Globally Optimal Monetary Policy

Christian Payne Hennigan

University of Colorado Boulder

November 9, 2021

Latest Version

Abstract

Optimal monetary policy is important on a practical level for central banks. Computational barriers, however, have limited research into important optimal monetary policy questions. With some important exceptions using computational techniques tailored to specific

1 Introduction

to move beyond simplifications themselves often have fatal issues, like being limited in incorporating an effective lower bound and having a hard time dealing with anything beyond a small number of variables in the model.[5] Even when looking at a small-scale New Keynesian Model, the degree of the nonlinearities involved and number of variables makes standard procedures difficult to apply. Deep learning techniques are convenient because they can approximate any unknown function[6], are less sensitive to good initial guesses, and largely avoid the curse of dimensionality (Kang et al 2020, Raissi 2018). This is crucial for them to be practical for other applications like the ones mentioned at the outset, since a central bank may have no idea what the function they are searching for looks like and often deal with large models. Other papers have used neural networks to help solve other economic models,[7] but to my knowledge this is the first paper to make use of these techniques for optimal monetary policy research.

This paper illustrates these techniques as applied to the question of how firm expectations and price distortions should jointly influence optimal monetary policy. Inspired by the literature on optimal policy in the lineage of Ramsey (1927) and Stokey and Lucas (1983), I will have the central bank maximize household welfare, constrained by the optimal choices made by firms and households in a decentralized economy. This allows analysis beyond a small neighborhood of the steady state, in a way synergistic with the computational approach discussed later.[8] To think of how firm expectations and price distortions may influence optimal monetary policy, I will use a small-scale New Keynesian model, since it quite elegantly includes both nominal rigidity through price dispersion and imperfect competition in the form of price staggering by monopolistically competitive firms.

What happens to these two distortions in applications? Usually there is an assumption of a production subsidy provided through a lump sum tax in order to eliminate the mark-up distortion from imperfect competition. When this is done, monetary policy is concerned with the only

[5] The most glaring examples of these problems are higher order pertubation methods being unable to accommodate inequality constraints like an effective lower bound (Swanson et al 2006) and discretation, as well as projection methods, succumbing to sensitivity to initial conditions and the curse of dimensionality.

[6] More specifically, any Borel measurable function, by the universal approximation theorem (Bach 2017)

[7] Recent applications include Scheidegger and Bilionis (2017), Duarte (2018), Maliar et al. (2019), Ferandez-Villaverde et al. (2020), and Azinovic et al. (2020).

[8] I also avoid convoluted construction inherent in traditional methods of deriving the objective function that makes comparisons even between very similar non-standard model environments and substandard policies dif-

remaining distortion, relative price dispersion. A significant amount of initial price dispersion is usually just assumed away (Gali 2003, Woodford 2003). With these assumptions, even with the countless extensions of the model, the base intuition of optimal monetary policy remains largely the same: the goal of policy should be to stabilize the price level to keep the economy at its natural level of output. In stabilizing the price level, a central bank stamps out the nominal distortions associated with sticky prices, reaching the flexible price equilibrium.[9] However, these are not so innocuous. They require us to already be near the long-run value, and to assume stabilization policy that we find optimal now to have already been used in the past. What happens though when, for whatever reason, we find ourselves stranded away from the long run? You could consider, for example that a central bank decides to change its inflation target. Even if everything else is optimal before and after besides the target, the bank has to decide the best transition path. You could also imagine a shock that is not predicted, such as a financial crisis or a large supply chain blockage, that shifts the environment enough that the bank has to think about transitioning back to "normal".

Some papers have broken these assumptions to study what would happen away from the long-run outcome. Yun (2005), for example, does not assume away initial price dispersion. Though he confirms the result that complete stabilization of the price level is optimal in the absence of initial price dispersion, he finds that optimal inflation targets respond to changes in the level of price distortion otherwise. Because price dispersion in a second order term, Yun cannot rely on a set of linearly approximated equilibrium conditions. He still maintains, however, a subsidy to eliminate the ine□ciency associated with monopolistic competition. By introducing an employment subsidy, he ends up with a one-to-one relationship between the growth rates of inflation and price dispersion. This drastically simplifies the first order conditions of the firm and he is able to drop all the forward looking parts of the model, which themselves evolve non-linearly. Yun then drastically simplifies the aggregate supply curve since only current costs matter. Although including price distortions while assuming an e□cient steady state gives clean analytical results even with non-linearity, including the forward-looking firm expectations that

cult (Benigno & Woodford, 2008).

[9] See, for example, Goodfried and King (1997), Rotemberg and Woodford (1997); Clarida, Gali, and Gertler

are necessary without that assumption makes this impossible.

To study both channels of distortion in this model, I directly include the nonlinear sources of ine□ciency, price dispersion and a distorted stead-state, as well as the forward-looking elements

and what follows was first done broadly Nakamura-Zimmerman 2020.[12] To get the optimal paths, I approximate every variable by a neural network with input being time (this is equivalent to thinking of the ODE system as a neural network with vector output, input still time). To train this neural network, I take all of the equilibrium conditions of the ODE, express them as $f(x) = 0$, and try to minimize with randomly chosen time points along the path. In other words, I try to get all of the neural network variables to run along the paths as close as possible to the path constraints while satisfying boundary conditions, slowly adjusting the networks to be more optimal with a batch of time points along the path. Because boundary conditions are not handled entirely well here (a common problem in the deep learning literature), I follow the procedure of Lagari et al (2020) to convert the problem into one of "hard boundaries". This is really just nesting the neural network into a temporary function of time to get the process started; it forces the boundary conditions to hold at all times.

I use and modify packages developed out of a team at Brown University (Deep XDE, Lu et al 2021) for solving ODE systems in the way I described. I get multiple paths by randomly selecting initial starting positions (you can think of these as different initial distortions). I collect 150 of these paths and I then train the interest rate as a neural network, this time as a function of the variables in the model. I am converting the open-form solution to a closed-form one. I do this by seeing what the relationships are between the paths. This, again, is similar to the procedure of Nakamura-Zimmerman (2020). Because I have commitment, this is valid, since nothing is a function of time (to use the language of optimal control theory, I have an autonomous solution). In other words, I am seeing what the interest rate would have to be in order to make the model evolve as it does. Again, all of this is necessary because it allows me to handle the non-linearities without a good initial guess. Other global methods fail here, since I have no idea what this value function looks like given that I do not make all the heavy assumptions on the analytical side. It also can handle broader problems because it is not held up by the curse of dimensionality in the same way. Even with my number of variables, things are very computational taxing.

I show, finally, that these two channels of inefficiency, price distortion and monopolistic com-

[12] They, however, focus on the linear-quadratic case so this also serves as an extension of their method.

petition, matter for optimal monetary policy, and, furthermore, that this latter channel can be expressed in terms of firm expectations of future costs. The extent to which one channel matters more than the other will depend on the parameters of the model and initial conditions. I find also that policy should be more aggressive when price dispersion increases. More broadly, in answering the question of how firm expectations and price distortions should influence optimal monetary policy, I am able to avoid the usual assumptions of a quadratic objective function and linear dynamics, without simplification for the sake of computation, and I move away from a view of monetary policy as stabilizing the economy around a long-run result in the face of small Gaussian shocks. In Section In qo22-15(he)-287(model)-286(and)-parea J 0 -23 on

2 Analytical Approach and Implications for Monetary Policy

2.1 The Model

In the following analysis, I will first illustrate the important elements and equilibrium conditions of the model environment, a nonlinear New Keynsian Model[13] making use of the staggered price-setting of Calvo (1983) with money only playing the role of a unit of account[14]. I will then use those equilibrium conditions to both set up the constraint set and rewrite the reward function in a convenient fashion for the central bank's problem. The policy prescription of the central bank - an optimal interest rate rule - will be inferred from the optimal dynamics of the decentralized economy. I will be freely taking from the results of the full underlying model, only reproducing the results as they are important for the central bank's problem. The full underlying model can be found in Appendix A. I set the modelling environments in continuous time, which helps me characterize much of the equilibrium dynamics analytically and to avoid needing to compute expectations even within a global solution.[15]

After the problem is posed, I will give the analytical results and their implication for monetary policy. I will show how the optimal dynamics should be thought of in terms of an interest rate policy that most optimally structures marginal cost as an equilibrium function of the state variables. Although the full result can only be determined numerically, I will demonstrate through analytical results that with a significant level of price dispersion that stabilization is not optimal in the short run and that a distorted stead state implies that an inflation target is not optimal, though I will confirm the general result of the monetary policy literature that stabilization around some value will be the optimal long run policy. Finally, I will numerically approximate the model solution using deep learning techniques. These techniques, though used for this specific case, can be much more widely applied than what is shown here.

13

2.2 Value Function

The objective function is exactly the consumer's utility function, though this no longer is expressed in terms of consumption and labor, but in terms rather of marginal utility and the product of price dispersion and marginal cost. These are derived from the equilibrium conditions of the underlying model. The central bank is thus trying to maximize:

$$\int_0^{\infty} e^{-rt} t \ln(c_t) - y \frac{n_t^{1+g}}{1+g}$$

Of course, in a decentralized economy consumption and labor are both functions of the underlying state variables. We will express them in that form. To do this, consider the first order conditions of the household. For any interior solution, and when $y > 0$:

$$1/c_t = l_t$$

$$y l_t^g = l_t w_t$$

$$y l_t^g c_t = w_t$$

We also have an expression for aggregate production:

$$y_t = \frac{A n_t}{v_t}$$

where v_t refers to price dispersion, defined mathematically as:

$$v_t = \int_0^1 \left(\frac{p_{it}}{p_t}\right)^{-\#} di$$

$$v_t \geq 1$$

at a much higher computational cost. See Achdou et al. (2017) for an explanation of the advantages of continuous time in this manner.

Note that $v_t = 1$ would imply e□ciency. Price dispersion acts as the point of ine□ciency flowing from staggered price setting. Price dispersion can also be thought of as a misalignment between decisions made on the basis of marginal cost and those made on the basis of marginal utility. In practical terms it acts as a wedge between production in terms of inputs and in terms of output after aggregation.

Combining together the first order and market conditions:

$$mc_t = y n_t^{1+g} / v_t$$

$$mc_t = y (A l_t)^{(1+g)} v_t^g$$

$$c_t = A\left(\frac{mc_t}{y v_t^g}\right)$$

where $mc_t = \frac{w_t}{A_t}$ refers to real marginal cost.

One important thing to take away here is that the policy functions of the representative agent - in this context, optimal consumption and labor as functions of the underlying state variables - can be analytically expressed in terms of of a function of price dispersion and marginal cost. This is a partial equilibrium result, and of course in the full New Keynesian model this will not be enough because firms also take into consideration expected future marginal costs as well as the current marginal cost when determining their prices. As we will see, the evolution of price dispersion is defined by a path constraint, but there is no such constraint on marginal cost. The equilibrium value for marginal cost will end up being an unknown function of the state variables. Really, marginal cost serves as a still unknown function with solves the equilibrium conditions and the maximized Bellman equation or Hamiltonian. (Fernandez-Villaverde et al (2012)).

The traditional approach is to linearize the equilibrium conditions and solve the system of

This is why our approach to solve for the general equilibrium values has been to augment the vector of state variables of the household's value function by the law of motions for expectations and price dispersion.

A quick summary of parameters:

r : Discounting Parameter

q : Rate of the Calvo Process, an exponential distribution

(1/ q gives the expected wait time until the next price change, q ¡ 0)

$\#$: Elasticity of Substitution

y : The Disutility of Labor

g : Inverse of Frisch Labor Supply Elasticity

My approach is di□erent from than of Yun (2005) because I do not subsidize employment. Because of this, I cannot ignore the forward-looking pricing equation of the firm:

$$\max_{p_{it}} E_t \int_t^{\infty} \frac{l_t}{l_t} e^{-q(t-t)} \left[\frac{p_{it}}{p_t} y_{it} - mc_t\, y_{it} \right] dt$$

The first order conditions of the firm is as follows. The ratio of the optimal new price, common across all firms able to reset their prices, and the prices of the final good, is given by:

$$\frac{p_{it}}{p_t} = \frac{\#}{\#\ 1}\frac{S_{Ct}}{S_{Rt}}$$

where:

$$S_{Rt} = \sum_{t}^{\infty} I_t e^{q}$$

that optimal structure, and then backing out the interest rate implied by the optimal dynamics of the system. It is important to remember that any other choice of interest rate not given by these implied dynamics will cause the dynamics to be different, and hence worse.

Let us break down equations (8) and (9) further. Equation (8) is the first order condition for inflation:

$$(1+p_t\frac{1-\theta}{q})^{\frac{1}{1-\theta}}\underbrace{m_p}_{①} = \underbrace{\lambda_{S_R}(\theta-1)S_{Rt}+\lambda_{S_C}\theta S_{Ct}}_{②}+\underbrace{\lambda_v\theta(1+p_t(1-\theta)/q)^{\frac{1}{1-\theta}}\lambda_v\theta v_t}_{③}$$

The weighted benefits from raising inflation ①, represented by the lagrange multiplier m_p, will have two effects. The first will be through the effect on the optimal firm pricing decision ②. Inflation causes expected real revenues and real costs to both decrease. Whether this will cause the reset price to go up or down will depend on the current values of S_R and S_C, which represent the weight on the reset price by expected future aggregate demand conditions and marginal costs, respectively. Inflation also has an impact on price dispersion ③. Generally, inflation causes price dispersion to increase, though the overall effect of this will depend on the current levels of inflation and price dispersion. Any model, including Yun (2005), which adds a subsidy to employment will lose the effects represented by ②.

Equation (9) is the first order condition for marginal cost:.209 c 3.42917 -6.209[(reset)-1813]TJ -2[(())]T

State variables:

$$\frac{\partial H}{\partial S_{Rt}} = \lambda_{S_R} + r\lambda_{S_R} = \lambda_{S_R}(q - (\# - 1)p_t) + \mu_p \frac{\#}{\# - 1}\frac{S_{Ct}}{(S_{Rt})^2} \tag{10}$$

$$\frac{\partial H}{\partial S_{Ct}} = \lambda_{S_C} + r\lambda_{S_C} = \lambda_{S_C}(q - \#p_t) - \frac{\mu_p}{S_{Rt}}\frac{\#}{\# - 1} \tag{11}$$

$$\frac{\partial H}{\partial v_t} = \lambda_v + r\lambda_v = \frac{mc_t}{1+g} + \lambda_v(\#p_t - q) - \mu_{mc}\gamma y(l_t A)^{-(1+g)}v_t^{g-1} \tag{12}$$

$$\frac{\partial H}{\partial l_t} = \lambda_l + r\lambda_l = \frac{1}{l_t} + \lambda_l(r - r_t + p_t) + \mu_{mc}(1+g)\gamma A^{-(1+g)}l_t^{-(2+g)}v_t^{g}] \tag{13}$$

As well as the transversality conditions, specified in Appendix D.1.

2.4 Determining Optimal Marginal Cost

From (7) and (13), we have:

$$\begin{aligned}\frac{1}{l_t} &= \mu_{mc}\gamma(1+g)v_t^g A(l_t A)^{-(2+g)} \\ 1 &= \mu_{mc}\gamma(1+g)v_t^g(l_t A)^{-(1+g)} \\ (l_t A)^{1+g} &= \mu_{mc}\gamma(1+g)v_t^g\end{aligned} \tag{14}$$

From (9) and the above, we have:

$$\begin{aligned}(l_t A)^{1+g} &= (\lambda_{S_C} + \frac{v_t}{1+g})\gamma(1+g)v_t^g \\ l_t &= \frac{1}{A}[(\lambda_{S_C} + \frac{v_t}{1+g})\gamma(1+g)v_t^g]^{\frac{1}{1+g}}\end{aligned} \tag{15}$$

Taking now the original equilibrium condition for marginal cost (5) and using the results above, we get:

$$mc_t = \frac{1}{(1+g)\lambda_{S_C} + v_t} \tag{16}$$

This is the most fundamental result of the model, and represents the optimal function for marginal cost to take in our system. The way to think about the optimal interest rate is precisely to *struc-*

ture the environment in such a way to make this marginal cost possible. In Ramsey's language, it will be the way to decentralize the solution. It says that marginal firm cost is the inverse of a linear combination of the shadow price on expectations of future cost and price dispersion. The shadow price on expected future costs is weighted by a measure of labor supply elasticity. This e□ectively shows us that the optimal marginal cost is related to the two ine□ciencies of the New Keynesian Model - ine□cient production due to monopolistic competition and price dispersion.

Compare the above equation with the results of Yun (2005):

$$mc_t = \frac{1}{(1+h)v_t} \tag{17}$$

where h is the optimal subsidy rate, which would be set to 1/ (# 1). Note that because of the employment subsidy, the long-run steady state value is now e□cient, Yun does not have to take into consideration the evolution of expectations in his model, and the weighted shadow price of future expectations of cost, $(1+g)L_S$

2.5 Determining the Interest Rate

Let us get an expression of the interest rate.

Taking the time derivative of I_t, we arrive at:

$$\dot{I} = \frac{y(1+g)}{A}[(L_{S_C} + \frac{v_t}{1+g})y(1+g)v_t^g]^{\frac{g}{1+g}}[(L_{S_C} + \frac{v_t}{1+g})gv_t^g \dot{v} + v_t^g(L_{S_C} + \frac{\dot{v}}{1+g})]$$

Dividing by I_t:

$$\frac{\dot{I}}{I_t} = \frac{(L_{S_C} + \frac{v_t}{1+g})gv_t^g \dot{v} + v_t^g(L_{S_C} + \frac{\dot{v}}{1+g})}{(L_{S_C} + \frac{v_t}{1+g})v_t^g}$$

$$= g\frac{\dot{v}}{v_t} + \frac{(1+g)L_{S_C} + v}{((1+g)L_{S_C} + v_t)} \tag{18}$$

Taking the time derivative of marginal cost:

$$\dot{mc} = \frac{(1+g)L_{S_C} + v}{((1+g)L_{S_C} + v_t)^2} \tag{19}$$

From (4), the equilibrium condition for $\frac{\dot{I}}{I_t}$, and (18), and (19), we arrive at the following identity defining the interest rate:

$$r_t = r + p_t \quad g\frac{\dot{v}}{v_t} + \frac{\dot{mc}}{mc_t} \tag{20}$$

Note that this is an equilibrium object, not a rule per se. However, we must remember that marginal cost is in our system a function of the underlying state variables in an unknown way. Inflation, also, is a function of state variables, though in a direct, observable way. We can see this through a re-write of the above as:

$$r_t = r + \frac{q}{1 \quad \#}[(\frac{\#}{\# \quad 1}\frac{S_{Ct}}{S_{Rt}})^{1 \quad \#} \quad 1] \quad g\frac{\dot{v}}{v_t} + \frac{(1+g)L_{S_C} + v}{((1+g)L_{S_C} + v_t)} \tag{21}$$

Keep in mind that L_{S_C} and L_{S_C} cannot be determined analytically and must be computed. I will later demonstrate machine learning techniques for solving this problem that can be applied to other uses. Before moving on to computation, however, let us analytically compare the dynamics of our result so far with a rule of the Taylor Rule variety.

2.6 Comparison of Dynamics with Taylor Rule

Consider for example the following ing ing2.6 Comparison of Dynamics with Taylor Rule

2.7 Comparison of Dynamics of Price Stability Rule

We can also determine a rule that seeks perfect price stability, in other words one that sets out to:

$$p_t = 0, \forall t$$

From this we get:

$$v = q(1 \quad v_t)$$

We also have the following relationship between expectation terms:

$$\frac{S_{Ct}}{S_{Rt}} = \frac{\# \quad 1}{\#}$$

This does not imply the two terms are always constant, but it *does* imply that their growth rates must be the same to maintain a constant proportion. From the equilibrium dynamics of each we get that:

$$S_R = qS_{Rt} \quad 1$$

$$S_C = qS_{Ct} \quad mc_t$$

Which means that:

$$qS_{Rt} \quad 1 = qS_{Ct} \quad mc_t$$

So we arrive at an expression of marginal cost:

$$mc_t = 1 + q(S_{Ct} \quad S_{Rt})$$

Taking the time derivative:

$$mc = q(S_C \quad S_R) = 0$$

the optimal interest rate is what I "back out" of the process controlling the costate term.

$$\lambda_{S_R} = (r - (g - (\# - 1)p_t))\lambda_{S_R} - \mu_p \frac{\#}{\# - 1}\frac{S_{Ct}}{(S_{Rt})^2} \tag{26}$$

$$\lambda_{S_C} = (r - (g - \#p_t))\lambda_{S_C} + \mu_p \frac{\#}{\# - 1}\frac{1}{S_{Rt}} \tag{27}$$

$$\lambda_v = (r - \#p_t + g)\lambda_v + \frac{mc_t}{1+g} + \frac{g}{1+g}\frac{1}{v_t} \tag{28}$$

$$S_R = (g - (\# - 1)p_t)S_{Rt} - 1 \tag{29}$$

$$S_C = (g - \#p_t)S_{Ct} - mc_t \tag{30}$$

$$v = g(1 + p_t(1 - \#)/g)^{\#/(1-\#)} + (\#p_t - g)v_t \tag{31}$$

s.t.

$$(1 + p_t\frac{1 - \#}{g})^{\frac{1}{1-\#}} = \frac{\#}{\# - 1}\frac{S_{Ct}}{S_{Rt}} \tag{32}$$

$$mc_t = \frac{1}{(1+g)\lambda_{S_C} + v_t} \tag{33}$$

$$\frac{\mu_p}{g}(1 + p_t(1 - \#)/g)^{\frac{\#}{1-\#}} = \lambda_{S_R}(\# - 1)S_{Rt} + \lambda_{S_C}\#S_{Ct} + \lambda_v[\#[(1 + (1 - \#)p_t/g)]^{\frac{1}{1-\#}} - \#v_t] \tag{34}$$

I will use this system in our numerical computation. The steady state values are also computationally determined. To see the steady state implied by a specified parameterization, see 3.4.1.

2.9 Non-Optimality of Zero-Inflation Long-Run Target

I will not solve here for the full steady-state allocation, but instead want to point to one result: the non-optimality of perfect price stability.

From the above, assuming $p_{ss} = 0$ we have first from equation (31):

$$0 = g(1 - v_{ss})$$

$$\tilde{n} \quad v_{ss} = 1$$

From equations (29) and (30), we have that:

$$S_{Rss} = \frac{1}{q}$$
$$S_{Css} = \frac{mc_{ss}}{q}$$

From equation (32) and the above results:

$$1 = \frac{\#}{\# \quad 1}\frac{S_{Css}}{S_{Rss}}$$
$$\tilde{n} \quad mc_{ss} = \frac{\# \quad 1}{\#}$$

Using these results, we can take equations (26) and (27) to arrive at:

$$L_{S_{Rss}} = \frac{qm_p}{(r \quad q)}$$
$$L_{S_{Css}} = \frac{\#qm_p}{(\# \quad 1)(r \quad q)}$$

Substituting into equation (34):

$$\frac{m_p}{q} = \frac{(\# \quad 1)qm_p}{q(r \quad q)} \quad \frac{\#(\# \quad 1)\#qm_p}{q(r \quad q)\#(\# \quad 1)}$$
$$\tilde{n} \quad \frac{1}{q} = \frac{(\# \quad 1)}{(r \quad q)} \quad \frac{\#}{(r \quad q)}$$
$$\tilde{n} \quad \frac{(r \quad q)}{q} = \quad 1$$
$$\tilde{n} \quad r \quad q$$

3 Numerical Solution: A Deep Learning Approach

In recent years deep learning has been applied to problems of PDEs (Raissi et al 2018). More recently it has been applied to explicit economics questions (Duarte 2018, Fernandez-Villaverde et al 2020). My approach follows most closely the method of Nakamura-Zimmerer et. al (2021a, 2021b), who use neural networks to approximate two-part boundary problems by first solving a system of ordinary di□erential equations, though in that paper the authors limit themselves to an LQ framework. I then use this data to train a neural network for approximating the relevant portions of the value function along the optimal path. The insight is that because I am only interested in the interest rate, I do not need to fully approximate the entire value function. In other words, I first numerically compute the optimal path, and then use the optimal path to train an additional model for the interrelation of state and control variables. This approach of solving the system of equations defining the monetary policy problem as a boundary problem and then using the results to train an additional model of the interest rate significantly aids with computational speed.

As mentioned in the introduction, there are advantages of this deep learning approach compared to more familiar methods. By the universal approximation theorem (Bach 2017), a neural network can approximate any unknown Borel measurable function, and neural networks are less sensitive to good initial guesses than collocation methods. A neural network method (largely) allows one to avoid the curse of dimensionality that define grid based methods, which forms the bulk of economic numerical methods.

I create a "layer" by stacking N_1 neurons on top of each other:

$$N(x;Q) \quad (n(x;Q_1),...,n(x;Q_{N1}))^T$$

3.2 Deep Learning: Application

Let us now look at how to actually implement the deep learning framework. To get a feel for the approach, consider a subset of our equilibrium conditions:

$$0 = \quad L_{S_R} + (r \quad (q \quad (\# \quad 1)p_t))L_{S_R} \quad m\frac{\#}{\# \quad 1}\frac{S_{Ct}}{(S_{Rt})^2}$$

$$0 = \quad L_{S_C} + (r \quad (q \quad \#p_t))L_{S_C} + m\frac{\#}{\# \quad 1}\frac{1}{S_{Rt}(}$$

ary conditions, I reformulate the neural network as a neural form to incorporate "hard boundaries", as shown in Lagari et al (2020).

A neural form is any construction that is built upon a neural network. For our purposes, consider for example the neural form associated with inflation:

$$(\frac{T - t}{T})p_0 + t(t - T)p(t, Q))) + (\frac{t}{T})p_{ss}$$

We can see that for the above at either boundary - the terminal steady state or the initial condition - the neural form is constructed by design to fit the boundary condition with complete accuracy. The central component, the actual neural network within the neural form, is what is trained to fit the path conditions.

The infinite-horizon variation is obtained with the limit $T \to \infty$. I will use the error above defined for a particular value of T, then extend that value as I solve if the terminal errors are above a certain tolerance.

The total loss is defined as:

$$\begin{aligned} loss(t; Q) = & \, err^2_{LS_R} + err^2_{LS_C} + err^2_{Lv} + err^2_{mc} + err^2_{mp} \\ & + err^2_{S_R} + err^2_{S_C} + err^2_{p} + err^2_{S_R,0} + err^2_{S_C,0} + err^2_{v,0} \\ & + err^2_{LS_R,T} + err^2_{LS_R,T} + err^2_{LS_R,T} \end{aligned}$$

To solve the model, I choose the parameter set Q to minimize the above global loss function over a set of time points.

$$\frac{1}{|D|} \sum_{i=1}^{|D|} loss(t_i; q)$$

The solution will be the open loop solution, in other words I will have the optimal path. Once I have the optimal path, then I can define another neural network to approximate the optimal

interest rate as a function of the state variables:

$$r(x) = NN(x;q)$$

Note that I had previously defined the optimal interest rate as:

$$r_t = r + p_t \quad g\frac{v}{v_t} + \frac{mc_t}{mc_t}$$

I thus already have the optimal interest rate defined on the optimal path. To train the neural network, I first obtain a set of optimal paths with randomly chosen initial points. I then take a set of points along these optimal paths and minimize the error to approximate the interest rate as a function of the state variables. In other words, I convert a set of open loop solutions to a closed loop one.

$$err_r(x,t) = \frac{1}{|D|}\sum_{i=1}^{|D|}[r(x;q) \quad r(t)]^2$$

3.3 Technical Details

All coding was done in python using the Tensorflow library. For the first round, I construct a fully connected neural network with 8 hidden layers of 120 neurons each. The library DeepXDE was used for the first round (Lu et al 2021). The sigmoid activation function and Adam stochastic gradient descent-type algorithm are adopted in the neural network. For the second

3.3.1 Parameterization and Steady State Values

Table 1: Parameterization		
g	1	Frisch labor supply elasticity
r	0.01	Subjective rate of time preference
y	1	Preference for leisure
q	0.65	Calvo parameter
#	25	Elasticity of substitution for intermediate goods

The following steady state values are computationally determined.

Table 2: Steady State Values		
$L_{S_{RSS}}$	-0.0200	Costate, Discounted Future Revenues
$L_{S_{CSS}}$	0.0208	Costate, Discounted Future Costs
L_{vSS}	-1.4844	Costate, Price Dispersion
S_{RSS}	1.5380	Discounted Future Revenues
S_{CSS}	1.4765	Discounted Future Costs
v_{SS}	$1.0+1.9472 \quad 10^{9}$	Price Dispersion
m_{pSS}	0.0197	Lagrange coe□cient, inflation and auxiliary variables
mc_{SS}	0.9600	Marginal Cost
p_{SS}	$-8.1181 \quad 10^{6}$	Inflation
r_{SS}	0.9992%	Interest Rate

3.4 Computational Results

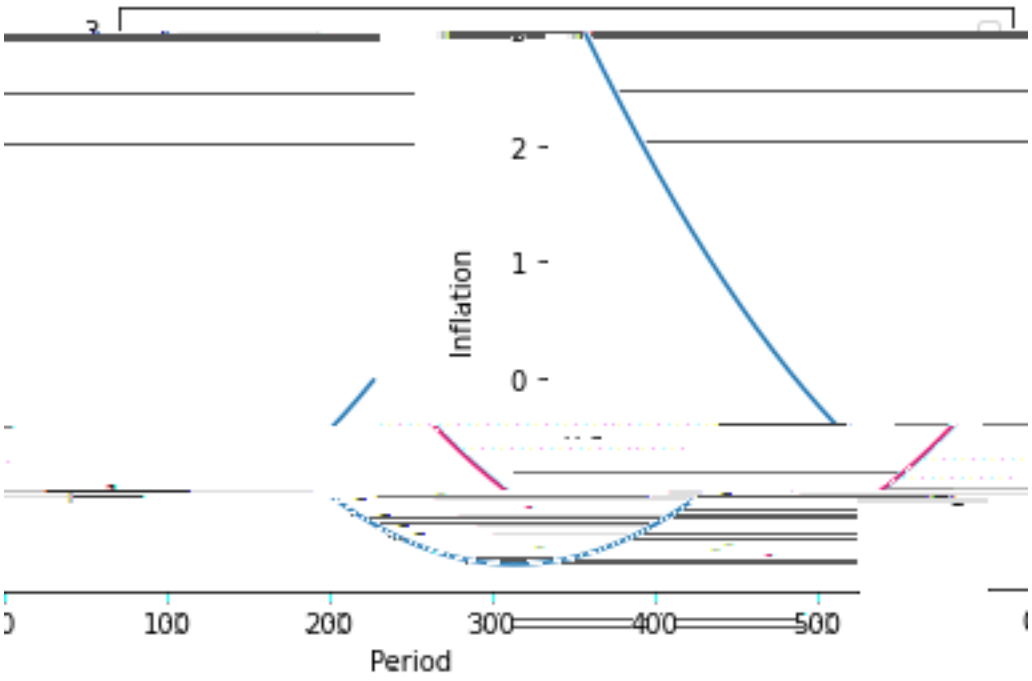
We arrive then at the computational results. First, I want to give a sample set of transition paths with a set of initial conditions. Next, I will show what the interest rate looks like in terms of the other variables of the model.

3.4.1 Sample Transition Paths

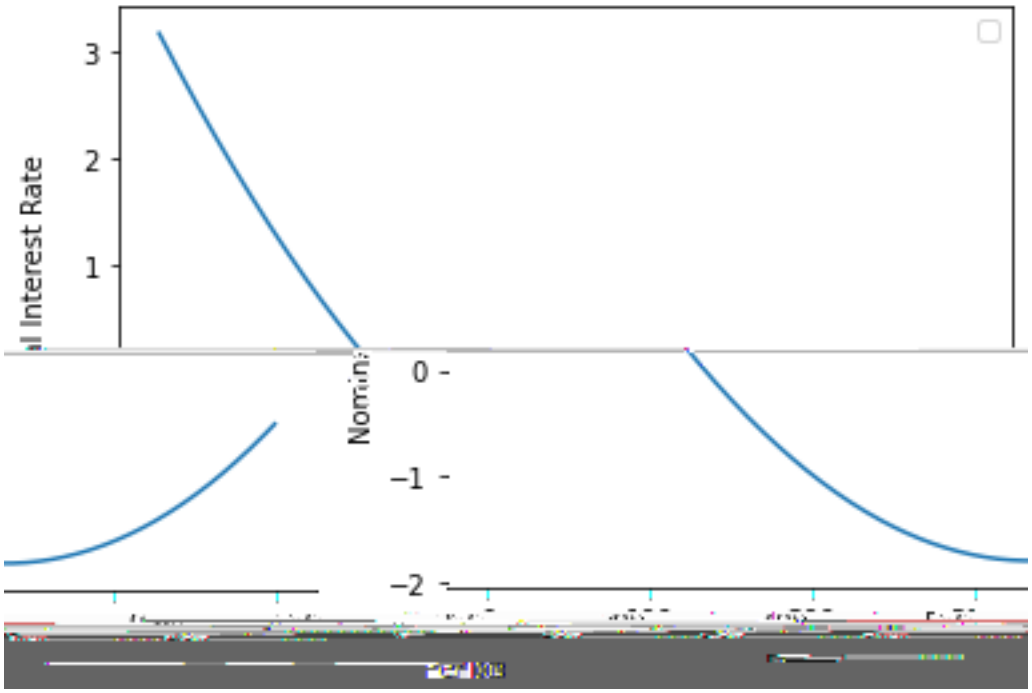
The initial conditions are as follows. Note that marginal cost and inflation are functions of the other three true state variables of the model. I include them for clarity.

Table 3: Initial Conditions

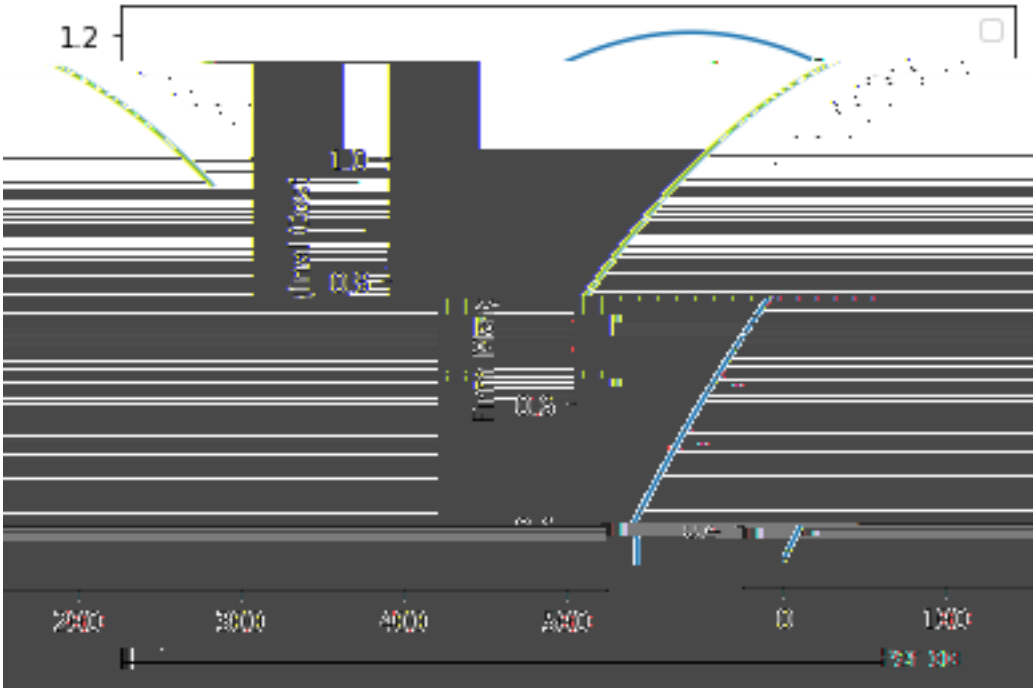
S_0	1.62	Discounted Future Revenues
S_0	1.61	Discounted Future Costs
v_0	1.62	Price Dispersion
mc_0	0.348	Marginal Cost
p_0	$2.96 \quad 10^{6}$	Inflation



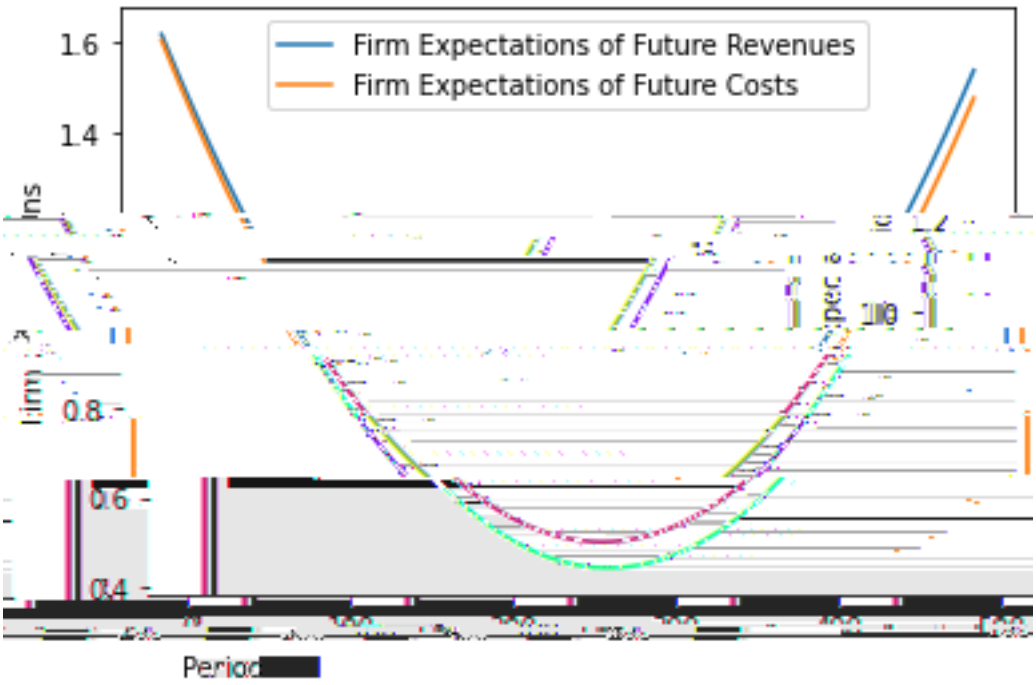
This is done by having the interest be relatively high at the outset, and then lowering it to guide inflation back to its long-run value.



You can see the e□ect through marginal cost as well, which rises before lowering to its long run value.

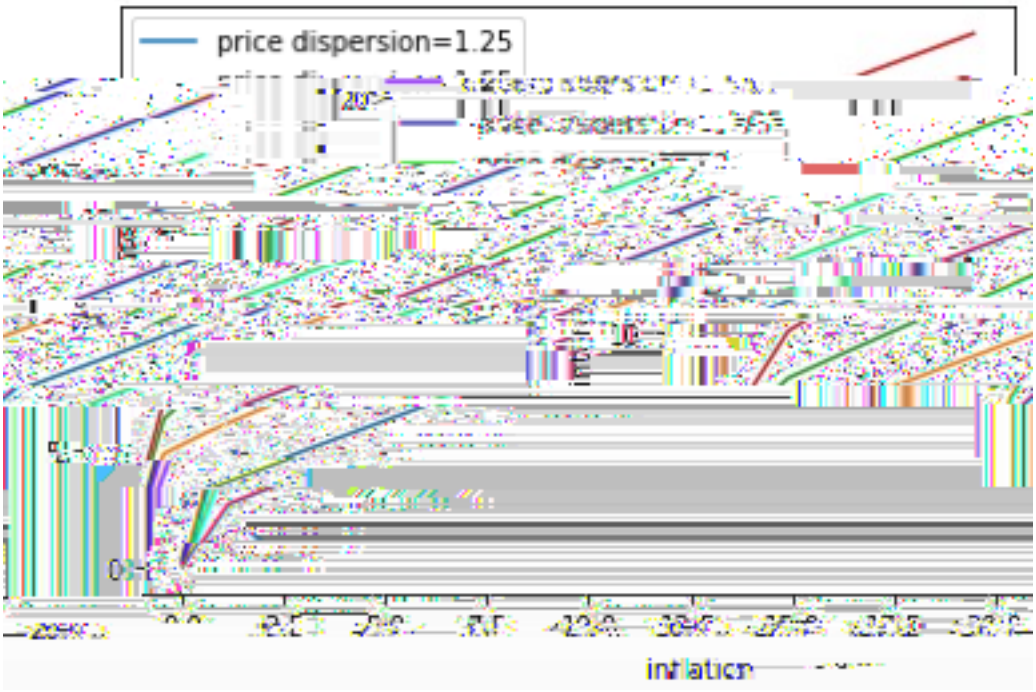


Finally, there are the dynamics of firm expectations. The central bank acts to at first have high expectations of revenues and costs, with a smaller di□erence between the two. This lowers economic output. Then the central bank causes both expectations to fall, though it does so to a greater extent with expectations of future costs. This causes the gap between the two to increase. These dynamics, again, are lost if linearization is doen to the model.



3.4.2 The Interest Rate

I illustrate next the interest rate as a function of relevant state variables. Note that here the interest rate is given as a function of inflation (p) at particular values of price dispersion (v). Remember that $v=1$ implies total e□ciency, with v being bounded below by 1.



A few things are of note here. First, is the confirmation that nonlinearities matter. It also supports the idea that the price dispersion term does in fact influence optimal monetary policy. Traditionally, this term is considered orthogonal to the policy decision, or at the very least it is discarded as being second order. The relationship is intuitive: as ine□ciencies associated with price dispersion become more pronounced, then the central bank should be more aggressive overall.