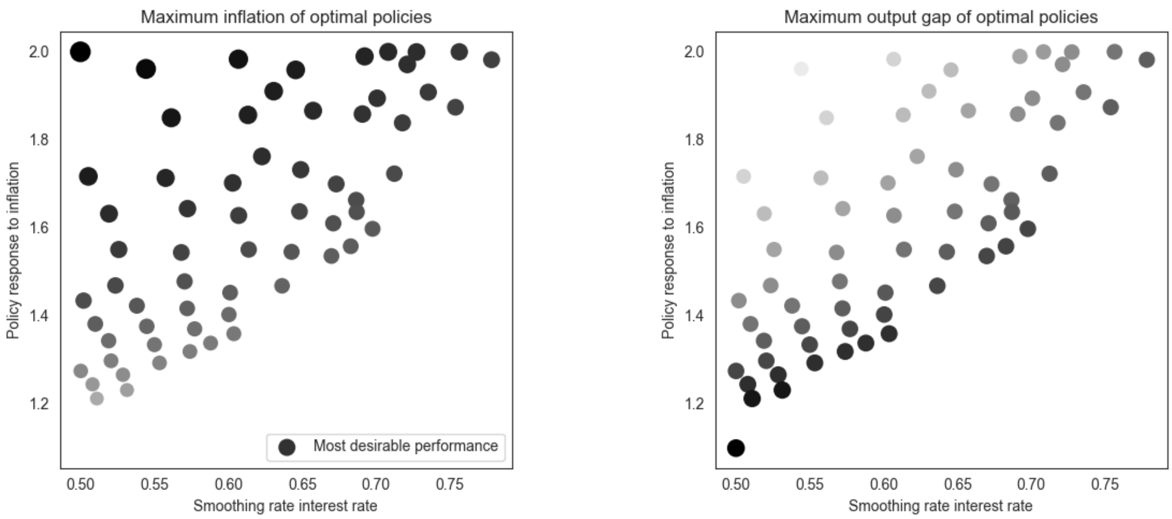


Figure 5.1: Parallel axis plot showing the performance of the optimal policies.

From this figure, we see that there is a trade-off between optimizing the inflation-related objectives, or optimizing the output gap-related objectives. This is characterized by the cross between 'largest change inflation' and 'y max'.

Having only two policy levers allows us to plot the performance of an objective against the different combinations of policy levers. This is done in Figure 5.2, which plots the maximum inflation (left) and the maximum output gap (right) for different optimal policies. The larger and darker dots are more desirable performances than the smaller and lighter dots. The trade-off is clearly shown, as the preferred policies for one objective are the least preferred policies for the other objective. In addition to this, we see that the optimal policies have formed a triangle; there are no Pareto optimal policies that have a relatively low policy response to inflation and a high smoothing rate of inflation.



(a) Maximum inflation for optimal policies

(b) Maximum deviation output gap for optimal policies

Figure 5.2: Scatter plots with objective on z-axis. We prefer a smaller value for both objectives. The value of the objective is indicated by the size of the circle (bigger circle is preferred option) and the color (darker color is preferred). We clearly see a trade-off as the colors and the sizes in the two figures are almost inverted.

Based on their performance in the objectives, we end up with 4 policies, besides the base policy

(Policy 0), with the following values:

Table 5.1: Chosen policies with the values for their levers

Policy	Description	Policy response to inflation	Smoothing rate interest rate
0	Baseline	1.5	0.8
1	All quarters inside inflation domain	1.961096	0.544366
2	Lowest inflation at time t = 12	1.981912	0.778116
3	Best inflation	2.000000	0.500000
4	Best output gap	1.098116	0.500231

We see that Policies 1 and 3 have values that are near each other. Policy 1, 2, and 3 are optimized for inflation objectives, while Policy 4 is optimized for the output gap. Whereas 1, 2, and 3, have a high response to inflation, this is much smaller for Policy 4. Both Policies 3 and 4 let the current interest rate be minimally influenced by the previous interest rate.

Figure 5.3 shows how these policies behave in (the same) 15000 scenarios. This figure shows the areas in which the outcomes for inflation (right) and output gap (left) are. The blue area shows the baseline policy, and the red area shows the area of that specific policy. The dashed lines represent the outer boundaries of 90% of the results. However, as stated before, we should be aware that the uncertainties distributions are not validated, reducing the value of this statistic 90%. Policies 1, 2, and 3 are optimized for one or more inflation objectives, whereas Policy 4 is optimized for output gap objectives. Consequently, this can be seen in the results in the Figure: whereas Policy 1, 2, and 3 have low inflation, Policy 4 has more inflation, but clearly has less deviation in output gap.

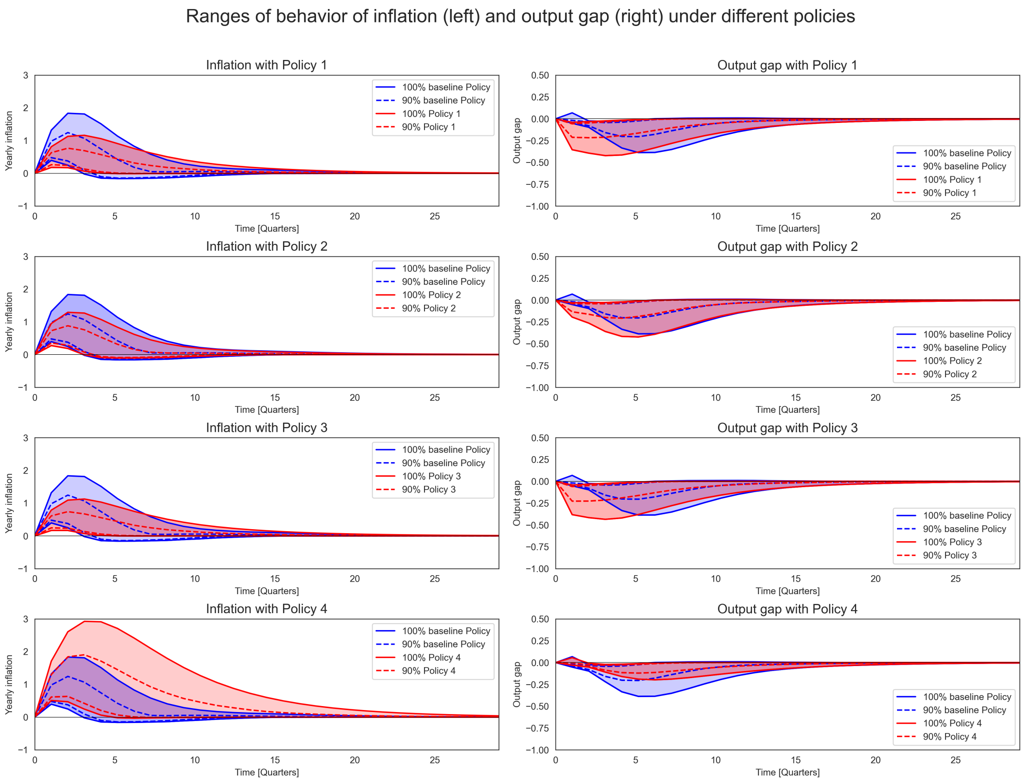


Figure 5.3: Ranges in which inflation and output gap fall for different policies under 15000 scenarios.

5.1.1. Robustness metrics

These selected policies are evaluated for robustness using robustness metrics.

Regret metrics

The scores of the policies for the regret metrics are shown in Table 5.2 and Table 5.3. The policy that responds strongly to inflation (Policy 3) has the lowest maximum inflation rate in at least 90& scenario. However, this policy scores the worst for the output gap objectives. The policy that optimizes the output gap objectives, Policy 4, has the least regret when it comes to output gap objectives and the most regret when it comes to inflation-related objectives. The baseline policy scores the best on minimizing the period-to-period change in inflation. Although Policy 2 results in the least regret for inflation at time t = 12, it performs poorly when it comes to minimizing maximum inflation and minimizing the period-to-period change in inflation. Compared to Policy 1, Policy 2 performs better on the output gap objectives but is outperformed on the maximum inflation and change in inflation.

Of the baseline policy, Policy 1 and Policy 2, none is clearly outperforming one of the others. If minimizing maximum inflation is considered the most important, Policy 1 scores the best, after Policy 3, followed by Policy 2. However, if the inflation rate at time t = 12 is the most important, Policy 2 performs best.

Table 5.2: Regret metrics for different policies part 1. Least regret per metric is shown in cursive.

Policy	Regret maximum inflation	Regret inflation at time t = 12	Adjusted regret inflation at time t = 12
Policy 0	0.468	0.012	0.012
Policy 1	0.029	0.022	0.022
Policy 2	0.152	*0.004*	*0.004*
Policy 3	*0.000*	0.024	0.024
Policy 4	1.057	0.157	0.157

Table 5.3: Regret metrics for different policies part 2. Least regret per metric is shown in cursive.

Policy	Regret change inflation	Regret maximum output	Regret standard deviation output gap
Policy 0	0.226	0.075	0.000
Policy 1	0.012	0.095	0.000
Policy 2	0.113	0.079	0.000
Policy 3	*0.000*	0.105	0.000
Policy 4	0.173	*0.000*	0.000

Satisficing metrics

In Table 5.4 we see the fractions of cases out of 15000 scenarios that satisfy certain thresholds. For quarters without deflation, the threshold is 0 quarters; for 12 quarters in the inflation domain, the threshold is 12 quarters; and for inflation at time t = 12 is the threshold whether inflation is between 2% and 4% annually. If the fraction in the column 'Quarters without deflation' and in the column '12 quarters within the inflation domain' is equal, it means that in the cases when it leaves the inflation domain, the inflation rate drops below 2%. Policy 3, the policy that responds strongly to inflation, scores best in all these cases. After Policy 3, Policy 1 performs best.

An important comment is that all these metrics are inflation-oriented. No thresholds are included for the output gap. Interestingly, Policy 4, which optimizes the output gap objectives, scores better on the 'inflation at time t = 12 within domain' metric than Policies 0, 2, and 3, which are optimized for inflation objectives.

Table 5.4: Fraction of cases out of 15000 scenarios that satisfy the satisficing metric. Cursive values indicate the highest percentages.

Policy	Quarters without deflation	12 quarters in inflation domain	Inflation at time t = 12 in domain
Policy 0	0.062	0.062	0.726
Policy 1	0.815	0.815	0.964
Policy 2	0.168	0.168	0.785
Policy 3	*0.930*	*0.930*	*0.990*
Policy 4	0.798	0.760	0.908

Descriptive metrics

For the descriptive metrics to be meaningful, it is necessary to validate the distributions of uncertainties. If the uncertainties follow a uniform distribution, we get the following descriptive metrics, as shown in Table 5.5 and Table 5.6. For the descriptive metrics, we consider the mean and standard deviation. For this metric, we see a bit more deviation in which policies are preferred (for all objectives, a low mean and a low standard deviation is preferred). If the central bank bases its decision largely on the inflation rate at time t = 12, Policy 2 would be the best policy. This policy not only has the lowest mean inflation at time t = 12, but also the smallest standard deviation. This would mean that implementing this policy is more likely to result in an inflation rate close to their expectations than under other policies.

Table 5.5: Descriptive statistics for output gap objectives. Preferred values are in cursive.

Policy	Inflation at time t = 12		Maximum inflation		Largest change inflation	
	Mean	Standard deviation	Mean	Standard deviation	pi_change_means	pi_change_std
Policy 0	0.010	0.017	0.759	0.240	0.302	0.025
Policy 1	0.016	0.026	0.471	0.162	0.126	0.028
Policy 2	*0.008*	*0.013*	0.562	0.171	0.216	*0.024*
Policy 3	0.017	0.027	*0.450*	*0.158*	*0.114*	0.028
Policy 4	0.065	0.108	1.143	0.397	0.249	0.022

Table 5.6: Descriptive statistics for output gap objectives. Preferred values are in cursive.

Policy	Maximum output gap		Deviation in output gap	
	Mean	Standard deviation	Mean [$\times 10^{-3}$]	Standard deviation [$\times 10^{-3}$]
Policy 0	0.105	0.052	0.373	0.176
Policy 1	0.120	0.053	0.424	0.190
Policy 2	0.106	0.054	0.389	0.174
Policy 3	0.127	0.055	0.436	0.197
Policy 4	*0.065*	*0.028*	*0.220*	*0.900*

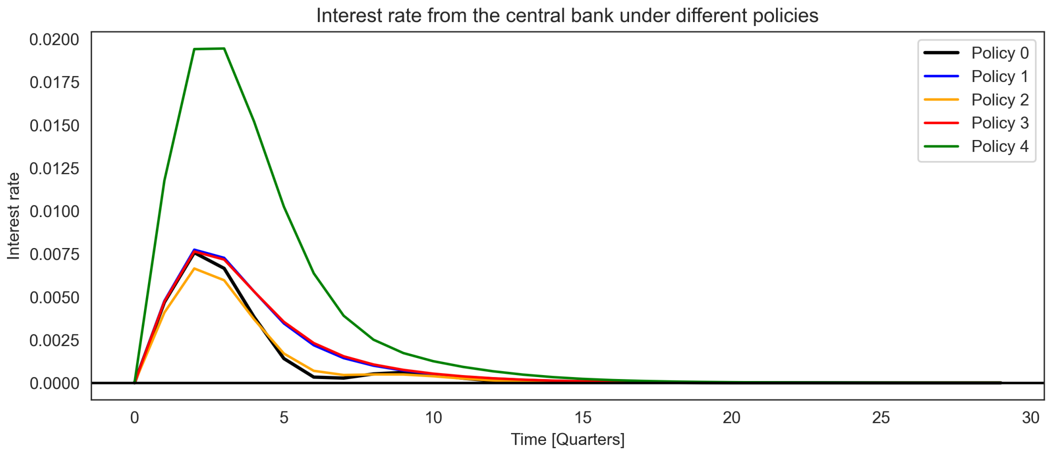


Figure 5.4: Interest rate set by the central bank under the baseline policy (Policy 0), policy optimizing inflation (Policy 3) and policy optimizing output gap (Policy 4)

5.2. Robust policy for inflation control

Based on our robustness metrics, no policy is clearly outperforming the other policies. Even though Policy 3 has the least regret for inflation-related objectives, it has the largest regret on output gap-related objectives. If the central bank focuses mainly on the inflation rate at time t = 12, policy 2 provides a low overall regret and has the lowest regret on inflation at time t = 12. However, this does result in a larger maximum inflation and period-to-period change in inflation than other policies. Important to note is that the inflation rate at time t = 12 is within an acceptable range for all policies. To maintain a stable inflation rate without hurting the output gap too much, Policy 1 is preferred. Policy 3 has the best performance on inflation-related satisficing metrics, followed by Policy 1.

If the central bank is solely interested in inflation control, they should respond heavily to the current inflation and minimize the influence of the previous interest rate.

In Figure 5.4, we see the interest rate time series under different policies for the baseline scenario. We see that Policy 1 and Policy 3 overlap a great deal. Only Policy 4, the output gap objectives optimizing policy, deviates significantly by setting a high interest rate.

In practice, the central bank does not take the interest rate as continuous rates but rather considers the interest rate as a set of discrete values. In this case, we can consider Policies 0 and 2, and Policies 1 and 3 as similar policies, and it is even likely that these four policies actually result in the same discrete values for the interest rate. If this is the case, we end up with two policies: inflation-focused and output gap-focused. As the central bank focuses on inflation control, it will focus on inflation.

6 Discussion

These results should be interpreted solely within the context of this thesis. As this research has been exploratory in nature, additional steps must be taken before a central bank or other institute can implement exploratory modeling and analysis (EMA) as part of their analyzes. In this chapter, we evaluate our research methods and our findings and make recommendations for future research. More specifically, we will discuss the following components of our research:

- the model (and the made assumptions) as a representation of (a simplification of) reality,
- the experiment set-up, including computational aspects of our research,
- limitations of the analysis methods we have used, both general limitations as well as limitations specified to our case, and
- the applicability of EMA on dynamic stochastic general equilibrium DSGE models.

Note that we only consider limitations that are relevant to the scope of the investigation. Even though we will touch upon critiques of the use of DSGE models, in this thesis, we accept the use of DSGE models and solely consider the limitations of applying EMA on these types of model.

6.1. The model and the assumptions as a representation of reality

Let us start by discussing the way we have treated this model in order to apply EMA. EMA requires us to think about features of the model that are not needed in the conventional methods that are applied to DSGE models. The most important features are (1.) identifying the ranges given to the uncertain parameters, (2.) the definition of the objectives, and (3.) defining the robustness metrics.

First, a major limitation of this research is how the target inflation rate of 2% annual inflation has been included. Several decisions in this research set-up have been made based on the assumption that the model has been calibrated around a zero inflation steady state, whereas the derivation of the equation has been a constant inflation rate steady state. The constant inflation rate that actually had to be assumed was 2%, which caused the inflation rate domain in the objectives to change to 2% - 4%. In addition, a domain of 2 percentage points is not in line with the central bank's interest, as they aim for a 2% rate. This also influenced how inflation is described in the text and in figures: as a zero inflation rate was assumed, the deviation from this inflation rate is equal to the inflation rate in that time step. However, because a 2% inflation rate should have been assumed, a deviation from this steady state is not equal to the inflation rate in that period. Although this significantly influences the interpretation of the subquestions, it does not necessarily influence the contribution that this research has. This research has tried to explore the use of EMA when assessing uncertainty in DSGE models and, even with this limitation in place, has succeeded in doing so.

The ranges of the uncertain parameters are based on the expertise of the central bank. We have set the distribution to a uniform distribution, based on the idea of conducting a wide exploration of the parametric uncertainty in DSGE models. For a more realistic implementation of EMA, other distributions could be used, such as the commonly used beta, gamma, or normal distribution (An & Schorfheide, 2007). Adding a distribution that is more in line with reality will also allow us to use the descriptive metrics as robustness metrics. Other articles, such as Xiao et al. (2018), included discrete values of parameters that are now outside our range. An important implication that must be taken into account is the effect of the ranges we set for the uncertainties on the validity of the model. Additional research is needed on how to ensure the validity of the model and its relationship with the given parameter ranges in EMA.

Much research can be found on estimating the values for the parameters in DSGE models using data, also known as the estimation of DSGE models. Using data we could have found estimates, or even ranges, we estimate the parameter values and create a range of 10% (reference) around these estimates. This would allow us to improve the quality of our analyzes and, consequently, the quality of the policy advice we could give using EMA. As an example of how the quality of the policy advice could be improved, let us consider the euro area. Although they share the same currency and are all affected by the interest rate set by the European Central Bank (ECB), the economy of each country could show significant differences and even differ in inflation rates. Muggenthaler et al. (2021) showed that European economies responded differently to the pandemic, while a common policy had to be set affecting all the euro countries. Stempel and Zahner (2022) find that the ECB has reacted more strongly to countries whose inflation rate exhibits greater deviations from their long-term trend. This might have negatively affected countries with smaller deviations. By using data from different countries, we could expand our analysis by evaluating how the interest rate affects the different (types of) economies in the euro area. EMA could help map which economies will be (dis)advantaged by certain policies and help the ECB design robust policies.

The objectives and robustness metrics are defined on the basis of the (assumed) goals of the central bank. Not only have we simplified reality by assuming that our DSGE models is the only model used to set the interest rate, the goals of the central bank are also assumed and might not be complete or in line with their actual practices. More common is the implementation of a welfare loss function, as is done in (Górajski et al., 2023). An advantage of a certain loss function is that it is easily comparable between policies. EMA requires us to interpret the results of the robustness metric and present which policy has which advantage. Although this might increase the quality of the analysis, it might be undesirable for the policy maker not to have a straightforward solution.

As the main goal of this thesis is to explore the use of EMA in DSGE models, some assumptions and modeling decisions are also motivated by showing the versatility of EMA and to assess which features of EMA are more suitable than others. Therefore, the interests and objectives are relatively general. We have ignored objectives related to the specific interest rate, such as the fact that the ECB changes its interest rate in steps of 0.25 or 0.5 percentage points (Heinemann & Ullrich, 2007). A logical next step is to take a real case and define case-specific objectives and robustness metrics based on concrete goals. This could result in new challenges, like goals that are not easily implemented in the objectives or robustness metrics, and new trade-offs. Other models include other economic variables as well, such as Christoffel et al. (2010), which include trade variables, employment, and the real exchange rate. Objectives related to these variables must be identified. Applying our exploration on a theoretical case is, however, considered an essential first step in implementing EMA in the field of macroeconomics. Having relatively general objectives and robustness metrics provides (macro)economists, who usually are not familiar with EMA, an accessible first impression of EMA.

We have considered only two theoretical cases (the occurrence of a supply shock and of a demand shock), whereas this basic DSGE model allows us to simulate a monetary shock and a technological shock as well. On one hand, applying the same methods to other cases can be considered redundant; however, on the other hand, applying our behavior-based scenario discovery using time-series clustering to our supply shock simulation had more challenges than applying this method to the demand shock. Xiao et al. (2018) considered technology, energy efficiency shocks, government expenditure shocks, and a tax rate shock. For each of these shocks, different challenges might arise. However, as long as the shocks are defined as a forced deviation from the steady state, EMA should be able to provide additional information over conventional methods.

Due to time constraints, choices must be made about what to do and what not to do. We considered applying other methods in our open exploration and directed search to have a larger contribution to our main research question than applying the same methods on different theoretical cases. However, it is still insightful to analyze the other shocks to find the best way EMA can be applied to each type of shock.

In the derivations of the equations of the DSGE model, an important assumption is made that allows us to log-linearize the model around a zero inflation steady state. This assumption requires that the deviations be small (Boneva et al., 2016). As we have seen, this is not the case for all experiments in our OE. Especially after a supply shock, we see large deviations from the steady state. Therefore, we cannot ensure the validity of the model in these cases. While the following assumption can be made for small deviations $x_t \equiv \log X_t - \log X = \log \frac{X_t}{X} = \log(1 + \%\text{change}) \approx \%\text{change}$, large deviations will be exaggerated under this assumption, because $\log(1 + \%\text{change}) < \%\text{change}$ if the change is large. For now, we can get away with mentioning this in the interpretation of our results. However, if EMA is applied to a real case, we should understand how to deal with this assumption and what to do with experiments that result in large deviations. Future research could look at how the results are affected by the shock size, how to set up the model and the experiments to ensure that the model is still valid, and how to identify experiments that result in invalid results.

Sack and Wieland (2000) showed that the uncertainty regarding the parameters could serve as an explanation for the smoothing of the interest rate. This contradicts our results in which several of our optimal policies had a minimal smoothing behavior of the interest rate. Based on the robustness metrics we considered, policies with minimal smoothing parameters were not outperformed by policies with higher degrees of smoothing. However, implementing more realistic distributions of the uncertainties might give different results.

6.2. Experiment set-up and computational features of our research

EMA requires a lot of computational power, as can be seen when calculating the Silhouette Index (SI) to determine which number of clusters would be best in our OE. In this research, we made an informed decision by calculating the SI for five sets of 2000 experiments and searching for a pattern between the number of clusters and the SI. Taking only 2000 experiments drastically reduced the computation time, and by doing this for 5 sets we assume that we can generalize our findings Implicitly, we assume that a larger set of experiments follows the same pattern as the 2000 experiments. Although another index, such as the Davies-Boulding index, might be less accurate, it is computationally less complex (Petrovic, 2006). This might allow us to qualify the clustering of a large data set, which makes the decision on number of clusters more straightforward.

Another improvement is by calculating the SI for multiple sets of slightly more experiments. However, as we considered calculating the SI for a set of 15000 experiments to be not achievable, there is an upper limit to the number of experiments we can do this for. The decision would still be an informed decision and not the exact solution. Making use of a supercomputer, which has a significantly better computational power, could solve this issue and provide us with the exact solutions, which is preferred over having to make an informed decision based on assumptions. It would also allow us to increase the number of experiments and, by this, to improve the quality of our research. Based on the results of this research, we do not have a reason to believe that the use of a regular computer is one of the main limitations. In fact, because this research can be considered an invitation to (macro)economists, who may not always have access to a supercomputer, to start using EMA in their research, we believe we should do this research in the same way it will most likely be done in practice. In this way, we were able to identify computational limitations of the methods and come up with a solution to deal with this. It could be interesting to conduct the same analysis on a supercomputer and compare the results to find out which added benefit a supercomputer could have.

6.3. Analysis methods

As we have made clear before, the basic idea behind EMA is that we run a model of a policy problem a large number of times and analyze the results. Just like the type of robustness metrics to use, the best method(s) for this analysis depends on the case. The time series clearly showed different types of behavior, which piqued our interest. Therefore, we decided to do scenario discovery using the behavior of the time series as a basis. To this end, we had to decide on a clustering algorithm and a way to validate how well the clustering algorithm worked with different numbers of clusters. In line with Steinmann et al. (2020), the Complexity Invariant Distance was used to create the clusters. This method has an advantage over other methods, such as the Euclidean distance, as it considers the 'complexity' of our time series when calculating the distance. Validation was done to decide what number of clusters to use, but not to qualify the quality of the clustering in an absolute manner. Addressing the quality could improve the interpretation of the analysis and be an indication that another clustering algorithm should be considered, or, in the worst case, this method is not suitable for this specific case. As an overall improvement of this method, an additional step could be integrated that improves the quality of the clustering. By applying different clustering algorithms and comparing the quality of the results of these algorithms using different time series, we make sure that we find the best possible time series clustering for that specific case. To be more efficient, the central bank could research which algorithm and cluster are best for the time series clustering for DSGE models and apply this without having to do the previously mentioned step for every analysis.

In our open exploration of the demand shock, the time series clustering approach we took provided us a proper basis for PRIM as we had multiple clusters that were large enough and of which the time series had distinctly different behavioral characteristics than the other clusters. Compared to the demand shock experiments, the experiments for the supply shock resulted in 1.) less 'complex' behavior of the time series, i.e. less (variety in) oscillation over time, and 2.) a larger magnitude of the extreme outcomes, i.e. the outliers had a larger deviation from the steady state throughout the whole simulation period. Due to the lack of (variety of) complexity in our time series, the CID will almost correspond to the Euclidean distance between time series. This makes it not surprising that our approach to the clustering algorithm did not provide a more useful basis for PRIM: we got one (or two) clusters with many experiments and several clusters containing the outliers. Therefore, since the clusters were too small and/or did not exhibit any unique behavioral characteristics, applying PRIM and interpreting it was less insightful. One of the options could have been to manually divide the large cluster manually, for example, into experiments resulting in a positive or a negative value at the first time step. Another way is to include the previously mentioned step to find the best clustering algorithm (and internal validation metric).

PRIM assumes that the uncertainties are not correlated, which may not be a valid assumption. Another approach is to transform the database before applying PRIM using Principal Component Analysis PRIM (Dalal et al., 2013). This method accounts for the parameters being correlated, but it decreases the ease of interpretation of our results.

Another approach we could have taken is a Multi-Scenario Multi-Objective Robust Optimization approach, which is a novel optimization approach. With this method, solutions that satisfy all constraints in all uncertainty scenarios are searched (Shavazipour et al., 2021). Under different values of the model uncertainties, the effect on the model objectives of the different policy levers is analyzed to find out how the trade-offs are affected by different states of the world. Compared to other approaches, the Multi-Scenario Multi-Objective Robust Optimization strikes a pragmatic balance between robustness considerations and optimality in individual scenarios, at reasonable computational costs (Bartholomew & Kwakkel, 2020). Thresholds could be established to avoid creating policies with undesirable economic outcomes, such as deflation.

6.4. Application of EMA in DSGE models

This model is significantly different from the standard models on which the EMA methods are applied. Whereas for example the lake model, the workhorse model for the EMA workbench, does not have a steady state that it forces itself to go to, the dynamic stochastic general equilibrium (DSGE) models have an equilibrium to go to (Kwakkel, 2017). In their paper, Maier et al. (2016) a large part of the uncertainty of the models they discuss is the existence of uncertain futures. However, this uncertainty does not exist in DSGE models, as it always converges to the steady state. In addition to this, there are no clear conflicting interests, resulting in the absence of a clear trade-off. Because this type of uncertainty and conflicting interests are not present in DSGE models, it is questionable whether EMA methods are the best approaches to assess uncertainty in the models and come up with robust decision policies. This does not mean that these methods are not suitable, but we should be cautious with the interpretation of the results and refrain from statements that these methods will result in a better applicability of DSGE models in real life, even after assessing the uncertainty. The outcomes of the models are not necessarily conflicting, and the DSGE models do not present a wicked problem, as we see, for example, in environmental sciences. EMA methods might be somewhat overpowered in assessing uncertainty in DSGE models and giving a false sense of robustness.

DSGE models are subject to criticism, as described in Storm (2021). Therefore, it is interesting to see what these methods can add to the academic research related to these models and might deal with some of the criticisms.

The fifth fallacy in Storm (2021) discusses the Lucas critique and states that 'Lucas-robust models do not exist, because – for reasons of performativity and reflexivity – individual rules of behavior may change in response to a change in policy regime.' The Lucas critique states that the parameters in the models change in response to policies and are not consistent over time Lucas Jr (1976). Lucas-robust models feature deep model parameters that are invariant to changes in the policy regime (Storm, 2021). As Hurtado (2014) has shown, the parameters in the DSGE models, including those assumed to be structural, change over time. Storm (2021) suggests that users of DSGE models should be aware of the dynamics and be cautious about the effect of this. This research only partly fills this awareness gap. In the methods applied in this research, the parameters are constant per experiment and do not change over time, whereas we should assume that some parameters might be dynamic over time. Although it is likely that the outcomes with these dynamic parameters are still within the ranges of our current analyses (assuming that the ranges of these dynamic parameters overlap with the ranges of our constant parameters), we cannot conclude anything about the behavior of our outcomes with dynamic parameters.

Another important limitation is, as called by L. P. Hansen and Sargent (2001), 'Wallace's conondrum'. The model assumes that the agents are rational and therefore know the behavior of a policy maker. This results in a lack of free variables for the policy maker to choose: 'their behavior rules are already known and responded to by private agents' (L. P. Hansen & Sargent, 2001; Sargent & Wallace, 1976). Taking the rational agent assumption in a somewhat looser way, we should also consider the case that, based on the behavior of the central bank, agents will adjust their behavior over time. If the central bank is very responsive, actors may behave differently than when the central bank is more reluctant to change the interest rate. This means that not for each policy, another distribution or range of uncertain parameters should be considered. This can be done a priori, by changing the distributions per policy when evaluating its performance, and using satsificing and descriptive metrics, or a posteriori by including this aspect in the interpretation of the results.

Before EMA can be applied in an extensive way, it is necessary to formalize how to properly interpret the results of EMA to form policy advice given the limitations of DSGE models. EMA does not increase the validity of DSGE models, and an appropriate interpretation is essential for it to be useful for policy makers.

7 Conclusion

In this thesis, we have analyzed shock and parametric uncertainty in a dynamic stochastic general equilibrium (DSGE) model. This is done through exploratory modeling and analysis (EMA), an approach that has not yet been used in relation to DSGE models. This research can be seen as an introduction to EMA in the world of (macro)economics. Compared to conventional methods, this introduces a novel way of evaluating the performance of policies. Whereas the use of a loss function is commonly done to evaluate and compare the performance of policies, EMA allows us to take (multiple) specific components of the model and evaluate and compare the performance of different policies on objectives that represent the different interests of the central bank. This maps out the trade-offs and could aid the central bank in making robust policy decisions. By applying EMA to two (theoretical) policy cases that are solved with the help of DSGE models, we explore the applicability of EMA on DSGE models, identify the added benefits of EMA, and find topics that might be of interest for future research.

DSGE models are used by central banks to set the interest rate to mitigate the economic consequences of an external shock in the economy. To do so, assumptions must be made about the values of the parameters in this model and about the characteristics of the shock. This introduces parametric and shock uncertainty into the model. EMA analyzes how the outcomes of the models are affected by these parametric uncertainties and helps design robust policies that will achieve the goals of the central bank, even though the exact state of the world is unknown.

In the first part of this thesis, we analyze how the different exogenous and endogenous parameters affect the results of the model after a demand and a supply shock, which is characteristic for the open exploration part of EMA. By doing so, we were able to answer the first subquestion, "*What influence do the uncertain parameters have on the behavior of inflation and the output gap over time after a supply shock and after a demand shock?*", which is answered through a behavior-based scenario discovery using time series clustering. By distinguishing different behaviors of the model variables inflation and output gap, the parameters that are most influential in causing these different types of behavior are found. Although it is not within the ability of the central bank to adjust the economy in such a way that they can influence these model variables to create the most desirable economic state, it is important for them to know which behavior they can expect to occur in the economy. Knowing which assumptions about what exogenous parameters they should be cautious about, as it might be critical in their predictions, will improve the quality of their analyses. In this case, the share of firms that can adjust their prices at every time step and the persistence of the shock turned out to be the most influential parameters for the behavior of the output gap after a supply shock. If the true value of (one of) these two parameters deviates from the value the central bank used in its predictions, the economy might behave substantially differently from its expectations. These two parameters also influence the behavior of inflation, as well as the degree to which consumers substitute between goods. Importantly, we cannot assign one behavioral characteristic to one parameter, ceteris paribus, but the combination of the parameters being in certain ranges causes distinguishable behavior. However, we get an idea of how the parameters influence the outcomes of the model. Low persistence, meaning that the duration of the shock is short, could result in more oscillation behavior for both the output gap and inflation after a supply shock; and a high degree of price stickiness, meaning that the economy is slower to adjust to a new economic state (think of response to a change in the interest rate), could result in less oscillating behavior as well. More importantly, we have seen that the two policy levers, the input parameters that the central bank can influence, also have a large influence on the behavior of the economy after a negative supply shock. Changing these two policy levers is, contrary to the previously two mentioned exogenous parameters, within the power of the central bank. These levers are the degree to which the current interest rate is influenced by (1.) the interest rate in the previous period (smoothing parameter of the interest rate) and (2.) the inflation in the same period. If the central bank decides not to have large changes in the interest rate between two periods, the output gap and inflation are more likely to have large deviations from the steady state. The effect of how much inflation influences the current interest rate is not clearly distinguishable. After a demand shock, the policy levers did not have a significant influence on the behavioral characteristics that we distinguished in this analysis. However, clustering of time series of economic variables after a demand shock provided only limited information on behavioral characteristics.

The fact that these two policy levers have a significant effect on the behavior of the output gap and inflation after at least one shock indicates the need for robust decision making, which is done in the next step of our research, as well. In addition to this, when formulating a policy problem, the central bank is advised to pay special attention to their assumptions regarding the price stickiness, the persistence of the shock, and the degree of substitution of the consumer, and how the interaction of these assumptions.

In the second case, different robust policies are designed in reaction to a theoretical demand shock (Directed Search). By doing so, the second subquestion is answered: ”*What are the optimal values for the smoothing parameter for the interest rate and the policy response to inflation the central bank should follow to robustly achieve stability in the economy after a supply shock?*” An important step in designing robust policies is the formulation of objectives and the robustness metrics. The objectives and robustness metrics should reflect the goals and interests of the central bank. In our theoretical case, the objectives and robustness metrics reflect the central bank’s desire to minimize the deviations of the output gap and inflation and minimize the period-to-period change of these deviations. Additionally, robustness metrics are included that reflect the central bank’s need to be accurate in their predictions. The fact that the central bank is focused on inflation control is reflected in the objectives and robustness metrics. Consequently, this is also reflected in the assessment of the policies. Because there is a trade-off between optimizing inflation-related objectives and optimizing output gap-related objectives, the policies that have the most desirable effect on inflation result in sub-optimal performances for the output gap and vice versa. We find that the policy with the minimizing the output gap and the policy minimizing the deviation in inflation from the steady state both minimize interest rate smoothing. The optimal policy for the output gap minimizes the response of the interest rate to current inflation. The optimal policy for minimizing the deviation in inflation maximizes the influence of current inflation on the current interest rate. These two policies tend to have either the best or the worst score on the robustness metrics. The policy that is considered to best maintain economic stability in a robust way puts little emphasis on the previous interest rate, and a large emphasis on the current inflation, but not to extreme measures.

Having applied EMA on DSGE models to answer the two subquestions that are representative for the two main branches of EMA (Open Exploration and Directed Search), we can answer the research question of this thesis: ” *What can we learn from applying exploratory modeling and analysis methods to dynamic stochastic general equilibrium models and how can the central bank implement this when deciding the interest rate?* In this research we have shown that EMA can provide useful insights into DSGE models and could help the central bank designing robust policies. However, a few more steps need to be taken before the central bank can incorporate EMA in its policy creation process. The cases in this research are solely theoretical cases using the most basic DSGE model, without any real-world data and simple assumptions about the central bank’s interests. However, the central bank might work with adjusted DSGE models and/or have other interests than we have assumed now, requiring other objectives and robustness metrics to be formulated. By applying EMA to a real policy problem and/or including real-world data to estimate the parametric input values and ranges, additional benefits or pitfalls of EMA can be identified. As stated, EMA is yet unknown in the world of economics, causing a knowledge gap for both economists and those familiar with EMA. Collaborative research is therefore advised combining the economists’ insights in the (practical) usage of DSGE models and the EMA modeler’s experience to find the best suitable algorithms and methods for the different policy problems and to optimally formulate this policy problem for applying EMA. There are several extensions of DSGE models that we can analyze with EMA. Future research is recommended to optimize which EMA methods and algorithms, such as adjusting the clustering algorithm for the time series after a supply shock, are best suited for DSGE models and their functionalities. When additional research is conducted, one of the focus should be ensuring the validity of the DSGE models when EMA is applied.

References

Altavilla, C., Burlon, L., Giannetti, M., & Holton, S. (2022). Is there a zero lower bound? the effects of negative policy rates on banks and firms. *Journal of Financial Economics*, *144*(3), 885–907.

An, S., & Schorfheide, F. (2007). Bayesian analysis of dsge models. *Econometric reviews*, *26*(2-4), 113–172.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern recognition*, *46*(1), 243–256.

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations research*, *41*(3), 435–449.

Bartholomew, E., & Kwakkel, J. H. (2020). On considering robustness in the search phase of robust decision making: A comparison of many-objective robust decision making, multi-scenario many-objective robust decision making, and many objective robust optimization. *Environmental Modelling & Software*, *127*, 104699.

Batista, G. E., Keogh, E. J., Tataw, O. M., & De Souza, V. M. (2014). Cid: An efficient complexity-invariant distance for time series. *Data Mining and Knowledge Discovery*, *28*, 634–669.

Baumeister, C. (2023). *Pandemic, war, inflation: Oil markets at a crossroads?* (Tech. rep.). National Bureau of Economic Research.

Ben-Haim, Y., & Van den End, J. W. (2019). Fundamental uncertainty about the natural rate of interest: Info-gap as guide for monetary policy. *De Nederlandsche Bank Working Paper*.

Bezerra, L. C., López-Ibánez, M., & Stützle, T. (2017). An empirical assessment of the properties of inverted generational distance on multi-and many-objective optimization. *International conference on evolutionary multi-criterion optimization*, 31–45.

Boehl, G. (2022). Efficient solution and computation of models with occasionally binding constraints. *Journal of Economic Dynamics and Control*, *143*, 104523.

Boivin, J., & Giannoni, M. (2006). Dsge models in a data-rich environment.

Boneva, L. M., Braun, R. A., & Waki, Y. (2016). Some unpleasant properties of loglinearized solutions when the nominal rate is zero. *Journal of Monetary Economics*, *84*, 216–232.

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, *77*(1), 34–49.

Buiter, W. H. (2003). Deflation: Prevention and cure.

Čapek, J., et al. (2009). *Sensitivity analysis of a dsge model* (tech. rep.). Working paper.

Carceller, M., & van den End, J. W. (2023). Robust monetary policy under shock uncertainty.

Christiano, L. J., Trabandt, M., & Walentin, K. (2010). Dsge models for monetary policy analysis. In *Handbook of monetary economics* (pp. 285–367). Elsevier.

Christoffel, K. P., Coenen, G., & Warne, A. (2010). Forecasting with dsge models.

Coello, C. A. C. (2007). *Evolutionary algorithms for solving multi-objective problems*. Springer.

Dalal, S., Han, B., Lempert, R., Jaycocks, A., & Hackbarth, A. (2013). Improving scenario discovery using orthogonal rotations. *Environmental Modelling & Software*, *48*, 49–64.

De Jong, D. N., & Dave, C. (2012). *Structural macroeconometrics*. Princeton University Press.

De Nederlandsche Bank. (2021). Ecb sets clear and simple inflation target. https://www.dnb.nl/en/general-news/dnbulletins-2021/ecb-sets-clear-and-simple-inflation-target/

Del Negro, M., & Schorfheide, F. (2013). Dsge model-based forecasting. In *Handbook of economic forecasting* (pp. 57–140). Elsevier.

Den Haan, W. J., Rendahl, P., & Riegler, M. (2018). Unemployment (fears) and deflationary spirals. *Journal of the European Economic Association*, *16*(5), 1281–1349.

DNB. (2020). Supply and demand shocks due to the coronavirus pandemic contribute equally to contraction in production. https://www.dnb.nl/en/general-news/dnbulletin-2020/supply-and-demand-shocks-due-to-the-coronavirus-pandemic-contribute-equally-to-contraction-in-production/

Dudek, A. (2020). Silhouette index as clustering evaluation tool. *Classification and Data Analysis: Theory and Applications 28*, 19–33.

Fernández-Villaverde, J., & Guerrón-Quintana, P. A. (2021). Estimating dsge models: Recent advances and future challenges. *Annual Review of Economics*, *13*, 229–252.

Ferroni, F., Grassi, S., & León-Ledesma, M. A. (2015). *Fundamental shock selection in dsge models* (tech. rep.). School of Economics Discussion Papers.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and computing*, *9*(2), 123–143.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: An introduction to the new keynesian framework and its applications*. Princeton University Press.

Górajski, M., Kuchta, Z., & Leszczyńska-Paczesna, A. (2023). Price-setting heterogeneity and robust monetary policy in a two-sector dsge model of a small open economy. *Economic Modelling*, *122*, 106227.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & Ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global environmental change*, *23*(2), 485–498.

Halim, R. A., Kwakkel, J. H., & Tavasszy, L. A. (2016). A scenario discovery study of the impact of uncertainties in the global container transport system on european ports. *Futures*, *81*, 148–160.

Hallegatte, S., Shah, A., Brown, C., Lempert, R., & Gill, S. (2012). Investment decision making under deep uncertainty–application to climate change. *World Bank Policy Research Working Paper*, (6193).

Hansen, L. P., & Sargent, T. J. (2001). Acknowledging misspecification in macroeconomic theory. *Review of Economic Dynamics*, *4*(3), 519–535.

Hansen, N., Toscani, R., & Zhou, J. (2023). Europe's inflation outlook depends on how corporate profits absorb wage gains. https://www.imf.org/en/Blogs/Articles/2023/06/26/europes-inflation-outlook-depends-on-how-corporate-profits-absorb-wage-gains

Heinemann, F., & Ullrich, K. (2007). Does it pay to watch central bankers' lips? the information content of ecb wording. *Swiss Journal of Economics and Statistics*, *143*, 155–185.

Herman, J. D., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2015). How should robustness be defined for water systems planning under change? *Journal of Water Resources Planning and Management*, *141*(10), 04015012.

Hoffman, F. O., & Hammonds, J. S. (1994). Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk analysis*, *14*(5), 707–712.

Homaei, S., & Hamdy, M. (2020). A robustness-based decision making approach for multi-target high performance buildings under uncertain scenarios. *Applied Energy*, *267*, 114868.

Hurtado, S. (2014). Dsge models and the lucas critique. *Economic Modelling*, *44*, S12–S19.

Jahan, S., & Mahmud, A. S. (2013). What is the output gap. *Finance & Development*, *50*(3), 38–39.

Jain, M., Jain, M., AlSkaif, T., & Dev, S. (2022). Which internal validation indices to use while clustering electric load demand profiles? *Sustainable Energy, Grids and Networks*, *32*, 100849.

Jiang, S., & Yang, S. (2016). Convergence versus diversity in multiobjective optimization. *International Conference on Parallel Problem Solving from Nature*, 984–993.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., & Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, *42*, 55–71.

Kollat, J. B., & Reed, P. M. (2007). A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. *Advances in Water Resources*, *30*(3), 408–419.

Krugman, P., & Wells, R. (2021). *Macroeconomics*. Worth Publishers, Macmillan Learning.

Kwakkel, J. H. (2017). The exploratory modeling workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling & Software*, *96*, 239–250.

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting and Social Change*, *80*(4), 789–800.

Kwakkel, J. H., Eker, S., & Pruyt, E. (2016). How robust is a robust policy? comparing alternative robustness metrics for robust decision-making. *Robustness analysis in decision aiding, optimization, and analytics*, 221–237.

Kwakkel, J. H., Haasnoot, M., & Walker, W. E. (2016). Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty. *Environmental Modelling & Software*, *86*, 168–183.

Kwakkel, J. H., & Jaxa-Rozen, M. (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling & Software*, *79*, 311–321.

Kwakkel, J. H., & Pruyt, E. (2013). Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting and Social Change*, *80*(3), 419–431.

Kwakkel, J. H., & Pruyt, E. (2015). Using system dynamics for grand challenges: The esdma approach. *Systems Research and Behavioral Science*, *32*(3), 358–375.

Kwakkel, J. H., Walker, W. E., & Haasnoot, M. (2016). Coping with the wickedness of public policy problems: Approaches for decision making under deep uncertainty.

Lane, P. R. (2023). Disinflation and monetary policy in the euro area.

Lempert, R. J. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis.

Lempert, R. J. (2019). Robust decision making (rdm). *Decision making under deep uncertainty: From theory to practice*, 23–51.

Lucas Jr, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester conference series on public policy*, *1*, 19–46.

Maier, H. R., Guillaume, J. H., van Delden, H., Riddell, G. A., Haasnoot, M., & Kwakkel, J. H. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental modelling & software*, *81*, 154–164.

Matheson, T. D. (2019). Chapter 15 interest rates and inflation. In *Brazil* (ch015). International Monetary Fund. https://doi.org/10.5089/9781484339749.071.ch015

McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *42*(1), 55–61.

Muggenthaler, P., Schroth, J., Sun, Y., et al. (2021). The heterogeneous economic impact of the pandemic across euro area countries. *Economic Bulletin Boxes*, *5*.

Petrovic, S. (2006). A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters. *Proceedings of the 11th Nordic workshop of secure IT systems*, *2006*, 53–64.

Ratto, M., Hohberger, S., & Burgert, M. (2016). New software tool for the evaluation of dsge models: An application with dynare.

Ratto, M. (2008). Analysing dsge models with global sensitivity analysis. *Computational Economics*, *31*, 115–139.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53–65.

Sack, B., & Wieland, V. (2000). Interest-rate smoothing and optimal monetary policy: A review of recent empirical evidence. *Journal of Economics and Business*, *52*(1-2), 205–228.

Salazar, J. Z., Reed, P. M., Herman, J. D., Giuliani, M., & Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. *Advances in water resources*, *92*, 172–185.

Sargent, T. J., & Wallace, N. (1976). Rational expectations and the theory of economic policy. *Journal of Monetary economics*, *2*(2), 169–183.

Shavazipour, B., Kwakkel, J. H., & Miettinen, K. (2021). Multi-scenario multi-objective robust optimization under deep uncertainty: A posteriori approach. *Environmental Modelling & Software*, *144*, 105134.

Slanicay, M. (2014). Some notes on historical, theoretical, and empirical background of dsge models. *Review of Economic Perspectives*, *14*(2), 145–164.

Smets, F., & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European economic association*, *1*(5), 1123–1175.

Statnikov, R. B., & Matusov, J. B. (2012). *Multicriteria optimization and engineering*. Springer Science & Business Media.

Steinmann, P., Auping, W. L., & Kwakkel, J. H. (2020). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change*, *156*, 120052.

Stempel, D., & Zahner, J. (2022). *Dsge models and machine learning: An application to monetary policy in the euro area* (tech. rep.). MAGKS Joint Discussion Paper Series in Economics.

Storm, S. (2021). Cordon of conformity: Why dsge models are not the future of macroeconomics. *International Journal of Political Economy*, *50*(2), 77–98.

Taylor, J. B., & Williams, J. C. (2010). Simple and robust rules for monetary policy. In *Handbook of monetary economics* (pp. 829–859). Elsevier.

The Economist. (2023). Wage-price spirals are far scarier in theory than in practice. https://www.economist.com/finance-and-economics/2023/06/15/wage-price-spirals-are-far-scarier-in-theory-than-in-practice

Tovar, C. E. (2009). Dsge models and central banks. *Economics*, *3*(1), 20090016.

Walker, W., Marchau, V., & Kwakkel, J. (2013). Uncertainty in the framework of policy analysis [Jaar van uitgave: 2013]. In W. Walker & W. Thissen (Eds.), *Public policy analysis : New developments* (pp. 215–261). Springer.

Xiao, B., Fan, Y., & Guo, X. (2018). Exploring the macroeconomic fluctuations under different environmental policies in china: A dsge approach. *Energy Economics*, *76*, 439–456.

Yen, G. G., & He, Z. (2013). Performance metric ensemble for multiobjective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, *18*(1), 131–144.

A

Additional information about model equations and parameters

A.1. Parameters and variables

There are several parameters that will be used. De Jong and Dave (2012) and Galí (2015) are used to find the parameter values, as well as expert knowledge from a central bank.

Output gap ($\tilde{y}$)

Variable

The output gap represents the difference between the actual level of output in the economy and its potential level. If the value is 0, the actual economy is at full capacity (as assumed by the conventional macroeconomic theory). A negative value of the output gap means that resources are underutilized. Positive values of the output gap mean that the demand for products is larger than the capacity to supply them, which usually results in inflation.

Inflation (π)

Variable

Inflation represents the rate of change in the general price level in an economy. It captures the average increase in prices over time and, in this research, is defined as a certain percentage price change per year.

Nominal policy interest rate (i)

Variable

The interest rate represents the nominal rate at which households and firms can borrow or lend money. Lower interest rates tend to stimulate consumption and investment as borrowing becomes cheaper and more attractive. Higher interest rates, on the other hand, can discourage borrowing and spending, which can lead to a slowdown in economic activity.

Discount factor (β)

Parameter

The discount factor represents the subjective rate at which households value future consumption relative to current consumption. The discount factor reflects the intertemporal demands of households and determines their savings and consumption decisions over time. The discount factor should have a value between 0 and 1, with higher values representing a stronger demand for future consumption.

Inverse marginal elasticity of substitution (σ)

Parameter

The inverse marginal elasticity of substitution represents the elasticity of substitution between consumption and leisure. It measures how individuals respond to changes in the relative price of consumption and leisure when making decisions about how much to consume and how much labor to supply. It usually has a value of 1.

Degree of price stickiness (κ)

Parameter

The level of price stickiness refers to the degree to which prices in an economy adjust in response to changes in economic conditions. Thus when the price stickiness increases, an exogenous shock is

expected to have a longer lasting effect as the economy is slower to adjust. This value depends on the Calvo parameter and the discount factor.
Calvo parameter (ω)
Parameter
The Calvo parameter represents the degree of nominal price stickiness or the frequency at which firms are able to adjust their prices. A higher Calvo parameter (closer to 1) results in more persistence of the consequences of a shock as firms are more unable to adjust their prices in the short run. A lower Calvo parameter indicates a faster adjustment.
Time demand parameter (γ)
The time demand parameter depends on the inverse Frisch elasticity of the labor supply, the targeted share of impatient agents, and the smoothing parameter.

It usually has a value of 2.5 but through the parametric uncertainty of the three previously mentioned parameters.
Inverse Frisch elasticity of labor supply (χ)
Parameter
The inverse Frisch elasticity of labor supply represents the responsiveness of labor supply to changes in the real wage rate. A higher value of the inverse Frisch elasticity results in bigger responses to changes in the real wage rate. This results in a faster adjustment after a shock. The value of the inverse Frisch elasticity is typically between 0 and 1.
Targeted share of impatient agents (z)
Parameter
The targeted share of impatient agents refers to the proportion of households or economic agents in the model that are assumed to have a relatively higher degree of impatience or discount the future more heavily when making consumption decisions. This is included in the model to make sure the economic agents do not have homogeneous demands but to include some differentiation. This parameter has a value of 0.33.
Real natural rate of interest (r^n)
Exogenous variable
The real natural rate of interest represents the interest rate in equilibrium that prevails in the absence of nominal friction or disturbances. It can be thought of as the real interest rate consistent with stable inflation and long-term full employment. It has a value of 0.
Demand shock (d)
Exogenous shock
The demand shock represents an exogenous change in the demands of households for consumption, leisure, or intertemporal substitution. A positive shock represents an increase in demand and thus results in a positive output gap. These shocks can differ in size and represent a theoretical shift from the steady state.
Supply shock (μ)
Exogenous shock
The supply shock refers to an exogenous shock that affects inflation in an economy. It represents a change in the input prices or production costs that firms face, influencing their decisions regarding prices, output, and employment. A positive exogenous shock leads to an increase in the potential level of an economy.
Auto-regressive term of demand/supply shock (ρ_d)
Parameter
The auto-regressive term of demand shock refers to the component of the shock that exhibits persistence over time. The assumption is that the demand shock does not fully dissipate immediately, but it takes a few time steps to disappear. This is usually assumed to be 0.8, which means that in the step after a full shock, we still observe 80% of the shock.
Expected value at time $t+1$ for parameter x ($E(x_{t+1})$
This model assumes that agents have expectations about future variables and incorporate those expectations into their current decision-making. The assumption is that the expected value is 50% the previous step and 50% of the value in the next step.
Policy response to inflation (φ_π)
Parameter, lever
The policy response to inflation is a measure set by the policy maker to what extent it responds to

inflation.
Policy response to output gap (φ_y)
Parameter, lever
The policy response to inflation is a measure set by the policy maker to determine to what extent it responds to the output gap. In our model, we keep this value constant at 0.5, since the primary focus of our policy maker is assumed to be inflation.
Smoothing parameter policy interest rate (ρ_r)
Parameter, lever The smoothing parameter is a parameter that determines the degree of smoothing of the interest rate in the monetary policy rule. The central bank prefers to make gradual adjustments, as it assumed that this will reduce volatility and uncertainty in the economy, allowing agents to make more informed decisions. This parameter is set by the policy maker itself and normally takes up a value of 0.8.

B
Shock size

Because the shock size is not based on data and is thus chosen relatively arbitrarily, we are interested in the effect of a change in shock size. Therefore, we consider a shock that is ten times as big.

The results show that the effect on the variables in the models, output gap and inflation, is also ten times as big. To check for the whole behavior in the model we have taken the 25%, 50%, 75% quantiles and the maximum values in timesteps t=1, t=4, t=8, t=12 and t=30 for the positive and the negative values separately, as well as the number of experiments with positive values in these time steps. See Figure B.1 for a schematic explanation of the values that are taken. The black lines correspond to the contours of the experiment run, thus are the maximum (positive) and minimum (negative) values. The blue (red) area is the positive (negative) area with the blue (red) lines showing the quantiles and maximums (minimums) we are interested in. Note that, for simplicity, this visualisation assumes a linear distribution. In Figure B.2 we see the difference between a linear and a nonlinear distribution. We check the quantiles so we have insights on the distribution of experiments within the run underneath the maximum value for the different shock sizes and how this is distributed.

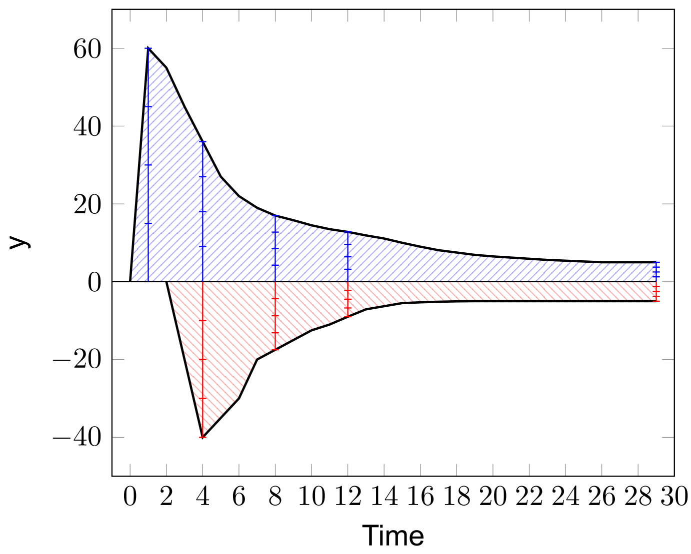
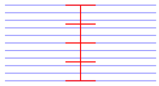
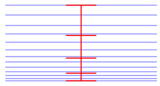


Figure B.1: A schematic contour of the outcomes of the experiment runs for a demand shock of size 0.01. For time t=1, t=4, t=8, t=12, and t=29, the 25, the first, second, third quantiles and maximum values are taken for both the positive values and negative values sepearately are taken. The experiments are, hypothetically taken as linear, resulting in equal distances between all quantiles.

The Tables B.1, B.3 show the results for, respectively, the demand shock and the supply shock. The ratios for most values are around 10, which indicates a 1-to-1 relation of the size of the shock and its



(a) Linear distribution of experiment time series



(b) Nonlinear distribution of experiment time series

Figure B.2: Linear and nonlinear distribution of the experiment time series. The blue lines represent the different time series and the red lines represent the lowest values, the 25% quantile, the 50% quantile, the 75% quantile and maximum value. We see that in the linear distribution, the red vertical lines are equally distributed, whereas for the nonlinear distribution the 25% quantile is smaller as the experiment time series are more dense to each other.

effect on the variables. For extremes, especially for supply shock, we see that the ratio is not always near 10. As can be seen in the results section, the extremes for these two shocks are clear outliers from the rest of the experiments. The maximum and minimum values take on the values of these outliers, resulting in a larger possibility of not having a ratio of 10.

Table B.1: Comparison of the output gap and inflation results after a small (0.01) and a big (0.1) demand shock and the accompanying ratios. The time-column reflects the time steps and either the positive (+) or the negative (-) values.

Time	Variable	Output gap			Inflation		
		Small shock	*Big shock*	*Ratio*	*Small shock*	*Big shock*	*Ratio*
1+	25% quantile	0.344	3.447	10.032	0.368	3.714	10.080
	50% quantile	0.493	4.916	9.970	0.829	8.2839	9.993
	75% quantile	0.617	6.169	9.999	1.579	15.606	9.882
	Maximum	0.891	8.979	10.075	5.680	60.838	10.711
4-	25% quantile	-0.028	-0.278	10.042	-0.168	-1.663	9.878
	50% quantile	-0.059	-0.583	9.913	-0.453	-4.434	9.790
	75% quantile	-0.113	-1.138	10.042	-0.947	-9.577	10.110
	Minimum	-0.285	-2.878	10.110	-3.830	-39.284	10.257
4+	25% quantile	0.020	0.210	10.714	0.153	1.540	10.065
	50% quantile	0.049	0.516	10.500	0.280	2.842	10.162
	75% quantile	0.094	0.949	10.123	0.472	4.697	9.948
	Maximum	0.451	4.465	9.892	4.907	31.871	6.495
8-	25% quantile	-0.007	-0.064	9.186	-0.036	-0.374	10.315
	50% quantile	-0.016	-0.159	9.717	-0.097	-0.989	10.156
	75% quantile	-0.039	-0.374	9.600	-0.220	-2.215	10.053
	Minimum	-0.179	-1.879	10.502	-1.248	-12.623	10.111
8+	25% quantile	0.004	0.037	10.130	0.045	0.455	10.158
	50% quantile	0.012	0.123	10.116	0.109	1.120	10.275
	75% quantile	0.032	0.335	10.520	0.217	2.158	9.927
	Maximum	0.117	1.328	11.380	2.376	18.351	7.725
12-	25% quantile	-0.001	-0.009	10.160	-0.008	-0.085	10.116
	50% quantile	-0.003	-0.036	10.631	-0.031	-0.309	9.837
	75% quantile	-0.010	-0.102	10.071	-0.086	-0.846	9.851
	Minimum	-0.156	-1.415	9.056	-0.611	-5.581	9.127
12+	25% quantile	0.001	0.011	10.209	0.010	0.105	10.483
	50% quantile	0.003	0.035	10.079	0.036	0.360	10.110
	75% quantile	0.010	0.100	10.244	0.094	0.948	10.129
	Maximum	0.083	1.008	12.202	1.743	14.555	8.350
29-	25% quantile	-0.000	-0.000	10.960	-0.000	-0.000	9.880
	50% quantile	-0.000	-0.000	10.940	-0.000	-0.001	9.374
	75% quantile	-0.000	-0.000	9.604	-0.001	-0.011	9.938
	Minimum	-0.056	-0.419	7.543	-0.082	-0.946	11.557
29+	25% quantile	0.000	0.000	9.831	0.000	0.000	10.175
	50% quantile	0.000	0.000	10.042	0.000	0.005	10.398
	75% quantile	0.000	0.003	9.623	0.007	0.074	10.626
	Maximum	0.028	0.346	12.512	0.538	5.481	10.183

Table B.2: Number of positive values for the output gap and inflation for time steps 1, 4, 8, 12, and 29 caused by a small and a big negative supply shock and the ratio of these two.

Period	Output gap			Inflation		
	Small shock	*Big shock*	*Ratio*	*Small shock*	*Big shock*	*Ratio*
1	15000	15000	1.000	15000	15000	1.000
4	5307	5209	0.982	5307	11409	0.996
8	7030	7019	0.998	7030	8261	1.004
12	7335	7435	1.014	7335	9367	1.000
29	7602	7658	1.007	7602	9354	0.993

Table B.3: Comparison of the output gap and inflation results after a small (0.01) and a big (0.1) supply shock and the accompanying ratios. The time-column reflects the time steps and either the positive (+) or the negative (-) values.

Time	Variable	Output gap			Inflation		
		Small shock	*Big shock*	*Ratio*	*Small shock*	*Big shock*	*Ratio*
1-	25% quantile	-1.445	-14.427	9.982	-1.389	-14.299	10.295
	50% quantile	-2.142	-21.550	10.061	-5.167	-47.251	9.145
	75% quantile	-3.464	-34.496	9.957	-18.164	-168.910	9.299
	Minimum	-15472.12	-61190.102	3.955	-16877.540	-86275.827	5.112
1+	25% quantile	0.717	6.753	9.422	3.005	30.571	10.172
	50% quantile	1.774	16.522	9.313	6.338	63.741	10.057
	75% quantile	4.138	38.376	9.274	10.309	102.322	9.925
	Maximum	63.657	773.009	12.143	56.440	543.896	9.637
4-	25% quantile	-0.750	-7.401	9.863	-0.567	-6.491	11.458
	50% quantile	-1.538	-15.463	10.051	-1.662	-16.883	10.159
	75% quantile	-3.060	-30.630	10.010	-5.640	-53.963	9.567
	Minimum	-7076.28	-24582.77	3.474	-27189.26	-117212.19	4.311
4+	25% quantile	0.288	2.875	9.966	2.040	20.116	9.863
	50% quantile	1.044	10.001	9.579	5.424	54.671	10.079
	75% quantile	3.665	30.168	8.231	12.939	125.043	9.664
	Maximum	259.148	2574.650	9.935	770.376	6859.637	8.904
8-	25% quantile	-0.235	-2.362	10.074	-0.106	-1.134	10.735
	50% quantile	-0.768	-7.749	10.084	-0.451	-4.408	9.771
	75% quantile	-2.204	-21.971	9.970	-1.437	-14.483	10.076
	Minimum	-162.924	-8449.364	51.861	-3251.097	-26960.212	8.293
8+	25% quantile	0.038	0.389	10.232	0.453	4.594	10.125
	50% quantile	0.185	1.976	10.695	1.911	18.358	9.608
	75% quantile	1.006	11.471	11.402	6.310	59.532	9.434
	Maximum	2556.103	11848.179	4.635	1895.606	15613.561	8.237
12-	25% quantile	-0.049	-0.502	10.149	-0.044	-0.481	10.845
	50% quantile	-0.292	-2.979	10.186	-0.229	-2.444	10.692
	75% quantile	-1.086	-10.815	9.958	-0.950	-9.394	9.890
	Minimum	-101.048	-855.173	8.463075	-660.245	-8888.853	13.463
12+	25% quantile	0.010	0.095	9.398	0.085	0.849	9.985
	50% quantile	0.037	0.388	10.506	0.553	5.347	9.669
	75% quantile	0.152	1.424	9.350	2.452	22.408	9.1399
	Maximum	1913.593	3830.050	2.001	5792.446	21844.945	3.771
29-	25% quantile	-0.000106	-0.001	10.159218	-0.000	-0.000	10.213
	50% quantile	-0.003	-0.032	9.345	-0.001	-0.007	9.506
	75% quantile	-0.062	-0.622	9.968	-0.008	-0.064	7.778
	Minimum	-24.022	-256.498	10.676	-38.734	-211.222	5.453
29+	25% quantile	0.000	0.000	9.299	0.000	0.002	10.261
	50% quantile	0.000	0.002	9.814	0.007	0.069	10.105
	75% quantile	0.002	0.024	11.064	0.139	1.316	9.452
	Maximum	87.306	130.281	1.492	232.483	794.519	3.418

C
Additional information objectives

C.1. Inflation objectives

In this thesis, the steady state inflation rate is assumed to be 2%. To explore the use of EMA, a desired inflation domain is added, which is assumed to be between 2% and 4%.

Central banks can adjust the interest rate to influence the interest rate. Although negative interest rates are possible, it is more desirable to stay in a positive range. For this, inflation is preferred to deflation. Because it is not conventional to have negative interest rates, deflation is extremely unwanted for the central bank; it cannot use its main tool, that is, the adjustment of the interest rate (Krugman & Wells, 2021). Because deflation in itself is unwanted, irrespective of its size, we consider the number of periods in which deflation occurs, and not its size. Therefore, high values for inflation, deflation, and large variances are assumed to be unwanted. This brings us to the following assessment criteria on which we have to assess our policies.

- Magnitude of (extreme) inflation
- Duration of unwanted positive inflation
- Duration of deflation
- Variance in inflation between periods
- Economic consequences at time $t = 12$

To assess these criteria, we have to create objectives we can give to our model. These are the following objectives:

- Inflation level at time $t = 12$ *(minimize)*
- Maximum inflation rate between time $t = 0$ and $t = 12$ *(minimize)*
- Number quarters with deflation *(minimize)*
- Quarters of inflation inside of desired domain *(maximize)*
- Largest change in inflation between two subsequent periods after time $t = 2$ *(minimize)*
-

It is possible that there is a trade-off between these objectives. Example

It is possible to think of other objectives as well, such as the largest change in inflation between two subsequent periods (large variances between periods are unwanted), or the number of periods in which inflation (or deflation) above a certain threshold occurs (by doing this, we create a policy that is within a certain domain).

Inflation rate at time $t = 12$

An important goal of the central bank is the level of inflation after three years, at time $t = 12$. When conducting our experiments, we aim to minimize inflation and keep it as close to 2% as possible. In this thesis, the ideal inflation rate is assumed to be between 2% and 4% annually, as both low inflation or deflation and high inflation are undesirable. Therefore, this is the threshold in the subsequent analysis.

Maximum inflation or deflation between time $t = 0$ and $t = 12$
Even if the inflation rate at time $t = 12$ is around 2%, it is undesirable to have extreme inflation or deflation, as this would create economic instability; therefore, we want to minimize this.

C.1.1. Number of quarters in which deflation occurs
Deflation is an unwanted phenomenon in the economy and could potentially limit the impact the central bank could have in achieving price stability. Therefore, one of the objectives we can consider is the number of periods in which low inflation or deflation occurs. Adding this objective shows that deflation is more undesirable than large inflation.

C.1.2. Quarters of inflation outside of desired domain
By determining a domain of inflation rates that is acceptable for the central bank, we can search for policies that have the least number of periods outside of this domain. A policy that has zero quarters outside of this domain has acceptable values for inflation, and its largest deviation between two subsequent periods is the difference between the lower and upper bounds of this domain. Once the inflation is outside of the domain, we, however, do not know how extreme this deviation is; for example, it could be outside of the domain with 0.01%, or with 5%. The threshold we will consider is between 2% and 4% yearly inflation.

Within the scope of this research, this is assumed to be a well-balanced set of objectives to create an optimal policy using the inflation time series. Depending on the needs of the central bank, these objectives can be adjusted. Think, for example, of summing the squared undesired deviations from the desired upper or lower limit at each time step, creating an objective that 'punishes' for large deviations (a 1 time deviation of 10 results in a value of 100, whereas a deviation of 1 occurring 10 times results in a value of 10).

C.2. Output gap objectives
The central bank we consider is less focused on the output gap than on inflation. However, large output gaps are also unwanted, both positive and negative, as well as large variances. For example, one of the aspects we can consider when we focus on output gap is employment. A negative output gap results in unemployment, which is considered undesirable. However, a positive output gap might result in wage-price spirals; because of the large demand for labor, wages might rise, resulting in higher prices, causing employees to ask for higher wages. Large variances in the output gap result in instability in the economy, such as uncertain employment and firms being hesitant with their investments, as well as low confidence in the economy. In this thesis, we evaluate both sides equally; there is no preference for having a positive or negative output gap. The goal of the central bank is to have a policy that is consistently close to the steady state, meaning small deviations and low variance.

We take the following assessment criteria:

- Magnitude of deviation from steady state
- Variance of deviation between $t = 0$ and $t = 12$

This results in the following objectives:

- Maximum deviation from the steady state up to and including time step $t = 12$
- Standard deviation of the time series

These two objectives should be minimized.

Again, including these two objectives is assumed to be sufficient for this thesis, in which we show the application of EMA methods for DSGE models. Based on the research statement or the policy goal, the objectives can be adjusted. For example, think of adding a threshold or taking the average of the three most extreme values (which should be either all positive or all negative).

Maximum deviation from the steady state up to and including time step $t = 12$
Taking the maximum deviation from steady state, positive or negative, we get an idea of the impact of the shock on the economy. Because we want this impact to be small, we search for a policy that minimizes this objective.

Standard deviation of the time series between $t = 0$ and $t = 12$

The standard deviation is a standard way to assess the variance of data. It is calculated by taking the distance between the value of the time series at each time step and the mean of the time series.

D Experiment set up explanation

In the EMA-workbench, we can set up uncertainties, levers, scenarios, and levers. Policies are different configurations of the levers, and scenarios are different configurations of the uncertainties.

When applying EMA, we have to give the model specific model characteristics. These characteristics are established through the XLMR-framework, as presented in Section 2.4.

- **Model uncertainties**: The input parameters that are uncertain fall under the model uncertainties. Per model uncertainty, a range is specified in which these parameters might fall. This way, the uncertainty space is created. In our open exploration, we also include levers in the model uncertainties, because we want to explore the effect of the levers on the outcomes of the model.
- **Uncertainty space**: The different model uncertainties create the uncertainty space: an (imaginative) space that represents all the SOW we assume are possible based on the uncertainty ranges. A visualization of the uncertainty space created by three uncertainties is presented in Figure D.1. The ranges we expect uncertainties u_1,u_2, and u_3 are in are, respectively, $(0, 3), (0, 3.5)$, and $(0, 4)$.
- **Scenario**: A scenario is a point in the uncertainty space. It is, thus, a combination of values for the model uncertainties. We can define scenarios, or randomize scenarios. Figure D.1 shows potential scenarios in the uncertainty space.
- **Policy levers**: The levers represent the actions that are in control of the decision maker. They can change these policy levers and, by that, affect the behavior of the model. Just like for the model uncertainties, we include a range of values per policy lever, creating a decision space.
- **Decision space**: The different policy levers create the decision space: an (imaginative) space that representing the ways the policy maker is able to act.
- **Policy**: A policy is a point in the decision space. It is a combination of values for the policy levers, which, like the scenarios, can be specified, or randomized.
- **Experiment**: An experiment is a combination of a scenario and a policy. If we want to analyse the effect of 1 policy for 1000 scenarios, 1000 experiments will be conducted.
- **Model outcomes**: The model outcomes is the data that we get out of the model. We have to specify what data we want to get out of the model to analyse, like the time series for inflation. It is possible to add whether you want to minimize or maximize outcomes, like the inflation rate at time $t = 12$. This is done in the directed search.
- **Outcome space**: The outcomes from all experiments together form the outcome space, bounded by the minimum and maximum values for each outcome. Unless logical boundaries exists for the outcomes (for example in case an outcome cannot be negative), the exact outcome space is unknown. In order to get the best estimate, a large number of experiments is required.
- **Model objectives**: The model objectives can be considered the quantification of the different goals existing in the policy problem. As the problems are usually complex, trade-offs might exist in the different model objectives.
- **Objective space**: The different objectives form the objective space.

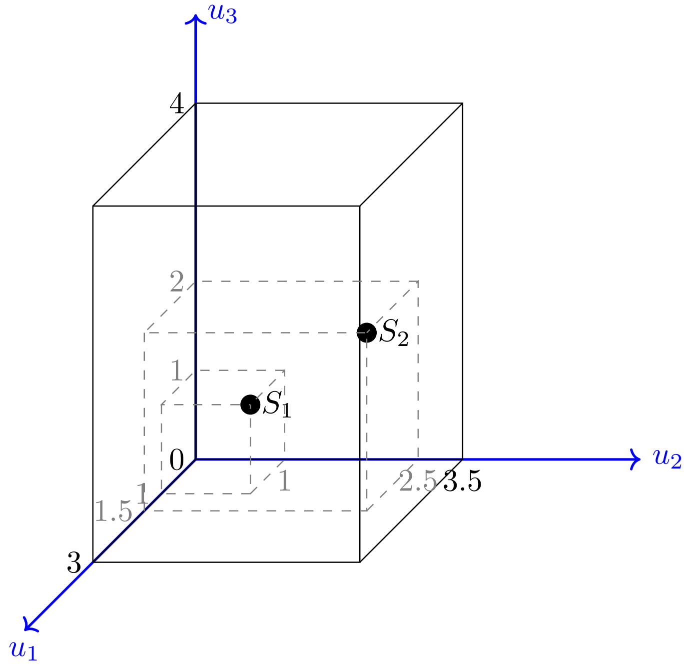


Figure D.1: Visualization of the uncertainty space created by three theoretical continuous uncertainties u_1 (range (0,3)), u_2 (range (0,3.5)), and u_3 (range (0,4)). Two possible scenarios are shown: $S_1(u_1 = u_2 = u_3 = 1)$, and $S_2(u_1 = 1.5; u_2 = 2.5; u_3 = 2.$

In Figure D.2 we include policies, whereas Figure D.3 shows the setup without policies.

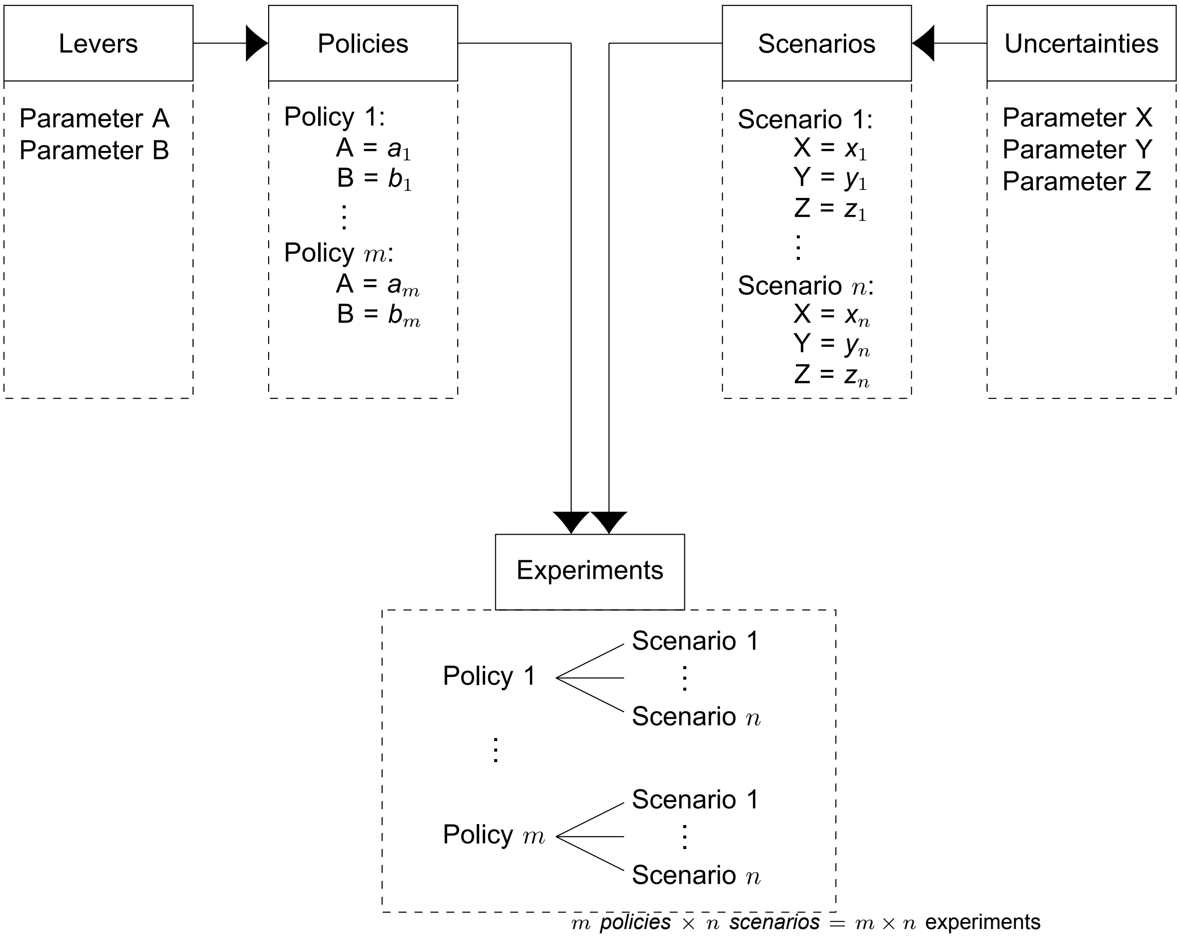


Figure D.2: A system with levers A and B and uncertainties X, Y, and Z. For the experiments, this configuration has m policies with different values for A and B and n scenarios with different values for X, Y, and Z. This results in $m \times n$ experiments.

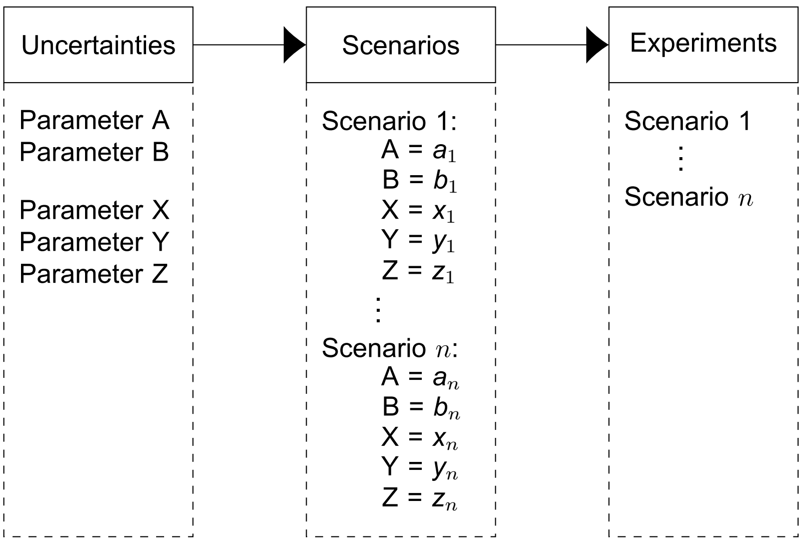


Figure D.3: No policies are defined, thus the levers (A and B) are considered to be uncertainties, like X, Y, and Z. n scenarios with n different values for A, B, X, Y, and Z result in n experiments.

E

Performance metrics multi-objective optimization algorithm

By evaluating proximity, convergence, diversity, and consistency, we can consider the fitness of the approximation sets (Salazar et al., 2016).

- **Proximity** refers to the (minimum) distance between solutions in an approximation set and the reference set.
- **Convergence** is the gap between the candidate solutions and the true Pareto Optimal Set, which we want to minimize (Jiang & Yang, 2016).
- **Diversity** of the approximate set of solutions represents the distribution of all Pareto solutions across the objective space. Ideally, the Pareto front would span across the whole objective space, representing the full set of trade-offs. We thus want to maximize diversity (Jiang & Yang, 2016).
- **Consistency** refers to Pareto approximate sets that capture all portions of trade-offs (Salazar et al., 2016).

These properties are quantified through performance metrics: generational distance, epsilon-indicator, hypervolume, inverted generational distance, and spacing, which are implemented in the EMW.

Salazar et al. (2016) has clearly explained these metrics, which is used for this thesis as well.

- **Generational distance** is a good metric for the *proximity* between the different reference set and the approximation set. It finds the minimum point in the average Euclidean distance vector between solutions in an approximation set and its corresponding nearest solutions in the reference set in the objective space. We want to minimize this distance, as a large value for the generational distance implies that none of the solutions in the approximation set is near the reference set. This general distance is conceptually easy, but comes with a few disadvantages: it only shows the distance and does not state anything about, for example diversity, and the generational distance can be heavily influenced by a single point if it is located distinctly closer to the reference set compared to the other points in the approximation set.
- **Inverted generational distance** is considered an improvement over the generational distance. Bezerra et al. (2017) states that the principle is similar to the generational distance, but the order of the fronts is reversed. They state that this metric is equal to the generational distance, but instead computes 'the distance between each objective vector in the reference front and its closest objective vector in the approximation front, averaged over the size of the reference front.'
- **epsilon-indicator** represents the value of the largest path the approximation set has to pass in the objective space to perform better on all objectives. Hence, it shows how well all trade-offs are presented in the approximation set, or, as previously described, the *consistency*. If the trade-offs are not as well presented as the reference case, thus, if a point in the objective space lies further away from the actual Pareto front than the reference set, the distance is large. A pitfall of this metric is in the case gaps in the Pareto approximation set (because we only consider (randomized) vectors and not define a plane in the objective space, it is possible that there is a gap between

vectors). In this case, the distance through which certain points in the approximation set have to pass is larger than it actually should be. The additive epsilon-indicator becomes larger than it would have been without the gap.

- **Hypervolume** is an indicator that can be used to evaluate the diversity and the proximity of the approximation set. It quantifies the volume of the objective space dominated by an approximation set. How this is done is explained further in the future/visualization. We want to maximize this metric as this would mean that the approximation set is largely distributed across the space.
- **Spacing** is, as described by Yen and He (2013), a measure for 'how evenly the non-dominated solutions are distributed along the approximation front.'

These three metrics should be either minimized (generational distance and additive epsilon-indicator) or maximized (hypervolume). As Figure E.1 shows, no abnormalities are seen in the performance metrics of the convergence.

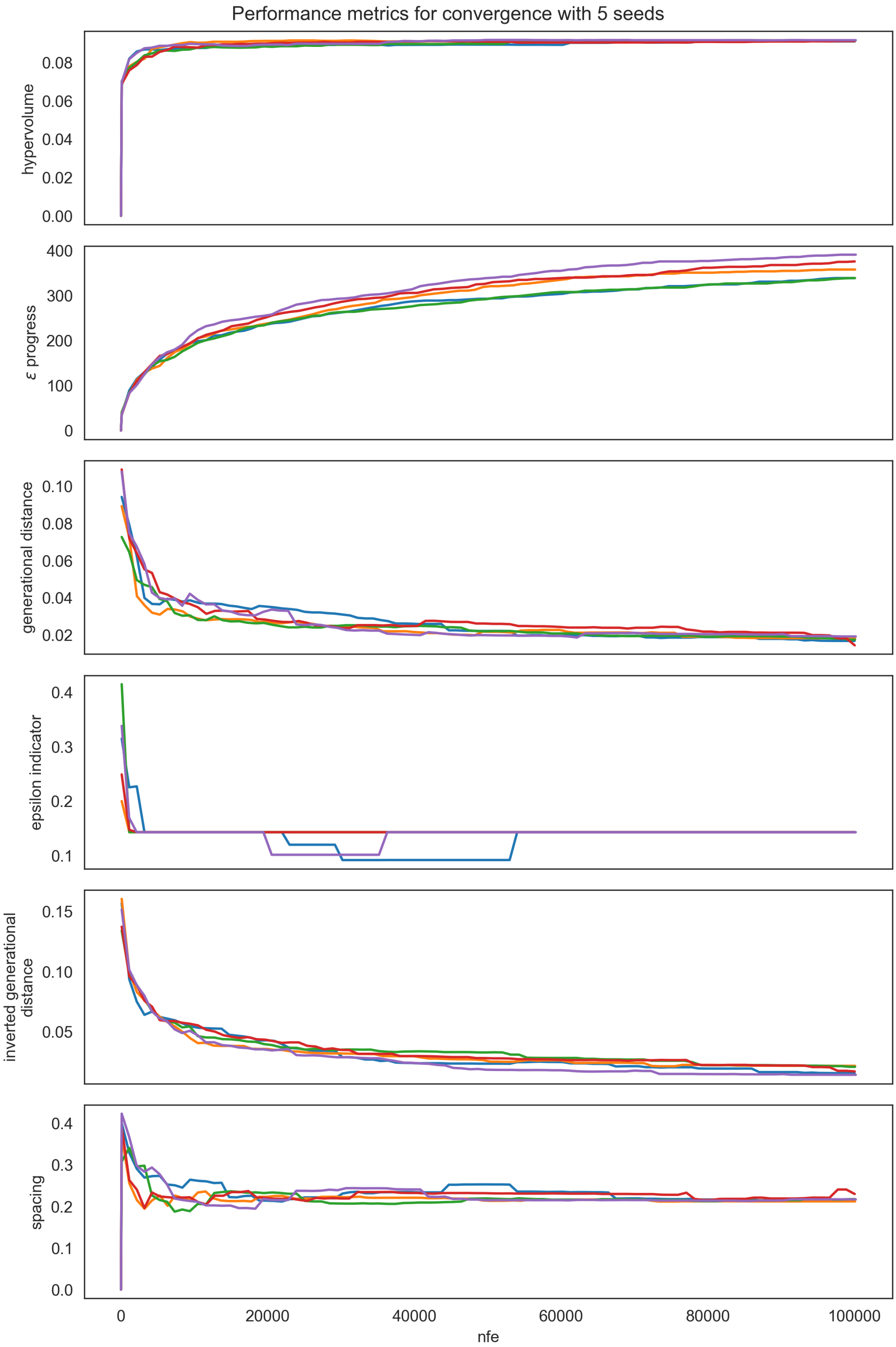


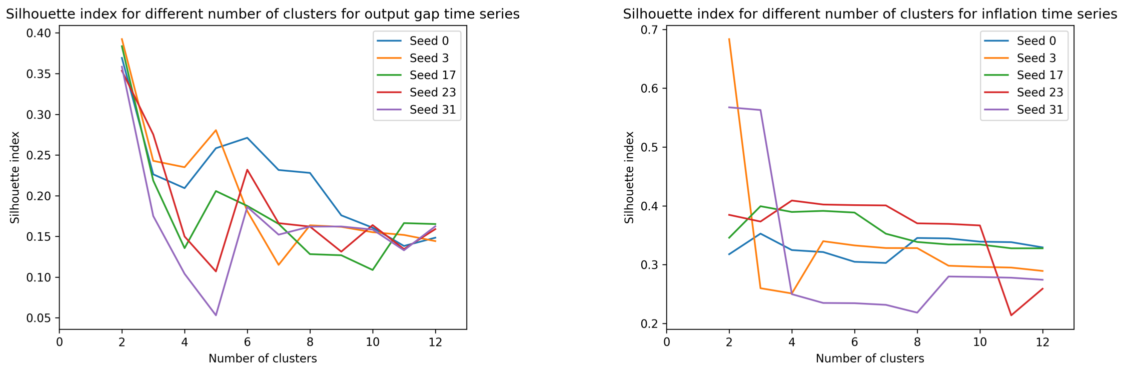
Figure E.1: Performance metrics for convergence

F

Results Open Exploration - extended

F.1. Behavior based scenario discovery using time series clustering

The first step in our behavior based scenario discovery using time series clustering, we have to decide upon the number of clusters we are interested in. This is done using the Silhouette Index (SI) to see which number of clusters is the best to do our analysis with. Because this cannot be done on our final experiment run with 15000 experiments, we take five experiment runs with 2000 experiments. Visually, in Figure F.1, we see that there is not a clear number of clusters that has the best SI.



(a) Silhouette indices of the output gap time series clustering.

(b) Silhouette indices of the inflation time series clustering.

Figure F.1: Silhouette scores for output gap time series and inflation time series for 5 (seeds 0, 3, 17, 23, 31) experiment runs with each having 2000 experiments. No clear

In Table F.1 and Table F.2 we see the number of experiments per cluster for each amount of clusters and the seed and the accompanying SI. In the range 4 to 12, we see that there is an increase in SI once the biggest cluster is divided into smaller clusters. Per seed the row with the highest SI in the range 4+ clusters is in cursive. This differs per seed. We therefore decide to check the number of experiments per cluster for different the numbers of clusters for our final experiment. We focus on the number of clusters in which the biggest cluster has clearly reduced compared to the biggest cluster with one less number of clusters. The number of experiments per cluster is shown in Table 4.1 for the output gap and in Table 4.4 for the inflation time series clustering.

F.2. Visualization of clusters of interest output gap - supply shock

These clusters of interest are visualized to see which cluster describes which types of behavior. In these figures, all experiments are shown with the grey time series and the cluster of interest is in a specific color. The order of the clusters is based on the number of experiments, which cluster 1 having the most experiments and the last one having the least amount of experiments in the cluster.

In Figure F.2, we see the 2 clusters when we cluster the time series into two clusters. The main difference between the two clusters is the size of the deviation. The time series in cluster 2 are have a smaller deviation from the steady state than cluster 1. There is not a distinct difference in types of behavior (like oscillation) between the two clusters.

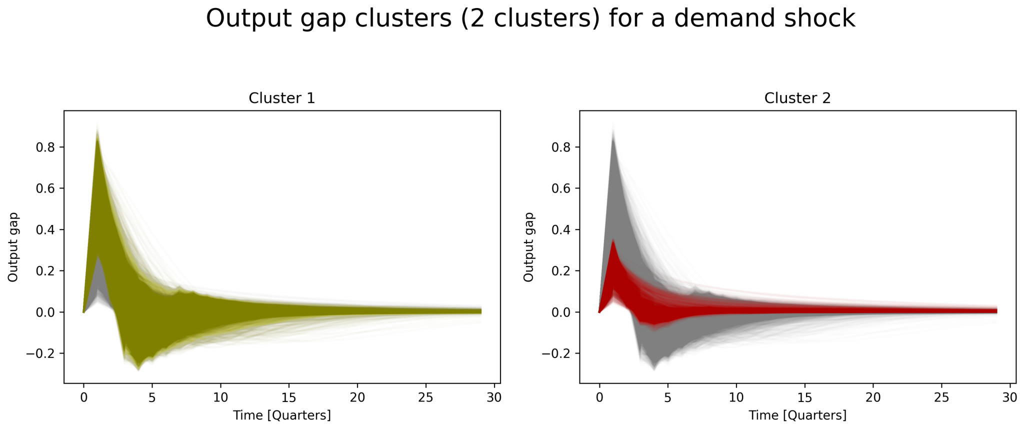


Figure F.2: Visualization of the 2 clusters of interest when the output gap time series is clustered into 2 clusters.

In Figure F.3 the 3 clusters of interest are shown when the output gap time series is clustered into 6 clusters. Like the 2 clusters in a 2 clustering approach, there is not a clear difference in type of behavior between the 3 clusters. The clusters differ mainly in the size of the deviation from the steady state with cluster 1 having the biggest deviations and cluster 3 having the smallest deviations.

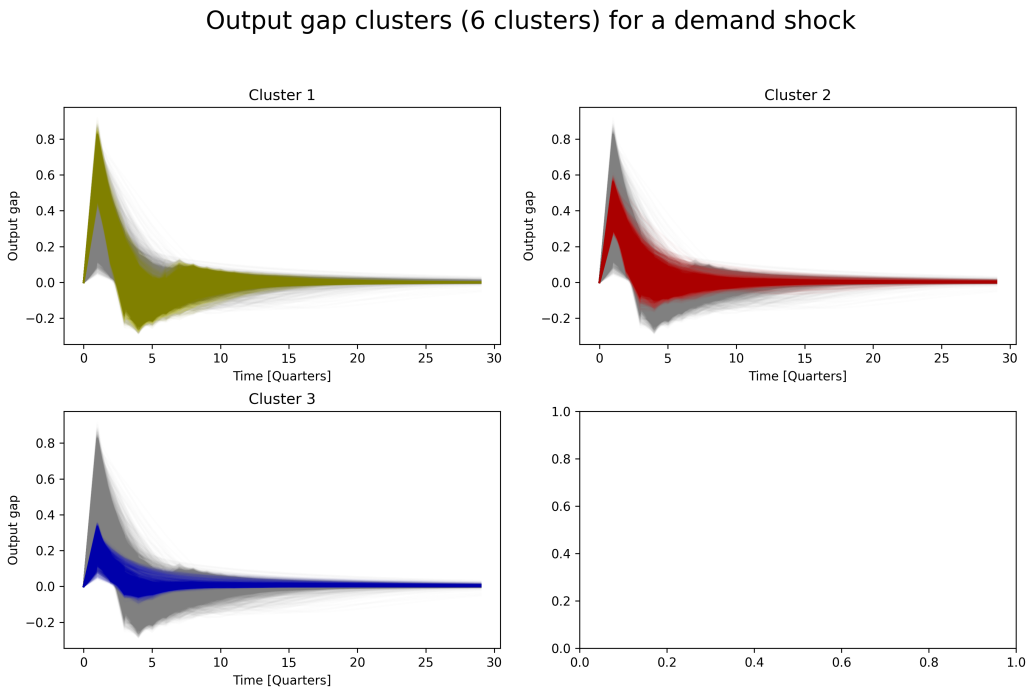


Figure F.3: Visualization of the 3 clusters of interest when the output gap time series is clustered into 6 clusters.

In Figure F.4 the 4 clusters of interest are shown when the output gap time series is clustered into 7 clusters. Compared to the 2 and 6 clusters approach, we now can distinguish differences in behavior between the clusters. Whereas clusters 1 to 3 are still relatively similar to the clusters in the clustering with a total of 6 clusters and mainly differ in their deviation from the steady state, cluster 4 clearly shows a different type of behavior. Cluster 4 includes the experiments that have the clearest oscillation: moving from the highest peak to the most negative deviation and oscilating back to the steady state.

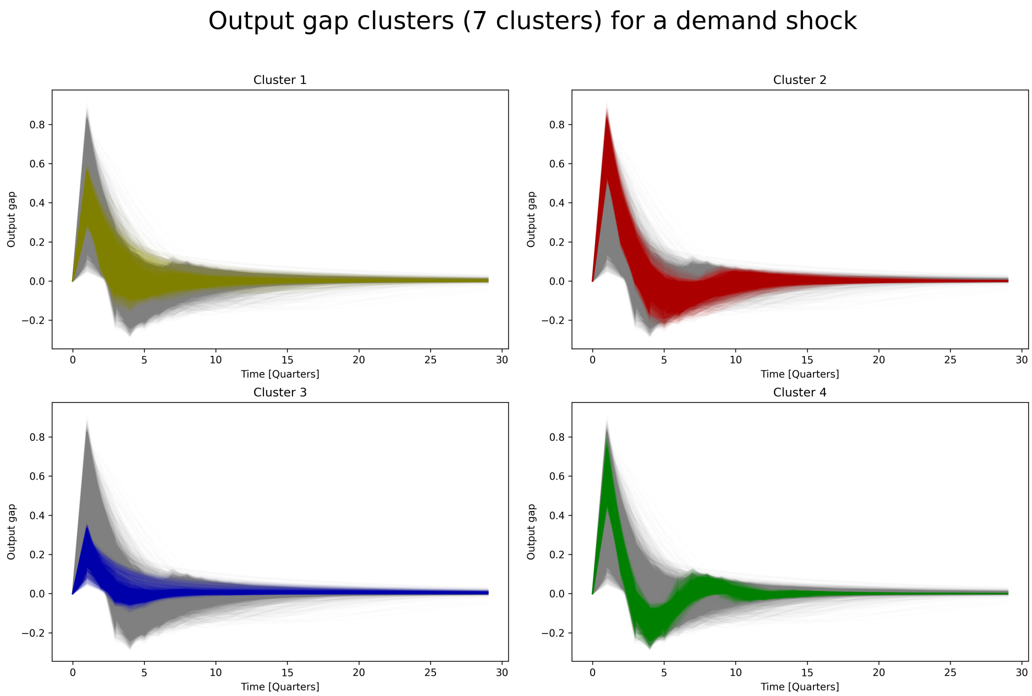


Figure F.4: Visualization of the 4 clusters of interest when the output gap time series is clustered into 7 clusters.

F.3. PRIM -output gap - demand shock

We check the peeling and pasting trajectory for each cluster and decide on the most suitable one. The boxes we are interested in are the ones with the highest density for each number of restricted dimensions. Every added number of restricted dimension means an extra parameter is used to describe the subset of experiments that are of interest. Considering the density, coverage, parameter range and the corresponding p-values, we decide on the best description for the each cluster. In these tables, the parameter is given and the range that is assigned to this parameter in the model. For each box, a range is given to the parameters in which the parameters are and asterisks are given to indicate the significance of the upper and lower bound. The asterisk next to the number of restricted dimensions indicates whether this is the box we decide to use or not. The coverage decreases as we increase the number of restricted dimensions, whereas the density increases as well. For the output gap clustering for the demand shock, the steps are elaborately described in the main text. For the other clustering, the main text will only include the final box descriptions for each cluster of interest.

- For cluster 1, experiments with the average behavior and medium deviation, Table F.3 describes the boxes of interest with 1 to 5 restricted dimensions. We see that the smoothing parameter is the first parameter to describe the box. The range of this smoothing parameter narrows down as more parameters are added The upper bound stays significant with a p-value smaller than 0.01. The step from 3 to 4 restricted dimensions clearly narrows down the ranges of the parameters that are seen in with 3 restricted dimensions. However, we also see a clear drop in coverage, going from 0.916 to 0.168, and an increase in density, going from 0.362 to 0.828. The added dimension with 5 restricted dimensions, does not provide extra insights and hardly change the coverage and density. The range given to the extra parameter, (the inverse Frisch elasticity of labor supply) fully coincides with the range that is given to the model. Therefore, we decide to go with the box description with 4 restricted dimensions to describe the experiments that are in cluster 1.

- For cluster 2, of which the boxes are described in Table F.4, we see a big drop in coverage when going from 3 to 4 restricted dimensions. The ranges of the parameters given for the 3 restricted dimensions box narrow down when going from 3 to 4 restricted dimensions. Because the density is already relatively high for 3 restricted dimensions, we decide to go with the box with 3 restricted dimensions for cluster 2.
- Table F.5 describes the boxes for cluster 3. The coverage decreases a lot when going from going 2 to 3 and from 3 to 4 restricted dimensions, and density increases a lot from 2 to 3 restricted dimensions. The ranges do narrow down when adding restricted dimensions, but this is less than we have seen for the boxes described in Table F.3. In this case we have to balance the drop coverage and the narrowed parameter ranges and increase in density when going from 3 to 4 restricted dimensions. Because the density is already relatively high and the ranges are not narrowing down enough to justify the drop in coverage, we decide to go with the box with 3 restricted dimensions to describe the experiment included in cluster 3.
- For cluster 4 of which the descriptions of the boxes are shown in Table F.6, we see the same effect as for the boxes describing cluster 3. However, in this case, the narrowing down of the ranges, the increase in density, and the fourth parameter having a significant upper bound that is not near the range given to the parameter in the model, justifies the drop in coverage. Therefore, in this case, we decide to go with the box with 4 restricted dimensions.
- In Table F.7 we can see that, for cluster 5, the two added parameters when going from 3 to 5 restricted dimensions not only do not have a significant upper or lower bound, they also fully coincide with the given range to the model. The ranges of the parameter do not change when we go from 3 to 5 restricted dimensions. Therefore, we assume the box with 3 restricted dimensions to be the best description of the experiments in cluster 5.
- For cluster 6, as described in Table F.8, the boxes with 1 to 3 restricted dimensions have a coverage of 1, but a density of less than 0.1. The box with 4 restricted dimensions is clearly the best, with ranges that do not almost fully coincide with given ranges.

This brings us to the final table, Table 4.3, which includes the box for each cluster that best describes the experiments in those clusters.

Table F.1: Number of experiments per cluster for the output gap time series clustering for different seeds and different number of clusters and the accompanying silhouette index (SI). The number of clusters (4+ clusters) of each seed with the highest SI are in italic.

Seed	nr of clusters	Cluster number												Total	SI
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*		
	2	1619	381											2000	0.372474
	3	1595	381	24										2000	0.195523
	4	1595	381	23	1									2000	0.182093
	5	1030	565	381	23	1								2000	0.216501
	6	1030	536	381	29	23	1							2000	0.171059
0	7	1030	536	381	29	19	4	1						2000	0.151692
	8	818	536	381	212	29	19	4	1					2000	0.165228
	9	*536*	*497*	*381*	*321*	*212*	*29*	*19*	*4*	*1*				*2000*	*0.222490*
	10	536	497	321	275	212	106	29	19	4	1			2000	0.190025
	11	536	497	321	275	111	106	101	29	19	4	1		2000	0.185141
	12	536	497	321	275	111	106	101	29	19	3	1	1	2000	0.167391
	2	1674	326											2000	0.338813
	3	1674	325	1										2000	0.278595
	4	1613	325	61	1									2000	0.091649
	5	*1194*	*419*	*325*	*61*	*1*								*2000*	*0.202379*
	6	1194	419	325	49	12	1							2000	0.161412
3	7	1050	419	325	144	49	12	1						2000	0.127980
	8	701	419	349	325	144	49	12	1					2000	0.161968
	9	701	349	325	216	203	144	49	12	1				2000	0.165315
	10	701	349	325	216	203	144	49	10	2	1			2000	0.159574
	11	701	325	222	216	203	144	127	49	10	2	1		2000	0.155243
	12	701	233	222	216	203	144	127	92	49	10	2	1	2000	0.146784
	2	1606	394											2000	0.353229
	3	1591	394	15										2000	0.176587
	4	807	784	394	15									2000	0.246232
	5	784	575	394	232	15								2000	0.211205
	6	575	400	394	384	232	15							2000	0.234864
17	7	575	400	384	269	232	125	15						2000	0.184547
	8	575	400	384	269	232	125	14	1					2000	0.179748
	9	*1855*	*120*	*12*	*5*	*2*	*2*	*2*	*1*	*1*				*2000*	*0.764531*
	10	1855	120	12	5	2	2	1	1	1	1			2000	0.764201
	11	1855	113	12	7	5	2	2	1	1	1	1		2000	0.752468
	12	1855	113	12	7	4	2	2	1	1	1	1	1		0.751845
	2	1364	636											2000	0.396938
	3	1343	636	21										2000	0.236056
	4	1343	470	166	21									2000	0.137964
	5	1343	465	166	21	5								2000	0.084584
	6	1343	465	166	20	5	1							2000	0.077679
23	7	1051	465	292	166	20	5	1						2000	0.140548
	8	1051	465	256	166	36	20	5	1					2000	0.129198
	9	*656*	*465*	*395*	*256*	*166*	*36*	*20*	*5*	*1*				*2000*	*0.151868*
	10	656	465	395	256	139	36	27	20	5	1			2000	0.133156
	11	656	465	395	256	139	36	27	14	6	5	1		2000	0.125104
	12	656	465	395	139	129	127	36	27	14	6	5	1	2000	0.123829
	2	1582	418											2000	0.351682
	3	1569	418	13										2000	0.163747
	4	1119	450	418	13									2000	0.254465
	5	1094	450	418	25	13	0							2000	0.165060
	6	1094	418	409	41	25	13							2000	0.156601
31	7	1094	418	409	41	25	12	1						2000	0.136859
	8	1094	409	306	112	41	25	12	1					2000	0.085971
	9	*598*	*496*	*409*	*306*	*112*	*41*	*25*	*12*	*1*				*2000*	*0.165409*
	10	598	496	406	306	112	41	25	12	3	1			2000	0.140961
	11	598	496	406	306	112	41	25	10	3	2	1		2000	0.138604
	12	598	496	306	270	136	112	41	25	10	3	2	1	2000	0.150573

Table F.2: Number of experiments per cluster for the inflation time series clustering for different seeds and different number of clusters and the accompanying silhouette index (SI). The number of clusters (4+ clusters) of each seed with the highest SI are in italic.

Seed	**nr of clusters**	**Cluster number**												**Total**	**SI**
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*		
0	2	1681	319											2000	0.569043
	3	1681	317	2										2000	0.564758
	4	*1129*	*552*	*317*	*2*									*2000*	*0.316775*
	5	1129	552	317	1	1								2000	0.314030
	6	1129	552	278	39	1	1							2000	0.308105
	7	1125	552	278	39	4	1	1						2000	0.281916
	8	1125	552	278	35	4	4	1	1					2000	0.272191
	9	1125	552	209	69	35	4	4	1	1				2000	0.265144
	10	641	552	484	209	69	35	4	4	1	1			2000	0.271760
	11	641	552	484	209	69	30	5	4	4	1	1		2000	0.265026
	12	641	552	484	209	52	30	17	5	4	4	1	1	2000	0.261105
3	2	1159	841											2000	0.484194
	3	1159	825	16										2000	0.474182
	4	1159	825	13	3									2000	0.466198
	5	*1159*	*672*	*153*	*13*	*3*								*2000*	*0.467893*
	6	1159	672	148	13	5	3							2000	0.460010
	7	672	602	557	148	13	5	3						2000	0.319112
	8	672	602	557	148	8	5	5	3					2000	0.316632
	9	612	602	557	148	60	8	5	5	3				2000	0.287723
	10	612	602	557	118	60	30	8	5	5	3			2000	0.279766
	11	612	602	557	118	60	30	8	5	4	3	1		2000	0.279078
	12	602	557	499	118	113	60	30	8	5	4	3	1	2000	0.295240
17	2	1982	18											2000	0.683804
	3	1384	598	18										2000	0.243034
	4	1383	598	18	1									2000	0.233823
	5	1383	598	17	1	1								2000	0.230224
	6	937	598	446	17	1	1							2000	0.334656
	7	937	598	422	24	17	1	1						2000	0.327103
	8	937	598	422	24	14	3	1	1					2000	0.321060
	9	*1922*	*58*	*10*	*4*	*2*	*1*	*1*	*1*	*1*				*2000*	*0.802994*
	10	1922	38	20	10	4	2	1	1	1	1			2000	0.775681
	11	1922	38	20	10	4	1	1	1	1	1	1		2000	0.775166
	12	1922	38	20	10	3	1	1	1	1	1	1	1	2000	0.774859
23	2	1999	1											2000	0.759391
	3	1009	990	1										2000	0.405291
	4	990	695	314	1									2000	0.408268
	5	990	685	314	10	1								2000	0.386190
	6	*990*	*685*	*284*	*30*	*10*	*1*							*2000*	*0.390803*
	7	990	685	218	66	30	10	1						2000	0.386645
	8	990	685	218	66	16	14	10	1					2000	0.383523
	9	685	594	396	218	66	16	14	10	1				2000	0.268770
	10	685	594	396	218	66	14	13	10	3	1			2000	0.265798
	11	600	594	396	218	85	66	14	13	10	3	1		2000	0.270646
	12	600	594	396	218	85	44	22	14	13	10	3	1	2000	0.264899
31	2	1261	739											2000	0.318368
	3	1132	739	129										2000	0.348788
	4	1132	739	127	2									2000	0.347616
	5	1132	739	126	2	1								2000	0.344194
	6	1132	738	126	2	1	1							2000	0.341287
	7	*738*	*602*	*530*	*126*	*2*	*1*	*1*	*0*					*2000*	*0.347909*
	9	738	579	495	126	35	23	2	1	1				2000	0.331933
	10	738	579	495	103	35	23	23	2	1	1			2000	0.324369
	11	738	579	495	103	35	23	21	2	2	1	1		2000	0.316852
	12	738	579	495	103	23	21	20	15	2	2	1	1	2000	0.310842

Table F.3: Output gap, cluster 1

Parameter	**Range**	**Nr of restricted dimensions**				
		1	*2*	*3*	*4**	*5*
Smoothing parameter	[0.5 - 1]	[0.50 - 0.95***]	[0.5 - 0.93***]	[0.5 - 0.91***]	[0.50 - 0.73***]	[0.50 - 0.73***]
Calvo parameter	[0.6 - 0.95]		[0.60 - 0.93***]	[0.60 - 0.92***]	[0.74***-0.87***]	[0.74***-0.87***]
Autoregressive factor of demand shock	[0.5 - 0.95]			[0.5 - 0.91***]	[0.58*** - 0.77***]	[0.58*** - 0.77***]
Policy reaction to interest	[1 - 2.5]				[1.3*** - 2.5]	[1.3*** - 2.5]
inverse Frisch elasticity of labor supply	[0.6 - 1.5]					[0.6 - 1.5]
Coverage		0.994	0.974	0.916	0.168	0.162
Density		0.289	0.313	0.362	0.828	0.835

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Table F.4: Output gap, cluster 2

Output gap, ed shock	Range	**Nr of restricted dimensions**			
		1	*2*	*3**	*4*
Calvo parameter	[0.6 - 0.95]	[0.69*** - 0.95]	[0.75*** - 0.95]	[0.79*** - 0.95]	[0.84*** - 0.95]
Autoregressive factor of demand shock	[0.5 - 0.95]		[0.50 - 0.80***]	[0.50 - 0.73***]	[0.50 - 0.68***]
Smoothing parameter	[0.5 - 1]			[0.63*** - 1]	[0.73*** - 1]
Inverse marginal elasticity of substitution	[1 - 2.5]				[1.3*** - 2.5]
Coverage		0.986	0.814	0.548	0.199
Density		0.347	0.56	0.815	0.996

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Table F.5: Output gap, cluster 3

Output gap, ed shock	**Range**	**Nr of restricted dimensions**			
		1	*2*	*3**	*4*
Calvo parameter	[0.6 - 0.95]	[0.60 - 0.86***]	[0.60 - 0.81***]	[0.60 - 0.80***]	[0.60 - 0.72***]
Smoothing parameter	[0.5 - 1]		[0.65*** - 1]	[0.81*** - 1]	[0.86*** - 1.0]
Autoregressive factor of demand shock	[0.5 - 0.95]			[0.5 - 0.80***]	[0.50 - 0.74**]
Policy reaction to interest	[1 - 2.5]				[1 - 2.4]
Coverage		0.999	0.946	0.486	0.283
Density		0.235	0.392	0.850	0.968

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Table F.6: Output gap, cluster 4

Output gap, ed shock	**Range**	**Nr of restricted dimensions**			
		1	*2*	*3*	*4**
Smoothing parameter	[0.5 - 1]	[0.5 - 0.93***]	[0.5 - 0.93***]	[0.5 - 0.93***]	[0.5 - 0.75***]
Autoregressive factor of demand shock	[0.5 - 0.95]		[0.52*** - 0.95]	[0.52*** - 0.95]	[0.65*** - 0.83***]
Policy reaction to interest	[1 - 2.5]			[1.3*** - 2.5]	[1.6*** - 2.5]
Calvo parameter	[0.6 - 0.95]				[0.60 - 0.74***]
Coverage		0.998	0.995	0.797	0.283
Density		0.173	0.181	0.314	0.731

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Table F.7: Output gap, cluster 5

Output gap, ed shock	**Range**	**Nr of restricted dimensions**				
		1	*2*	*3**	*4*	*5*
Calvo parameter	[0.6 - 0.95]	[0.63*** - 0.95]	[0.63*** - 0.95]	[0.85*** - 0.95]	[0.85*** - 0.95]	[0.87*** - 0.95]
Autoregressive factor of demand shock	[0.5 - 0.95]		[0.67*** - 0.95]	[0.77*** - 0.90***]	[0.77*** - 0.90***]	[0.77*** - 0.90***]
Smoothing parameter	[0.5 - 1]			[0.5 - 0.89***]	[0.5 - 0.89***]	[0.5 - 0.89***]
Discount Factor	[0.95 - 1]				[0.95 - 1]	[0.95 - 1]
Share of populaten (z)	[0.2 - 0.5]					[0.2 - 0.5]
Coverage		1.00	0.998	0.505	0.488	0.426
Density		0.0955	0.151	0.667	0.678	0.729

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

Table F.8: Output gap, cluster 6

Output gap, ed shock	**Range**	**Nr of restricted dimensions**			
		1	*2*	*3*	*4**
Calvo parameter	[0.6 - 0.95]	[0.6 - 0.92***]	[0.6 - 0.92***]	[0.6 - 0.92***]	[0.6 - 0.81***]
Policy response to inflation	[1 - 2.5]		[1.1 - 2.5]	[1.1 - 2.5]	[1.5*** - 2.5]
Smoothing parameter	[0.5 - 1]			[0.5 - 0.89***]	[0.5 - 0.75***]
Autoregressive factor of demand shock	[0.5 - 0.95]				[0.84*** - 0.95]
Coverage		1.00	1.00	1.00	0.639
Density		0.063	0.0663	0.0857	0.721

* indicates a p-value smaller than 0.05
** indicates a p-value smaller than 0.025
*** indicates a p-value smaller than 0.01

F.4. inflation - demand shock

Clustering into 2 clusters gives us the 2 clusters as shown in Figure F.5. We see that the behavior of cluster 2 is more extreme.

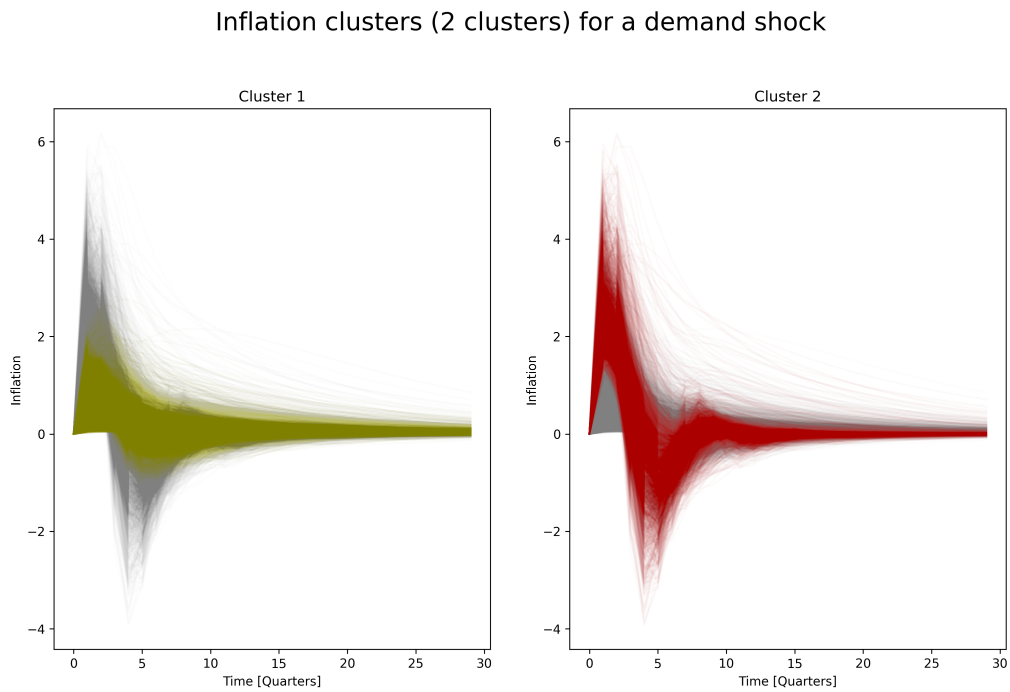


Figure F.5: Visualization of the 2 clusters when the inflation time series of a demand shock is clustered into 2 clusters.

F.5. Supply shock

Figure F.6 shows the clusters when we cluster the output gap into 2 clusters. We see that this is not yet insightful.

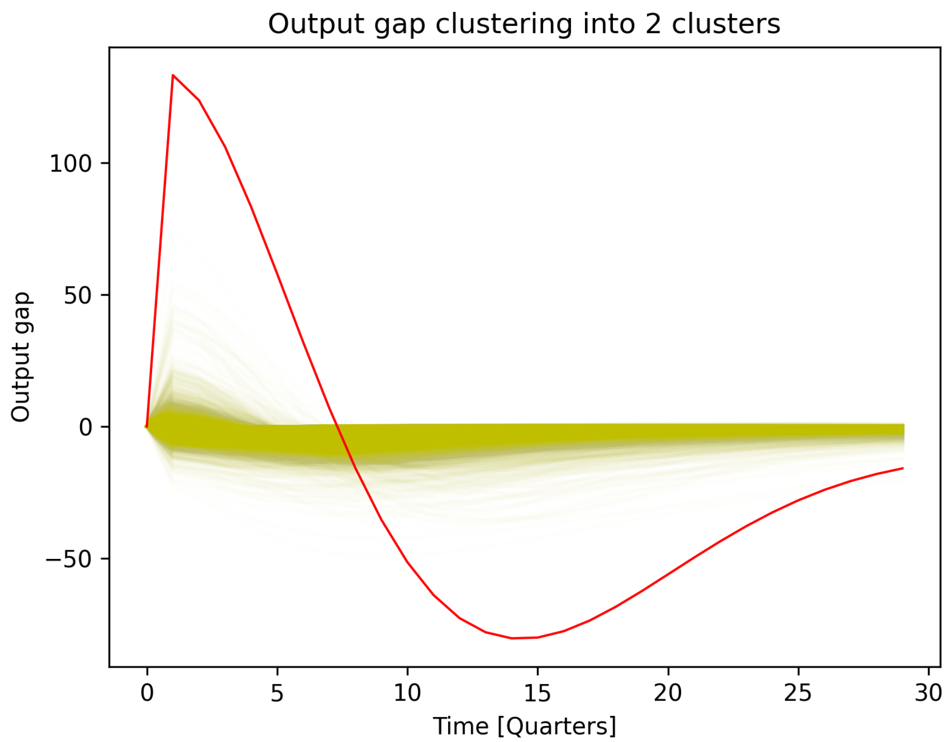


Figure F.6: Clusters when output gap time series are clustered into two cluster. The yellow time series is the large cluster and the red line is the single time series in the second cluster.

When the inflation time series are clustered into 2 clusters, the clusters shown in Figure F.7 come out. The second cluster shows distinct behavior from cluster 1 by having the highest peaks.

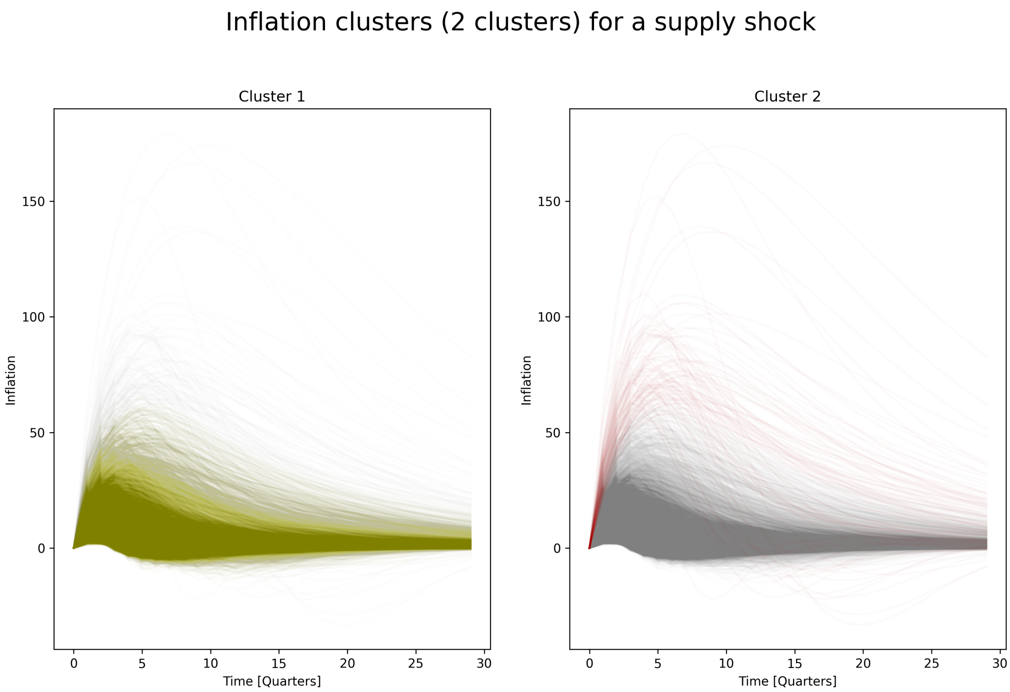


Figure F.7: The 2 clusters when inflation time series are clustered into 2 cluster.

F.6. Clustering on an adjusted dataset

Because the clustering algorithm sticks with one large cluster, we removed some of the outliers. The first time series that are dropped, are the experiments that are clustered in the 5 smallest clusters when we cluster the time series into 12 clusters. Note that this results in a different adjusted dataset for the output gap and inflation.

The number of experiments per cluster that for this adjusted dataset are shown in Table F.9 for the output gap time series and in Table F.10 for the inflation time series.

Table F.9: Number of experiments per cluster for output gap time series clustering for the experiment run with 14993 experiments

Number of clusters	Cluster number												Total
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	14542	451											14993
3	14542	426	25										14993
4	14542	267	159	25									14993
5	14542	233	159	34	25								14993
6	11668	2874	233	159	34	25							14993
7	11668	2874	233	159	34	17	8						14993
8	11668	2874	233	159	28	17	8	6					14993
9	11668	2874	233	159	28	17	6	5	3				14993
10	11668	2874	233	132	28	27	17	6	5	3			14993
11	11668	2874	233	132	28	27	10	7	6	5	3		14993
12	11668	2874	233	132	28	27	10	7	6	4	3	1	14993

Table F.10: Number of experiments per cluster for inflation time series clustering for the experiment run with 14993 experiments

Number of clusters	Cluster number												Total
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	14889	104											14993
3	14889	83	21										14993
4	14889	76	21	7									14993
5	12443	2446	76	21	7								14993
6	12443	2446	76	13	8	7							14993
7	12443	2446	49	27	13	8	7						14993
8	12443	2001	445	49	27	13	8	7					14993
9	12443	2001	445	49	27	13	7	7	1				14993
10	12443	2001	445	41	27	13	8	7	7	1			14993
11	12443	1892	445	109	41	27	13	8	7	7	1		14993
12	12443	1892	445	109	41	27	13	8	7	6	1	1	14993

The adjusted clustering still shows that we end up with a big cluster and at most a second cluster of interest. We adjust the clustering again, this time focusing on just the biggest cluster when we cluster into 2 clusters.

This resulted in the following number of experiments per cluster.

Table F.11: Number of experiments per cluster for inflation time series clustering for the experiment run with 14889 experiments

Number of clusters	**Cluster number**												**Total**
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	11668	2874											14889
3	11668	2504	370										14889
4	11668	2165	370	339									14889
5	11668	2165	370	310	29								14889
6	11668	2165	343	310	29	27							14889
7	11668	2165	314	310	29	29	27						14889
8	11668	2165	314	214	96	29	29	27					14889
9	11668	2165	222	214	96	92	29	29	27				14889
10	11668	2165	222	214	96	96	29	27	19	10			14889
11	11668	1450	715	222	214	92	92	29	27	19	10		14889
12	11668	1450	715	222	214	92	76	29	27	20	19	10	14889

Table F.12: Number of experiments per cluster for output gap time series clustering for the experiment run with 14542 experiments

Number of clusters	**Cluster number**												**Total**
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	
2	11668	2874											14542
3	11668	2504	370										14542
4	11668	2165	370	339									14542
5	11668	2165	370	310	29								14542
6	11668	2165	343	310	29	27							14542
7	11668	2165	314	310	29	29	27						14542
8	11668	2165	314	214	96	29	29	27					14542
9	11668	2165	222	214	96	92	29	29	27				14542
10	11668	2165	222	214	96	92	29	27	19	10			14542
11	11668	1450	715	222	214	96	92	29	27	19	10		14542
12	11668	1450	715	222	214	92	76	29	27	20	19	10	14542

Also with this second adjustment, our biggest clusters include most of the experiments and the other clusters are not likely to result in significant results.

For the output gap, the visualisation of the 5 biggest clusters are shown in Figure 4.11. We do see different behaviors for these clusters. However, when PRIM is applied, the density is only above 0.8 for the first cluster.

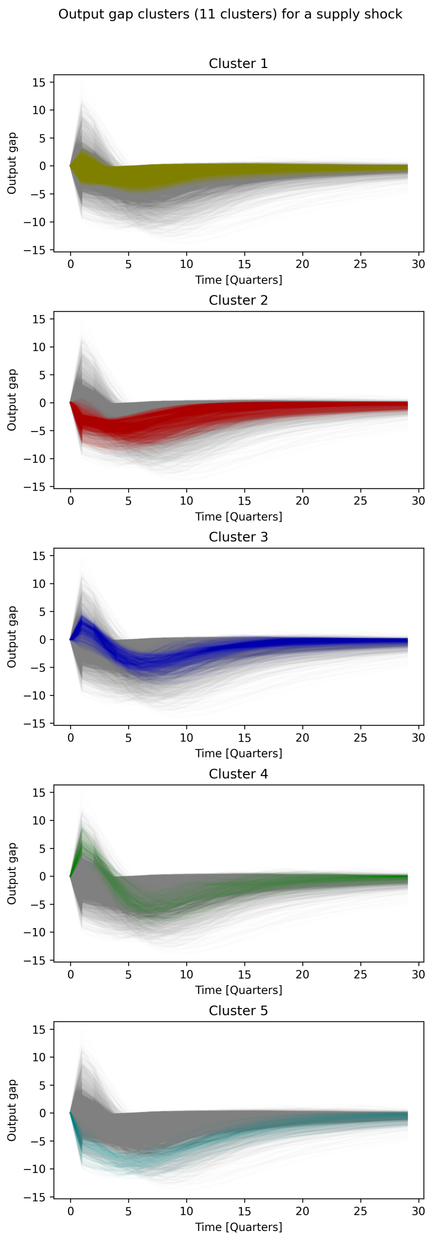


Figure F.8: 4 clusters of interest for adjusted output gap time series when clustering into 11 clusters.