ESSAYS ON MACROECONOMIC EXPECTATIONS

Laura Gáti

A dissertation

submitted to the Faculty of

the department of Economics

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College

Morrissey College of Arts and Sciences

Graduate School

May 2021

ESSAYS ON MACROECONOMIC EXPECTATIONS

Laura Gáti

Advised by: Associate Professor Ryan Chahrour

Professor Susanto Basu

Professor Peter Ireland

Abstract

This dissertation consists of three independent explorations of the interplay between expectations and macroeconomic activity. I investigate economic dynamics and policy issues concerning the management of expectations both from the lens of business cycles and medium-term fluctuations, embracing a rational expectations approach as well as venturing into the wilderness of bounded rationality.

The first chapter, "Monetary Policy & Anchored Expectations - An Endogenous Gain Learning Model," investigates how a concern to anchor expectations affects the conduct of monetary policy. The chapter first proposes a novel model of expectations which provides a notion for unanchored expectations. In this model, expectations are the more unanchored the higher the sensitivity of long-run inflation expectations to short-run fluctuations. I then embed the expectations model in a general-equilibrium model of the business cycle, and estimate the extent of unanchoring using data on inflation expectations. Within the context of the thus calibrated model, I derive the Ramsey-optimal monetary policy. The main result is that it is optimal for a central bank to anchor inflation expectations to the inflation target. The way the central bank can achieve this is by responding very aggressively with its interest rate tool to fluctuations in long-run inflation expectations.

The observation that motivates the second chapter, "Talking in Time - Dynamic Central Bank Communication," is that the management of expectations by the mon-

etary authority is a dynamic problem in two ways. Firstly, in a dynamic economy, a central bank needs to decide *when* to communicate. Secondly, every time the central bank does talk to the public, it also has to choose *what* to talk about: the present or the future? The chapter thus extends existing macroeconomic research on various dimensions of optimal central bank communication by asking what the implications of dynamics are for the optimal information provision of the central bank to the public. To this end, I analyze a Bayesian Persuasion game between a central bank and the private sector in a static and a dynamic setting, in which the private sector tracks one economic variable, while the central bank wishes it to track a second variable instead. Importantly, the two problems are identical except for the correlation structure between the two variables, which is either cross-sectional or temporal. This way, I isolate the role of dynamics for the optimal communication policy.

The main result is that in the dynamic setting, the prior beliefs of the private sector become endogenous to central bank communication and dampen the effectiveness of the central bank's communication. Therefore, the central bank faces a new tradeoff: it needs to push against priors in two ways. Relatively to the static solution, the central bank talks more about the economic variable it wants the private sector to learn about, and it also talks with less clarity in order to render the private sector's beliefs sufficiently responsive to its messages at each point in time.

In the third chapter, "ICT-Specific Investment Shocks and Economic Fluctuations - Evidence and Theory of a General-Purpose Technology," joint with Marco Brianti, I explore empirically the role of information and communication technologies (ICT) for medium-run economic fluctuations. The first set of results demonstrate the hump-shaped dynamics of total factor productivity after an innovation in ICT, as identified in a structural vector autoregression context. I interpret the hump-shaped impulse response as a consequence of the slow diffusion of ICT technologies, and test this hypothesis using an estimated two-sector growth model. The second set of results document that, viewed through the lens of the model, the data favor the interpretation of innovations in the ICT-sector playing the role of a general-purpose technology. In other words, the slow buildup of the overall effect on productivity stems from the gradual diffusion of ICT in the economy.

Contents

List of Tables

List of Figures

What would happiness be that was not measured by the immeasurable grief at what is?

Theodor Adorno, *Minima Moralia*

1 Monetary Policy and Anchored Expectations - An Endogenous Gain Learning Model

1.1 Introduction

> Inflation that runs below its desired level can lead to an unwelcome fall in *longer-term inflation expectations*, which, in turn, can pull actual inflation even lower, resulting in an adverse cycle of ever-lower inflation and inflation expectations. [...] *Well-anchored inflation expectations* are critical[.]
>
> Jerome Powell, Chairman of the Federal Reserve[1]
>
> *(Emphases added.)*

There is broad consensus among policymakers that anchoring inflation expectations is central to the modern conduct of monetary policy. As Powell emphasizes, policymakers think of expectations as anchored when long-run expectations do not fluctuate systematically with short-run inflation surprises. If long-run expectations did respond to short-term inflation surprises, policymakers fear this would result in an adverse cycle of self-enforcing movements in expectations.

This paper studies the conduct of monetary policy in a New Keynesian model where the degree of expectations anchoring responds endogenously to economic conditions. The main contribution is providing the first theory of monetary policy when expectations can exhibit varying degrees of unanchoring. I propose a novel, quantitatively realistic model of expectation formation which captures the time-varying sensitivity of long-run expectations to short-run conditions. I take the model to the data to quantify my novel anchoring mechanism, and use it to characterize optimal policy analytically and numerically.

In order to provide a convincing analysis of optimal policy, I need a qualitatively and quantitatively realistic model of expectation formation. To this end, I build on new

[1] "New Economic Challenges and the Fed's Monetary Policy Review," August 27, 2020, Jackson Hole.

work by Carvalho et al. (2019), who model the anchoring of expectations as a discrete gain function that determines whether agents learn about long-run inflation using a small or large weight on past expectations errors. My methodological contribution is to extend this work along two dimensions. First, I embed an anchoring model of expectation formation in a general equilibrium New Keynesian model, allowing me to formally consider the monetary policy problem. Second, I use a continuous gain function, capturing the time-varying degrees of unanchoring in the data.

Fig. 1 provides some preliminary evidence that a model of a smoothly varying gain function may be appropriate. It shows the regression coefficient of a rolling window regression of long-run inflation expectations on short-run forecast errors. Letting $\bar{\pi}_t^i$ denote the 10-year inflation expectation of forecaster i from the Survey of Professional Forecasters (SPF), and $f_{t|t-1}^i \equiv \pi_t - \mathbb{E}_{t-1}^i \pi_t$ denote forecaster i's one-year ahead forecast error, I run

$$\Delta \bar{\pi}_t^i = \beta_0^w + \beta_1^w f_{t|t-1}^i + \epsilon_t^i, \tag{1}$$

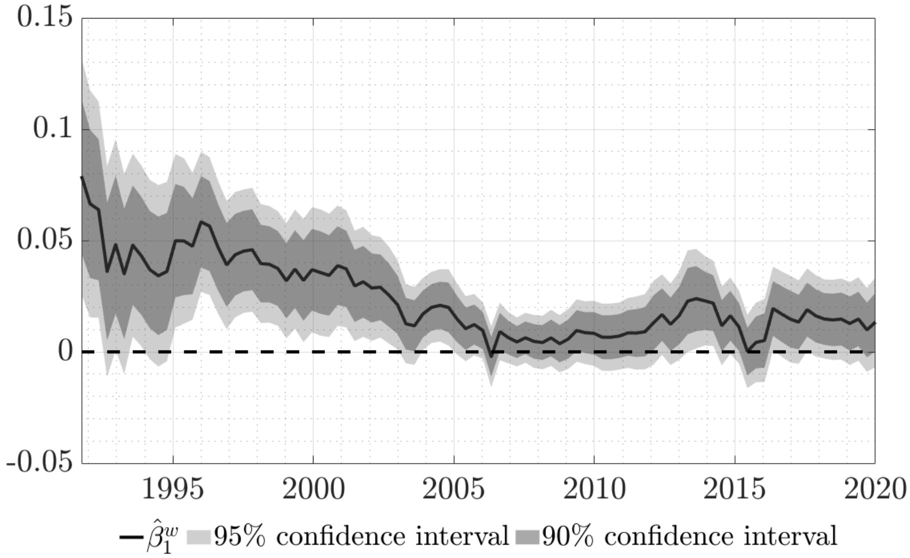
where w indexes windows of 20 quarters. Fig. 1 plots the time series of the estimated coefficient $\hat{\beta}_1^w$, along with 90% and 95% confidence intervals.[2]

As seen on Fig. 1, the relationship between 10-year inflation expectations and one-year surprises, $\hat{\beta}_1^w$, is time-varying. Up to about 2005, $\hat{\beta}_1^w$ is positive and significant, with varying magnitude. Between 2005 and 2010, by contrast, it is not significantly different from zero. In this period, then, long-run expectations were unresponsive to short-run conditions. Around 2010, this changed once more. A significantly positive $\hat{\beta}_1^w$ since 2010 indicates that on several occasions in the last ten years, short-run inflation surprises were passed through to long-run expectations.

As an empirical contribution, I bring my smooth gain function to the data to explore

[2] Appendix A.1 presents robustness checks. Appendix A.2 discusses alternative measures of long-run inflation expectations.

Figure 1: Time series of responsiveness of long-run inflation expectations to inflation surprises



the functional form describing the anchoring of expectations. Calibrating the parameters of the New Keynesian core of my model, I employ a simulated method of moments estimation strategy (Duffie and Singleton (1990), Lee and Ingram (1991), Smith (1993)) to back out a piecewise linear approximation to the true functional form in the data. The estimation results show that my novel gain function has several reasonable properties. First, I estimate an implied gain time series with a median of 0.098, very close to the values obtained by Milani (2014) and Carvalho et al. (2019) (0.082 and 0.145 respectively). This number means that agents in the model rely only on the 10 most recent quarters of data when updating their long-run expectations. Most importantly, the expectations process is nonlinear: larger forecast errors in absolute value lead to a higher sensitivity of long-run expectations. In line with other studies, like Hebden et al. (2020), I also find that a forecast error of a particular size leads to higher gains if its sign is negative than if it is positive. In other words, negative inflation surprises unanchor expectations more than same-sized positive mistakes.

Having calibrated the model to the data, and establishing that it provides a realistic description of expectations, I perform my analysis of policy in the model in three stages. First, I present an analytical characterization of the Ramsey problem of the monetary authority. I demonstrate that the first-order condition of the Ramsey problem is a target criterion for the policymaker. The target criterion prescribes that the central bank should smooth out the effects of shocks by taking advantage of the current and future expected degree of unanchoring.

Second, I solve the optimality conditions of the Ramsey problem numerically. Because the endogenous gain function renders the model nonlinear, I rely on global methods to obtain the optimal policy function. Like the degree of unanchoring, the optimal interest-rate setting is time-varying. Since the interest rate responds to movements in long-run expectations, the monetary authority acts aggressively when expectations unanchor. When expectations are well-anchored, by contrast, the central bank accommodates fluctuations in inflation because in this case, long-run expectations do not respond to short-run conditions.

Third, I consider the class of simple, time-invariant Taylor rules that constitute the most common specification of monetary policy in practice. I solve for the optimal response coefficient of the interest rate to inflation numerically in the case of both the anchoring and the rational expectations versions of my model. It turns out that the central bank finds it optimal to respond less to inflation under anchoring than under rational expectations. The reason is that positive feedback between inflation expectations, interest-rate expectations and inflation induces too high volatility in response to aggressive interest rate movements. Moreover, although the Taylor rule is not fully optimal in the model, an optimally chosen inflation coefficient does a good job in eliminating most of the volatility coming from unanchored expectations.

The model I use to study the interaction between monetary policy and anchoring is a behavioral version of the standard New Keynesian (NK) model of the type widely used for monetary policy analysis. Monetary policy in the rational expectations (RE) version of this model has been studied extensively, for example in Clarida et al. (1999) or Woodford (2003), whose exposition I follow. The formulation of a target criterion to implement optimal policy is in the tradition of Svensson (1999).

The behavioral part of the model is the departure from rational expectations on the part of the private sector. Instead, I allow the private sector to form expectations via an adaptive learning scheme, where the learning gain - the parameter governing the extent to which forecasting rules are updated - is endogenous. The learning framework represents an extension to the adaptive learning literature advocated in the book by Evans and Honkapohja (2001). This literature replaces the rational expectations assumption by postulating an ad-hoc forecasting rule, the perceived law of motion (PLM), as the expectation-formation process. Agents use the PLM to form expectations and update it in every period using recursive estimation techniques. My contribution to this literature is to study optimal monetary policy in a learning model with an endogenous gain.

Adaptive learning is an attractive alternative to rational expectations for several reasons. First, many studies document the ability of adaptive learning models to match empirical properties of both expectations and macro aggregates. Adaptive learning models imply that forecast errors are correlated with forecast revisions, a feature of expectations documented by Coibion and Gorodnichenko (2015). In fact, the prediction of adaptive learning models concerning the response of expectations to shocks exactly aligns with new evidence by Angeletos et al. (2020), suggesting that in response to shocks, expectations initially underreact, and then overshoot. As for the macro evidence, Milani (2007) demonstrates that estimated constant gain learning models match the persistence of inflation without recourse to backward-looking elements in the Phillips curve. Eusepi and Preston (2011) show how a calibrated adaptive learning version of the real business cycle (RBC) model outperforms the rational expectations version. In particular, even with a small gain, the learning model leads to persistent and hump-shaped responses to iid shocks, resolving the long-standing critique of RBC models of Cogley and Nason (1993).

Secondly, having an endogenous gain improves the empirical properties of adaptive learning models further. Milani (2014) documents that endogenous gain models can generate endogenous time-varying volatility in models without any exogenous time-variance. Additionally, Carvalho et al. (2019) estimate the evolution of the endogenous gain for the last fifty years in the US. Not only does their model display excellent out-of-sample forecasting performance in terms of matching long-run expectations, but the estimated gain time series invites a reinterpretation of the Great Inflation as a period of unanchored expectations.

Thirdly, an extensive experimental literature in the spirit of Anufriev and Hommes (2012) demonstrates that simple adaptive learning rules provide the best fit among competing models to how individuals form expectations in controlled lab settings. In fact, the idea that individuals in the economy do not know the true underlying model and thus rely on simple, estimated heuristics to form expectations is intuitive from an economic perspective. Economists do not know the true model of the economy, so why should firms and households? And just like an econometrician estimates statistical models to form forecasts of relevant variables, it is reasonable to suppose that the private sector does so too.

My work is also related to the literature attempting to explain features of expectations data using departures from the full information rational expectations (FIRE) paradigm. This literature consists of two main lines of attack. The first is questioning the assumption of full information. In this body of work, information is either not fully or not symmetrically available, or information acquisition is costly (Mankiw and Reis (2002), **?**, Mackowiak and Wiederholt (2009), Angeletos and Pavan (2009)). The second strand of this literature, to which this paper belongs, instead emphasizes that expectation formation departs from rational expectations. Arguably the leading candidate in this literature, Bordalo et al. (2018), complements the empirical results of Coibion and Gorodnichenko (2015) with new evidence to contend that the pattern of over- and underreaction to news is consistent with diagnostic expectations instead of dispersed information. As opposed to these papers, my objective is not to find the model that best rationalizes stylized facts of expectations in the data. Instead, my contribution to this literature is to investigate how anchoring, a little-studied feature of the expectations data, affects the conduct of monetary policy.

Adaptive learning models have been the object of scrutiny in the past, perhaps most prominently in Sargent (1999). The main interest of Sargent (1999) is exploring whether a story in which the Federal Reserve learns the natural rate hypothesis adaptively can better match the Great Inflation than an alternative where the Fed understands that there is no exploitable tradeoff between inflation and unemployment. My focus is different. Based on the reduced-form empirical evidence of Fig. 1, I entertain an anchoring expectation formation on the part of the private sector and ask what the implications for optimal monetary policy are.

The paper is structured as follows. Section 1.2 introduces the model. Section 1.3 describes the learning framework and spells out the anchoring mechanism. Section 1.4 estimates the anchoring function. Section 1.5 presents the results in three parts. First, Section 1.5.1 discusses an analytical characterization of the Ramsey policy. Second, Section 1.5.2 solves for the interest rate sequence that implements the optimal Ramsey allocation using global methods. Third, Section 1.5.3 investigates the optimal choice of response coefficients if monetary policy is restricted to follow a Taylor rule. Section 1.6 concludes.

1.2 The model

Apart from expectation formation, the model is a standard New Keynesian (NK) model with nominal frictions à la Calvo (1983). The advantage of having a standard NK backbone to the model is that one can neatly isolate the way the anchoring mechanism alters the behavior of the model. Since the mechanics of the rational expectations version of this model are well understood, I only lay out the model briefly and pinpoint the places where the assumption of nonrational expectations matters.[3]

1.2.1 Households

The representative household is infinitely-lived and maximizes expected discounted lifetime utility from consumption net of the disutility of supplying labor hours:

$$\hat{\mathbb{E}}_t^i \sum_{T=t}^{\infty} \beta^{T-t} \left[U(C_T^i) - \int_0^1 v(h_T^i(j)) dj \right]. \tag{2}$$

In the above, the only non-standard element is the expectations operator, $\hat{\mathbb{E}}^i$. As detailed in Section 1.3, this operator captures non-rational expectations and is assumed to satisfy the law of iteration expectations, so that $\hat{\mathbb{E}}_t^i \hat{\mathbb{E}}_{t+1}^i = \hat{\mathbb{E}}_t^i$. The remaining elements are familiar from the rational expectations version of the NK model. $U(\cdot)$ and $v(\cdot)$ denote the utility of consumption and disutility of labor, respectively, and β is the discount factor of the household. I am using standard CRRA utility of the form $U(c_t) = \frac{c_t^{1-\sigma}}{1-\sigma}$. $h_t^i(j)$ denotes the supply of labor hours of household i at time t to the production of good j, and the household participates in the production of all goods j. Similarly, household i's consumption bundle at time t, C_t^i, is a Dixit-Stiglitz composite of all goods in the economy:

$$C_t^i = \left[\int_0^1 c_t^i(j)^{\frac{\theta-1}{\theta}} dj \right]^{\frac{\theta}{\theta-1}}, \tag{3}$$

where $\theta > 1$ is the elasticity of substitution between the varieties of consumption goods. Denoting by $p_t(j)$ the time-t price of good j, the aggregate price level in the economy can then be written as

$$P_t = \left[\int_0^1 p_t(j)^{1-\theta} dj \right]^{\frac{1}{\theta-1}}. \tag{4}$$

[3] For more details on the NK model, see Woodford (2003).

The budget constraint of household i is given by

$$B_t^i \leq (1+i_{t-1})B_{t-1}^i + \int_0^1 w_t(j)h_t^i(j) + \Pi_t^i(j)dj - T_t - P_tC_t^i, \tag{5}$$

where $\Pi_t^i(j)$ denotes profits from firm j remitted to household i, T_t taxes, and B_t^i the riskless bond purchases at time t.[4]

The only difference to the standard New Keynesian model is the expectations operator, $\hat{\mathbb{E}}^i$. This is the subjective expectations operator that differs from its rational expectations counterpart, $\mathbb{E}$, in that it does not encompass knowledge of the model. In particular, households have no knowledge of the fact that they are identical. By extension, they also do not internalize that they hold identical beliefs about the evolution of the economy. This implies that while the modeler can suppress the index i, understanding that $\hat{\mathbb{E}}^i = \hat{\mathbb{E}}^j = \hat{\mathbb{E}}$, households cannot do so. As we will see in Section 1.2.3, this has implications for their forecasting behavior and will result in decision rules that differ from those of the rational expectations version of the model.

1.2.2 Firms

Firms are monopolistically competitive producers of the differentiated varieties $y_t(j)$. The production technology of firm j is $y_t(j) = A_t f(h_t(j))$, whose inverse, $f^{-1}(\cdot)$, signifies the amount of labor input. Noting that A_t is the level of technology and that $w_t(j)$ is the wage per labor hour, firm j profits at time t can be written as

$$\Pi_t^j = p_t(j)y_t(j) - w_t(j)f^{-1}(y_t(j)/A_t). \tag{6}$$

[4]For ease of exposition I have suppressed potential money assets here. This has no bearing on the model implications since it represents the cashless limit of an economy with explicit money balances.

Firm j's problem then is to set the price of the variety it produces, $p_t(j)$, to maximize the present discounted value of profit streams

$$\hat{\mathbb{E}}_t^j \sum_{T=t}^{\infty} \alpha^{T-t} Q_{t,T} \left[\Pi_t^j(p_t(j)) \right], \tag{7}$$

subject to the downward-sloping demand curve

$$y_t(j) = Y_t \left(\frac{p_t(j)}{P_t} \right)^{-\theta}, \tag{8}$$

where

$$Q_{t,T} = \beta^{T-t} \frac{P_t U_c(C_T)}{P_T U_c(C_t)} \tag{9}$$

is the stochastic discount factor from households. Nominal frictions enter the model through the parameter α in Equation (7). This is the Calvo probability that firm j is not able to adjust its price in a given period.

Analogously to households, the setup of the production side of the economy is standard up to the expectation operator. Also here the model-consistent expectations operator $\mathbb{E}$ has been replaced by the subjective expectations operator $\hat{\mathbb{E}}^j$. This implies that firms, like households, do not know the model equations, nor do they internalize that they are identical. Thus their decision rules, just like those of the households, will be distinct from their rational expectations counterparts.

1.2.3 Aggregate laws of motion

The model solution procedure entails deriving first-order conditions, taking a loglinear approximation around the nonstochastic steady state and imposing market clearing conditions to reduce the system to two equations, the New Keynesian Phillips curve and IS curve. The presence of subjective expectations, however, implies that firms and households are not aware of the fact that they are identical. Thus, as Preston

(2005) points out, imposing market clearing conditions in the expectations of agents is inconsistent with the assumed information structure.[5]

Instead, I prevent firms and households from internalizing market clearing conditions.[6] As Preston (2005) demonstrates, this leads to long-horizon forecasts showing up in firms' and households' first-order conditions. As a consequence, instead of the familiar expressions, the IS and Phillips curves take the following form:

$$x_t = -\sigma i_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} \beta^{T-t}((1-\beta)x_{T+1} - \sigma(\beta i_{T+1} - \pi_{T+1}) + \sigma r_T^n), \tag{10}$$

$$\pi_t = \kappa x_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t}(\kappa\alpha\beta x_{T+1} + (1-\alpha)\beta\pi_{T+1} + u_T). \tag{11}$$

Here x_t, π_t and i_t are the log-deviations of the output gap, inflation and the nominal interest rate from their steady state values, and σ is the intertemporal elasticity of substitution. The variables r_t^n and u_t are exogenous disturbances representing a natural rate shock and a cost-push shock respectively.

The laws of motion (10) and (11) are obtained by deriving individual firms' and households' decision rules, which involve long-horizon expectations, and aggregating across the cross-section. Importantly, agents in the economy have no knowledge of these relations since they do not know that they are identical and thus are not able to impose market clearing conditions required to arrive at (10) and (11). Thus, although the evolution of the observables (π, x) is governed by the exogenous state variables (r^n, u) and long-horizon expectations via these two equations, agents in the economy are unaware of this. As I will spell out more formally in Section 1.3, it is indeed the equilibrium mapping between states and jump variables the agents are attempting to

[5]The target of Preston (2005)'s critique is the Euler-equation approach as exemplified for example by Bullard and Mitra (2002). This approach involves writing down the loglinearized first-order conditions of the model, and simply replacing the rational expectations operators with subjective ones. In a separate paper, I demonstrate that the Euler-equation approach is not only inconsistent on conceptual grounds as Preston (2005) shows, but also delivers substantially different quantitative dynamics in a simulated New Keynesian model.

[6]There are several ways of doing this. An alternative to Preston (2005)'s long-horizon approach pursued here is the shadow price learning framework advocated by Evans and McGough (2009).

learn.

To simplify notation, I gather the exogenous state variables in the vector s_t and observables in the vector z_t as

$$s_t = \begin{bmatrix} r_t^n \\ \bar{i}_t \\ u_t \end{bmatrix} \qquad \text{and} \qquad z_t = \begin{bmatrix} \pi_t \\ x_t \\ i_t \end{bmatrix}, \tag{12}$$

where $\bar{i}_t$ is a shock to the interest rate that only shows up in the model for particular specifications of monetary policy.[7] This allows me to denote long-horizon expectations by

$$f_{a,t} \equiv \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} z_{T+1} \qquad \text{and} \qquad f_{b,t} \equiv \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\beta)^{T-t} z_{T+1}. \tag{13}$$

As detailed in Appendix A.4, one can use this notation to reformulate the laws of motion of jump variables (Equations (10), (11) and (41)) compactly as

$$z_t = A_a f_{a,t} + A_b f_{b,t} + A_s s_t, \tag{14}$$

where the matrices $A_i, \ i = \{a, b, s\}$ gather coefficients and are given in Appendix A.4. Assuming that exogenous variables evolve according to independent AR(1) processes, I write the state transition matrix equation as

$$s_t = h s_{t-1} + \epsilon_t \qquad \text{with} \qquad \epsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma), \tag{15}$$

where h gathers the autoregressive coefficients ρ_j, ϵ_t the Gaussian innovations ε_t^j, and η the standard deviations σ_t^j, for $j = \{r, i, u\}$. $\Sigma = \eta\eta'$ is the variance-covariance matrix

[7] For generality, I treat the exogenous state vector as three-dimensional throughout the paper, even when the monetary policy shock is absent.

of disturbances.[8]

1.3 Learning with an anchoring mechanism

The informational assumption of the model is that agents do not know the equilibrium mapping between states and jumps in the model. Without knowing the form of the observation equation, then, they are not able to form rational expectations forecasts. Instead, agents postulate an ad-hoc forecasting relationship between states and jumps and seek to refine it in light of incoming data. In other words, they act like an econometrician: they estimate a simple statistical model and constantly attempt to improve their model as new data arrive.

1.3.1 Perceived law of motion

I assume agents consider a forecasting model for the endogenous variables of the form

$$\hat{\mathbb{E}}_t z_{t+1} = a_{t-1} + b_{t-1} s_t, \tag{16}$$

where a and b are estimated coefficients of dimensions 3×1 and 3×3 respectively. This perceived law of motion (PLM) reflects the assumption that agents forecast jumps using a linear function of current states and a constant, with last period's estimated coefficients. Summarizing the estimated coefficients as $\phi_{t-1} \equiv \begin{bmatrix} a_{t-1} & b_{t-1} \end{bmatrix}$, here 3×4, I can rewrite Equation (16) as

$$\hat{\mathbb{E}}_t z_{t+1} = \phi_{t-1} \begin{bmatrix} 1 \\ s_t \end{bmatrix}. \tag{17}$$

I also assume that

$$\hat{\mathbb{E}}_t \phi_{t+k} = \phi_t \quad \forall\, k \geq 0. \tag{18}$$

[8]For the sake of conciseness, I have suppressed the expressions for these in the main text. See Appendix A.4.

This assumption, known in the learning literature as anticipated utility (Kreps (1998)), means that agents do not internalize that they will update the forecasting rule in the future. Clearly, this poses a higher level of irrationality than not knowing the model and using statistical techniques to attempt to learn it. Nevertheless, because it has been demonstrated not to alter the dynamics of the linearized model (Sargent (1999)), anticipated utility has become a standard assumption in the adaptive learning literature in order to simplify the algebra.

Since the states s_t are exogenous, I assume that agents know Equation (15), the equation governing the evolution of s_t.[9] Then, the PLM together with anticipated utility implies that k-period ahead forecasts are constructed as

$$\hat{\mathbb{E}}_t z_{t+k} = a_{t-1} + b_{t-1} h^{k-1} s_t \quad \forall k \geq 1. \tag{19}$$

The timing assumptions of the model are as follows. In the beginning of period t, the current state s_t is realized. Agents then form expectations according to (16) using last period's estimate ϕ_{t-1} and the current state s_t. Given exogenous states and expectations, today's jump vector z_t is realized. This allows agents to evaluate the most recent forecast error

$$f_{t|t-1} \equiv z_t - \phi_{t-1} \begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix} \tag{20}$$

to update their forecasting rule. The estimate is updated according to the following

[9]This is another common simplifying assumption in the adaptive learning literature. In an extension, I relax this assumption and find that it has similar implications as having agents learn the Taylor rule: initial responses to shocks lack intertemporal expectation effects, but these reemerge as the evolution of state variables is learned.

recursive least-squares algorithm:

$$\phi_t = \left(\phi'_{t-1} + k_t R_t^{-1} \begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix} f'_{t|t-1}\right)', \tag{21}$$

$$R_t = R_{t-1} + k_t\left(\begin{bmatrix} 1 \\ s_{t-1} \end{bmatrix} \begin{bmatrix} 1 & s_{t-1} \end{bmatrix} - R_{t-1}\right). \tag{22}$$

Here R_t is the 4×4 variance-covariance matrix of the regressors and k_t is the learning gain, specifying to what extent the updated estimate loads on the forecast error. Clearly, a high gain implies high loadings and thus strong changes in the estimated coefficients ϕ_t. A low gain, by contrast, means that a given forecast error only has a small effect on ϕ_t.

1.3.2 Endogenous gain as anchoring mechanism

The vast majority of the learning literature specifies the gain either as a constant, $\bar{g}$, or decreasing with time, so that $k_t = t^{-1}$, where t indexes time. Instead, to capture the notion of anchoring, I follow Carvalho et al. (2019) in allowing the gain to fluctuate in a time-varying way in response to short-run forecast errors. I use the following endogenous gain specification: I assume the gain evolves as

$$k_t = \mathbf{g}(k_{t-1}, f_{t|t-1}), \tag{23}$$

where $\mathbf{g}(\cdot)$ is a smooth function. I refer to $\mathbf{g}(\cdot)$ as the anchoring function. Notice that this function is an extension to the conventional decreasing or constant gain specifications, as it nests both as special cases. A decreasing gain implies $\mathbf{g}(k_{t-1}, f_{t|t-1}) = t^{-1}$, while a constant gain $\mathbf{g}(k_{t-1}, f_{t|t-1}) = \bar{g}$. Furthermore, in both special cases, $\mathbf{g}_1 = \mathbf{g}_2 = 0$.

By contrast, my novel specification entails the following assumptions on the function

$\mathbf{g}(\cdot)$:

$$\mathbf{g}_1 \geq 0, \tag{24}$$

$$\mathbf{g}_{22} \geq 0. \tag{25}$$

Equation (24) says that the current value of the gain depends positively on its past value. In a similar vein, Equation (25) suggests that $\mathbf{g}(\cdot)$ is convex in its second argument. In other words, the gain is increasing in the absolute value of forecast errors.

The novelty of specifying the evolution of the gain using (23) with the properties in (24) and (25) is that it offers a time- and state-dependent model of private sector learning. In both of the standard, exogenous gain schemes the gain is divorced from the current environment. It either decreases deterministically ($k_t = t^{-1}$), or is a constant ($k_t = \bar{g}$). This means that the private sector's learning is static and not state-dependent in the sense that a volatile environment results in the same learning pattern as a stable one. My specification, instead, allows the private sector to learn in a time-varying and state-dependent way. When high volatility results in large forecast errors in absolute value, the anchoring function outputs a larger gain than when forecast errors are close to zero. Thus the learning coefficients a_t and b_t respond more to forecast errors in volatile states of the world than in stable ones. The time-varying nature of the gain therefore generates periods of well-anchored expectations (low gain) as well as highly unanchored episodes (high gain). The model thus interprets the gain as a metric of the extent of unanchoring.

My anchoring mechanism also displays desirable empirical features. First, as demonstrated by Milani (2014), endogenous gain learning models can match time-varying volatility in the data, a feature that constant gain learning models cannot account for. More importantly, since a_t is my model's expression for long-run expectations, the fact that the gain is a convex function of forecast errors allows the model to match the time-varying sensitivity of long-term inflation expectations documented in Fig. 1.

I am not the first to postulate an endogenous gain. In fact, Milani (2014)'s estimation is based on an endogenous gain scheme introduced by Marcet and Nicolini (2003). Furthermore, Carvalho et al. (2019) are the first to lay out an endogenous gain specification as a model of anchored expectations. The novelty of my endogenous gain model is that, as opposed to Carvalho et al. (2019), the anchoring function $\mathbf{g}(\cdot)$ is a smooth function. In Carvalho et al. (2019)'s specification, discussed in detail in Appendix A.6, $\mathbf{g}(\cdot)$ is a discrete function of the endogenous component of forecast errors. In their partial equilibrium model, firms use a constant gain if the endogenous component of forecast errors exceeds a given threshold, otherwise they use a decreasing gain.

Assuming instead that $\mathbf{g}(\cdot)$ is a smooth function is useful for two reasons. First, since the derivatives of $\mathbf{g}(\cdot)$ exist, this formulation allows me to consider the Ramsey problem of the monetary authority. This is thus a necessary condition to be able to study optimal monetary policy in the model, which is the main contribution of this paper. Second, the smoothness of $\mathbf{g}(\cdot)$ renders the size of the gain a meaningful metric of the degree of unanchoring. By contrast, in Carvalho et al. (2019)'s model, expectations are either fully anchored or unanchored. My model thus provides more flexibility regarding the extent of unanchoring, in line with the evidence on varying degrees of responsiveness of long-run expectations to short-run mistakes.

My smooth specification of $\mathbf{g}(\cdot)$ also allows me to take the anchoring function to the data. As discussed in Section 1.4, I employ a semi-nonparametric approach to back out the functional form of $\mathbf{g}(\cdot)$ from the data. I then use the estimates to calibrate my model to carry out the numerical analysis of monetary policy.

It is intuitive why the central bank might care whether expectations are anchored or not. When expectations are strongly unanchored at time t, the private sector believes that the true DGP involves a different mapping between states and jumps than they previously maintained. Private sector forecasts will thus drift in the direction of the update, implying that the observables will also shift in the same direction owing to the law of motion (14). From the perspective of the central bank, stabilization of the observables therefore implies stabilization of expectations. Indeed, the contribution of this paper is to analyze formally the nature of the monetary policy problem when expectations are formed using the anchoring mechanism in Equation (23).

1.3.3 Actual law of motion

To complete the model, I now use the specifics of the anchoring expectation formation to characterize the evolution of the jump variables under learning. Using the PLM from Equation (16), I write the long-horizon expectations in (13) as

$$f_{a,t} \equiv \frac{1}{1-\alpha\beta} a_{t-1} + b_{t-1}(I_3 - \alpha\beta h)^{-1} s_t \qquad \text{and} \qquad f_{b,t} \equiv \frac{1}{1-\beta} a_{t-1} + b_{t-1}(I_3 - \beta h)^{-1} s_t. \tag{26}$$

Substituting these into the law of motion of observables (Equation (14)) yields the actual law of motion (ALM):

$$z_t = g^l_{t-1} \begin{bmatrix} 1 \\ s_t \end{bmatrix}, \tag{27}$$

where g^l is a 3×4 matrix given in Appendix A.5. Thus, instead of the state-space solution of the RE version of the model (Equations (15) and (A.13)), the state-space solution for the learning model is characterized by the pair of equations (15) and (27), together with the PLM (19), the learning equations (21) and (22), as well as the anchoring function (23).

1.3.4 Simplifying assumptions

To simplify the analytical work in Section 1.5.1, I make two assumptions that I maintain for the rest of the paper.

Assumption 1. *$e_i' a_t = 0, \; i = 2, 3, \forall t$ and $b_t = g \; h \; \forall t$.*

Here e_i is a 3×1 vector of zeros with 1 as its i^{th} element. Assumption 1 amounts to restricting the intercepts in the forecasts of the output gap and the interest rate and the slope coefficients of all forecasts to what they would be under rational expectations. This means that instead of learning the intercept and slope parameters for all three endogenous variables, the private sector only learns the intercept of the inflation process.

The rationale behind this assumption is that it is the smallest possible deviation from rational expectations that at the same time has enough flexibility to study the unanchoring of inflation expectations. In particular, since the inflation intercept is learned using my novel anchoring mechanism, the formulation is able to capture the time-varying sensitivity of long-run inflation expectations to short-run forecast errors in inflation. In extensions, I relax this assumption and find that it does not change the qualitative or quantitative implications of the model.

With Assumption 1, the updating equations (21) and (22) simplify to a single equation:

$$\bar{\pi}_t = \bar{\pi}_{t-1} + k_t f_{t|t-1}, \tag{28}$$

where, using the notation that $b_1 \equiv e_1 b$, the k-period-ahead inflation forecast is given by

$$\hat{\mathbb{E}}_t \pi_{t+k} = \bar{\pi}_{t-1} + b_1^k s_t, \quad \forall k \geq 1. \tag{29}$$

Lastly, the one-period-ahead forecast error in inflation simplifies to

$$f_{t|t-1} = \pi_t - (\bar{\pi}_{t-1} + b_1 s_t). \tag{30}$$

Assumption 2. $\mathbf{g}_1 = 0$.

The interpretation of this assumption on the anchoring function is that the current gain depends only on forecast errors, not on its own past level. The motivation is to simplify the analytical solution of the Ramsey problem in Section 1.5.1. Appendix

A.9 provides a treatment of the Ramsey problem with the general specification of the anchoring function, relaxing Assumption 2. As seen in the Appendix, this does not change the mechanics nor the intuition behind the solution of the Ramsey problem.

1.4 Quantification of learning channel

The numerical analysis of monetary policy requires a functional specification for the anchoring function $\mathbf{g}(\cdot)$ of Equation (23). For this reason, I back out the functional form of $\mathbf{g}(\cdot)$ from data. Apart from its usefulness for numerical analysis, the form of the anchoring function is interesting in its own right because it describes empirical properties of the anchoring of expectations. In particular, central bankers may want to know how much forecast errors of a particular sign and magnitude unanchor expectations.

I carry out the estimation in two steps. I first calibrate the parameters of the underlying New Keynesian model. Conditional on these parameter values, I estimate the anchoring function by simulated method of moments à la Lee and Ingram (1991), Duffie and Singleton (1990) and Smith (1993). I target the autocovariance structure of the Baxter-King filtered observables of the model and expectations. The observables are CPI inflation from the Bureau of Labor Statistics (BEA), the output gap and the federal funds rate from the Board of Governors of the Federal Reserve System.[10] For expectations, I rely on 12-month-ahead CPI inflation forecasts from the Survey of Professional Forecasters (SPF). The dataset is quarterly and ranges from 1981-Q3 to 2020-Q1. Appendix A.7 contains a detailed description of the estimation methodology.

1.4.1 Calibration

For the calibration of the NK backbone of the model, I split the parameters into two subsets. The first is calibrated using values from the literature, while the second is calibrated to match the moments outlined above. Tables 1 and 2 show the two subsets

[10] The output gap measure is constructed as the difference between real GDP from the Bureau of Economic Analysis (BEA) and the Congressional Budget Office's (CBO) estimate of real potential output.

of calibrated parameters respectively. As Table 1 depicts, I adopt standard parameters from the literature where possible. In particular, for β, σ and other parameters underlying κ, the slope of the Phillips curve, I rely on the parameterization of Chari et al. (2000), advocated in Woodford (2003).

The composite parameter κ is given by $\kappa = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}\zeta$, where ζ is a measure of strategic complementarity in price setting. Assuming specific factor markets, constant desired markups with respect to output levels and no intermediate inputs, $\zeta = \frac{\omega+\sigma^{-1}}{1+\omega\theta}$. Here θ is the price elasticity of demand and ω is the elasticity of the marginal cost function with respect to output. Chari et al. (2000)'s calibration involves $\theta = 10, \sigma = 1, \omega = 1.25, \beta = 0.99$, so that together with my choice of α, κ is pinned down. Note that I lower β slightly (0.98 instead of Chari et al. (2000)'s 0.99). This allows the model to better match the autocovariance structure of the output gap because it lowers the pass-through of long-horizon expectations in the IS curve.

The probability of not adjusting prices, α, is set to match an average price duration of two quarters, which is slightly below the numbers found in empirical studies.[11] In the literature, the average price duration is a little above 7 months. My number for α corresponds to 6 months. I choose the lower end of this spectrum in order to allow the learning mechanism, and not price stickiness, to drive the bulk of the model's persistence.

To simplify the numerical analysis as well as interpretation, I restrict the shocks to be iid. The volatilities of the disturbances and, where applicable, the output-coefficient of the Taylor rule, are set to match the above-mentioned moments. As shown in Table 2, this implies standard deviations of 0.01 for the natural rate and monetary policy shock and 0.5 for the cost-push shock. In sections of the paper where I assume a Taylor rule, the moment-matching exercise results in a 0.3 coefficient on the output gap. The last

[11] On the lower end of the empirical values for α, Bils and Klenow (2004) find a mean duration of 4.3 months and Klenow and Malin (2010) 6.9 months. Klenow and Kryvtsov (2008) and Nakamura and Steinsson (2008) agree on between 7-9 months, while Eichenbaum et al. (2011)'s number is 10.6 months.

Table 1: Parameters calibrated from literature

β	0.98	stochastic discount factor
σ	1	intertemporal elasticity of substitution
α	0.5	Calvo probability of not adjusting prices
κ	0.0842	slope of the Phillips curve
ψ_π	1.5	coefficient of inflation in Taylor rule*
$\bar{g}$	0.145	initial value of the gain

Table 2: Parameters set to match data moments

ψ_x	0.3	coefficient of the output gap in Taylor rule*
σ_r	0.01	standard deviation, natural rate shock
σ_i	0.01	standard deviation, monetary policy shock*
σ_u	0.5	standard deviation, cost-push shock

*Parameters with an asterisk refer to sections of the paper where a Taylor rule is in effect.

parameter of the New Keynesian core is the inflation coefficient of the Taylor rule, ψ_π. Unless otherwise specified, I set ψ_π to 1.5, the value recommended by Taylor (1993). Note that the asterisks in Tables 1 and 2 demarcate parameters that pertain to the Taylor rule and thus only to sections of the paper which assume that a Taylor rule is in effect.

I initialize the value of the gain at $\bar{g} = 0.145$. This is Carvalho et al. (2019)'s estimate for the case of unanchored expectations in their discrete anchoring function model. In the case of a discrete anchoring function, as in Carvalho et al. (2019), this parameter has important implications for model dynamics because by construction, the gain takes on this value very frequently. However, since my specification for $\mathbf{g}(\cdot)$ is smooth, $\bar{g}$ does not have a strong bearing on model dynamics.

1.4.2 Estimation

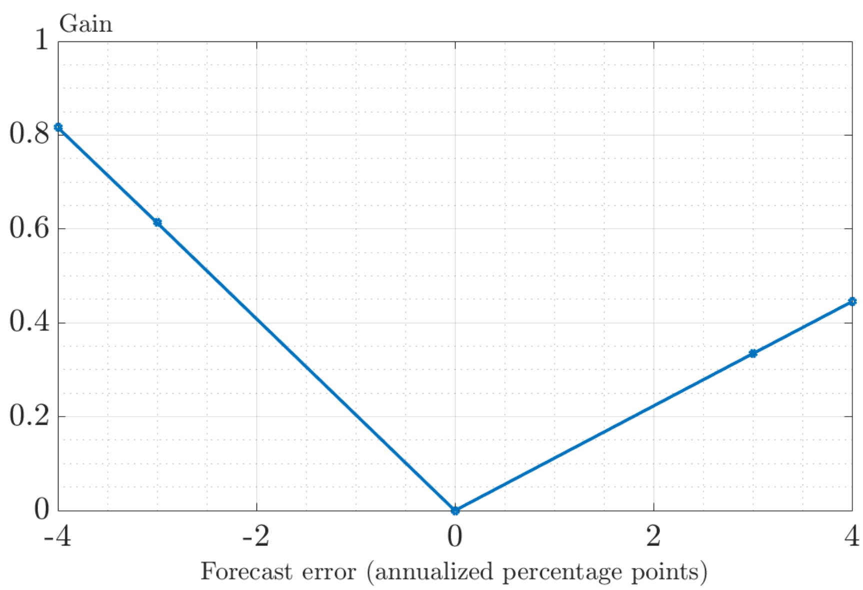
In inferring the functional form of the anchoring function $\mathbf{g}(\cdot)$, I proceed as follows. First, since the analysis in Section 1.5 relies on Assumptions 1 and 2, I similarly impose these on the estimation. In other words, I seek to back out the functional form of the relationship between the size of the gain and forecast errors in inflation, $k_t = \mathbf{g}(f_{t|t-1})$.

To be as close as possible to a non-parametric estimate of $\mathbf{g}(\cdot)$, while at the same time preserve the shape of the function, I employ a piecewise linear approximation of the form:

$$\mathbf{g}(f_{t|t-1}) = \sum_i \gamma_i b_i(f_{t|t-1}). \tag{31}$$

Here $b_i(\cdot)$ is a piecewise linear basis and γ is a vector of approximating coefficients. The index i refers to the breakpoints of the piecewise linear approximation. As explained above, I estimate γ by simulated method of moments, targeting the autocovariance structure of the observables of the model and expectations.

Figure 2: Estimated $\hat{\alpha}$: k_t as a function of $f_{t|t-1}$



Estimates for 5 knots, cross-section of size $N = 1000$

Fig. 2 presents the estimated coefficients: $\hat{\gamma} = (0.82;\ 0.61;\ 0;\ 0.33;\ 0.45)$. The circles indicate the nodes of the piecewise linear approximation. The interpretation of the elements of $\hat{\gamma}$ is the value of the gain the private sector chooses when it observes a forecast error of a particular magnitude. For example, a forecast error of -4 pp in inflation is associated with a gain of 0.82.

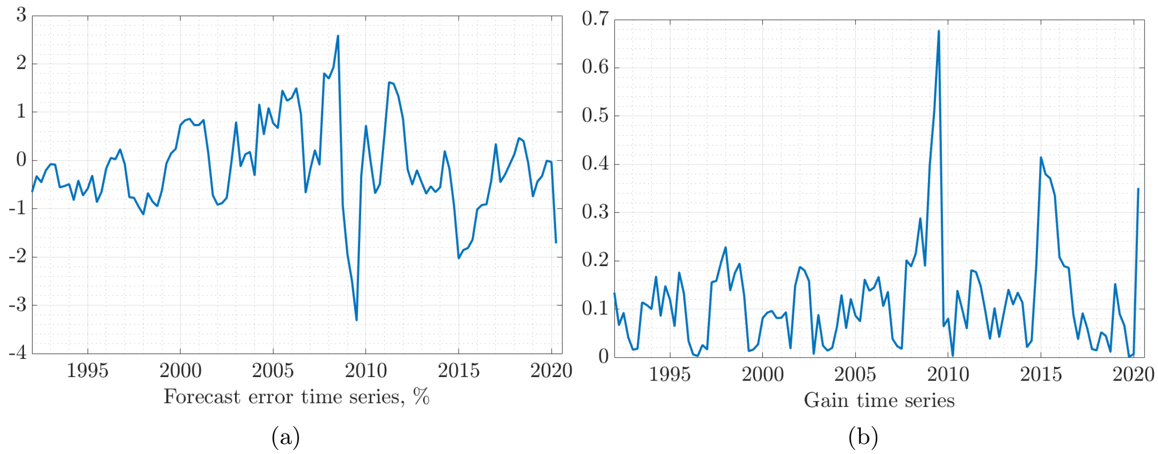
Before interpreting the estimated anchoring function, it is helpful to work out what amounts to a large gain. In particular, since the gain is the model's metric of the extent of unanchoring, from what size of the gain should one consider expectations to be significantly unanchored? The consensus in the literature on estimating learning gains is that if the true model is one with constant gain learning, then the gain lies between 0.01-0.05. On the higher end of the empirical estimates, Branch and Evans (2006) obtain a constant gain on inflation of 0.062. Milani (2007) finds 0.0183. The estimates of the maximal value for endogenous gains lie somewhat higher, at 0.082 in Milani (2014) and at 0.145 in Carvalho et al. (2019). Calibrated models tend to use the consensus values between 0.01-0.05, with the number 0.05 having attained particular prominence. However, Eusepi and Preston (2011) find that the value of 0.002 is sufficient to significantly alter the dynamic behavior of a standard RBC model.

An intuitive interpretation of the gain is that its inverse gives the number of past observations the private sector uses to form its current forecasts. Eusepi and Preston (2011)'s number, 0.002, thus implies that firms and households rely on the last 125 years of data. By contrast, the consensus number of 0.05 translates to using five years of data.

Seen from both of these perspectives, a gain of 0.05 corresponds to a sizable amount of unanchoring. With this benchmark number in mind, the message of Fig. 2 seems stark: the estimated coefficients are very large. The highest gain value in my estimation, 0.82, can be interpreted as seeing forecast errors of 4 pp in absolute value prompting the private sector to discount any observations older than about 5 months. That is a very short time.

Presumably, such large forecast errors are rare, however. In fact, Fig. 3 confirms that this is indeed the case in the SPF data since the 1990s. As shown on the left panel, in this time period, forecast errors were always below 4 percentage points in absolute value, the largest being an error of -3.31 pp. The right panel depicts the time series for the gain implied by the model and the estimated $\hat{\gamma}$. The mean and median values

Figure 3: Forecast errors and implied gain in the SPF data



Forecast errors are in annualized percentage points.

of the gain, 0.121 and 0.098 respectively, suggest that there is substantial sensitivity of long-run expectations to forecast errors overall in the data.

Clearly, the main feature of the gain is that it is time-varying. Periods of near-zero values are frequently followed by high values around or above 0.1. The negative inflation surprises brought on by the Great Recession and the Covid-19 crisis unsurprisingly result in very large gains, almost 0.7 and around 0.35 respectively, yielding highly unanchored expectations. More surprising is the fact that the decade-long boom between 2010 and 2020 also involves a very unanchored episode. In fact, the gain in 2015 is above 0.4, exceeding its 2020 value. Moreover, following 2015, the gain remained elevated for almost five years. By contrast, after the Great Recession, expectations recovered fast and by 2010 looked very well-anchored.

On the one hand, this suggests that exogenous shocks such, for example shocks to credit or to oil prices, have similar effects as the Fed missing its inflation target. On the other hand, a potentially more alarming implication is that failing to get inflation back up to target seems to go hand-in-hand with persistently unanchored expectations. Thus central bank misses may result in unanchoring that is more difficult to combat than that following exogenous shocks.

The three sizable unanchoring episodes since 2008 may also stem from a change in the pattern of forecast errors. Up to 2008, most forecast errors were positive and small in magnitude. This reversed in 2008. Not only were most post-2008 forecast errors negative, but they were also larger than the positive surprises. This connects with the finding that unanchoring is asymmetric. Notice that, as seen on Fig. 2, negative inflation surprises raise the gain about twice as much as positive ones do. Consistent with the findings of Hebden et al. (2020), this indicates that long-run expectations are more sensitive to negative than to positive surprises. This implies, then, that not only is the magnitude of forecast errors after 2008 responsible for the spikes in the gain, but also that inflation consistently turned out to be lower than what the public had expected.

1.5 Monetary policy and anchoring

This section sets up and solves the optimal monetary policy problem in the model with the anchoring expectation formation. In Section 1.5.1, I begin by analyzing the Ramsey problem of determining optimal paths for the endogenous variables that policy seeks to bring about. While the anchoring mechanism introduces substantial nonlinearity into the model, it is possible to derive analytically an optimal target criterion for the policymaker to follow. As we shall see, the optimal rule prescribes for monetary policy to act conditionally on the stance of expectations, and will thus be time-varying and state-dependent. In particular, whether expectations are anchored or not matters for the extent to which there is a tradeoff between inflation and output gap stabilization, and also for the volatility cost of getting expectations anchored.

I then turn to the question of how to implement optimal policy. Section 1.5.2 uses global methods to solve for the interest rate sequence that implements the target criterion. I then discuss the properties of the optimal interest rate policy and why it is successful in stabilizing both inflation expectations and inflation.

The optimal interest rate policy will thus be a nonlinear function of all the states in the model - a complicated object to compute. In practice, however, monetary policy is most commonly modeled using time-invariant rules like the Taylor rule that are simple to compute and to communicate to the public. In Section 1.5.3, I therefore restrict attention to Taylor-type feedback rules for the interest rate. I solve for the optimal Taylor-rule coefficient on inflation numerically and investigate how this choice affects the anchoring mechanism. As I expand upon further in Section 1.5.3, a Taylor rule with given coefficients involves higher fluctuations in the anchoring model than under rational expectations because it does not allow the central bank to respond to long-run expectations. At the same time, it involves responding to inflation even in periods when expectations are anchored, causing excess volatility. For this reason, the optimal Taylor-rule inflation-coefficient is lower than under rational expectations.

1.5.1 The Ramsey policy under anchoring

I assume the monetary authority seeks to maximize welfare of the representative household under commitment. As shown in Woodford (2003), a second-oder Taylor approximation of household utility delivers a central bank loss function of the form

$$L^{CB} = \mathbb{E}_t \sum_{T=t}^{\infty} \{(\pi_T - \pi^*)^2 + \lambda_x (x_T - x^*)^2\}, \tag{32}$$

where λ_x is the weight the central bank assigns to stabilizing the output gap. In the rational expectations New Keynesian model, λ_x is a function of deep parameters: $\lambda_x = \frac{\kappa}{\theta}$. Just as under rational expectations, one can show that it is optimal to set the central bank's targets, π^* and x^*, to zero. The central bank's problem, then, is to determine paths for inflation, the output gap and the interest rate that minimize the loss in Equation (32), subject to the model equations (10) and (11), as well as the PLM (19), the learning equations (21) and (22), and the anchoring function (23). The full statement of the Ramsey problem under Assumptions 1 and 2 is as follows:

$$
\begin{aligned}
&\min_{\{\pi_t, x_t, i_t, \bar{\pi}_{t-1}, k_t\}_{t=t_0}^{\infty}} \mathbb{E}_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} (\pi_t^2 + \lambda_x x_t^2) \qquad \text{s.t.} \\
&x_t = -\sigma i_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} \beta^{T-t} ((1-\beta) x_{T+1} - \sigma(\beta i_{T+1} - \pi_{T+1}) + \sigma r_T^n), \\
&\pi_t = \kappa x_t + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} (\kappa\alpha\beta x_{T+1} + (1-\alpha)\beta\pi_{T+1} + u_T), \\
&\hat{\mathbb{E}}_t \pi_{t+k} = \bar{\pi}_{t-1} + b_1^k s_t \quad \forall k \geq 1, \\
&\bar{\pi}_t = \bar{\pi}_{t-1} + k_t f_{t|t-1}, \\
&k_t = \mathbf{g}(f_{t|t-1}).
\end{aligned} \tag{33}
$$

1.5.1.1 Optimal Ramsey policy as a target criterion

As foreshadowed above, the nonlinearity of the model due to the anchoring function prevents a full analytical solution to the Ramsey problem. Therefore I now characterize the first-order conditions of the problem analytically, and proceed in Section 1.5.2 to solve the full problem numerically. The details of the derivations are given in Appendix A.8, which also illustrates how the endogeneity of the gain introduces nonlinearity into the model.

The solution of the Ramsey problem is stated in the following proposition.

Proposition 1. *Target criterion in the anchoring model*

The targeting rule in the simplified learning model with anchoring is given by

$$
\pi_t = -\frac{\lambda_x}{\kappa} x_t + \frac{\lambda_x}{\kappa} \frac{(1-\alpha)\beta}{1-\alpha\beta} \Big(k_t + f_{t|t-1} \mathbf{g}_{\pi,t} \Big) \Big(\mathbb{E}_t \sum_{i=1}^{\infty} x_{t+i} \prod_{j=0}^{i-1} (1 - k_{t+1+j} - f_{t+j|t+j-1} \mathbf{g}_{\bar{\mathbf{ß}},\mathbf{t+j}}) \Big). \tag{34}
$$

Proved in Appendix A.8. For a general target criterion without Assumption 2, see Appendix A.9.

Here, $\mathbb{E}_t$ is the central bank's expectation. The notation reflects the assumption that the central bank has model-consistent expectations and observes the private sector's expectations. This assumption, which Gaspar et al. (2010) refer to as "sophisticated central banking," is quite strong. In practice, it is likely that the central bank's measure of private sector expectations is noisy at best. Nevertheless, it is a useful case that can serve as a benchmark for future research.

The interpretation of Equation (34) is that the *intra*temporal tradeoff between inflation and the output gap due to cost-push shocks is complemented by two *intertemporal* tradeoffs: one due to learning in general, and one due to anchoring in particular. The first intertemporal effect comes from the current level of the gain, k_t, which captures how far learning is from converging to rational expectations. The second intertemporal tradeoff is manifest in the derivative of the anchoring function today, $\mathbf{g}_{\pi,t}$, as well as in all expected levels and changes in the gain in the future in the expression $(1 - k_{t+1+j} - f_{t+j|t+j-1}\mathbf{g_{\bar{\beta},t+j}})$ in the second bracket on the right-hand side. These expressions say that the presence of anchoring qualify the first intertemporal tradeoff because now the degree and direction in which the gain changes today and is expected to change in the future matter too. In other words, the central bank needs to consider whether its chosen interest rate sequence contributes to anchoring expectations in future periods, or whether it actually serves to unanchor them.

Let me investigate these channels in isolation. To see exactly what the role of anchoring is in the target criterion, consider first the special case of exogenous gain adaptive learning, for simplicity with a constant gain specification. In this case the anchoring function and the forecast error are irrelevant (since $\mathbf{g_i} = 0, i = \pi, \bar{\pi}$) and (34) boils down to

$$\pi_t = -\frac{\lambda_x}{\kappa}x_t + \frac{\lambda_x}{\kappa}\frac{(1-\alpha)\beta}{1-\alpha\beta}k\left(\sum_{i=1}^{\infty} x_{t+i}(1-k)^i\right), \tag{35}$$

which is the analogue of Gaspar et al. (2010)'s Equation (24).[12] This result, found also by Molnár and Santoro (2014), suggests that already the presence of learning by

[12] Gaspar et al. (2010) provide a parsimonious summary of Molnár and Santoro (2014).

itself is responsible for the first intertemporal tradeoff between inflation and output gap stabilization. The fact that the central bank now has future output gaps as a margin of adjustment means that it does not have to face the full tradeoff in the current period. Learning allows the central bank to improve the current output gap without sacrificing inflation stability today; however, this results in a worsened tradeoff in the future. In other words, adaptive learning by itself allows the central bank to postpone the current tradeoff to later periods.

Intuitively, this happens because adaptive expectations are slow in converging to rational expectations. In the transition, the private sector's expectations do not adjust to fully internalize the intratemporal tradeoff. This gives the monetary authority room to transfer the tradeoff to the future.

Contrasting Equations (35) and (34) highlights the role of my novel anchoring channel. With anchoring, the extent to which policy can transfer the intratemporal tradeoff to future periods depends not only on the stance of the learning process, as in (35), but also on whether expectations are anchored, and in which direction they are moving. In fact, not only the current stance and change of anchoring matters, but also all expected future levels and changes.

Anchoring, however, complicates the possibility of transferring today's tradeoff to the future. One can see this in the fact that forecast errors and the derivatives of the anchoring function are able to flip the sign of the second term in (34). This means that anchoring can alleviate or worsen the intertemporal tradeoff. To see the intuition, consider the equation system of first-order conditions from solving the Ramsey problem. While the full system is presented in Appendix A.8, let us focus solely on Equation (A.32), the equation governing the dynamics of observables in the model:

$$2\pi_t = -2\frac{\lambda_x}{\kappa}x_t + \varphi_{5,t}k_t + \varphi_{6,t}\mathbf{g}_{\pi,t}. \tag{36}$$

The Lagrange multipliers $\varphi_5 \geq 0$ and $\varphi_6 \geq 0$ are the multipliers of the updating equation (21) and the anchoring function respectively. This equation, upon substitution of the solutions for the two multipliers, yields the target criterion. It is therefore easy to read off the intuition at a glance. First, since $\varphi_{5,t} k_t > 0$, one immediately obtains the above-discussed conclusion that as long as the adaptive learning equation is a constraint to the policymaker ($\varphi_{5,t} > 0$), the central bank has more room to transfer the contemporaneous tradeoff between inflation and the output gap to the future.[13]

However, whether the anchoring equation alleviates or exacerbates the inflation-output gap tradeoff depends on the sign of $\mathbf{g_{\text{ß},t}}$. If the derivative is positive, the effect is the same as above, and the central bank has more leeway to postpone the tradeoff to the future. By contrast, if the derivative is negative, that is expectations are becoming anchored, the intratemporal tradeoff is worsened.

Why do unanchored expectations give the central bank the possibility to postpone its current inflation-output gap tradeoff? The reason is that when expectations become unanchored, the learning process is restarted. A not-yet converged learning process implies, as discussed above, that postponing the tradeoff is possible. Restarting the convergence process thus unlocks this possibility.

This seems to suggest that from a smoothing standpoint, the central bank should prefer to have unanchored expectations. As will be shown in Sections 1.5.2-1.5.3, volatility considerations will suggest otherwise. But in fact, even the smoothing viewpoint involves some ambiguity on whether expectations should be anchored from the perspective of the central bank. Clearly, the central bank prefers to face a learning process that on the one hand has not yet converged, and on the other is converging only slowly. A high gain under unanchored expectations implies both a sizable distance from con-

[13]Strictly speaking, φ_5 and φ_6 are never zero in this model. The reason is that exogenous disturbances induce unforecastable variation throughout the lifetime of the economy. Thus, at any point in time, forecast errors can emerge that will unanchor expectations, restarting the learning process. This is in stark contrast with Carvalho et al. (2019), where the gain only depends on the endogenous component of the forecast error. Therefore, in their model, absent regime switches, expectations can never become unanchored once learning has converged.

vergence as well as faster learning and thus faster convergence. Therefore, ideally the central bank would like to have expectations anchored but the gain far from zero; a contradiction. Once the gain approaches zero, only unanchored expectations can raise it again to restart the learning process. But once the gain is large, the only way to slow down learning is to anchor expectations, that is, to lower the gain.

1.5.1.2 Time-consistence of optimal plans under adaptive learning

Now simplify the target criterion further, assuming that learning has converged, $k_t = \mathbf{g}_{\mathbf{ß}} = 0$. We are left with

$$\pi_t = -\frac{\lambda_x}{\kappa} x_t, \tag{37}$$

which corresponds to the optimal discretionary solution for rational expectations in Clarida et al. (1999). This is formalized in the following proposition.

Proposition 2. *Coincidence of commitment and discretion under adaptive learning*

In an adaptive learning model with exogenous or endogenous gain, the optimal Ramsey policies under commitment and discretion coincide. The optimal Ramsey plan is more akin to discretion than to commitment as it does not involve making promises about future policy actions. Optimal policy is thus not subject to the time-inconsistency problem of Kydland and Prescott (1977).

To illustrate this result in a parsimonious manner, consider a simplified version of the model. The planner chooses $\{\pi_t, x_t, e_t, k_t\}_{t=t_0}^{\infty}$ to minimize

$$\mathcal{L} = \mathbb{E}_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} \Big\{ \pi_t^2 + \lambda x_t^2 + \varphi_{1,t}(\pi_t - \kappa x_t - \beta e_t + u_t)$$
$$+ \varphi_{2,t}(e_t - e_{t-1} - k_t(\pi_t - e_{t-1})) + \varphi_{3,t}(k_t - \mathbf{g}(\pi_t - e_{t-1})) \Big\},$$

where the IS curve, $x_t = \mathbb{E}_t x_{t+1} + \sigma e_t - \sigma i_t + \sigma r_t^n$, is a non-binding constraint, and is therefore excluded from the problem. φ_i are Lagrange-multipliers and $\mathbb{E}_t x_{t+1}$ is rational. In this simplified setting, e_t is a stand-in variable capturing inflation expectations and evolves according to a recursive least squares algorithm. The anchoring function $\mathbf{g}(\cdot)$ specifies how the gain k_t changes as a function of the current forecast error according to Assumption 2. Note that the problem involves commitment because the monetary authority internalizes the effects of its actions both on the evolution of expectations and on that of the gain.

After some manipulation, first-order conditions reduce to:

$$2\pi_t + 2\frac{\lambda_x}{\kappa}x_t - \varphi_{2,t}(k_t + \mathbf{g}_{\mathbf{\beta}}(\pi_t - e_{t-1})) = 0, \tag{38}$$

$$-2\beta\frac{\lambda_x}{\kappa}x_t + \varphi_{2,t} - \mathbb{E}_t\,\varphi_{2,t+1}(1 - k_{t+1} - \mathbf{g}_{\mathbf{f}}(\pi_{t+1} - e_t)) = 0. \tag{39}$$

Inspection of this system reveals that, unlike the rational expectations case ($e_t = \mathbb{E}_t\,\pi_{t+1}$), the optimal solution does not involve lagged multipliers. This implies that the monetary authority cannot condition the optimal time path of inflation and the output gap on the past; optimal policy is not history-dependent.

This result is not specific to endogenous gain models, but generalizes to any adaptive learning. It has been demonstrated in the case of constant gain learning by Molnár and Santoro (2014), and for decreasing gain models by Mele et al. (2019).

The intuition behind this result is easy to see if one compares rational expectations and adaptive learning in an infinitely repeated game setting. Under rational expectations, the lagged multiplier appears in the solution because expected inflation is a jump variable. This reflects that expectations fulfill a form of optimality, rendering the private sector a strategic player. The adjustment of expectations under rational expectations enables the central bank to make promises about future policy that are incorporated into expectations.

Not so for adaptive expectations. Learning agents look exclusively to past data to form expectations. Their expectations thus cannot incorporate the policymaker's promises about the future course of policy. In fact, a private sector with adaptive expectations has a pre-specified, non-strategic expectation formation. Therefore, households and firms act as an automaton, leaving the central bank unable to make promises that have any effect on expectations.

Should we be surprised that adaptive learning involves no distinction between discretion and commitment? Not at all if we recall that the rational expectations revolution had as one of its aims to remedy this feature of expectations. The seminal Lucas-critique, for instance, emphasizes that reduced-form regressions are not ideal guiding principles for policy precisely because they miss the time-varying nature of estimated coefficients due to model-consistent expectations that incorporate promises about policy action (Lucas (1976)).

One might be concerned that a model of anchoring that is not immune to the Lucas-critique may not be a desirable normative model for policy. But recall that the anchoring model is a description of the economy in transition, not of one at its ergodic mean. It characterizes expectation formation en route to becoming model-consistent as the private sector learns the underlying DGP.

Moreover, data suggest that in terms of positive implications, learning models fare much better than rational expectations models do. Since the seminal work by Coibion and Gorodnichenko (2015), rational expectations as a structural model for expectations has been rejected in numerous empirical papers. At the same time, the literature cited in the Introduction demonstrates the success of learning models in fitting both the properties of expectations in the data and in improving the business cycle dynamics of other model variables. To the extent that adaptive expectations offer an alternative structural model of the expectations process that captures main features of the data, in particular the feature of unanchoring, investigating their role for policy is a valuable exercise.

There are also avenues to address the Lucas-critique in learning models. One option is to reintroduce a sense of optimality to expectation formation directly, as in the literature on central bank reputation. (See Cho and Matsui (1995) and Ireland (2000)). Another possible remedy is to retain a sufficient degree of forward-looking expectations as in the finite-horizon planning approach advocated by Woodford (2019). A third possibility is to model communication by the central bank in the form of news shocks that enter the state vector, and thus show up in the information set of agents, as in Dombeck (2017). All in all, Proposition 2 does not render the advances of the rational expectations revolution void. Instead, it points to the fact that as long as the private sector extracts information from data to form beliefs, expectations will quantitatively resemble adaptive expectations. Thus, the policy predictions of adaptive expectations can complement our understanding of monetary policy from rational expectations with novel, empirically relevant insights.

1.5.2 Implementing the Ramsey policy: the optimal interest rate sequence

Having a characterization of optimal policy in the anchoring model as a first-order condition, the next relevant question is how the central bank should set its interest rate tool in order to implement the target criterion in (34). In other words, we would like to know what time-path of interest rates implements the optimal sequence of inflation and output gaps. As emphasized in Section 1.5.1, the nonlinearity of the model does not admit an analytical answer to this question. I therefore solve for the optimal interest rate policy numerically using global methods. I rely on the calibration presented in Tables 1 and 2 and the estimated parameters of the expectations process in Section 1.4. Furthermore, I set the central bank's weight on output gap fluctuations, λ_x, to 0.05, the value estimated by Rotemberg and Woodford (1997).

Appendix A.10 outlines my preferred solution procedure, the parameterized expectations approach, while Appendix A.11 gives the details of the parametric value function iteration approach I implement as a robustness check. The main output of this procedure is an approximation of the optimal interest rate policy as a function of the vector of state variables. Due to Assumption 2, the relevant state variables are expected mean inflation and the exogenous states at time t and $t-1$, rendering the state vector five-dimensional:

$$X_t = (\bar{\pi}_{t-1}, r_t^n, u_t, r_{t-1}^n, u_{t-1}). \tag{40}$$

As a first step, I plot how the approximated policy function depends on $\bar{\pi}_{t-1}$, while keeping all the other states at their mean. The result, depicted on Panel (a) of Fig. 4, suggests that optimal interest-rate setting responds linearly and very sensitively to the stance of expectations, $\bar{\pi}_{t-1}$. If expected mean inflation decreases by 5 basis points, the interest rate drops by about 250 basis points.[14]

This is a large response. Clearly, optimal policy involves subduing unanchored expectations by injecting massive negative feedback to the system. One may then wonder why optimal policy is so aggressive on unanchored expectations when the analysis of the target criterion in Section 1.5.1 suggested that learning can alleviate the stabilization tradeoff between output and inflation.

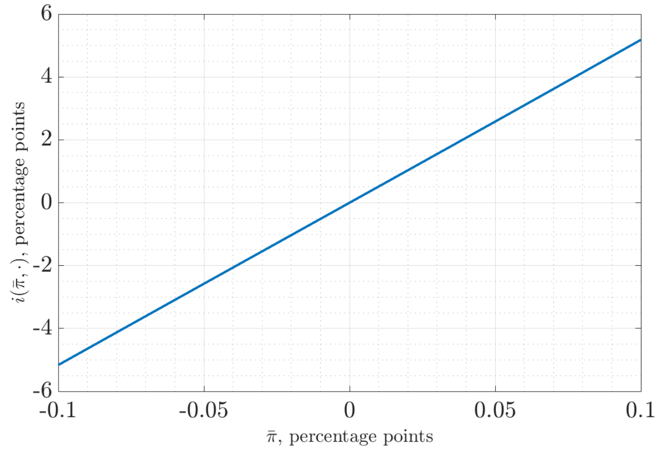
The reason is that the anchoring expectation formation introduces another intertemporal tradeoff to monetary policy: a volatility tradeoff. One can see this on Fig. 5, portraying the dynamics of the system following a two-standard-deviation inflationary cost-push shock, conditional on a Taylor rule with baseline parameters. The figure contrasts the rational expectations version of the model with a well-anchored, weakly anchored and strongly unanchored scenario in the anchoring model.

The same shock that under rational expectations completely vanishes by the second period, triggers a large, persistent and oscillatory response in the anchoring model.[15]

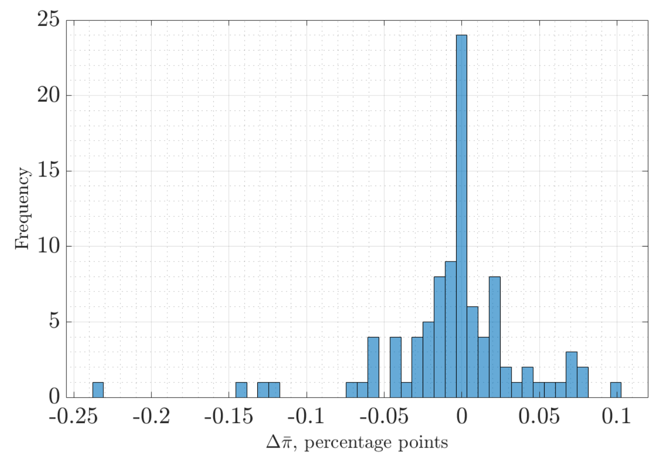
[14]This result is qualified if one relaxes Assumption 1. If the private sector learns about all observables, not just inflation, interest rate expectations play a major role in determining forecasts of future inflation and thus add a stabilizing channel that is absent from the current specification. In the more general case, then, an order of magnitude smaller responses in the current interest rate are sufficient.

[15]As periodically noted in the literature, adaptive learning models tend to produce impulse responses that exhibit damped oscillations. Authors making explicit note of this phenomenon include Evans and

Figure 4: Policy function and implied volatility in long-run expectations



(a) $i(\bar{\pi}, \text{all other states at their means})$

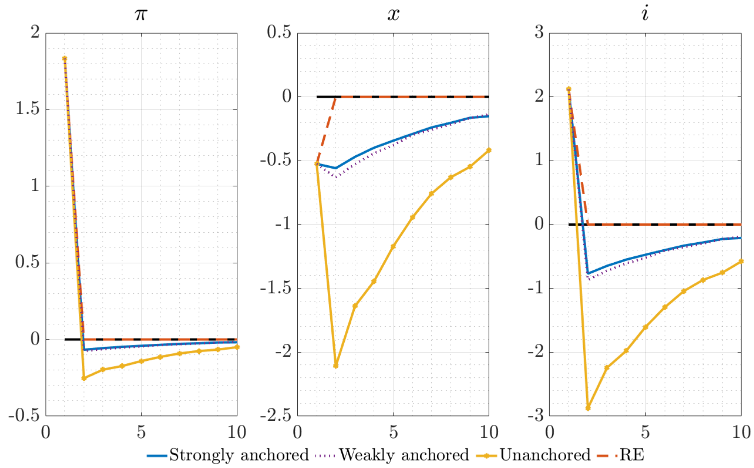


(b) Histogram of changes in $\bar{\pi}$

Clearly, the size and persistence of the shock, as well as the magnitude of oscillations increase the more unanchored expectations are. This comes from the fact that if expectations are anchored, stable expectations lower the pass-through between shocks and observables. Instead, if expectations are unanchored, they become volatile, passing through the shocks and amplifying them. Having unanchored expectations, then, comes at a volatility cost in the central bank's target variables. This volatility cost dictates that, in the long run, the central bank wishes to have expectations anchored.

Honkapohja (2001), Evans et al. (2013) and Anufriev and Hommes (2012). The reason is that under an adaptive learning framework, forecast errors following an impulse are oscillatory. In fact, the higher the learning gain, the higher the amplitude of forecast error oscillations. Appendix A.12 presents a simple illustration for why this is the case.

Figure 5: Impulse responses after a cost-push shock



Shock imposed at $t = 25$ of a sample length of $T = 400$ (with 5 initial burn-in periods), cross-sectional average with a cross-section size of $N = 100$. The remark on whether expectations are anchored or not refers to whether the gain is below the 10^{th}, in the neighborhood of the 50^{th}, or above the 90^{th} percentile of simulated gains at the time the shock hits.

From the viewpoint of the central bank, this problem is amplified by the fact that anchoring expectations itself comes at a convex volatility cost. Anchoring expectations requires an aggressive interest response because by these means the central bank can introduce negative feedback to the system. But innovations to the interest rate surprise the private sector, raising forecast errors. The more unanchored expectations are, the more volatility the interest rate movement inflicts on the economy.

We can see from the policy function how optimal policy resolves this tradeoff: it reacts extremely aggressively to movements in long-run expectations. This way, the central bank hopes to avoid even larger interventions that would become necessary were expectations to unanchor further. To avoid having to pay so high a price, the central bank is extra aggressive in the short run to prevent massive unanchoring from ever materializing. Thus the optimal response to the volatility tradeoff is to temporarily increase volatility in order to reduce it in the long-run. In this way, the central bank's aggressiveness in the model is driven by the desire to prevent upward (downward) drifting long-run expectations from becoming a self-fulfilling inflationary (deflationary) spiral.

The large interest-rate responses resemble the idea advocated by Goodfriend (1993) that the central bank moves to offset "inflation scare" (or "deflation scare") episodes. As Goodfriend shows, it was historically not uncommon to move the interest rate by hundreds of basis points to subdue inflation scares. For example, in March 1980, the Fed raised the interest rate by 230 bp to convince the public that it would not tolerate high inflation. The optimal policy function prescribes that this is exactly what the policymaker should do to fight unanchored expectations.

What Fig. 4 also suggests is that acting aggressively in the short run indeed delivers the long-run benefits of stabilizing economic fluctuations. As Panel (b) depicts, employing the optimal policy implies that realized changes in long-run inflation expectations are very small. As seen on the histogram of $\Delta\bar{\pi}$, with the optimal policy in place, the model spends most of its time in the region of minuscule fluctuations in $\bar{\pi}$. The mode of the distribution is a change of 0.3 basis points in absolute value, implying that in normal times, the central bank only needs to raise or lower the interest rate by 15 basis points. In other words, the aggressive nature of optimal policy allows the central bank to keep expectations anchored or quickly reanchor them following shocks. In this way, the monetary authority eliminates as much volatility stemming from unanchored expectations as it possibly can.

1.5.3 Optimal Taylor rule under anchoring

Monetary policy is often formulated using a Taylor rule. Proponents of such a characterization, like Taylor (1993) himself, emphasize the benefits of having a simple, time-invariant and easily verifiable rule. Also in the anchoring model, a policymaker may thus be interested in using a Taylor-type approximation to optimal policy in order to combine the benefits of having a simple, yet near-optimal rule.[16] Therefore I

[16] Recall from Woodford (2003) that even under rational expectations, a standard Taylor rule is not fully optimal because its purely forward-looking nature precludes the use of promises of future policy. Since Proposition 2 tells us that in the anchoring model there is no distinction between commitment and discretion, this point may be less relevant here than for rational expectations.

now consider the restricted set of Taylor-type policy rules and ask what value of the time-invariant Taylor-rule coefficient on inflation is optimal in the case of the anchoring model.

In this section, I thus restrict attention to a standard Taylor rule:

$$i_t = \psi_\pi(\pi_t - \pi^*) + \psi_x(x_t - x^*) + \bar{i}_t, \tag{41}$$

where ψ_π and ψ_x represent the responsiveness of monetary policy to inflation and the output gap respectively. Lastly, $\bar{i}_t$ is a monetary policy shock. I also assume that when the Taylor rule is in effect, the central bank publicly announces this. Thus Equation (41) is common knowledge and is therefore not the object of learning. In an extension, I consider the case where the Taylor rule is not known (or not believed) by the public and therefore is learned together with the relations (10) and (11). This only changes model dynamics for a short period of time as long as the Taylor rule has not yet been learned. In the approximately five quarters where the public's learning is in progress, the model's responses are dampened; afterwards, the model dynamics are identical to those of the baseline.

I compute the optimal Taylor rule coefficient on inflation numerically by minimizing the central bank's expected loss in a cross-section of $N = 100$ simulations of both the rational expectations and learning versions of the model. I continue to use the calibration of Tables 1 and 2 and to parameterize the anchoring function using the estimated $\hat{\gamma}$ from Section 1.4.

Table 3 presents the optimal Taylor rule coefficient ψ_π for the rational expectations and anchoring models. The table also compares the baseline parameterization with an alternative in which the central bank attaches no weight to output gap stabilization. One notices that if the central bank has no concern to stabilize the output gap ($\lambda_x = 0$), ψ_π is infinity for RE, but strictly below infinity for the anchoring model. For the rational expectations version of the model, this is because if the central bank suffers no loss upon output variation, then the fact that the divine coincidence is violated does not pose a problem. An infinite inflation coefficient then allows the authority to eliminate inflation fluctuations altogether.

Table 3: Optimal coefficient on inflation, RE against anchoring for alternative weights on output

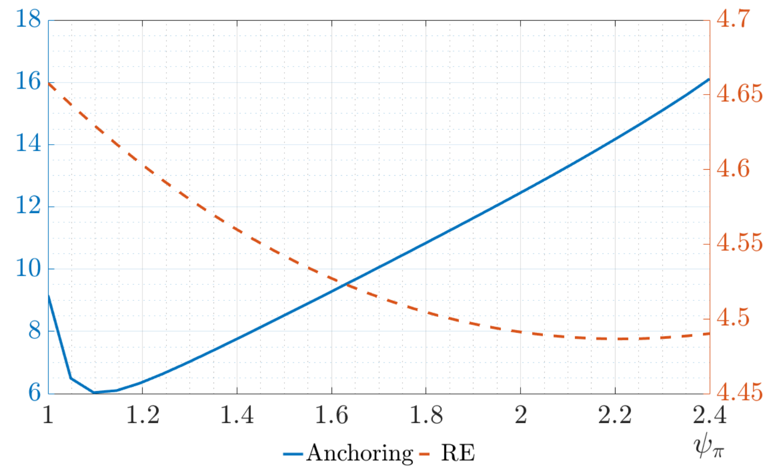
	$\psi_\pi^{*,RE}$	$\psi_\pi^{*,Anchoring}$
Baseline ($\lambda_x = 0.05$)	2.2101	1.1083
$\lambda_x = 0$	∞	1.4421

Not so for the anchoring model. Even for $\lambda_x = 0$, the optimal inflation coefficient is below infinity. Also for the baseline calibration, the monetary authority finds it optimal to choose a significantly lower ψ_π in the anchoring model than under RE. Why this is the case can be gleaned from Fig. 6, which depicts the central bank's loss as a function of the inflation coefficient ψ_π for the baseline calibration.

As seen on Fig. 6, both models predict a sharp increase in the loss as ψ_π is lowered for low values of ψ_π. This is intuitive: the lower ψ_π, the more inflation fluctuation the central bank tolerates. This leads to losses in both models. In the anchoring model, the loss is further increased by the fact that inflation volatility leads to higher forecast errors. This implies higher fluctuations in expectations, which in turn feed back into inflation. In this manner, the positive feedback loop in the anchoring model in general leads to higher inflation fluctuations in the anchoring model than in the RE version, resulting in a higher loss as well.

But Fig. 6 has a more surprising implication too. As opposed to RE, the loss is strongly convex in the anchoring model. In fact, this remains the case also if the central bank has no concern for output gap stabilization ($\lambda_x = 0$). In the anchoring model, then, the losses caused by tolerating too much inflation volatility cannot be eliminated by raising ψ_π infinitely. This means that beyond a threshold, raising ψ_π

Figure 6: Central bank loss as a function of ψ_π



Sample length is $T = 100$ with a cross-section of $N = 100$.

actually increases inflation volatility, instead of decreasing it. For this reason, as seen in Table 3, the optimal choice involves a much smaller inflation coefficient under anchoring than under RE (1.1 instead of 2.2). How does this come about?

The mechanism is the following. As the IS curve of Equation (10) makes explicit, the private sector relies on expectations of not just future inflation and output gaps, but also of interest rates when choosing its actions today. A high ψ_π, together with the assumption that the private sector knows the Taylor rule, means that if conditions today cause the private sector to expect high inflation in the future, the private sector will internalize future policy responses and therefore also expect high interest rates down the line. This implies a shift in the entire term structure of expectations, adding fuel to the positive feedback between expectations and outcomes. But not only that. The fact that agents can anticipate future policy actions induces oscillatory expectations, as adverse shocks today are expected to be offset by expansionary policy, et vice versa.[17] In the case of the private sector internalizing the policymaker's reaction function, then, anticipation effects of future interest rates play a key role in determining the conduct

[17] In a more general specification of the model where the private sector does not initially know the Taylor rule, impulse responses take time to become oscillatory because as long as the Taylor rule is not yet learned, intertemporal anticipation effects cannot play out. But eventually as the Taylor rule is internalized, impulse responses oscillate exactly as in the present version of the model.

of policy.

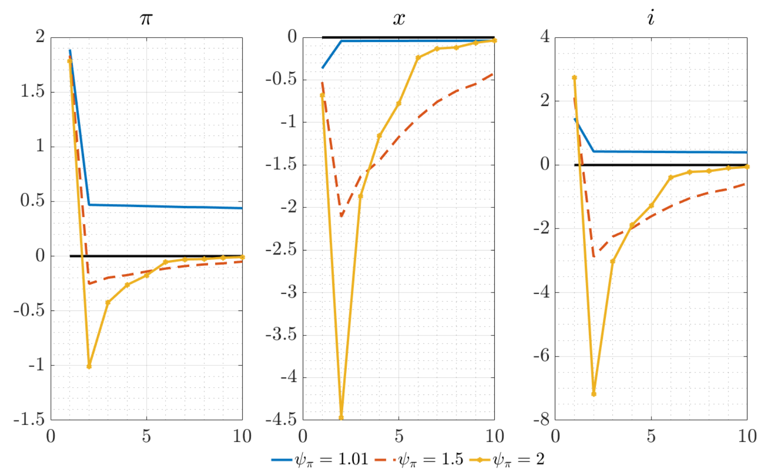
We can see how this plays out by contrasting impulse responses of the model for various levels of inflation aggressiveness ψ_π, depicted on Fig. 7, again for an inflationary cost-push shock. A small value of $\psi_\pi = 1.01$ imposes too little negative feedback to get inflation back to its pre-shock value fast. A too high value of ψ_π has the contrary effect: the impulse responses overshoot. In fact, the higher ψ_π, the larger the overshooting. This highlights the anticipation effects of future interest rates. When the contractionary monetary policy shock hits, inflation and the output gap fall on impact. By the next period, expectations update, initially implying lower future inflation and output gaps. But an internalized Taylor rule means that the private sector connects low future inflation with expansionary policy going forward. This expectation causes the overshoot.

The model dynamics here resemble the predictions of Ball (1994) of expansionary disinflation. Ball (1994) observes that an overlooked implication of rational expectations New Keynesian models is that disinflations are expansionary. Since the prediction that disinflation is costless, indeed expansionary, is at odds with data, Ball concludes that central bank announcements must suffer from credibility issues that render expectations unresponsive to the announcement of a coming disinflation.

Note that in the present context, as seen on Fig. 5, the amplitude of the oscillations depends positively on the size of the gain. In other words, the more unanchored expectations are, the more impulse responses look as Ball (1994) predicts: we obtain an expansionary disinflation. This feature indicates that the key difference between Ball's interpretation and my model is what channel generates the unresponsiveness of expectations. Thus I arrive at a different conclusion than Ball (1994); instead of credibility issues, it is anchored expectations that are responsible for the absence of expansionary disinflations.

In contrast to the conventional wisdom in the adaptive learning literature, as ex-

Figure 7: Impulse responses for unanchored expectations for various values of ψ_π



Cost-push shock imposed at $t = 25$ of a sample length of $T = 400$ (with 5 initial burn-in periods), cross-sectional average with a cross-section size of $N = 100$.

emplified for example by Orphanides and Williams (2004), I thus find that monetary policy specified as a Taylor rule should be *less* aggressive on inflation than what would be optimal under rational expectations. Technically speaking, the difference stems from two sources. The first is whether one models the aggregate relationships of the model as corresponding to the learning assumptions concerning individual agents. As suggested in Section 1.3, the majority of the papers in the learning literature consider Euler-equation learning. This implies that long-horizon expectations do not enter the IS and Phillips curves.

By contrast, relying on Preston (2005)'s long-horizon learning approach reintroduces the full term structure of expectations into the IS and Phillips curves and thus assigns a role to interest-rate expectations. Eusepi et al. (2018) is the first to use the long-horizon approach to investigate the interaction between long-horizon learning and interest-rate expectations. Unsurprisingly, they reach the same conclusions as I do: anticipation effects coming from interest-rate expectations force the central bank to be less aggressive on inflation than it would be under rational expectations.

Secondly, the anchoring mechanism is responsible for the lower optimal value of the inflation coefficient. Standard specifications of the gain in the adaptive learning tradition involve no channel through which monetary policy could affect the learning of agents. Whether the gain is constant or deterministically decreasing, there is no link between the interest-rate setting of the central bank and the gain.

Here, however, the anchoring mechanism relates the gain to forecast errors. Forecast errors, in turn, arise as a function of central bank action, in particular of whether the central bank is able keep inflation at the private sector's perceived mean. Movements in the interest rate, then, are a double-edged sword: in contributing to get inflation to the target, they can close forecast errors and thus stabilize long-run expectations. But any central bank intervention also results in adjustments in expectations (either because it takes the private sector by surprise, or because the private sector understands its implications for future policy), increasing the volatility of observables.

This highlights that the time-varying nature of optimal policy is its key characteristic. The ability to take the stance of anchoring into account is what enables the central bank to be exceedingly aggressive if and only if expectations are about to unanchor, and maintain a dovish stance otherwise. Thus it seems reasonable to expect that if the central bank were allowed to select time-varying Taylor-rule coefficients, it would choose low coefficients when expectations are anchored and high ones when expectations show signs of unanchoring. Such a central bank, as well as one following the optimal policy, would appear to the econometrician as time-varying, echoing the findings of Lubik and Matthes (2016).

1.6 Conclusion

Central bankers frequently voice a concern to anchor expectations, that is, to render expectations of long-run inflation unresponsive to short-run economic conditions. The contribution of this paper is to lay out a simple behavioral model which captures the time-varying sensitivity of long-run expectations to short-run conditions and to investigate how this affects the conduct of monetary policy. I quantify my novel anchoring mechanism by estimating the form of the function that determines the degree of unanchoring. I use the model to characterize monetary policy both analytically and numerically.

The simulated method of moments estimation establishes that the anchoring function provides a realistic description of expectations. On the one hand, expectations become more sensitive to forecast errors when the private sector made larger mistakes in predicting inflation in the past. On the other hand, like in Hebden et al. (2020), negative mistakes unanchor expectations more than positive ones of the same magnitude do. Moreover, I use the estimated anchoring function to back out an implied time series for the gain, the model's metric for the degree of unanchoring. The median gain, 0.098, is in line with estimates in the literature such as Milani (2014) and Carvalho et al. (2019), and implies that agents in the model discount observations older than 10 quarters when updating their long-run expectations.

Using the thus quantified model, I provide three sets of results on monetary policy. I first consider the Ramsey policy of the central bank, deriving an analytical target criterion that prescribes how the monetary authority should respond to shocks. I show that the presence of my novel anchoring channel makes it desirable and feasible to smooth out shocks over time. However, the extent this is feasible varies over time in tandem with the current and expected future degree of unanchoring.

Second, I use global methods to solve the nonlinear system of first-order conditions of the Ramsey problem numerically. I thus obtain an approximation to the optimal policy function, providing the fully optimal path of interest rates conditional on the sequence of exogenous disturbances. Like in Goodfriend (1993)'s account of US monetary policy, the optimal policy involves responding aggressively when expectations unanchor in order to suppress the volatility that high degrees of unanchoring cause. By contrast, the central bank need not intervene when expectations are anchored, resulting in accommodating inflation fluctuations in this case.

Lastly, I explore the implications of the model for the most common specification of monetary policy, the Taylor rule. As my numerical solution for the optimal inflation coefficient demonstrates, the central bank should be less aggressive on inflation in my model than under rational expectations. The first reason is that a time-invariant Taylor rule involves the same response to a given movement in inflation regardless of the degree of unanchoring. Second, as in Eusepi et al. (2018), interest-rate expectations impose limits on how aggressive the policymaker can be. Therefore, the central bank induces additional volatility into the model if it responds too aggressively to fluctuations in inflation. An optimal inflation coefficient in my model is able to eliminate the majority of such additional volatility.

A number of interesting questions emerge from the analysis of monetary policy and the anchoring expectation formation. One may wonder whether the central bank can use tools other than its leading interest rate to anchor expectations. Especially concerns around a binding zero lower bound would motivate the use of alternative monetary policy tools. Thus the interaction between anchoring and central bank communication, in particular forward guidance, would be worthwhile to examine. In future work, I plan to make explicit the communication policy of the central bank to investigate whether the anchoring expectation formation could help to resolve the forward guidance puzzle.

This, however, requires overcoming the implication of Proposition 2 that adaptive expectations are not able to incorporate any information that is not embedded in the current state vector. One option is to model central bank communication similarly to news shocks in the sense of Beaudry and Portier (2006). In this case, the anchoring model is likely to deliver differing predictions regarding the effectiveness of Delphic versus Odyssean forward guidance (Campbell et al. (2012)) because sharing the central bank's forecasts would not constitute a questioning of the interest rate reaction function, while committing to a future interest rate path would.

From this perspective, it is desirable to extend models of adaptive expectations in general to allow the private sector to internalize promises of the policymaker. As discussed in Section 1.5.1, the inability to incorporate communication stems from the fact that adaptive expectations involve a sub-optimal expectation formation. Viewed from the lens of an infinitely repeated game, the private sector is an automaton. One would thus need to reintroduce some notion of optimality into expectation formation that would render the private sector a strategic player. Such an extension would likely be akin to the reputation-building literature à la Cho and Matsui (1995) or to the finite planning-horizon model of Woodford (2019).

Presumably, there are variables other than long-run inflation where expectations exhibit varying degrees of unanchoring. For instance, fixed exchange rate regimes might become subject to speculative attacks the moment when the peg is no longer believed by the public. Exploring the causes and timing of currency crises is thus a natural application of my model which I plan investigate in future research.

In general, extensions to the anchoring expectation formation proposed here would be of interest. The determination of the gain could be endogenized using approaches that allow the private sector to choose its forecasting behavior in an optimizing fashion, perhaps by selecting among competing forecasting models as in Branch and Evans (2011) or by picking the size of the gain to minimize the estimated forecast error variance.

Lastly, a foray into the empirics of anchoring expectation formation is important to get a clearer idea of the expectation formation process of the public. In practice, it is likely that this process is heterogenous, not just across households and firms, but also within various demographic groups or sectors of the economy. If so, then monetary policy would need to collect and monitor a host of long-run expectations time series to manage the challenge of keeping expectations anchored.

2 Talking in Time

- Dynamic Central Bank Communication

2.1 Introduction

Central bankers and researchers have been emphasizing that regular, recurring communication with the public about the current and future economic outlook is fundamental for the transmission of monetary policy. However, the implications of dynamics for central bank communication are less well understood. Where there is a distinction between the current and future economic stance, how should the monetary authority weight the future versus the present in its communication? And in what way does a central banker talk differently when it is understood that she will talk again in the future?

The novelty of this paper is that it formulates a dynamic central bank communication problem and contrasts it with its static analogue. The static and dynamic models involve a Bayesian persuasion-type communication game between a central bank (CS) and the private sector (PS) about two economic fundamentals. The central bank tracks the first fundamental, while the private sector tracks the second. The central bank sends the private sector a noisy signal which is a weighted sum of the two fundamentals. Importantly, the two models are identical up to the correlation structure between the two states: while in the static model, this is a cross-sectional correlation between otherwise unrelated variables, in the dynamic model, instead, the two states are temporally dependent.

This allows me to isolate the role of dynamics for optimal communication design. In particular, I investigate the effect of dynamics on two dimensions of communication. I first ask how strongly the central bank should weight the current against the future fundamental in its signal - a dimension of communication I refer to as "targetedness." Secondly, I explore whether the optimal precision of the dynamic signal differs from that of the static one. In other words, does dynamic communication involve a different amount of noise than static communication?

I find that the main difference between the static and the dynamic problems is the endogeneity of the private sector's prior beliefs in the dynamic setting. The endogenous priors serve to dampen the persuasiveness of the central bank's signal in two crucial ways. First, in terms of targetedness, this implies that the optimal weight on the central bank's preferred state is always larger in the dynamic than in the static model. The key is that in a dynamic setting, the private sector uses its prior in two ways: firstly, to form a prior forecast of its preferred state, and secondly to filter out irrelevant information from the signal it receives today. From the perspective of the central bank, the private sector's optimal choice of how to use the prior always results in the prior dampening the effect of the central bank's signal. In particular, when the correlation between fundamentals is negative, the private sector relies more on the forecasting function of the prior, while when the correlation is positive, it leans more on the filtering function. Therefore the central bank has to weight the signal more heavily in the desired direction.

Second, in the dynamic model, prior beliefs can be too tight from the perspective of the central bank. This lowers the effectiveness of central bank communication because the private sector views the same signal as less informative when its priors are tight. Thus, to increase its persuasive power, the central bank optimally loosens the private sector's priors, and it does so by introducing more noise into the signal than it would in the static model. Optimal precision is thus lower in the dynamic problem than in the static one as the central bank smoothes the information it provides the public over time.

2.1.1 Related literature

The paper is related to two main strands of literature. First, my main point of reference is the global games literature in the vein of Morris and Shin (2002), Svensson (2006), Angeletos and Pavan (2007) and Hellwig and Veldkamp (2009). Several papers have used this literature as a starting point to study particular dimensions of communication. Chahrour (2014), for example, uses the rational inattention literature à la Sims (2003) to investigate the optimal amount of central bank communication.

To my knowledge, only few papers consider the time dimension in some form. One is Gaballo (2016), who analyzes Delphic forward guidance, the communication of the central bank about its own information set, in an overlapping generations (OLG) global games model. However, Gaballo only introduces dynamics in the evolution of the fundamental; central bank communication in his model is simply revealing information about the central bank's one-period-ahead forecast. In this sense, central bank communication in Gaballo (2016) maps one-to-one to the standard static problem of Morris and Shin (2002). Reis (2011) analyzes the optimal timing decision of an authority that knows about a future policy change the public is unaware of. The problem of the authority is to decide when to publicly announce the future policy change. Hansen and McMahon (2016) instead examine how the voting behavior of monetary policy members changes over time as a result of dynamically changing signaling incentives.

My work is also related to the Bayesian persuasion literature in the wake of Kamenica and Gentzkow (2011). Bayesian persuasion is a signaling game where a sender designs his communication so as to persuade a receiver to take a sender-preferred action. Such a setting has been widely adopted in many applications such as stress tests (Goldstein and Leitner (2018) and Inostroza and Pavan (2017)) or even central bank communication (Eunmi Ko's 2019 job market paper[18]). Importantly, the sender has access to a full-commitment technology, and thus persuasion in this setting works through signal design, not through untruthfulness.

Viewed from the lens of commitment, my work ties in closer with the Bayesian persuasion literature than with the macroeconomic literature on discretionary monetary policy and cheap talk such as Moscarini (2007) and Frankel and Kartik (2018). While

[18] This paper is available on Ko's website; see sites.google.com/site/iameunmiko/research.

the discretionary monetary policy literature has an explicit concern for dynamics, the central question in this literature is not how to provide information to the public, but how to design central bank action over time.

I choose to focus on a full-commitment communication problem in the Bayesian persuasion sense exactly because of the Barro and Gordon (1983)-result that reputation arises in equilibrium as a commitment device. With a rational receiver, lying is unsustainable in a dynamic game unless the sender does not care about his reputation. For central banks, where reputation is key for the proper transmission of monetary policy, this is clearly not the case. For this reason, my solution concept is that of a Perfect Bayesian Equilibrium (PBE) instead of a Markov Perfect Equilibrium (MPE).

Lastly, I focus on central bank communication about economic fundamentals, not about future policy. To use the terminology of Campbell et al. (2012), I thus model Delphic, and not Odyssean, forward guidance. This way I can isolate the effect of pure communication instead of considering the implications of the central bank tying its hands concerning the evolution of future policy.

The paper is outlined as follows. Section 2.2 develops the dynamic communication game in parallel with its static analogue. Sections 2.3 and 2.4 describe the targetedness and precision dimensions of the optimal communication policy respectively. Section 2.5 concludes.

2.2 The dynamic model and the static analogue

Suppose there are two correlated random variables. Throughout the paper, I will interpret these as correlated economic fundamentals, for example two realizations of GDP. Let θ_1, θ_2 denote the two random variables in the static environment, and θ_t, θ_{t+1} be the corresponding states in the dynamic world. I assume that

$$(\theta_1, \theta_2) \sim \mathcal{N}(0, \mathbf{V}) \quad \text{with} \quad \mathbf{V} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{42}$$

$$\theta_{t+1} = \rho\theta_t + \varepsilon_{t+1} \quad \text{with} \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad \text{and} \quad \sigma_\varepsilon^2 = 1 - \rho^2. \tag{43}$$

Thus in both static and dynamic models, the two states have zero mean and unit variance. The only difference between the static and dynamic setups is the interpretation of ρ, the correlation between the two variables. While in the static model, ρ captures correlation between two otherwise unrelated states, in the dynamic environment, ρ is a temporal correlation linking the two fundamentals to one another across time.

In this setting, consider a Bayesian persuasion game à la Kamenica and Gentzkow (2011), a communication game between a central bank (CB) and the private sector (PS), in which the CB has the opportunity to design its signals to persuade the PS to choose its actions in a way that maximizes the expected payoff of the CB. Like Kamenica and Gentzkow (2011), I assume that no lying is allowed, so that all signals are truthful.

Suppose the PS is interested in tracking one state, but the CB would like it to track the other. The example underlying this setup is one where the PS embarks on an investment project and is thus interested in the state relevant for the returns on investment. The CB instead has a mandate to smooth out business cycle fluctuations, and is therefore concerned with the state that captures the current stance of the business cycle. Letting I denote the PS' choice of investment, the loss functions of the two players are

$$\mathcal{L}^{PS,static} = \mathbb{E}^{PS}(I - \theta_2)^2, \tag{44}$$

$$\mathcal{L}^{CB,static} = \mathbb{E}^{CB}(I - b\theta_1)^2, \tag{45}$$

in the static model, and

$$\mathcal{L}^{PS,dynamic} = \mathbb{E}_t^{PS}(I_t - \theta_{t+1})^2, \tag{46}$$

$$\mathcal{L}^{CB,dynamic} = \mathbb{E}_0^{CB}\sum_{t=0}^{\infty}\beta^t(I_t - b\theta_t)^2, \tag{47}$$

in the dynamic model, with $b \in [0,\infty)$ denoting the weight which the CB places on tracking θ_1 and θ_t respectively, and $\beta \in (0,1)$ being the central bank's discount factor.

Apart from the temporal structure, the two problems are identical. In both setups, the PS seeks to set investment according to the state that determines the returns on investment, while the CB wishes to follow its mandate of stabilizing the business cycle. The only difference is that the dynamic setting lends this problem a temporal dimension, as the PS' and the CB's target states, θ_{t+1} and θ_t respectively, can be interpreted as different realizations of the same economic fundamental over time. To fix ideas, I will refer to θ_t and θ_1 as "today's state," and θ_{t+1} and θ_2 as "tomorrow's state."

I argue that this temporal dimension is a very natural element of any central bank's communication problem. Firstly, economic fundamentals evolve over time, introducing a distinction between the current and the future economic outlook. This distinction motivates the need for the CB to differentiate what time period it is talking about. Secondly, in a one-shot game it is not intuitive why the PS and the CB would have misaligned preferences. In a dynamic game, instead, the CB's mandate for business cycle stabilization motivates its interest in the current stance of the business cycle, while rationality dictates that the PS pays attention to the future state, since that is what determines the returns on investment. Lastly, in practice the monetary policy decision is a cyclical process, recurring several times per year, every time accompanied by communication about the central bank's outlook on the current and expected future stance of the economy. Thus, in practice, central bankers' communication is dynamic in two ways: it is both *about* a dynamic object - economic fundamentals -, as well as *itself* a temporal phenomenon.

To capture communication that involves both current and future economic fundamentals, I assume that the central bank sends a signal to the private sector of the following form

$$s = \theta_1 + \frac{1}{\psi}\theta_2 + v \qquad v \sim \mathcal{N}(0, \sigma_v^2), \tag{48}$$

$$s_t = \theta_t + \frac{1}{\psi}\theta_{t+1} + v_t \qquad v_t \sim \mathcal{N}(0, \sigma_v^2), \tag{49}$$

where the formulation with the time indices corresponds to the dynamic model.

There are two dimensions to the CB's communication problem embedded in this signal structure. The first choice variable, ψ, represents how strongly the CB down-weights the state the PS tracks. I refer to this dimension of dynamic communication as "targetedness," by which I mean the degree to which the signal is driven by one state versus the other. Targetedness is minimal when $\psi = 1$, so that the weight on both states in the signal is equal. I refer to such a signal as a "confounding signal." By contrast, when $\psi \to \infty$ or $\psi \to 0$, targetedness goes to infinity. As the last two cases make clear, a highly targeted signal can either be targeted toward today's state ($\psi \to \infty$) or toward tomorrow's ($\psi \to 0$).

The second choice variable of the CB is σ_v^2, the variance of the noise term v in the signal. This corresponds to the precision dimension of communication. A low σ_v^2 renders a signal with given targetedness more precise, et vice versa. Most papers in the central bank communication literature focus on this dimension of communication, asking how precise information the CB should optimally provide to the public. The targetedness dimension, instead, is a fundamentally new dimension to dynamic communication.[19]

For simplicity, I assume the CB observes the full history of states perfectly, including the one-period-ahead state, while the PS' information set only includes the history of

[19] Appendix B.1 considers an alternative formalization of the targetedness dimension via a signal structure comprising two independent signals, where the choice variables of the CB are the variances of the noise terms in each signal.

the signal. Formally, the information sets of the two players in a particular period are

$$\mathcal{I}^{CB} = \{\theta_1, \theta_2\}, \qquad \mathcal{I}^{PS} = \{s\}, \tag{50}$$

$$\mathcal{I}_t^{CB} = \{\theta_{t+1}, \theta_t, \ldots, \theta_0\}, \quad \mathcal{I}_t^{PS} = \{s_t, s_{t-1}, \ldots, s_0\}, \tag{51}$$

where again the expressions with time indices correspond to the dynamic model.

Noting that the PS' problem implies that the PS' investment choice corresponds to its expectation of tomorrow's state, the CB's problem can be stated as minimizing its loss function subject to optimal PS inference. Since the signal is linear with Gaussian noise, the optimal PS forecast in the static model is given by the optimal linear projection formula

$$\theta_{2|s} = \phi s, \tag{52}$$

and by the Kalman filter formula in the dynamic model

$$\theta_{t+1|t} = m_1\theta_{t|t-1} + m_2\theta_t + m_3\theta_{t+1} + m_4 v_t. \tag{53}$$

Here $\theta_{2|s} := \mathbb{E}^{PS}[\theta_2|s]$, while the optimal projection is given by the OLS formula $\phi = \frac{Cov(\theta_2,s)}{Var(s)}$. Similarly, $\theta_{t+1|t} := \mathbb{E}_t[\theta_{t+1}|\mathcal{I}_t^{PS}]$, and m_i, $i = 1, \ldots, 4$ are given by the Kalman filter, as derived in Appendix B.2. Then the CB's problem is

$$\min_{\psi,\sigma_v} \mathbb{E}^{CB}(\theta_{2|s} - b\theta_1)^2 \quad \text{s.t.} \quad \theta_{2|s} = \phi s \tag{54}$$

in the static model, and

$$\min_{\psi,\sigma_v} \mathbb{E}_0^{CB} \sum_{t=0}^{\infty} \beta^t(\theta_{t+1|t} - b\theta_t)^2 \quad \text{s.t.} \quad \theta_{t+1|t} = m_1\theta_{t|t-1} + m_2\theta_t + m_3\theta_{t+1} + m_4 v_t \tag{55}$$

in the dynamic model. Having set up the CB's problem, I close this section by providing the definition of equilibrium in the static and dynamic models.

Definition 1. *Let $\mu_X(x)$ be the probability distribution of a variable X induced by the PS' beliefs. A Perfect Bayesian Equilibrium is an action rule I_t, belief system μ and a communication policy (ψ^*, σ_v^*) such that*

- $I_t = \arg\min \mathcal{L}_t^{PS}(I_t, \theta_{t+1}) \quad s.t. \quad \mathbb{E}_t^{PS}(\theta_{t+1}|s_t)$,
- $(\psi^*, \sigma_v^*) = \arg\min \mathcal{L}^{CB}(\{I_t, \theta_t\}_{t=0}^{\infty}) \quad s.t. \quad \mathbb{E}_t^{PS}(\theta_{t+1}|s_t) \quad and$

  $s_t = \theta_t + \frac{1}{\psi}\theta_{t+1} + v_t \quad with \quad v_t \sim \mathcal{N}(0, \sigma_v^2)$,
- *PS beliefs $\mathbb{E}_t^{PS}$ come from μ $\forall t$, and for the dynamic model, μ is consistent with Bayes' rule:*

$$\mu_{\Theta|S=s}(\theta) = \frac{\mu_{S|\Theta=\theta}(s)\mu_\Theta(\theta)}{\mu_S(s)}.$$

2.3 Optimal targetedness

Integrating out the states from the central bank's loss function, one obtains $\mathcal{V}$, the CB's expected loss function. In the static model, this takes the following form

$$\mathcal{V}^{static} = Var(\theta_{2|s}) - 2bCov(\theta_{2|s}, \theta_1) + b^2, \tag{56}$$

where I have utilized the fact that $Var(\theta_1) = 1$. Analogously, the expected loss function in the dynamic model is given by

$$\mathcal{V}^{dynamic} = \frac{1}{1-\beta}\left(Var(\theta_{t+1|t}) - 2bCov(\theta_{t+1|t}, \theta_t) + b^2\right), \tag{57}$$

where I have similarly made use of the fact that $Var(\theta_t) = 1$. On the one hand, these expressions show that the structure of the static and dynamic communication problems is the same. One sees this on the fact that the expected loss function in both models depends on the variance of the PS' beliefs minus the weighted covariance between PS beliefs and the state the CB tracks. On the other hand, the optimal dynamic evolution of beliefs renders the variances and covariances of beliefs much more involved in the dynamic model than in the static one. This explains why the evaluated expected loss function in the dynamic model, provided in Appendix B.3, is a much longer expression than the one in the static model. For this reason, while an analytical solution for ψ exists in the static model, it cannot be obtained for the dynamic model. Therefore I now present the analytical solution for ψ in the static model, and contrast it with a numerical solution for the dynamic model.

Proposition 3. *In the static model, optimal targetedness is given by*

$$\psi^*_{static} = \frac{\sqrt{(-\rho^2+\sigma_v^2+1)\left((1-2b\rho)^2\sigma_v^2-(\rho^2-1)\left(4b(b-\rho)+1\right)\right)}-\rho\left(2b\sigma_v^2+\rho\right)+\sigma_v^2+1}{2\left(b\rho^2+b-\rho\right)\sigma_v^2-2\left(\rho^2-1\right)(b-\rho)}. \tag{58}$$

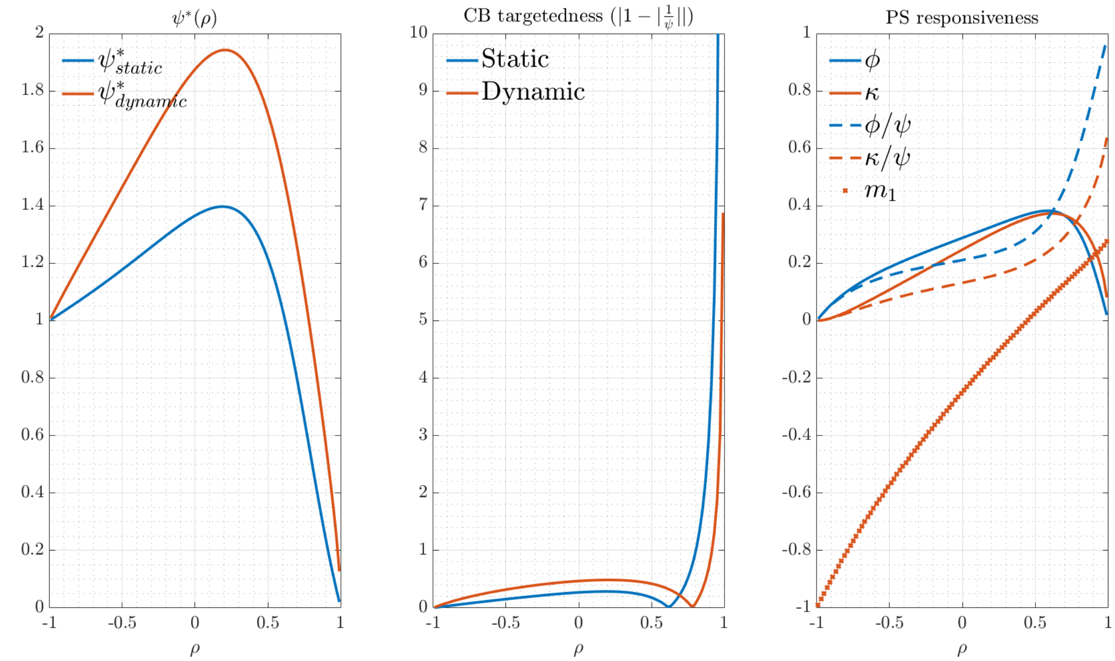
Figure 8 illustrates the optimal ψ in both models for a benchmark calibration of $b = 1$, $\beta = 0.99$ and $\sigma_v = 1$. I will use this figure throughout the section to analyze the optimal targetedness policy. The left panel depicts ψ^* as a function of ρ. The middle panel shows $|1 - |\frac{1}{\psi}||$, a metric of CB targetedness that evaluates the absolute distance between the weights the signal places on today's versus tomorrow's fundamental. The right panel plots various measures of PS responsiveness as a function of ψ^* and ρ: responsiveness to today's state and to noise (ϕ and κ in the static and dynamic models respectively), responsiveness to tomorrow's state ($\frac{\phi}{\psi}$ and $\frac{\kappa}{\psi}$) and, in the dynamic model only, to past beliefs (m_1). Before turning to the differences in optimal targetedness between the static and dynamic models, let me first characterize the commonalities between the two solutions for ψ. This is formalized in the following proposition.

Proposition 4. *Shared features of optimal targetedness policy*

Both in the static and dynamic model, $\psi^*(\rho)$ has the following properties:

- $\psi^* \geq 0 \; \forall \rho$.

Figure 8: Optimal targetedness and PS responsiveness as a function of ρ



The left panel of the figure shows the optimal targetedness, ψ^*. The middle panel shows the absolute distance between the weights on today's and tomorrow's state in the signal, an alternative measure of targetedness. In these two panels, the blue lines correspond to the static, and the red lines to the dynamic model. The right panel depicts the PS' responsiveness to today's state (ϕ and κ in static and dynamic models) and to tomorrow's state (ϕ/ψ and κ/ψ in static and dynamic models), as well as m_1, the PS' loading on past beliefs $\theta_{t|t-1}$ (dynamic model only)

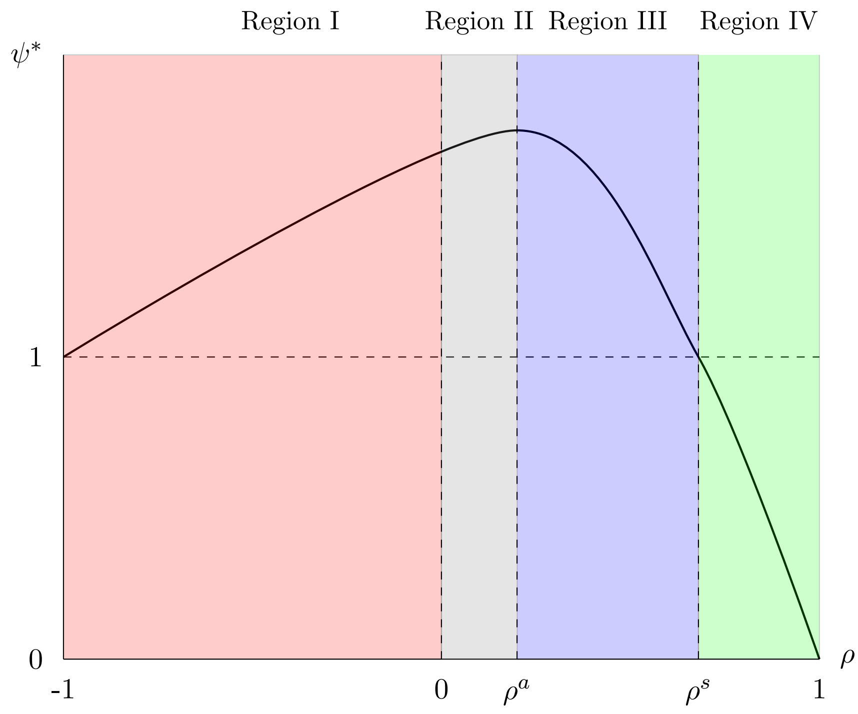
- *When* $\rho \to -1$, $\psi^* \to 1$ *from above.*
- *When* $\rho \to 1$, $\psi^* \to 0$ *from above.*
- $\exists\, \rho^a$ *such that for* $\rho < \rho^a$, ψ^* *is increasing in* ρ, *but for* $\rho > \rho^a$, ψ^* *is decreasing in* ρ.
- $\exists\, \rho^s > \rho^a$ *such that* $\psi^*(\rho^s) = 1$, *and for a small* $\epsilon > 0$, $\psi^*(\rho^s - \epsilon) > 1 > \psi^*(\rho^s + \epsilon)$.

I refer to ρ^a *and* ρ^s *as the alignment and switching threshold respectively.*

Figure 9 explains Proposition 4 by replotting the left panel of Figure 8 schematically, splitting the ρ-space into four different regions. In Region I, shown in red, $\rho < 0$, rendering the preferences of the PS and the CB severely misaligned. This means that there exists a $\bar{\psi}$ such that if $\psi > \bar{\psi}$, $\text{sign}(\kappa) = \text{sign}(\phi) = -\text{sign}(\psi)$. In other words, if the CB downweights tomorrow's state too much in the signal, the PS understands that the signal is primarily driven by today's state, and responds in the opposite direction.

One can see an example of this on the leftmost panel of Figure 10, which plots ϕ and κ, the PS' responsiveness to the signal in the static and dynamic models respectively, as a function of ψ. When $\rho = -0.99$, κ and ϕ switch sign when ψ is barely above 1. This constrains how strongly the CB can target the state it cares about, today's state, and therefore it selects the highest permissible $\psi < \bar{\psi}$. In the limit, when $\rho = -1$, $\bar{\psi} = 1$ and thus the best the CB can do is to send a perfectly confounding signal, $\psi = 1$, resulting in the PS ignoring the signal completely ($\kappa = \phi = 0$, as seen on the right panel of Figure 8). As ρ approaches zero from below, $\bar{\psi}$ rises, loosening the constraint on how strongly the CB can target θ_t, its preferred state. Therefore, throughout Region I, ψ^* is increasing in ρ.

Figure 9: Optimal targetedness policy across the two models



Once $\rho \geq 0$, we enter Region II, the gray-shaded region on Figure 9. Here, since $\rho \geq 0$, there is no $\bar{\psi}$ that constrains the CB in raising ψ and thus targeting its object of interest, today's state, more strongly. And it is desirable to raise ψ for the CB because ρ is sufficiently close to zero that the CB's and the PS' preferences are still misaligned.

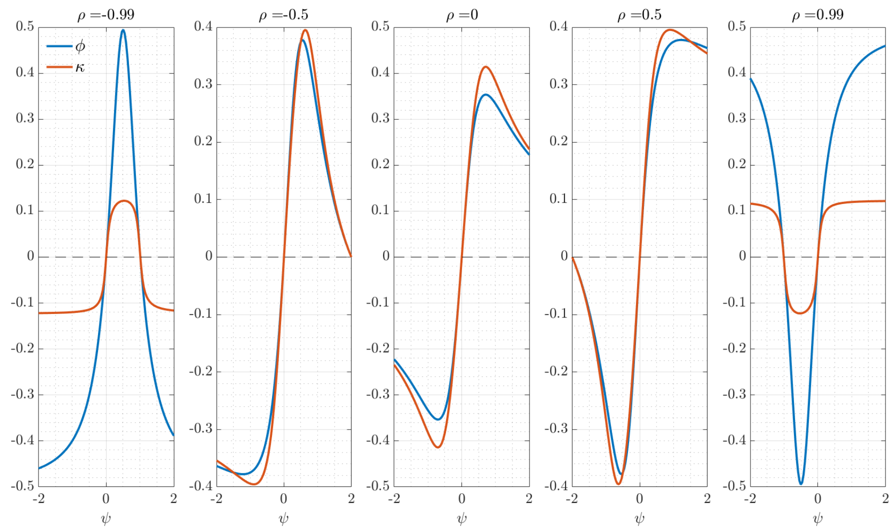
However, the CB cannot raise ψ indefinitely, since responsiveness falls the higher ψ is. For example, when $\rho = 0$, the middle panel of Figure 10 shows that $\frac{\partial \kappa}{\partial \psi} < 0$ for any $\psi > 1$.

The blue-shaded region, Region III, begins once $\rho > \rho^a$. At this point, ρ is high enough for the two players' preferences to become more aligned. Thus the CB becomes interested in raising the PS' responsiveness to tomorrow's state instead of today's. As shown in Appendix B.2, the PS' responsiveness to tomorrow's state is κ/ψ in the dynamic, and ϕ/ψ in the static model. Thus in both models the CB can raise this by lowering ψ. A lower ψ not only lowers the denominator, but also raises the numerator, as κ and ϕ are decreasing in ψ if $\psi > 1$ and ρ is positive and not too close to 1 (see third and fourth panels of Figure 10). Thus in this region, the CB maximizes responsiveness to a weighted average of today's and tomorrow's state.

Once $\rho > \rho^s$, the two states are so strongly correlated that the CB's and the PS' preferences are strongly aligned. In this region, shaded green and marked Region IV on Figure 9, the CB's primary concern is the noise in the signal. Since ρ is approaching 1, the CB is indifferent to whether the PS responds to today's or tomorrow's state, and the only thing that can get in the way of that is the noise. Thus as $\rho \to 1$, the CB shrinks $\psi^* \to 0$. As seen on the rightmost panels of Figures 8 and 10, κ and ϕ decrease as a result, while κ/ψ and ϕ/ψ shoot up. This way, the PS' responsiveness to tomorrow's state increases, while its responsiveness to the signal in general, and thus to noise, is kept in check. Thus once ρ crosses the switching threshold ρ^s, the CB becomes indifferent to which fundamental the PS responds to, and therefore switches the targetedness of the signal from θ_t to θ_{t+1}.

In this manner, ρ, the correlation between the two states, determines the optimal choice of targetedness in both the static and the dynamic model. In the limit, severe misalignment results in a perfectly confounding signal that does not target any of the two states, while the opposite extreme of perfect alignment leads the CB to target

Figure 10: PS responsiveness to the signal as a function of ψ for various values of ρ



The figure shows the PS' responsiveness to the signal, ϕ in the static model (blue) and κ in the dynamic model (red), as a function of of targetedness, ψ, for various levels of correlation between the states (ρ).

the PS' preferred state, θ_{t+1}, infinitely. Between the limiting cases, the CB balances the need for a high level of θ_t-targetedness if alignment is low against the shrinking responsiveness of the PS to the signal if it is too targeted toward today's state.

However, as Figure 8 indicates, ψ^* is quantitatively different across the two models. Moreover, the threshold ρ^s after which the CB only aims to drive down the responsiveness to noise arrives later in the dynamic model than in the static one. The next proposition states these results.

Proposition 5. *Higher weight on the central bank's target in the dynamic model*

$$\psi^*_{dynamic} \geq \psi^*_{static} \quad \forall \rho. \tag{59}$$

In the dynamic model, the central bank finds it optimal to target its own preferred state more than it would in the static model for any level of alignment.

Corollary 5.1.

$$\rho^s_{dynamic} > \rho^s_{static}. \tag{60}$$

The switching threshold is higher in the dynamic model than in the static model.

Why is it optimal for the central bank to target today's state more in the dynamic model? Recall that the comparison of the expected loss functions in Equations (56) and (57) revealed that in both models, the CB's loss function consists of a term that captures the variance of the PS' beliefs of tomorrow's state, minus a term capturing the covariance between those same beliefs and today's state. Thus the structure of the CB's objective is the same, implying that the difference in the solutions must come from differences in the evolution of beliefs.

Let us therefore investigate the laws of motion of beliefs in the two models. Substituting in the signal in the optimal forecast in the static model from Equation (54), and writing out the Kalman filter equation in the dynamic model, Equation (55), with the help of Appendix B.2, one can write the evolutions of the beliefs that enter the CB's loss function in the following way:

$$\text{Static:} \quad \theta_{2|s} = \phi\theta_1 + \frac{\phi}{\psi}\theta_2 + \phi v, \tag{61}$$

$$\text{Dynamic:} \quad \theta_{t+1|t} = m_1\theta_{t|t-1} + \kappa\theta_t + \frac{\kappa}{\psi}\theta_{t+1} + \kappa v_t. \tag{62}$$

This neatly illustrates two sources of difference between static and dynamic. The most obvious difference is the presence of m_1 in the dynamic belief equation, signifying that in a dynamic world, current beliefs are informed to some degree by past beliefs. In a static setting, by contrast, the loading on past beliefs is clearly zero. Yet even if the loading on past beliefs were zero in the dynamic model, the evolutions of beliefs would not coincide, because κ is not necessarily equal to ϕ. In fact, as Figure 10 indicates, in general $\kappa \neq \phi$. Let us look at each of these channels in turn.

2.3.1 Responsiveness in static versus dynamic

One can compute $\phi = \frac{Cov(\theta_2,s)}{Var(s)}$ easily to contrast it with the expression for κ from Appendix B.2:

$$\phi = \frac{\rho + \frac{1}{\psi}}{1 + \frac{1}{\psi^2} + 2\frac{\rho}{\psi} + \sigma_v^2}, \qquad \kappa = \frac{\rho p_4 + \frac{1}{\psi} p_1}{p_4 + \frac{1}{\psi^2} p_1 + 2\frac{\rho}{\psi} p_4 + \sigma_v^2}, \tag{63}$$

where $p_i,\ i = 1,4$ are the two diagonal elements of the 2×2 forecast-error-variance matrix P. These correspond to the prior variances of the two fundamentals at the steady state: p_1 is the tightness of the prior belief about θ_{t+1}, while p_4 the tightness of the prior on θ_t. Letting $\pi(\theta_T)$ and $p(\theta_T, s_t)$ denote prior and posterior variances of θ_T, where $T = t, t+1$ in the dynamic model, and $T = 1,2$ in the static model,

$$\pi(\theta_T) \coloneqq \mathbb{E}[(\theta_T - \theta_{T|t-1})^2], \tag{64}$$

$$p(\theta_T, s_t) \coloneqq \mathbb{E}[(\theta_T - \theta_{T|t})^2]. \tag{65}$$

For the dynamic model it follows that

$$p_1 = \pi(\theta_{t+1}), \tag{66}$$

$$p_4 = \pi(\theta_t). \tag{67}$$

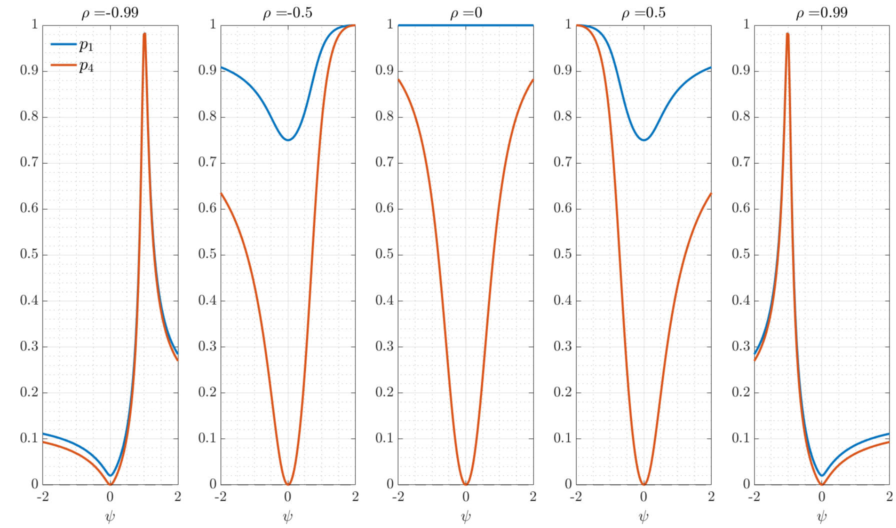
One observes that if $p_1 = p_4 = 1$, then $\kappa = \phi$. In this case, the prior variance of today's and tomorrow's state exactly equals the variance of the state, 1, signifying that the PS has not updated its beliefs after receiving signals from the CB. This suggests that the main difference between ϕ and κ is that the latter, through the tightness of the priors p_1 and p_4, captures the PS learning about θ_t and θ_{t+1} from observing the signal over time. Another way to say this is that the static problem is a special case of the dynamic problem in which the PS' prior beliefs have a constant, unit variance. Constant priors are unaffected by the CB's signal and thus drop out of the problem.

From comparing the analytical expressions for ϕ and κ, it is not obvious how the priors affect responsiveness, or how the CB's targetedness affects the priors. To get a better sense of this, consider Figure 11, which plots p_1 in blue and p_4 in red as a function of ψ, for different values of ρ. Firstly, it is clear that $(p_1, p_4) \in [0, 1]$. Recalling that $Var(\theta_T) = 1 \; \forall T \geq 0$, one can interpret a unit prior variance as a perfectly loose prior since even absent any information, the variance of the state is an upper bound on the variance of the PS' beliefs. A zero prior variance, by contrast, corresponds to a perfectly tight prior.

Secondly, $p_1 \geq p_4 \; \forall \; \psi$. This is most noticeable when $\rho = 0$, plotted on the middle panel of the figure. In this case, the PS cannot enter any given period t with anything but perfectly flat priors on θ_{t+1} because when the states are unrelated, there is no way to receive any information on θ_{t+1} in period $t-1$. The reason that even in this case of uncorrelated states, p_4 is still below 1 is that, in contrast with θ_{t+1}, the PS enters every period t already having received information about θ_t in the previous period. The signal structure in Equation (49) means that the PS effectively receives two signals about θ_t, as both s_t and s_{t-1} contain information about it. This is the reason why $p_4 \leq p_1$; unlike in the static model, in any given period in the dynamic model, the PS has more information about today's state than tomorrow's.

However, we see from Figure 10 that the magnitudes and the comparative statics of ϕ and κ are qualitatively and quantitatively similar, at least in the intermediate regions of ρ. This suggests that however the CB affects the tightness of the priors p_1 and p_4, these effects end up balancing each other out. Looking at Figure 11, we see that this comes from the fact that unless $\rho = 0$, p_1 and p_4 always respond to ψ in the same way. If decreasing ψ tightens the prior about θ_{t+1}, it also tightens the prior on θ_t, et vice versa. Thus manipulating the tightness of the priors through the choice of ψ does not do much for the CB in terms of affecting the PS' responsiveness to the signal. So this is not the primary wedge between the static and dynamic models. But if this is not it,

Figure 11: Prior variances of θ_{t+1} and θ_t (p_1 and p_4 respectively) as a function of ψ for various ρ



The figure plots the prior variances of the two states as a function of targetedness, ψ, for various levels of correlation between the states, ρ. The blue line, $p_1 = \pi(\theta_{t+1})$, corresponds to the tightness of the prior on tomorrow's state, while the red line, $p_4 = \pi(\theta_t)$, corresponds to the tightness of the prior on today's state.

then what is?

2.3.2 Loading on past beliefs

To answer this question, I now turn to the second dimension of difference between the static and the dynamic models: m_1, the loading on the past belief $\theta_{t|t-1}$. Let us start by investigating what this loading is.

$$m_1 = \rho - \kappa\left(\frac{\rho}{\psi} + 1\right). \tag{68}$$

Observing that $\frac{\rho}{\psi} + 1 = Cov(\theta_t, s_t)$, one can decompose m_1 into two elements. The first is ρ, the temporal dependence between the two states. From this, the PS subtracts the responsiveness-weighted covariance of the signal with today's state, $\kappa(\frac{\rho}{\psi} + 1)$. I will refer to the first component as the "AR-component" given that it comes from the autoregressive coefficient of the θ-process. Since the PS knows that $\theta_{t+1} = \rho\theta_t + \varepsilon_{t+1}$, the optimal forecast of θ_{t+1} given knowledge of θ_t is $\rho\theta_t$. Since in this environment, whether static or dynamic, the states are never fully revealed, the PS uses its best guess of θ_t in this optimal forecast: θ_t is replaced by $\theta_{t|t-1}$.

But the PS can do better than simply extrapolating θ_{t+1} from past beliefs using the AR-component $\rho\theta_{t|t-1}$. This insight is the starkest when $\rho = 0$. In this case, since the states are serially uncorrelated, one might think that past beliefs are uninformative and should thus be completely ignored by the PS. However, even if $\rho = 0$, $m_1 = -\kappa \neq 0$, so that the dynamic model does not coincide with the static model even in this case.

What is going on here is that when $\rho = 0$, the AR-component of m_1 is shut off, since the autoregressive transmission between the states is gone. But even in this case, the covariance of s_t with θ_t is not 0, but 1. So the PS can use its knowledge that today's signal was to a known extent ($Cov(\theta_t, s_t) = 1$) reflective of today's state instead of the state the PS cares about, θ_{t+1}. So $m_1 = -\kappa$ in this case can be interpreted as the PS filtering out its best guess of θ_t from its beliefs of θ_{t+1}. This is possible purely because at any given time, the CB's signal mixes today's and tomorrow's state, and therefore s_t is already the second signal the PS receives about θ_t.

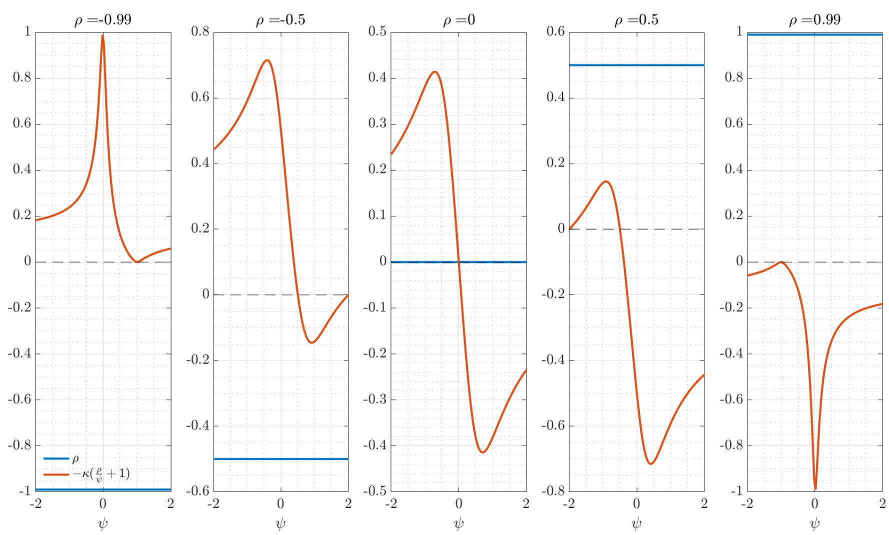
Thus we can interpret the $\kappa(\frac{\rho}{\psi}+1)$ part of m_1 as a component that the PS subtracts from the AR-component in order to filter out new information about θ_t arriving at time t. Therefore I will refer to this part as the "filter-component." Of course, this component is in action when $\rho \neq 0$ as well. And glancing at the right panel of Figure 8, we see from the evolution of ρ, ψ^* and κ why m_1 is negative for most of the ρ-space. In Region I, when ρ is negative and $\kappa(\psi^*)$ is small, the AR-component of m_1 dominates, resulting in $m_1 < 0$. In the limit of $\rho \to -1$, $\kappa \to 0$, and so $m_1 \to -1$, pushing the PS' beliefs exceedingly strongly in the opposite direction to what the CB intends. Thus, in this region, the AR-component of m_1 makes the CB's life difficult as it exerts influence on the PS' beliefs that works against the CB's signal.

When $\rho > 0$ but small, the AR-component no longer pushes against the CB's signal, but now the filter-component dominates, so that m_1 is still negative, and thus still causes the CB headaches as it dampens the effectiveness of the CB's signal. Only once

$\rho > \kappa(\frac{\rho}{\psi} + 1)$ does m_1 become positive, stopping to cancel out the effect of the CB's signal.

The main difference between the static and dynamic problem, then, from the perspective of the CB's choice of targetedness, is how to deal with the influence of the PS' priors on current beliefs. This influence is captured in the loading m_1, and, as we have seen, for most of the ρ-space, it goes the opposite way as the CB wants it to. Therefore, the new element of the CB's optimal choice of ψ in the dynamic model is aiming to cancel out the loading on the PS' prior beliefs if it goes the wrong way. Out of the two components that make up m_1, the CB can however only affect the filter-component. To understand how the central bank's choice of targetedness affects the filter-component, consider Figure 12, which illustrates how the filter-component that is subtracted from ρ changes when ψ changes, for various values of ρ. I also plot ρ in blue, so that m_1 is just the sum of the two lines.

Figure 12: The components of m_1 as a function of ψ for various values of ρ



The figure decomposes m_1, the PS' loading on past beliefs, $\theta_{t|t-1}$, into two components, the AR-component (ρ) in blue and the filter-component ($\kappa(\frac{\rho}{\psi}+1)$) in red, and plots how these components evolve as a function of targetedness, ψ, for various levels of autocorrelation, ρ. Note: the red line is the negative of the filter-component, so that the sum of red and blue lines gives m_1.

The main message of the figure is that because the sign of the filter-component is usually the opposite of the sign of the AR-component, a higher ψ^* in the dynamic model is the result of the CB trying to decrease m_1 when it has the wrong sign, or increase it when the sign is correct. When m_1 pushes against the CB's signal, the CB thus tries to raise the filter-component to cancel out m_1. When instead m_1 acts as a tailwind to the CB's signal, the CB tries to prevent the AR- and filter-components cancelling out, and thus lowers the filter-component. The figure illustrates why in both of these cases, the optimal thing to do is to select a larger ψ than what would be optimal in the static model.

When $\rho \ll 0$, for instance $\rho = -0.99$ on the leftmost panel of the figure, the AR-component is working strongly against the CB's signal. Intuitively, what is happening here is that because of the strong negative correlation, the PS' prior beliefs about θ_t are large and, from the standpoint of the CB, have the wrong sign. Since the PS tracks θ_{t+1} and $\rho < 0$, it uses its priors to conclude that θ_{t+1} is strongly in the opposite direction of the signal. In such a situation, the CB wishes that the filter-component of m_1 were stronger, so that the PS would filter out more of its priors, in other words, would update its beliefs of θ_t stronger after observing s_t. This leads the CB to want to increase the absolute value of the filter-component. As the leftmost panel of Figure 12 points out, the CB achieves this by raising ψ.

In the opposite extreme, when $\rho \to 1$, depicted on the rightmost panel of Figure 12, the AR-component goes the right way, but is now dampened by the PS' filtering out incoming information regarding θ_t. Thus in this case the CB wishes that the PS would filter out less of its priors, and therefore wants to minimize the absolute value of the filter-component. Now from the rightmost panel of Figure 12, we see that to do so, the CB again needs to raise ψ.

The endogenous prior beliefs of the private sector, and the way those priors influence current beliefs through the loading m_1, thus serve as a dampening force to how the central bank can influence beliefs and thus exercise persuasion on the PS. When misalignment is severe, the CB cannot affect the loading on the prior and regrets having informed the PS about θ_t in period $t-1$, as those prior beliefs push posteriors away from the CB's desired direction. By contrast, when alignment is very strong, the tail-wind effect of the priors is dampened by the PS' rational update from today's signal. In order to work against the dampening effect of the prior on the PS' beliefs, the CB finds it optimal to always push today's state more strongly in its signal than under the static model, either to work against undesirable prior beliefs, or to brake the updating away from desirable priors.

2.4 Optimal precision

Now I turn to the question of how precise the central bank's signal should be for a given targetedness. Thus, holding ψ fixed, the central bank now chooses σ_v to minimize the loss function in Equation (57). The next proposition gives an analytical characterization of the solution in the static and dynamic models.

Proposition 6. *Corner versus interior solutions*

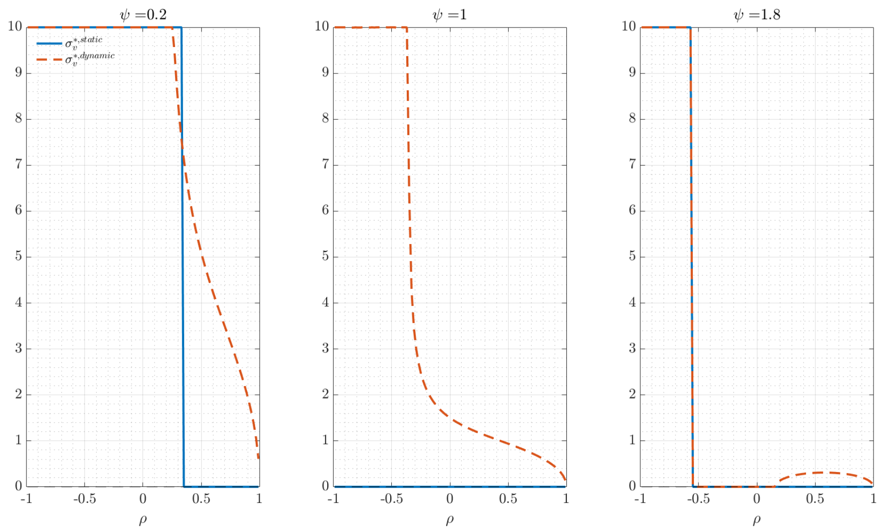
In the static model, optimal precision is a corner solution: σ_v^* is either 0 or ∞. In the dynamic model, the roots of the first order condition are

$$\begin{pmatrix} 0 \\ \pm\frac{\sqrt{1-\rho^2}}{\sqrt{\rho}\sqrt{\psi}} \\ \pm\sqrt{\frac{(\rho-1)(\rho+1)\psi^2(\rho-2b)(2b(\rho^2(\psi^2+1)+3\rho\psi+\psi^2)-\psi(\rho\psi+1))+\sqrt{2}\sqrt{b(\rho^2-1)^2\psi^4(2b-\rho)(\rho+\psi)^2(\rho\psi+1)(-2b(\rho+\psi)+\rho\psi+1)^2}}{\rho\psi^4(\rho-2b)(2b(\rho^3-2\rho-\psi)+\rho\psi+1)}} \\ \pm\sqrt{\frac{(\rho-1)(\rho+1)\psi^2(\rho-2b)(2b(\rho^2(\psi^2+1)+3\rho\psi+\psi^2)-\psi(\rho\psi+1))-\sqrt{2}\sqrt{b(\rho^2-1)^2\psi^4(2b-\rho)(\rho+\psi)^2(\rho\psi+1)(-2b(\rho+\psi)+\rho\psi+1)^2}}{\rho\psi^4(\rho-2b)(2b(\rho^3-2\rho-\psi)+\rho\psi+1)}} \end{pmatrix}, \tag{69}$$

and both corner and interior solutions exist.

This result suggests a qualitative difference between optimal precision in the static and dynamic models: whereas in the static model, the solution for σ_v is bang-bang, this is not necessarily the case in the dynamic model. Figure 13 provides a visual illustration of this point by showing $\sigma_v^{*,static}$ in blue and $\sigma_v^{*,dynamic}$ in red dashed, both as a function of ρ and for three different values of ψ: 0.2, 1 and 1.8. The ψ-values are selected to encompass a case where the CB's signal is targeted toward θ_{t+1} ($\psi < 1$), when the signal is not targeted toward any state ($\psi = 1$), and lastly when it is targeted toward θ_t ($\psi > 1$). Also, I continue to use the calibration with $b = 1$ and $\beta = 0.99$ throughout the section.

Figure 13: Optimal precision σ_v^* as a function of ρ for various values of ψ



2.4.1 Corner solutions in the static model

Focusing first on the blue line on Figure 13, which corresponds to σ_v^* in the static model, one observes the bang-bang nature of optimal static precision. When $\psi = 0.2 < 1$, so that the CB targets tomorrow's state in its signal, shown in the left column of Figure 13, the CB sets $\sigma_v^* = 0$ only if $\rho \geq \rho^b(\psi)$, that is, if ρ is higher than a particular threshold which I will refer to as the bang-bang threshold, and which is a function of ψ. Note that if $\psi < 1$ and $\sigma_v = 0$, the CB's signal in Equation (49) is an invertible moving average process, so in this case sending an infinitely precise signal ($\sigma_v = 0$) would mean that the CB reveals the θ-process fully. Therefore, in this case, when preference alignment is high enough ($\rho \geq \rho^b(\psi)$), the CB in the static model prefers to fully reveal the two fundamentals, and otherwise, it sends an infinitely imprecise signal.

A similar situation arises in the static model when $\psi > 1$, so that the signal is targeting the CB's preferred state, θ_t. Again, there exists a bang-bang threshold $\rho^b(\psi)$ such that for $\rho > \rho^b(\psi)$, the CB's signal becomes infinitely precise, and is infintely imprecise otherwise. Since in this case the signal weights the CB's preferred state more strongly, the bang-bang threshold is lower than it is when the signal is targeted toward the PS' preferred state: for $\psi^h > 1 > \psi^l > 0, \quad \rho^b(\psi^h) < \rho^b(\psi^l)$. This reflects that the CB reverts to communicating infinitely precisely already for a lower level of alignment if it gets to skew its signal towards θ_t.

The only ψ for which the CB chooses $\sigma_v^* = 0$ for any ρ is $\psi = 1$. The reason is that in this case, the signal is not targeting either θ_t or θ_{t+1}, and is thus perfectly confounding. Therefore the information provided by the CB improves the PS' expectations without pushing them in the wrong direction, even if $\rho \to -1$.

2.4.2 Tightness of prior beliefs in the dynamic model

Let us now examine the optimal precision σ_v^* in the dynamic model, shown by the red dashed line on Figure 13. The main qualitative difference is that, in contrast to the static model, the solution is not generally bang-bang. While there exist (ρ, ψ) pairs for which σ_v^* is corner (0 or ∞), for many (ρ, ψ) pairs the optimal precision is interior.

When $\psi = 0.2$ (left panel of Figure 13), so that the CB's signal is targeted toward θ_{t+1}, the PS' preferred state, σ_v^* shifts from ∞ to an interior value when ρ exceeds $\rho^i(\psi)$, a threshold value I will call interiority threshold. Optimal noise only shrinks to zero when $\rho = 1$. When $\psi = 1$, so that the weight on the two states in the CB's signal is equal, the same pattern obtains. For $\rho < \rho^i(\psi)$, $\sigma_v^* = \infty$, and as $\rho > \rho^i(\psi), \quad \sigma_v^* < \infty$, converging to zero for $\rho \to 1$. The only difference between the θ_{t+1}-targeted and the untargeted signal is that for $\psi^h = 1 > \psi^l > 0$, $\rho^i(\psi^h) < \rho^i(\psi^l)$. This resembles the intuition in the static model that for a signal that is more targeted towards the CB's preferred state (has a higher value of ψ), the CB pivots away from infinite imprecision already for a lower level of ρ.

Something entirely different happens when the CB's signal is strongly targeted towards θ_t, the CB's preferred state. This situation, depicted on the right panel of Figure 13, involves an initial bang-bang as σ_v^* transitions from ∞ to 0 when ρ hits a particular bang-bang threshold $\rho^b(\psi) < 1$. Afterwards, σ_v^* remains 0 as ρ increases, until ρ exceeds an interiority threshold. Here, σ_v^* becomes interior *from below*, surprisingly increasing as alignment between the two states grows. Finally, once alignment becomes very strong, σ_v^* reverses course and decreases back down to asymptote to 0 as $\rho \to 1$.

What is going on? Why is optimal precision often interior in the dynamic model, while it never is in the static model? And if σ_v^* passes from ∞ to 0 through an interiority region when $\psi \leq 1$, why does this interiority region disappear when $\psi > 1$, only to reappear for much higher ρ-values? To understand why all of this happens, it is helpful to investigate the tightness of prior and posterior beliefs in the two models. Recall the notation of $\pi(\theta_T)$ for prior and $p(\theta_T, s_t)$ for posterior variances of θ_T, introduced in Equation (65) of Section 2.3 and reproduced here for convenience:

$$\pi(\theta_T) \coloneqq \mathbb{E}[(\theta_T - \theta_{T|t-1})^2],$$

$$p(\theta_T, s_t) \coloneqq \mathbb{E}[(\theta_T - \theta_{T|t})^2].$$

Let us also define

$$I(\theta_T, s_t) \coloneqq \pi(\theta_T) - p(\theta_T, s_t), \tag{70}$$

as the decrease in uncertainty in θ_T given s_t. With a slight abuse of terminology, I will refer to $I(\theta_T, s_t)$ as mutual information on θ_T, i.e. the extent to which the signal is informative about θ_T.[20] As demonstrated in Appendix B.5, the PS' prior on any of the two states in the static model is equal to the unconditional variance of the fundamentals, 1. In the dynamic model, instead, in any period $t = \tau$, the posterior

[20] While, formally, mutual information is the expected difference between the log posterior and the log prior distribution, I simply compare the level difference between the second moments of the prior and posterior.

variance $p(\theta_{t+1}, s_t)$ is carried over to the next period, becoming the endogenous prior $\pi(\theta_t)$ in period $t = \tau + 1$. On the one hand, this has the important consequence that in the static model

$$I(\theta_T, s) = Var(\theta_{T|s}), \qquad T = 1, 2, \tag{71}$$

implying that mutual information is a monotonically decreasing function in σ_v. It is easy to verify that this is true for mutual information on θ_{t+1} in the dynamic model as well, as

$$I(\theta_{t+1}, s_t) = (1 - \rho^2) Var(\theta_{t+1|t}). \tag{72}$$

On the other hand, it clearly does not hold for θ_t:

$$I(\theta_t, s_t) = Var(\theta_{t|t}) - Var(\theta_{t|t-1}). \tag{73}$$

In other words, the dynamic model has the property that mutual information on θ_t is non-monotonic in σ_v. The following proposition and corollary summarize this insight, together with its implications for the optimal precision choice.

Proposition 7. *Tight priors dampen informativeness in the dynamic model*
Mutual information on θ_t is non-monotonic in σ_v because the endogenous prior carries over information from period $t-1$ to period t. Let s^h denote a high-precision signal, while s^l a low-precision one. If a precise signal in the previous period, s^h_{t-1}, tightens the prior in t sufficiently, then the reduction in uncertainty at time t due to the signal pair $\{s_t, s^h_{t-1}\}$ is lower than for the pair $\{s_t, s^l_{t-1}\}$.

Corollary 7.1. *Information smoothing in the dynamic model*
In the dynamic model, a lower degree of precision serves to keep priors from getting too tight; the optimal precision choice involves the smoothing of information provision.

The non-monotonicity of $I(\theta_t, s_t)$ comes from the signal structure in Equation (49), which implies that in any given period t, the current signal s_t is not the first to provide the PS with information about θ_t. If all signals have sufficiently high precision, then by the time period t arrives, the PS has already learned enough about θ_t from s_{t-1}, and will therefore find the current signal less informative than if it had acquired less information in the past.

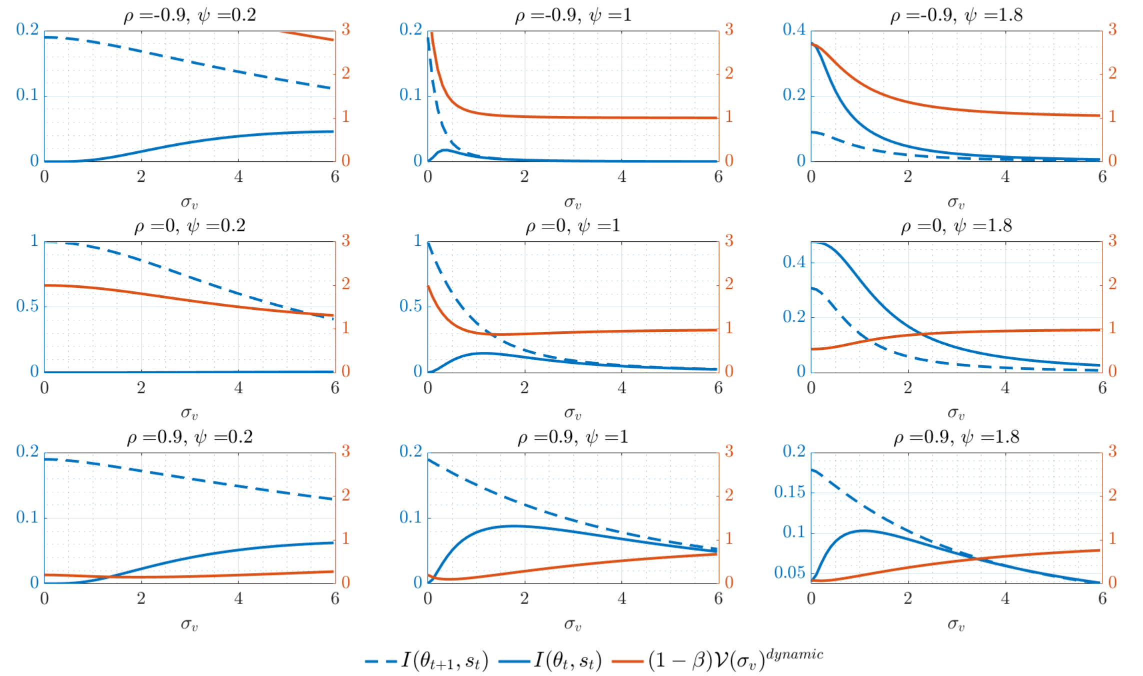
Figure 14 illustrates Proposition 7, plotting the reduction in uncertainty in θ_{t+1} in dashed blue against that in θ_t in solid blue, both as a function of σ_v for various (ρ, ψ) pairs, and with values on the left y-axis. The red line corresponds to the $(1-\beta)$-weighted central bank loss, $\mathcal{V}$, with values on the right y-axis. Just like Equation (72) suggests, $I(\theta_{t+1}, s_t)$ indeed decreases monotonically in σ_v. $I(\theta_t, s_t)$, instead, is hump-shaped in the majority of cases, assuming a maximum at a precision level I will denote by σ_v^I:

$$\sigma_v^I \coloneqq \arg\max I(\theta_t, s_t). \tag{74}$$

The CB's loss function $\mathcal{V}$, shown in red on Figure 14, lends support to Corollary 7.1. The interior solutions for σ_v are visible as the interior minima of $\mathcal{V}$. Whenever the optimal precision is interior, σ_v^* is in the neighborhood of σ_v^I. Thus σ_v^I pulls σ_v^* away from zero. In other words, the CB finds it optimal not to shrink the signal noise to zero because this allows it to push $I(\theta_t, s_t)$ up. This lets the CB ensure that in any period t, the current signal contributes to the PS' information set and thus exerts persuasion. In other words, the CB wants the PS to receive approximately equal amounts of information in each period, a phenomenon I refer to as the smoothing of information provision.

Thus it is the non-monotonicity of mutual information on θ_t that is behind the interior solutions in σ_v in the dynamic model. Intuitively, this non-monotonicity captures the effect of the tightness of the PS' priors on the CB's communication problem. Clearly, there are parameter configurations for which it is optimal to communicate in-

Figure 14: Dynamic loss $\mathcal{V}$ and mutual information $I(\theta_T, s_t)$ as a function of σ_v for various (ρ, ψ) pairs, $T = t, t+1$



The figure plots the $(1-\beta)$-weighted expected loss function of the CB, $\mathcal{V}$, in the dynamic model in red on the right axis, against mutual information on θ_{t+1} and on θ_t in blue dashed and blue solid on the left axis. Note: for my purposes, I define mutual information on θ_T as $I(\theta_T, s_t) := \pi(\theta_T) - p(\theta_T, s_t)$, so that it is the level difference between prior and posterior variances.

finitely precisely. But for many (ρ, ψ) pairs, precise communication in period $t-1$ would tighten the PS' prior in period t too much, rendering the PS less responsive to the CB's period t signal. Therefore the key driver of the difference between static and dynamic optimal precision is the varying tightness of the prior.

But it is not quite the only driver. If it were, then the CB would find it optimal to set $\sigma_v^* = \sigma_v^I$ for all (ρ, ψ) pairs. This is however clearly not the case, as one can see if one compares the maxima of $I(\theta_t, s_t)$ with the minima of the loss in the figure: sometimes $\sigma_v^* > \sigma_v^I$, sometimes $\sigma_v^* < \sigma_v^I$, and for the pair $(\rho, \psi) = (0, 1)$, $\sigma_v^* = \sigma_v^I = 0$. So to find out what other forces are behind the optimal precision choice, let us now investigate how σ_v^* depends on first ρ and then ψ.

Fixing ψ and increasing ρ means that for a fixed column, one moves down through the rows of Figure 14. When $\rho \ll 0$, $\sigma_v^* = \infty$. This tells us that when preference misalignment is strong, the CB wishes to minimize both $I(\theta_{t+1}, s_t)$ and $I(\theta_t, s_t)$, because in this case, any information in the signal will be misused by the PS in the sense that its investment choice will be of opposite sign to θ_t. When alignment rises above the interiority threshold, $\rho^i(\psi)$, for example when ρ rises from -0.9 to 0 in the $\psi = 1$ column of Figure 14, the CB seeks to maximize $I(\theta_t, s_t)$, i.e. to provide as much information about θ_t as possible. This requires both $\sigma_v < \infty$ and $\sigma_v > 0$ because of the non-monotonicity of $I(\theta_t, s_t)$. At the same time, depending on the extent of alignment, the CB sometimes sees it necessary to also minimize the information it reveals about θ_{t+1}. This is the case in the $(\rho, \psi) = (0, 1)$ panel of the figure, pushing σ_v^* above σ_v^I. As alignment becomes near perfect, for instance in the last row of the figure, where $\rho = 0.9$, the CB no longer simply maximizes $I(\theta_t, s_t)$ because providing information about θ_{t+1} pushes the PS in the same direction as information about θ_t. Now instead it is optimal to maximize $I(\theta_t, s_t) + I(\theta_{t+1}, s_t)$, the sum of information provision about both states. This pushes σ_v^* below σ_v^I toward zero.

Let us now hold alignment fixed and investigate how targetedness affects the optimal precision choice. For a fixed ρ, increase ψ by moving right along on any row of Figure 14. Except for the first row, where $\rho \ll 0$ and thus $\sigma_v^* = \infty \ \forall \psi$, a higher ψ leads to a strictly lower σ_v^*. For example, in the middle row of the figure, where $\rho = 0$, $\sigma_v^* = \infty$ when $\psi = 0.2$. As ψ increases to 1, σ_v^* takes on an interior value around 1.4. Finally, when ψ rises to 1.8, σ_v^* drops to zero. Intuitively, what is happening here is that the CB renders those signals precise that target its favored state strongly. What the example of the middle row of Figure 14 also makes clear is the effect of ψ on the interiority threshold $\rho^i(\psi)$. For the same alignment, $\psi = 0.2 < 1$ leads to $\sigma_v^* = \infty$, while a higher ψ, $\psi = 1$, results in an interior solution. Thus, as foreshadowed by Figure 13, the interiority threshold $\rho^i(\psi)$ is decreasing in ψ.

There is something special about $\psi > 1$, however. The case when the CB's signal is targeted towards θ_t, the CB's preferred state, leads to different behavior of optimal precision than $\psi \leq 1$. As the right panel of Figure 13 recalls, not only does this involve the only bang-bang σ_v^*-behavior in the dynamic model, so that σ_v^* switches from ∞ to 0 once ρ exceeds the bang-bang threshold $\rho^b(\psi)$, but σ_v^* rises above 0 again for ρ high enough. We can read off why this happens from the right column of Figure 14. The cases of $\rho = -0.9$ or 0 are the only instances on the figure where $I(\theta_t, s_t) > I(\theta_{t+1}, s_t)$. This is because for a combination of sufficiently low alignment and sufficiently high θ_t-targetedness, the CB's signal is very informative about θ_t, but this informativeness does not carry over to θ_{t+1}. In this case, the CB can max out informativeness about θ_t by setting $\sigma_v^* = 0$. If alignment is too low, however, as in the top right panel of Figure 14, where $\rho = -0.9$, the informativeness of the signal about θ_t is misused by the PS to deduce clues about θ_{t+1}, and thus the CB can do no better than to shut off precision and set $\sigma_v^* = \infty$. If alignment is instead too high, then it is no longer the case that the signal is more informative about θ_t than about θ_{t+1}, and therefore adding a little noise is optimal to push $I(\theta_t, s_t)$ up and $I(\theta_{t+1}, s_t)$ down a bit.

The optimal precision choice of the CB is thus a much more complicated object in the dynamic model than in the static one. It is mainly driven by the tightness of the private sector's prior, which, in the dynamic world, is endogenous to central bank communication. A too tight prior is a constraint to the CB as it renders current communication ineffective. Therefore adding noise to the signal serves the function of loosening up the prior and distributing the information provided to the PS evenly across time.

2.4.3 Approximating the indirect utility function of the central bank

As we have seen, different levels of alignment and targetedness modulate the ways in which the PS' prior becomes a constraint to the CB. That this is a complicated interaction can be grasped from Figure 15, which plots the optimal precision σ_v^* as a function of ρ and ψ. Note that Figure 13 is a cross-section of Figure 15 for three values of ψ, and that Figure 15, therefore, gives a more complete picture of the dependence of σ_v^* on ρ and ψ. Figure 15 suggests that on average, (ρ, ψ) pairs in which both ρ and ψ are high involve higher optimal precision than pairs with low ρ and ψ. It also shows that the full precision region ($\sigma_v^* = 0$) is concentrated in the neighborhood of low autocorrelation ($\rho \approx 0$) and high targetedness toward θ_t ($\psi \gg 1$).

Lastly, the surprising interiority region ($0 < \sigma_v^* < \infty$) in the area with high θ_t-targetedness and high alignment ($\rho \gg 0$) is also clearly visible. To summarize these insights as a simple behavioral rule, I now propose an indirect utility function to qualitatively capture the interplay between alignment, targetedness and the tightness of priors in the CB's optimal precision choice.

Proposition 8. *Indirect utility of the optimal precision choice*

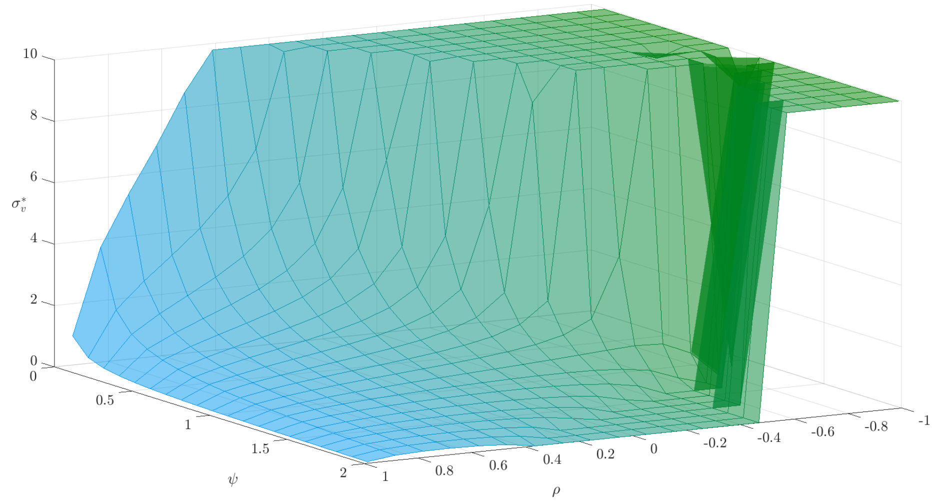
The central bank's optimal choice of signal precision can be approximated using the following rule

$$\max_{\sigma_v} \mathbb{I}_{\rho \geq \rho^i(\psi)} \Big((1 - w(\rho,\psi)) I(\theta_t, s_t) + w(\rho,\psi) I(\theta_{t+1}, s_t) \Big) - (1 - \mathbb{I}_{\rho \geq \rho^i(\psi)}) \Big(I(\theta_t, s_t) + I(\theta_{t+1}, s_t) \Big), \tag{75}$$

where $\mathbb{I}_{\rho \geq \rho^i(\psi)}$ *is an indicator function that takes on the value 1 if* $\rho \geq \rho^i(\psi)$ *and zero otherwise,* $\rho^i(\psi)$ *is the interiority threshold, and* $w(\rho, \psi) \in [0, 1]$ *is a weight function, increasing in* ψ *and with* $w(1, \psi) = 1$*.*

One can think of Proposition 8 as an algorithm specifying how a central banker should decide how clearly to communicate with the public. In the first step, the CB checks whether, given ψ, ρ exceeds the interiority threshold $\rho^i(\psi)$. If not, $\mathbb{I}_{\rho \geq \rho^i(\psi)} = 0$, and the second term of Equation (75) applies. This says that since for the given targetedness, alignment is not high enough, informative signals would be misused by the PS, and the CB had better give up on precise communication. It is thus optimal to minimize $I(\theta_t, s_t) + I(\theta_{t+1}, s_t)$, resulting in $\sigma_v^* = \infty$.

Figure 15: Optimal precision σ_v^* as a function of ρ and ψ



If instead $\rho > \rho^i(\psi)$, then $\mathbb{I}_{\rho \geq \rho^i(\psi)} = 1$, and the first term of Equation (75) becomes relevant. Since alignment is sufficiently high, the optimal thing to do is to maximize a weighted average of informativeness about θ_t and θ_{t+1}. The second step of the decision algorithm is to decide on those weights. Let us start with a limiting case: $\psi = 2, \rho = 1$, the closest tip of Figure 15. Since $w(1, \psi) = 1$ by definition, the first term of Equation (75) simply equals $I(\theta_{t+1}, s_t)$. What this is saying is that in the limit that the two states are perfectly correlated, the CB can simply maximize informativeness about θ_{t+1}, the PS' preferred state, because the two states coincide anyway. Since $I(\theta_{t+1}, s_t)$ is monotonically decreasing in σ_v, it is always maximized at $\sigma_v = 0$.

But when ρ is even marginally below 1, the weight $w(\rho, \psi)$ drops marginally below 1 also. Thus the weight on $I(\theta_t, s_t)$ in the maximization problem (75), $1 - w(\rho, \psi)$, exceeds zero, implying that as long as there is even a tiny misalignment between the two states, the CB has an incentive to push $I(\theta_t, s_t)$ up, i.e. to loosen the PS' prior beliefs. This is the crucial wedge that explains the interior solutions $\sigma_v^* > 0$ even when both alignment and θ_t-targetedness are high, and this is why the σ_v^*-function in the high-ρ, high-ψ region of Figure 15 is concave.

The indirect utility function of Proposition 8 thus summarizes how autocorrelation and targetedness affect the CB's precision choice. But the key determinant of that choice is the decision whether to provide information about θ_t or θ_{t+1}, and, consequently, how much loosening of the PS' prior is necessary. In the static model, where the CB cannot influence the prior, there is no possibility to temporally smooth information provided to the PS. In the dynamic model, instead, the (ρ, ψ) pairs determine both how much smoothing is desirable and how much is feasible. Therefore, the optimal precision choice in a dynamic world is ultimately a problem of finding the right tightness of the private sector's prior.

2.5 Conclusion

Central bank communication is a dynamic problem: in a world where economic fundamentals evolve over time and the monetary authority communicates repeatedly, central bankers have to provide information about both the present and the future. This paper is the first to investigate the implications of dynamics for central bank communication. In particular, I explore how strongly monetary policy should target current versus future fundamentals in its communication and how much noise it should introduce into its messages over time.

Analyzing two models identical up to the interpretation of the correlation between two states allows me to isolate the effect of temporal dependence between the states on the communication problem. It turns out that the fact that the private sector's prior beliefs become endogenous in the dynamic model drives a wedge between the static and dynamic communication policies. Both the optimal response of the private sector's beliefs to the prior as well as the tightness of the prior dampen the persuasiveness of the central bank's signal. This has implications both for optimal targetedness and for optimal precision.

Optimal targetedness in the two models differs due to the dampening channel of endogenous priors mattering for current beliefs. In particular, the dynamic model involves a higher weight on the central bank's preferred state in the signal than the static model for any correlation. For negative or positive but low correlation, this comes about because the private sector updates its priors too little, while for positive and high correlation, because it updates its priors too much.

Optimal precision is always weakly lower in the dynamic model than in the static one. This is driven by the central bank's desire to loosen the private sector's priors, which, if too tight, render the central bank's signal less persuasive. Adding noise to the signal flattens the information profile of the private sector in the sense that it does not receive too much information too early. The optimal precision choice thus involves the smoothing of information provision.

The dynamic communication policy studied in this paper has natural applications either to the everyday of the subset of the public that follows central bank announcements closely, or to unusual circumstances where the entire private sector pays a lot of attention to central bank announcements. An example for the former are stock-market participants. Since such individuals seek to maximize returns on investment, they expand vast amounts of resources at any given time to correctly infer the future stance of the economy from clues from the central bank. The analysis in this paper provides conditions for when the central bank should provide such investors with precise information, and when it is preferable to communicate in noisy "Fedspeak" that leaves them some residual uncertainty around the current and future economic outlook.

Times of crisis, on the other hand, are an example for when the general public listens closely to messages from central bankers. In such scenarios, current and future fundamentals are often misaligned, since even at the depth of the crisis, when the current output gap is negative, the future output gap is likely to be positive. For instance, as the US economy recovers from the Covid-19 induced recession, should the Federal Reserve attribute more weight to the current, negative output gap in its communication, or should it instead signal that good times are ahead? The present model suggests that as long as the Fed's mandate pertains to stabilizing the current output gap, its communication should focus more heavily on that, and this conclusion is stronger than it would be in a world where the private sector's beliefs are detached from priors.

3 ICT and Future Productivity: Evidence and Theory of a General-Purpose Technology

3.1 Introduction

Although there is large consensus on the importance of productivity as a driver of economic performance, there is less agreement on the underlying sources of productivity growth. For several years, much of the business-cycle literature decided to avoid such questions by proxying movements in productivity by exogenous shocks.[21] However, the robust empirical evidence of the slowdown in productivity right before the Great Recession has led recent literature to take a step back and devote more attention to the drivers of medium-term productivity growth.[22]

Along with Comin and Gertler (2006), theoretical contributions rationalize endogenous productivity dynamics by incorporating features of endogenous growth models in standard models of business cycles. Following Romer (1990), most of these papers augment final-good production functions with an expanding composite of intermediate goods invented by the Research & Development (R&D) sector in order to allow for an endogenous rate of adoption of new technologies.[23] Guided by the prediction of such theoretical work that R&D developments matter for growth, other papers attempt to provide empirical evidence of a slowdown in the productivity of the R&D sector. Specifically, they show that although research effort keeps rising, the rate of new ideas and discoveries is slowing down.[24]

[21] Kydland and Prescott (1982) and Long Jr and Plosser (1983) are among the first papers which consider productivity shocks in general equilibrium models.

[22] See Cette et al. (2016) and Byrne et al. (2016) among others.

[23] Bianchi et al. (2014), Anzoategui et al. (2016), Moran and Queralto (2017) and Guerron and Jinnai (2015) use similar techniques to endogenize growth. Bianchi et al. (2014) augment a DSGE model using a quality ladders model in the vein of Grossman and Helpman (1991). Anzoategui et al. (2016), Moran and Queralto (2017) and Guerron and Jinnai (2015), similarly to Comin and Gertler (2006), use a model of expanding variety à la Romer (1990).

[24] Jones (2009) and Bloom et al. (2017) are two contributions that highlight these facts.

Motivated by this wave of research, this paper follows a different path and argues that Information and Communication Technology (hereafter ICT) plays an important role in driving medium-term productivity in sectors that are ICT users. Our contribution is twofold. First, we provide robust empirical evidence to show that current rises in ICT investment explain significant and persistent increases in future Total Factor Productivity (hereafter TFP). Second, we analyze a standard theoretical framework in order to both rationalize our empirical results and to draw conclusions concerning the nature of the ICT contribution to future productivity.

In the empirical section, we identify technological shocks specific to the ICT sector in a Structural VAR context.[25] Our multivariate system includes three key variables: TFP, ICT investment (hereafter ICTI), and relative prices (hereafter RP). ICTI is defined as the total expenditure in equipment and computer software meant to be used in production for more than a year. Thus, an increase in ICTI is ICT capital deepening. RP is the ratio between the price level of ICT durable goods and the price level in the overall economy.

We use two main identifying restrictions in order to back out an ICT technology shock: (i) it must be orthogonal to current total factor productivity and (ii) maximize the impact effect on ICT investment. First of all, Restriction (i) is justified because the share of the ICT sector accounts for a negligible part of the whole economy implying that ICT shocks should have approximately zero impact effect on TFP. In other words, although TFP embodies ICT technical changes by construction, we argue that the direct impact effect of an ICT technological shock on TFP is so tiny to be approximately neglected in our identifying restriction. Moreover, we follow results by Greenwood et al. (1997) and Fisher (2006) (among others) to support Restriction (ii). In particular, a vast body of literature on sector-specific technological changes argues that a sectoral

[25] An interesting paper related to our empirical work is Jafari Samimi and Roshan (2012). The authors identify ICT shocks as a potential driver of the Iranian business cycle using a completely different identification strategy and obtaining qualitatively different results.

technology shock should boost quantities in the underlying sector implying a current large impact effect on ICT investment.[26]

In response to this shock, ICTI rises on impact and remains significant for several quarters. RP persistently and significantly declines for more than two years, suggesting that we are indeed identifying the correct sectoral shock. Our main result is that TFP, restricted not to respond on impact, rises after few quarters and remains significant and stable for at least 5 years, despite the tiny size of the ICT sector relative to the economy. Finally, variance decomposition analysis suggests that ICT shocks explain a remarkable size of economic fluctuations over the medium and long run. These shocks drive about 40% and 33% of the overall GDP and TFP fluctuations over a 10-year horizon, respectively.

Although our results are robust over different specifications, a critique to our empirical strategy is reverse causality coming from news on future TFP. As suggested by the news-shock literature, the positive reaction of ICTI on impact may be triggered by signals related to future increases in TFP and not to contemporaneous ICT technological improvements. In other words, our identification strategy may confound our shock of interest with a news shock which contemporaneously enhances investment in ICT capital goods. We address this issue by providing a series of alternative identification strategies which we show deliver the same time series of ICT innovations as our initial specification. The heart of these robustness checks is sequential identification of news and ICT shocks: we first identify a news shock in the spirit of Barsky and Sims (2011), and subsequently we identify our sectoral ICT shock using the previous identification strategy. Encouragingly, controlling for signals regarding future movements in TFP does not affect our results. We view this as strengthening the causality relation from current ICT investment to future TFP.

[26] For theoretical reasons discussed in more detail in Section 3.2, we do not impose any restriction on RP in our main specification; instead, we let the ICT shock maximize the impact response of ICTI. As pointed out by both Greenwood et al. (2000) and Basu et al. (2010), letting the the direction of RP be the sole identifying restriction does not properly capture technological changes between sectors. This is the main reason why we do not impose the direction of RP as a direct identifying condition.

In order to understand the economics behind our empirical results, we then analyze a two-sector DSGE model which allows ICT to be the general-purpose technology (hereafter GPT) of the whole economy. There are two main justifications for interpreting ICT as a GPT. On the one hand, there is a vast literature that makes a case for the general-purpose nature of ICT capital.[27] On the other hand, the tiny share of the ICT sector both in overall investment and overall output makes our results of a strong and persistent TFP increase after an ICT shock hard to interpret in absence of an additional force such as an externality coming from the general-purpose property of ICT capital.

In our model, both sectoral production functions are fed with three inputs: (i) labor, supplied by households, (ii) hard capital, produced by the final sector, and (iii) ICT capital, produced by the ICT sector. As explained by both Basu et al. (2003) and Basu and Fernald (2007), a GPT should be able to enhance accelerations in productivity in sectors that are users of the underlying technology. We then interpret ICT as the GPT of the last 30 years of the U.S. economy assuming that exogenous technological changes in the ICT sector are able to affect economy-wide productivity above the direct effect coming from the technology increase itself.[28] In particular, when an ICT technology shock arrives, both sectors accumulate ICT capital since it is easier to produce and cheaper to buy. This ICT capital deepening consequently enhances the productivity of both sectors by means of a spillover coming from the accumulated ICT capital stock.

What is the intuition behind our assumption of ICT as a GPT showing up in the economy as a spillover? To better understand this externality assumption, consider the special nature of ICT capital. Since the purpose of ICT capital is to improve information acquisition and sharing, the quality and speed of communication fundamentally depends on the diffusion of these technologies among agents. As a simple example, owning a

[27] See for example Oliner and Sichel (2000) and Stiroh (2002) amongst others.

[28] A clarification is in place here. In a two-sector model, total factor productivity is a convolution of the two exogenous productivities. Thus sectoral productivity changes trivially show up in TFP as a direct effect. Our assumption of ICT capital being a GPT means that overall productivity responds much more to an ICT-sector-specific technology shock than warranted by this shock directly. We address this question in detail in the main text in Section 3.2.

mobile phone enables one to contact another person instantaneously only if the other person is also endowed with the same technology. As a result, the effectiveness of ICT capital is intrinsically related to its own diffusion. This line of thought is what leads us to augment the production function of each sector with a spillover effect capturing the diffusion of ICT capital. Having a spillover arise from a state variable is also consistent with both the GPT literature above and with our empirical results, since it also leads to model dynamics in which the accumulation of ICT capital is a slow process and the benefits of an ICT technology shock show up in the production functions of ICT-users with lags.[29]

As a last step, we use both our empirical and theoretical results to estimate the key parameters of the model via impulse-response matching to an ICT technology shock. The key parameters are (i) the elasticity of productivity to ICT capital diffusion, namely the parameter which governs the spillover effect, and (ii) the standard deviation and (iii) persistence of ICT technology shocks. Results consistently obtain a positive spillover effect of ICT capital deepening on TFP, confirming that within this class of theoretical models, data supports the existence of spillovers from ICT capital. Thus, our theoretical model suggests to interpret the responses obtained in the empirical section in light of ICT as a general-purpose technology which enhances productivity of ICT capital users through a spillover effect related to its own diffusion.

The paper is structured as follows. We present empirical results and main robustness checks in Section 3.2. We then present and analyze the two-sector DSGE model in Section 3.3. We estimate via impulse-response matching key parameters of the model and assess its empirical reliability in Section 3.4. Section 3.5 concludes.

[29] Notice that differently to Basu et al. (2003) and Basu and Fernald (2007), we interpret the general-purpose nature of ICT in the spirit of an endogenous growth model.

3.2 Empirics

In this section we present our main empirical results. Our attempt is to properly identify technological shocks which are specific to the ICT sector in a Structural VAR context and to analyze their impact on key macroeconomic variables.

3.2.1 Main Specification

We start by illustrating our main specification where we impose minimal discipline on the model. In subsequent sections, we try vastly different empirical strategies. It turns out that the set of results presented here are consistent with the different robustness checks.

3.2.1.1 Data Our first-step specification is the following five-variable VAR

$$\begin{pmatrix} TFP_t \\ ICTI_t \\ GDP_t \\ C_t \\ RP_t \end{pmatrix} = B(L) \begin{pmatrix} TFP_{t-1} \\ ICTI_{t-1} \\ GDP_{t-1} \\ C_{t-1} \\ RP_{t-1} \end{pmatrix} + i_t \tag{76}$$

where TFP_t is the log-level of Fernald total factor productivity at time t, $ICTI_t$ is the log-level of real information and communication technology investment at time t,[30] GDP_t is the log-level of real gross domestic product at time t, C_t is the log-level of real consumption at time t, and RP_t is the log-deviation of the ratio between prices of ICT goods and services and the consumer price index (CPI).[31] [32] Our dataset is quarterly

[30] $ICTI_t$ is defined as the total expenditure at time t in equipment and computer software meant to be used in production for more than a year.

[31] Except for RP_t, which is not cointegrated with the remaining variables, we opt for estimating the VAR in levels since this produces consistent estimates of the impulse responses and is robust to cointegration of unknown forms. In particular, as suggested by Hamilton (1994), when there is uncertainty regarding the nature of common trends, estimating the system in levels is considered the conservative approach. The alternative approach is to estimate the system as a vector error correction model (VECM); our results remain very similar also in this case.

[32] Using the PCE deflator instead of the CPI yields similar results.

and covers the U.S. economy from 1989:Q1 to 2017:Q1.[33] $B(L)$ is a (5×5) matrix of lag-operator functions of the same order. Following the Bayesian Information Criterion (BIC), we choose one lag, which implies that we regress variables at time t with their own lagged values at $t-1$.[34] Finally, i_t is a (5×1) vector of correlated innovations where $\Sigma = i_t' i_t$.

3.2.1.2 Empirical Strategy

Our simplest identification strategy assumes that an ICT technological shock (hereafter ICT shock) has (i) no impact effect on TFP and (ii) maximal impact effect on ICTI. We justify these two assumptions with both empirical and theoretical arguments. First of all, as data released in April, 2018 by the Bureau of Economic Analysis (BEA) shows, the real value added of the information sector on real GDP is slightly below 5% for the underlying quarter.[35] Thus, since the ICT sector accounts for a negligible part of the whole economy, we think it is safe to assume that an ICT shock is orthogonal to current TFP.

In addition, the theory of multi-sector models, as presented by Greenwood et al. (1997), predicts that a productivity increase in one sector should lead to sectoral output becoming cheaper. Thus, we expect that an ICT shock should enhance sector-specific investment since ICT goods are less costly to produce and to buy. As a result, we expect an ICT shock to have a maximal impact effect on ICTI.

Using a notation similar to Barsky and Sims (2011), we implement our identification strategy as follows. Let y_t be the (5×1) vector of observables of length T presented

[33] All data are from the BEA except for TFP, which is the series constructed by John Fernald, available on his website.

[34] The Hannan-Quinn Criterion (HQ) suggests to use two lags. Results remains consistent following this second criterion.

[35] There are many ways to quantify the size of the ICT sector, both in terms of what share one considers and how one defines the ICT sector. The number in the text refers to the value added share of the ICT sector, defined as in Section 76. The statement that the ICT sector is tiny is however not sensitive to the definition of the sector or to the choice of share. Even with broader definitions, one obtains numbers in the range of 5%.

above. The reduced-form VAR takes the following form,

$$y_t = B(L)y_{t-1} + i_t$$

Assume now that there exists a linear combination that maps innovations i_t to structural shocks s_t,

$$A_0 s_t = i_t$$

which implies the structural-form VAR

$$A_0^{-1} y_t = C(L)y_{t-1} + s_t$$

where $C(L) = A_0^{-1}B(L)$ and $s_t = A_0^{-1}i_t$. The impact matrix A_0 must be such that $\Sigma = A_0 A_0'$. Notice that for any arbitrary orthogonalization $\tilde{A}_0$ such that $\Sigma = \tilde{A}_0 \tilde{A}_0'$, impact matrix A_0 can be back out by the appropriate orthonormal matrix D ($DD' = I$) where $A_0 = \tilde{A}_0 D$.

Thus, the matrix of impact responses to all shocks is:

$$\Pi(0) = \tilde{A}_0 D$$

and is specifically formed by the following elements, denoting the responses of variable i to shock j,

$$\Pi_{i,j}(0) = e_i' \tilde{A}_0 D e_j$$

where e_k is a selector column vector of the same dimension as $\tilde{A}_0$ with 1 in the kth element and zero elsewhere. In particular, notice that e_j is selecting a specific column of D. Let this column be denoted by γ_j. Using this notation, $\tilde{A}_0 \gamma_j$ is the vector of impact responses of all the variable to shock j.

Observe from System 76 that TFP_t is ordered first and $ICTI_t$ second. Our identification strategy is then the solution to the following problem:

$$\max_{\gamma_j} \Pi_{2,j}(0) = e_2' \tilde{A}_0 \gamma_j \tag{77}$$

subject to

$$\Pi_{1,j}(0) = e_1' \tilde{A}_0 \gamma_j = 0, \quad \text{and} \tag{78}$$

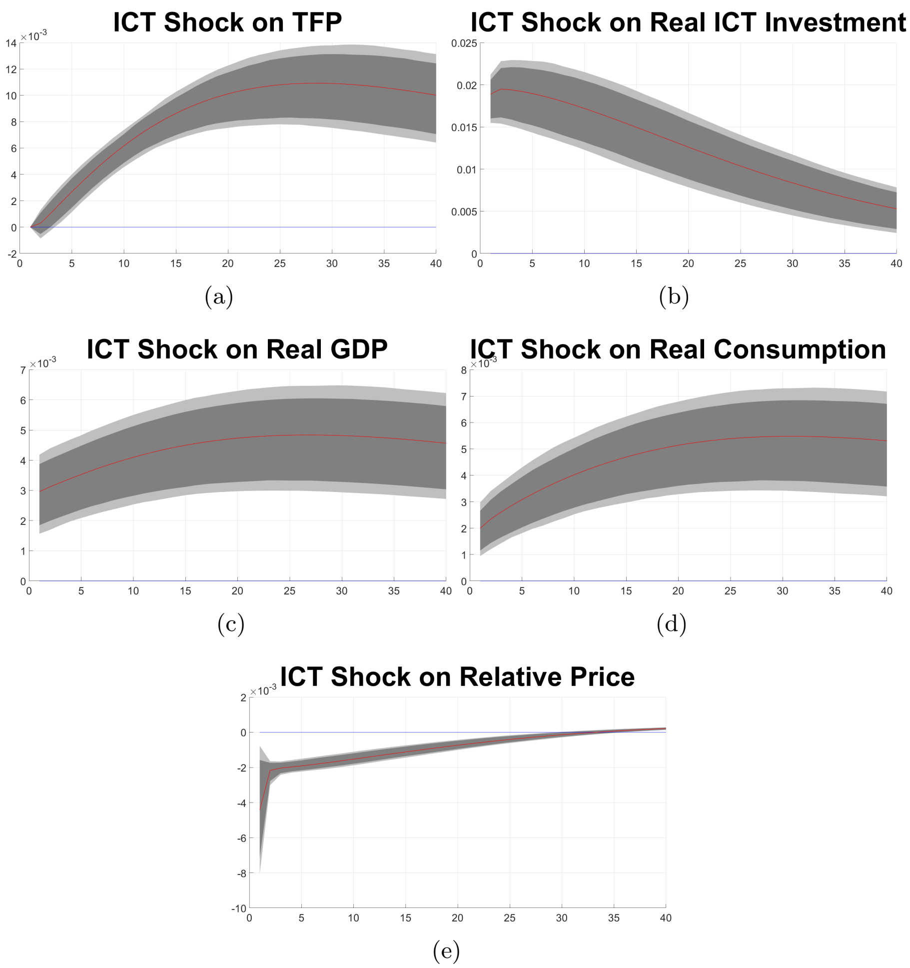
$$\gamma_j' \gamma_j = 1 \tag{79}$$

where j represents the arbitrary position of the ICT shock. Then, in order to ensure that this identification belongs to the space of possible orthogonalizations of Σ, the problem is formulated in terms of choosing γ_j conditional on any orthogonalization, $\tilde{A}_0$. Objective function 77 imposes that an ICT shock has a maximal impact effect on ICT investment. Constraint 78 orthogonalizes current TFP to ICT shocks and Constraint 79 satisfies the condition that γ_j is derived from an orthogonal matrix D.

3.2.1.3 Main Set of Results Figure 16 shows the estimated impulse responses of System 76 to the identified ICT shock. The shaded gray areas are the 90% and 95% confidence bands from the bias-corrected bootstrap procedure of Kilian (1998) using 2000 simulations. Our main interest, panel (a) of the figure, shows the impulse response of TFP to an ICT shock. TFP takes around 4 quarters before displaying a positive and significant effect and reaches its peak of 1.2% after 24 quarters. Real ICT investment has a large and positive impact response of almost 2% that gets even larger after a quarter. Then, it slowly starts to decay, remaining significant for more than 40 quarters. Real GDP has a significant impact response of 0.3% and reaches its peak of almost 0.5% approximately at the same time as TFP. Similarly, real consumption has an impact effect of 0.2% with a delayed peak of 0.5%. Finally, panel (e) depicts that response of relative prices. Despite imposing no restriction here, the response of RP is broadly in line with multi-sector theory. Relative prices drop significantly on impact by

0.4% and remain persistently below their own steady state value for almost 9 years.

Figure 16: Empirical impulse response of TFP, real ICT investment, real GDP and relative prices to an ICT shock.



The red solid lines are the estimated impulse responses to our ICT shock. The shaded dark gray area and the shaded light gray area are the 90% and 95% confidence intervals, respectively, from 2000 bias-corrected bootstrap replications of the reduced-form VAR. The horizontal axes refer to forecast horizon and the units of the vertical axes are percentage deviations.

In addition, Table 4 presents forecast error variance decompositions for the ICT shock on each variable. The table shows decompositions on impact, and at a one-, two-, four-, six-, and ten-year horizon. These results are very interesting in their own right, as they hint that the ICT sector may have a particular role for the overall productivity of the economy. In line with the impulse responses, ICT shocks, which are orthogonal to current productivity, explain nothing of current TFP fluctuations. At a ten-year horizon, however, this fraction is over a third! This crucial result is a strong sign that the ICT sector matters for overall productivity over and beyond the direct contribution of its sectoral productivity (which, as we have seen, is marginal). At the same time, ICT shocks drive almost the whole variation in ICT investment on impact, with a slow decay over time to below 50% after 10 years. Interestingly, both output and consumption have a remarkable reaction on impact: 26% and 19%, respectively. Moreover, this effect tends to increase, reaching 40% in both cases at the maximal horizon. Finally, ICT shocks only account for a small fraction of movements in relative prices. The fraction of fluctuations explained on impact is only 6% with a peak of 14% between four and six years. Overall, these variance decompositions all seem to suggest that the ICT shock leads to complex forms of reallocation and accumulation that take place over the medium run in the economy and that seem to result in productivity growth. Our structural model of Section 3.3 provides a theoretical framework to think about these dynamics and to give them an economic interpretation.

Table 4: Variance decomposition

	$h=1$	$h=4$	$h=8$	$h=16$	$h=24$	$h=40$
TFP	0	0.0023	0.0194	0.1088	0.2273	0.3382
ICT Investment	0.9997	0.9038	0.7964	0.6320	0.5310	0.4371
Real GDP	0.2620	0.3061	0.3486	0.3936	0.4046	0.3881
Real Consumption	0.1952	0.2638	0.3219	0.3931	0.4188	0.4064
Relative Prices	0.0618	0.0967	0.1276	0.1511	0.1516	0.1467

The letter h denotes the forecast horizon. The numbers refer to the fraction of the forecast error variance of each variable at various forecast horizons to the identified ICT shock

3.2.2 Controlling for News Shocks

In this section, we present a set of robustness checks aimed to show that our previous results are not driven by future signals regarding exogenous productivity. The main concern against the specification of 3.2.1 is reverse causality coming from the presence of news about future TFP that is not accounted for. Indeed, the news-shock literature warns that the positive reaction of ICTI on impact may be triggered by signals related to future increases in productivity and not to a contemporaneous ICT shock. In other words, our identification strategy may confound our shock of interest with a news shock which contemporaneously enhances investment in ICT capital goods. We address this issue by providing two main alternative identification strategies which we show deliver the same time series of ICT shocks as our initial specification.

As a first-pass test, we check if our results are robust once we remove all the forward-looking variables whose fluctuations may be unrelated to sector-specific technological changes: consumption and output. Technically speaking, consumption and output may Granger-cause movements in future TFP for reasons which are orthogonal to ICT shocks. For example, the forward-looking nature of consumption may provide the VAR with information regarding future changes in TFP not related to an ICT shock, which our identification strategy is not able to filter out.

Our second and main robustness test is running a VAR in which we identify both ICT shocks and news shocks. In particular, we apply a sequential identification where we first identify a news shock in the spirit of Barsky and Sims (2011), and subsequently we identify our sectoral ICT shock using our original identification strategy. This second strategy has the specific purpose of directly filtering out all the current movements in forward-looking variables related to future fluctuations of TFP which are not related to current ICT shocks, as such fluctuations are all captured by the identified news shock.

As is illustrated in the rest of this section, both our robustness checks recover the same series for structural ICT shocks as our initial specification, and thus leave our results unaffected. We therefore conclude that controlling for signals regarding future movements in productivity does not affect our identified ICT shocks, and thus confirm the causal relation between current ICT investment and future TFP.

3.2.2.1 Removing Forward-Looking Variables We now show our first-pass test in detail. As discussed by Sims (2012), Forni and Gambetti (2014) and Barsky et al.

(2015), the presence of forward-looking variables in the VAR is a necessary condition to correctly identify a news shock. In particular, since the univariate TFP process cannot predict its own future values,[36] a news shock can only be identified if there are variables that Granger-cause TFP inside the VAR. For example, Beaudry and Portier (2006) use current movements in stock prices to predict future productivity, but also contemporary fluctuations in consumption, hours, investment, may implicitly reflect information regarding future technical change.

Thus, as a simple test, we redo the exercise described in Section 3.2.1, omitting the two forward-looking macroeconomic variables not directly related with the ICT sector: consumption and output.[37] If we are actually confounding news shocks with our ICT shock series then we would expect to see some differences in this second estimate, because absent the two forward-looking variables, our VAR is deprived of a source of information useful to recover a news shock and should thus no longer be able to capture it appropriately.

Without consumption and output, we thus estimate the following three-dimensional system

$$\begin{pmatrix} TFP_t \\ ICTI_t \\ RP_t \end{pmatrix} = B(L) \begin{pmatrix} TFP_{t-1} \\ ICTI_{t-1} \\ RP_{t-1} \end{pmatrix} + i_t \tag{80}$$

where TFP_t, $ICTI_t$, and RP_t are defined and treated as in System 76. As before, the dataset is quarterly and covers the U.S. economy from 1989:Q1 to 2017:Q1. Finally, analogously to before, we follow the Bayesian Information Criterion (BIC) to choose one lag. Using the same identification strategy presented in 3.2.1.2, we recover a series of ICT shocks. The newly recovered series is correlated at 95% with the one estimated in the five-dimensional system, indicating that our main specification is not contaminated

[36]It is well-accepted that TFP growth can be approximated as a white noise process.

[37]Note that output implicitly reflects the behavior of hours and investment once we control for consumption.

by news. Figure 26 in Appendix C.1 shows the ICT shock series for both the three-dimensional system and the five-dimensional one. In line with the very high correlation coefficient, the two series overlap perfectly, suggesting that both procedures recover the same structural shock series. Thus, our first-pass test indicates that our initial result is not driven by the forward-looking power of consumption and output on TFP.

However, although this result is encouraging, the first-pass test has weak power because the jump of ICTI on impact may itself be triggered by exogenous signals on future TFP. In this case, ICTI would be itself the source to Granger-cause TFP and our ICT shock would be upward biased by news. To address this deeper concern, we proceed to employ a more complicated strategy aimed at directly capturing and filtering out news.

3.2.2.2 Sequential Identification Strategy We now turn to an alternative procedure to rule out the presence of news from our recovered ICT shock. In this second test we run a VAR where we identify both ICT shocks and news shocks. In contrast to our first-pass test, this approach aims to directly identify and filter out future signals unrelated to current ICT shocks.

We apply a sequential identification which involves first identifying a news shock in the spirit of Barsky and Sims (2011), and subsequently identifying the ICT shock using our original strategy. Specifically, in the first step the news shock is identified as the shock orthogonal to current TFP that best explains its future movements. Then, as a second step, the ICT shock is identified as the shock orthogonal to current TFP that maximizes the impact effect on ICT investment.

However, in order to correctly employ this procedure, we need to take care of a second issue. As Barsky and Sims (2011) warns,

> A general objection to our emprical approach would be that a number of structural shocks, which are not really news in the sense defined in the lit-

> erature, might affect a measure of TFP in the future without impacting it immediately. Among these shocks might be research and development shocks, investment specific shocks, and reallocative shocks. Our identification (and any other existing VAR identifications) would obviously confound any true news shock with these shocks.
>
> Barsky and Sims (2011), p. 278.

In other words, by naively employing the procedure presented above, any ICT shock would be interpreted as a news shock since it is orthogonal to current TFP and explains its future movements. Put plainly, to get the identification of the two shocks right, we need to add a restriction that is able to distinguish between them.

Thus, to avoid confounding ICT shocks with news shocks, we need to augment the first step with an extra restriction. The restriction we will use here is derived from a two-sector structural model in the tradition of Greenwood et al. (1997).[38] Here, we refrain from presenting the model in its entirety; the reader is referred to Section 3.3. Instead, Proposition 9 presents the key property of the model that gives rise to our identification assumption.

Proposition 9. *In a two-sector economy, if both production functions display identical input shares, then any economy-wide productivity shocks do not affect relative prices between the two sectors.*

Proof. See Appendix C.2. □

Proposition 9 implies that an anticipated future sector-neutral technological change (the so-called news shock) never affects relative prices between sectors. Obviously, this statement is correct under the extreme assumptions that (i) the shock is expected to have

[38] While our model belongs to the class of models along the lines of Greenwood et al. (1997), our exposition is closer to that of Oulton (2007). We discuss this choice in detail in Section 3.3.

the same impact on the two sectors and (ii) production functions have identical input shares. If those assumptions hold in the data, then the correct additional restriction would be to restrict the impulse response of relative prices to a news shock to hard zero at all horizons. However, this restriction would obviously penalize the identification of news shocks since it is unlikely that assumptions (i) and (ii) perfectly hold in the data. We therefore opt for a more moderate strategy where we impose that the impulse response of relative prices to news should be zero on some specific horizons rather than at all horizons. While results are robust to different horizon restrictions, our favorite specification is the one which sets relative prices to have zero response in the first three periods.

Data Motivated by Section 3.2.2.1, we augment our initial VAR system with stock prices in order to provide an additional forward-looking variable to the system. Thus, our current specification is the following six-variable VAR,

$$\begin{pmatrix} TFP_t \\ SP_t \\ ICTI_t \\ GDP_t \\ C_t \\ RP_t \end{pmatrix} = B(L) \begin{pmatrix} TFP_{t-1} \\ SP_{t-1} \\ ICTI_{t-1} \\ GDP_{t-1} \\ C_{t-1} \\ RP_{t-1} \end{pmatrix} + i_t \tag{81}$$

where TFP_t, $ICTI_t$, GDP_t, C_t, and RP_t are defined and treated identically to System 76. SP_t represents the log-transformation of Standard & Poor's 500 stock price index. As before, the dataset is quarterly and covers the U.S. economy from 1989:Q1 to 2017:Q1 and following the BIC and HQ criteria we choose one lag.

Empirical Strategy Using the notation of Section 3.2.1.2, assume for simplicity that $B(L) = B$ allows for only one lag. Define the impulse response of variable i to the identified shock j at time t as,[39]

$$\Pi_{i,j}(t) = e_i' B^t \tilde{A}_0 D e_j = e_i' B^t \tilde{A}_0 \gamma_j$$

and its variance decomposition up to time h as

$$\Omega_{i,j}(h) = \frac{\sum_{t=0}^{h} e_i' B^t \tilde{A}_0 D e_j e_j' D' \tilde{A}_0' B'^t e_i}{e_i' (\sum_{\tau=0}^{H} B^t \Sigma B'^t) e_i} = \frac{\sum_{t=0}^{h} e_i' B^t \tilde{A}_0 \gamma_j \gamma_j' \tilde{A}_0' B'^t e_i}{e_i' (\sum_{\tau=0}^{H} B^t \Sigma B'^t) e_i}$$

Given this and recalling that TFP is ordered first, ICTI second and RP sixth, the identification strategy can be summarized as follows,

Step 1 - Identification of γ_{news}

$$\max_{\gamma_{news}} \Omega_{1,news}(h) = \frac{\sum_{t=0}^{h} e_1' B^t \tilde{A}_0 \gamma_{news} \gamma_{news}' \tilde{A}_0' B'^t e_1}{e_1' (\sum_{\tau=0}^{H} B^t \Sigma B'^t) e_1}$$

subject to

$$\Pi_{1,news}(0) = 0,$$

$$\Pi_{6,news}(0) = \Pi_{6,news}(1) = \Pi_{6,news}(2) = 0, \quad \text{and}$$

$$\gamma_{news} \gamma_{news}' = 1.$$

where the first constraint represents the zero-impact restriction of news on TFP. The second constraint is the restriction derived from Proposition 9 to have the response of relative prices to news relatively small.[40] Finally, the last constraint imposes that γ_{news} should be a column derived from the orthogonal matrix D.

[39] We also suppress the constant to simplify the notation. Of course, our identification strategy can be applied for $B(L)$ allowing for more than one lag and with the constant.

[40] Our identification strategy is robust to different restrictions to relative prices. Results hold if we set the zero restriction on other horizons.

Step 2 - Identification of γ_{ICT}

$$\max_{\gamma_{ICT}} \Pi_{2,ICT}(0) = e_2' \tilde{A}_0 \gamma_{ICT}$$

subject to

$$\Pi_{1,ICT}(0) = 0,$$

$$\gamma_{news}\gamma_{ICT}' = 0, \quad \text{and} \quad \gamma_{ICT}\gamma_{ICT}' = 1.$$

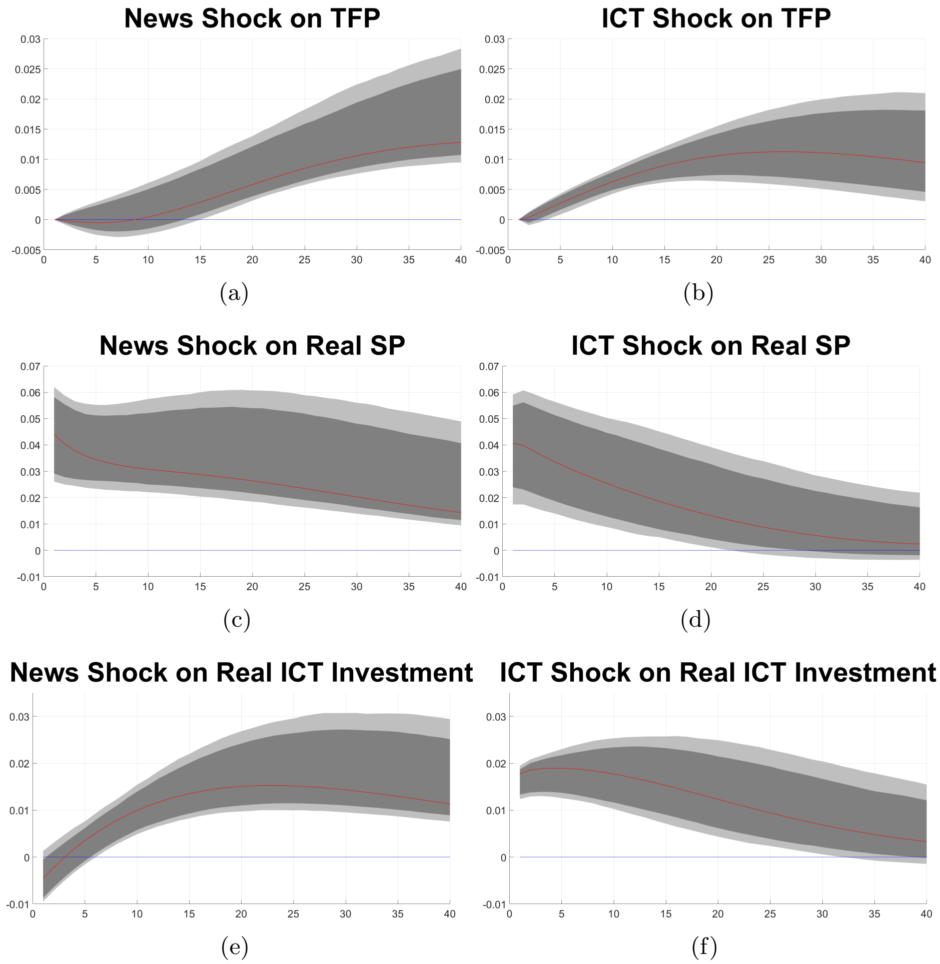
where the first constraint represents the zero-impact restriction of ICT on TFP and the last constraint imposes that both γ_{news} and γ_{ICT} should be two different columns derived from the orthogonal matrix D.

Results Figures 17 and 18 show estimated impulse responses to both news and ICT shocks. Figure 17 focuses on the responses of TFP, stock prices, and ICT investment, while Figure 18 shows the responses of GDP, consumption and relative prices. The shaded gray areas are the 90% and 95% confidence bands from the bias-corrected bootstrap procedure of Kilian (1998) using 2000 simulations.

A first comforting indication is that the responses to an ICT shock are almost identical to the ones from our main identification of Section 3.2.1. After few lags, TFP displays a positive and significant response and reaches its peak of 1.2% after 6 years. Interestingly, stock prices jump on impact and the effect remains significant over the medium run. Real ICT investment has a large and positive impact response and slowly starts to decay, remaining significant for a long period of time. Both real GDP and real consumption have a significant impact response and reach their peak approximately after 6 years. The response of relative prices is negative and persistent as we have previously discussed. Finally, the ICT shocks explains 27% of TFP fluctuations over a 10-year horizon. Also this result is completely in line with the previous results presented in Table 4.

In addition, we show the impulse responses to a news shock to confirm that our

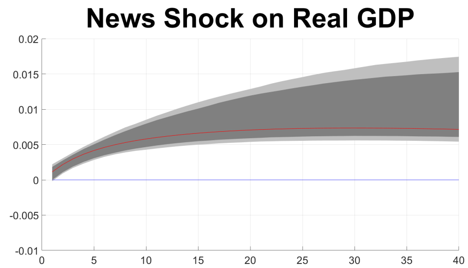
Figure 17: Empirical impulse responses of TFP, Real Stock Prices, and Real ICT Investment to a news shock and an ICT shock.



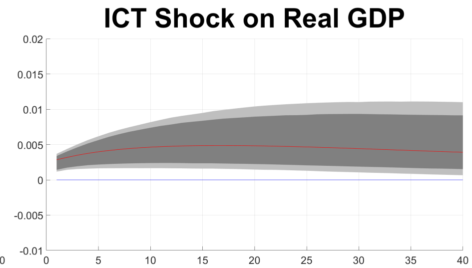
The red solid lines are the estimated impulse responses to our ICT shock. The shaded dark gray area and the shaded light gray area are the 90% and 95% confidence intervals, respectively, from 2000 bias-corrected bootstrap replications of the reduced-form VAR. The horizontal axes refer to forecast horizon and the units of the vertical axes are percentage deviations.

identification strategy is picking up news shocks correctly (left panel of the same figures). As the figures depict, constraining the response of relative prices to news does not qualitatively affect the behavior of news shocks; our identified news shock generates responses broadly in line with those found by the news shock literature. In particular, in line with Barsky and Sims (2011), it takes more that 2 years after the news shock for TFP to display a positive, significant and strongly persistent effect. Stock prices have a strong forecasting power and show a 4% impact effect which is persistent and significant

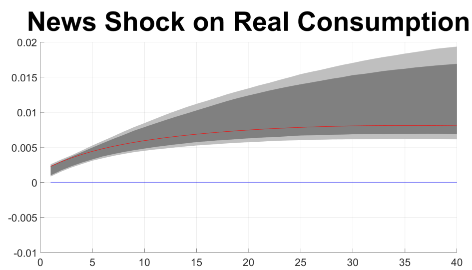
Figure 18: Empirical impulse responses of Real GDP, Real Consumption, and Relative Prices to a news shock and an ICT shock.



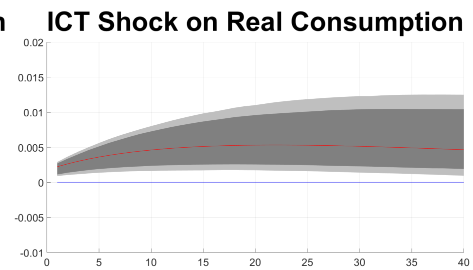
(a)



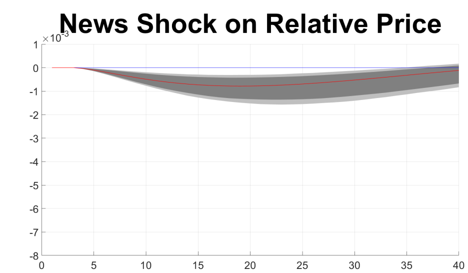
(b)



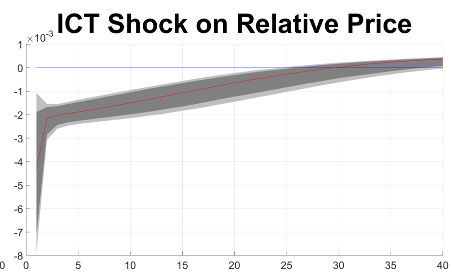
(c)



(d)



(e)

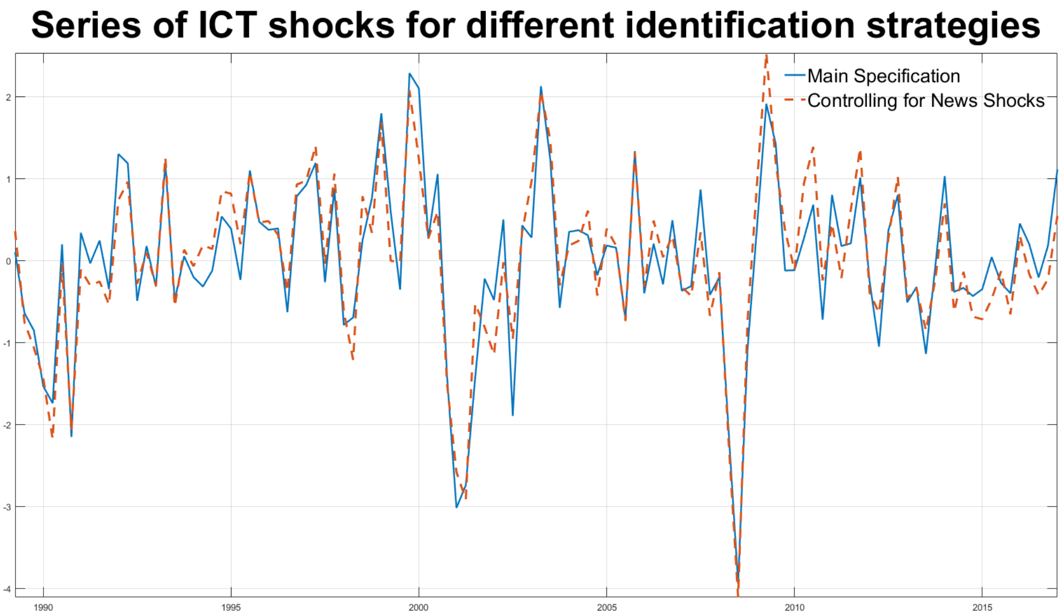


(f)

The red solid lines are the estimated impulse responses to our ICT shock. The shaded dark gray area and the shaded light gray area are the 90% and 95% confidence intervals, respectively, from 2000 bias-corrected bootstrap replications of the reduced-form VAR. The horizontal axes refer to forecast horizon and the units of the vertical axes are percentage deviations.

in the long run. Real ICT investment does not show a positive impact response but it slowly starts to increase, getting significant after almost 2 years. As found by Barsky and Sims (2011) in their main specification, real GDP has an insignificant impact effect, but it rises fast with all the variables of the system. Moreover, as was expected, considering the forward-looking nature of consumption, real consumption has a significant impact response and reaches its peak once the TFP response becomes significant. Finally, relative prices are initially constrained to have a zero response, and then show a small, negative and persistent deviation for at least five years which eventually dies out.

Figure 19: ICT shock series using two different empirical strategy.



Blue solid line represents shock series for the 5-dimension system 76 presented in 3.2.1. Red dotted line represents shock series for the 6-dimensional system 81 presented in 3.2.2.2 where we employ a 2-step identification strategy in order to control for news shocks.

Thus, not only are the responses to the ICT shock in line with those of our main specification, but the identified news shock also seems to be capturing the right thing: anticipated future technology changes. Having thus correctly filtered out news components from our ICT shock, we now check how this newly recovered ICT shock series compares to the series backed out by our main specification of Section 3.2.1.2. It turns out that the correlation between the two shock series is 94%! Figure 19 portrays the two ICT shock series. Again, the two recovered shocks are nearly identical, visually reinforcing the very high correlation between them. Thus we can confidently state that our initial result is not confounding any signal regarding future TFP with our estimated ICT shock.

3.3 Model

In the previous section, we have documented results regarding the relation of current ICT investment and future productivity. Clearly, to rationalize the empirical responses of aggregate macroeconomic variables to ICT shocks, we need a two-sector approach that distinguishes ICT goods from other output. In this section, we present a simple two-sector model that rationalizes our structural VAR estimates. We also attempt to give an economic explanation for why a sector that accounts for a minuscule fraction of the entire economy can give rise to medium-run responses of the magnitude we documented in the previous section. As discussed below in more detail, we therefore entertain the possibility that there are spillover effects coming from the stock of ICT capital, and we confront the model with data to see if this explanation can be supported empirically.

The model presented has several common features with the two-sector environment of Greenwood et al. (1997) where one sector displays a faster technology than the other. Three main assumptions distinguish our model from the one by Greenwood et al. (1997). First of all, we focus specifically on the ICT capital sector, rather than equipment, as the sector with faster technological improvement. This assumption will be necessary if we want to fit the fact that relative prices between ICT investment and the overall economy display a downward sloping trend. A second difference is that we consider two production functions - one for each sector. As Oulton (2007) shows, the two formulations are isomorphic, so our modeling choice has no implication for the resulting dynamics. The reason we choose this approach is because it allows for definitions of GDP and TFP in the model that are consistent with the way these variables are constructed in the national income accounts.[41] Finally, and crucially, we assume that the total ICT capital stock has a positive spillover effect on the production function of both sectors.

What considerations lead us to augment both production functions with a spillover

[41] See also Whelan (2003).

from ICT capital? First of all, the large, significant and persistent impulse responses of aggregates like TFP, GDP and consumption to an ICT shock are hard to square with the fact that the ICT sectoral share is less than 5%. This seems to suggest that there is an additional, indirect effect coming from the ICT sector. A spillover is a natural way of capturing such an indirect effect.

Secondly, a long line of literature going back to the early 1990s postulates that ICT may be a general-purpose technology.[42]

Following the classical definition of Bresnahan and Trajtenberg (1992), GPTs are technologies that alter the process of economy-wide production, and thus affect the productivity of the entire economy. This effect can come about through several channels. For example, investment in ICT goods may trigger accumulation of unobserved complementary capital, as suggested by both Basu et al. (2003) and Basu and Fernald (2007). Conversely, as advocated by Bresnahan et al. (2002), new ICT goods may change managerial practices, thus affecting the production process. We take a broader view in which we simply consider the diffusion of ICT goods to be the crucial element of ICT's general-purpose nature.

To understand what we mean by ICT diffusion, consider again the example of purchasing a mobile phone from the Introduction. Clearly, an individual's productivity improves when acquiring a mobile phone, since she gains access to many functionalities she couldn't enjoy previously. However, the full productivity boost can only materialize when others also use mobile phones. In fact, the more mobile phones there are in the economy, the more productivity gain the individual experiences. Thus, the higher the diffusion of ICT, the higher is its effect on productivity. This is the effect our spillover assumption means to capture. The key exercise, then, of our theoretical model will be to verify via impulse-response matching whether data support the existence of a spillover

[42] Oliner and Sichel (2000), Basu et al. (2003), and Basu and Fernald (2007) are some examples. In addition, we are somehow related to the so-called Gordon-Mokyr debate where the two authors present opposing views of future innovation. Among the series of published papers we highlight the followings two: Gordon (2012) and Mokyr (2014).

and thus the hypothesis of ICT as a GPT in the three decades following 1989 in the US economy.

3.3.1 Household

The household in our economy is completely standard. The economy is inhabited by a representative agent who maximizes the expected value of his lifetime utility as given by

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t U(c_t, l_t)\right] \tag{82}$$

with

$$U(c_t, l_t) = \log(c_t) - \frac{1}{1+\frac{1}{\chi}} l_t^{1+\frac{1}{\chi}}$$

where c_t and l_t represent consumption and labor at time t.

3.3.2 Final-Good Producers

The production side of the final-good sector consists of a set $[0,1]$ of firms. In each period, final-good firm $j \in [0,1]$ produces $y_t^c(j)$ using the services of labor l_t, and two types of capital: hard capital, k_t^c and ICT capital, k_t^i. In addition, the production function is augmented with (i) neutral productivity common across sectors, (ii) sector-specific productivity, and (iii) a positive spillover effect related to the diffusion of ICT capital in the whole economy. Production of final output of firm j is undertaken in accordance with

$$y_t^c(j) = A_t^c \; (k_t^c(j))^a \; (k_t^i(j))^b \; (l_t(j))^{1-a-b}, \quad 0 < a, b < 1 \tag{83}$$

where productivity is equal to $A_t^c = \eta_t \ \theta_t^c \ (k_t^i)^\gamma$. η_t is the neutral technological component which moves along the following deterministic trend

$$\eta_t = (\Gamma^\eta)^t e^{\varepsilon_t}, \tag{84}$$

$$\varepsilon_t = \rho_\nu \varepsilon_{t-1} + \nu_t \tag{85}$$

where $\rho_\nu \in (0,1)$ and $\nu_t \sim N(0, \sigma_\nu^2)$ represents a neutral technology shock.[43] θ_t^c is the sector-specific technological component, which moves along the following deterministic trend

$$\theta_t^c = (\Gamma^c)^t. \tag{86}$$

Then, γ represents the parameter that governs the size of the spillover effect of ICT capital in the production function. Moreover, $k_t^c(j)$ is the part of hard capital k_t^c used by firm j in oder to produce final-good output. Similarly, $k_t^i(j)$ and $l_t(j)$ are the parts of ICT capital k_t^i and labor l_t used by firm j in oder to produce final-good output, respectively.

Since all the final-good producers are identical, the following conditions hold,

$$\int_0^1 k_t^c(j)dj = k_{1,t}^c, \tag{87}$$

$$\int_0^1 k_t^i(j)dj = k_{1,t}^i, \tag{88}$$

$$\int_0^1 l_t(j)dj = l_{1,t}, \tag{89}$$

[43] In this case, a news shock would be a current signal regarding innovations to the future values of η_t.

where $k^c_{1,t}$ is the aggregate part of hard capital k^c_t devoted to the final-sector, i.e. sector 1. Moreover, $k^i_{1,t}$ and $l_{1,t}$ are the aggregate parts of ICT capital k^i_t and labor l_t devoted to the final-good sector, respectively.

Aggregate final-good output is defined as

$$y^c_t = A^c_t \ (k^c_{1,t})^a \ (k^i_{1,t})^b \ (l_{1,t})^{1-a-b}, \quad 0 < a, b < 1 \tag{90}$$

and it can be used for two purposes: consumption, c_t and investment in hard capital, i^c_t,

$$y^c_t = c_t + i^c_t \tag{91}$$

Notice that elements in the equation above are normalized such that both output and investment are measured in units of consumption. Moreover, hard structures can be produced from final output on a one-to-one basis. The stock of hard capital evolves according to

$$k^c_{t+1} = (1 - \delta^c)k^c_t + i^c_t \tag{92}$$

where $0 < \delta^c < 1$ represents the period depreciation of hard capital.

3.3.3 ICT-Good Sector

Production of ICT capital goods is symmetric to final-output production. It is populated by a set $[0, 1]$ of firms. Similarly, ICT-good firm $q \in [0, 1]$ produces $y^i_t(q)$ with labor l_t, hard capital, k^c_t and ICT capital, k^i_t. Firm q produces in accordance with

$$y^i_t(q) = A^i_t \ (k^c_t(q))^a \ (k^i_t(q))^b \ (l_t(q))^{1-a-b}, \quad 0 < a, b < 1 \tag{93}$$

where $A^i_t = \eta_t\ \theta^i_t\ (k^i_t)^\gamma$. Analogously to the first sector, production function is augmented by neutral productivity η_t, sector-specific productivity θ^i_t, and the spillover effect $(k^i_t)^\gamma$. In particular, θ^i_t is the sector-specific technological component which moves along the following deterministic trend

$$\theta^i_t = (\Gamma^i)^t e^{\zeta_t}, \tag{94}$$

$$\zeta_t = \rho_\iota \zeta_{t-1} + \iota_t \tag{95}$$

where $\rho_\iota \in (0,1)$ and $\iota_t \sim N(0, \sigma^2_\iota)$ represents an ICT technology shock. In line with the first sector, $k^c_t(q)$ is the part of hard capital k^c_t used by firm q in oder to produce ICT goods. Similarly, $k^i_t(q)$ and $l_t(q)$ are the parts of ICT capital k^i_t and labor l_t used by firm q in oder to produce ICT goods, respectively.

Since we assume homogeneity across ICT producers, following conditions hold

$$\int_0^1 k^c_t(q)dq = k^c_{2,t}, \tag{96}$$

$$\int_0^1 k^i_t(q)dq = k^i_{2,t}, \tag{97}$$

$$\int_0^1 l_t(q)dq = l_{2,t}, \tag{98}$$

where $k^c_{2,t}$ is the aggregate part of hard capital k^c_t devoted to the second sector. Similarly, $k^i_{2,t}$ and $l_{2,t}$ are the aggregate parts of ICT capital k^i_t and labor l_t devoted to the ICT-good sector, respectively.

Aggregate ICT-good output is defined as

$$y_t^i = A_t^i\ (k_{2,t}^c)^a\ (k_{2,t}^i)^b\ (l_{2,t})^{1-a-b}, \quad 0 < a, b < 1 \tag{99}$$

where $A_t^i = \eta_t\ \theta_t^i\ (k_t^i)^\gamma$ and y_t^i can be only investment in ICT capital, i_t^i,

$$y_t^i = i_t^i \tag{100}$$

Also here the elements in the equation above are normalized such that both output and investment are measured in the same units. Moreover, ICT output is converted to ICT goods on a one-to-one basis. The law of motion of ICT capital is described as follows,

$$k_{t+1}^i = (1 - \delta^i)k_t^i + i_t^i \tag{101}$$

where $0 < \delta^i < 1$ represents the period depreciation of ICT capital.

3.3.4 Market Clearing across Sectors

Since the economy is closed, input demands must equal supplies,

$$k_t^c = k_{1,t}^c + k_{2,t}^c, \tag{102}$$

$$k_t^i = k_{1,t}^i + k_{2,t}^i, \tag{103}$$

$$l_t = l_{1,t} + l_{2,t}. \tag{104}$$

Taking as given Equations 83-104 and assuming that factor markets are perfective competitive, the representative household maximizes intertemporal utility under the following constraint,

$$c_t + k^c_{t+1} + p_t k^i_{t+1} = \left(1 - \delta^c + r^c_t\right) k^c_t + \left(1 - \delta^i + \frac{r^i_t}{p_t}\right) p_t k^i_t + w_t l_t \tag{105}$$

where p_t represents the price of ICT capital expressed in consumption units. Since the price of consumption is normalized to be 1, p_t also represents the relative price between final output and ICT output.

3.3.5 Competitive Equilibrium and Solution Method

Competitive equilibrium is represented by (i) Equations 82-104, (ii) the household's first order conditions with respect to consumption c_t, labor l_t, hard capital k^c_t, ICT capital k^i_t, (iii) final-good producers' first order conditions with respect to $k^c_t(j)$, $k^i_t(j)$, and $l_t(j)$, and (iv) ICT-good producers' first order conditions with respect to $k^c_t(q)$, $k^i_t(q)$, and $l_t(q)$. In order to simulate responses to (i) neutral technology shocks, ν_t and (ii) ICT shocks, ι_t we employ a standard first-order approximation of the log-transformation of the system around the detrended (stationarized) steady state.[44]

3.4 Model Assessment

We now turn to describing the parameter values and testing the empirical performance of the model. Following Christiano et al. (2005), we partition the model parameters into two groups. We calibrate parameters of the first group using steady-state relations or results from previous works. We estimate the second group by minimizing the distance between the model-implied impulse responses and the empirical results presented in Section 3.2.1. In particular, the second group focuses on parameters specifically related to the effect of ICT investment on future TFP: (i) the standard deviation σ_ι of the ICT shock ζ_t, (ii) the persistence ρ_ι of the ICT shock ζ_t and (iii) the size of the spillover effect of ICT capital on productivity γ.

[44] The stationarized system as well as code to replicate our results are available on request.

3.4.1 Parameterization

We calibrate the model at a quarterly frequency. We set the discount factor β to $0.97^{\frac{1}{4}}$ to match the average yield of 3% of the 3-month Treasury bill using data from 1989:Q1 to 2017:Q1. We set the ICT share b equal to 0.0310 following Oulton (2012) and hard capital share a equal to 0.3. Moreover, adapting Oulton (2012) accordingly to our additional assumptions, we derive steady state growth rates of final-sector technology Γ^c and ICT-sector technology Γ^i from solving the system,

$$\begin{cases} \frac{p_t}{p_{t-1}} = \frac{\Gamma^c}{\Gamma^i} \\ \\ \frac{c_t}{c_{t-1}} = (\Gamma^c)^{\frac{1-b-\gamma}{1-a-b-\gamma}} (\Gamma^i)^{\frac{b+\gamma}{1-a-b-\gamma}} \end{cases} \tag{106}$$

where p_t/p_{t-1} and c_t/c_{t-1} are derived from the simple mean of their empirical related values using quarterly data from 1989:Q1 to 2017:Q1. This implies $\Gamma^c = 1.0034$ and $\Gamma^i = 1.0160$.[45][46] Finally, to avoid an over-identification problem we set the growth rate of sector-neutral technology Γ^η equal to one.

In addition, we set the Frisch elasticity of labor supply χ to be equal to 1. Moreover, following Bureau fo Economic Analysis (BEA) depreciation estimates, we set the quarterly depreciation rates for hard capital δ^c to be equal to 0.0206 in order to match an annual depreciation rate of 0.08. Similarly, according to BEA depreciation estimates (i) communication equipments and (ii) computer and electronic products depreciate at an annual rate of 0.15 and 0.14 respectively. Thus, we set the quarterly depreciation rates for ICT capital δ^i to be equal to 0.0398 in order to match the highest of the two values.

[45] Notice that the value of γ is unknown at this step since it is going to be evaluated using impulse-response matching. This implies that the solution for Γ^c, Γ^i, and γ should be a fixed point where given γ^* we obtain $(\Gamma^i)^*$ and $(\Gamma^c)^*$ solving 106. Given $(\Gamma^i)^*$ and $(\Gamma^c)^*$ we obtain γ^* by minimizing the impulse-response matching. At this step, however, we are simply solving the system assuming γ equal to zero, but a more refined solution will be provided in future.

[46] Other different systems might be use in order to pin down the steady state growth rates of both sectors. At this first stage, we opt for System 106 because both relative prices and aggregate consumption are well observable variables in the data.

3.4.2 Impulse-Response Matching

Now we estimate the model's key parameters governing the effect of ICT investment on future TFP by imposing theoretical impulse responses of an ICT shock (ζ_t) to be as close as possible to the empirical ones. Using empirical impulse responses of TFP, ICT investment, consumption, and relative prices, we estimate three parameters: the variance of an ICT technology shock, σ_ι^2, the persistence of the same shock, ρ_ι, and the size of the spillover effect of ICT capital on TFP, γ. We estimate $\Omega = (\gamma,\ \sigma_\iota^2,\ \rho_\iota)$ by minimizing,

$$\min_{\Omega} [\hat{\Psi} - \Psi(\Omega)]' \Lambda [\hat{\Psi} - \Psi(\Omega)] \tag{107}$$

where $\Psi(\Omega)$ denotes the mapping from Ω to the theoretical impulse responses, and $\hat{\Psi}$ denotes the empirical impulse responses of an ICT shock to TFP, ICTI, consumption, and relative prices. Λ is a diagonal weighting matrix, which is the inverse of the variance matrix of the empirical impulse responses. These elements serve as weights to impose more precision on some estimates of the optimization problem. Finally, consistent with our choice of horizons presented above, we use a horizon of 40 periods for each response.

Table 5: Estimated parameters

Symbol	Economic Interpretation	Estimated Value
σ_ι^2	Variance of ICT technological shock	0.01^2
ρ_ι	Persistence of ICT technological shock	0.9
γ	Size of spillover of ICT capital on TFP	0.5881

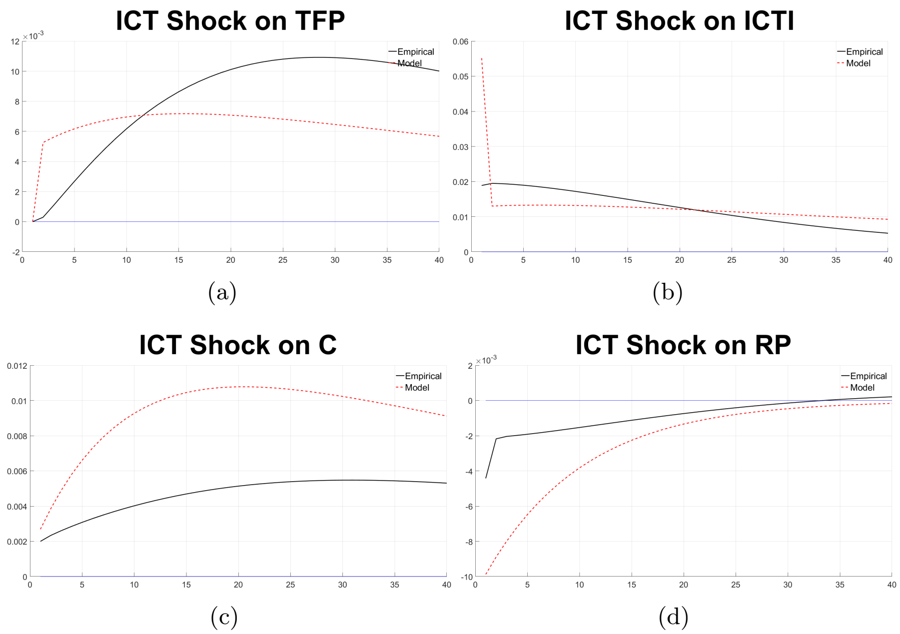
Parameters are estimated by imposing that the model's impulse responses of an ICT shock are as close as possible to the empirical ones. Theoretical estimates are impulse responses to an ICT-specific technological shock, ζ_t which are functions of key parameters: σ_ι^2, ρ_ι, and γ. Empirical impulse responses of an ICT shock to TFP, ICTI, consumption, and relative prices are derived from Specification 76 using the empirical strategy presented in Section 3.2.1.2. We use a horizon of 40 periods for each response.

Table 5 shows results of Problem 107. The estimate of the spillover effect of ICT capital on TFP, γ, is 0.58, while estimates of variance and persistence of an ICT-specific technological shock σ_ι^2 and ρ_ι are 0.01 and 0.9 respectively.[47] These results convey a number of interesting messages regarding the fit of our two-sector model. Most importantly, the key question of our model was whether the presence of ICT shocks alone can account for the empirical responses we observe in data, or whether ICT diffusion plays a crucial role in this dynamics. This amounts to asking whether $\gamma = 0$ is supported by the data. Clearly, our estimate of 0.58 rejects this hypothesis strongly. Thus, the data favor an interpretation of our model in which there are considerable spillovers to ICT capital. Accordingly, ICT shocks do not need to be extremely large in magnitude to have sizable effects on TFP, GDP and consumption. This is indicated by the value we obtain for σ_ι^2 of 0.01. Confronted with data, the model is clearly telling a story in which there are small (relative to the overall economy) but persistent ($\rho_\iota = 0.9$) ICT shocks that lead considerable medium-run effect on TFP and consumption because of large spillovers from ICT capital accumulation.

Using the estimated parameter values in Table 5, Figure 20 displays empirical impulse responses along with theoretical ones. Our simple model is able to mimic empirical movements in TFP, ICT investment, consumption, and relative prices. In both the model and the data, a sector-specific technological shock ζ_t has a delayed positive effect on TFP which tends to last over the medium and long run. Moreover, in both cases, ICT investment has a large initial peak and then tends to remain persistently positive over the 40-quarter horizon. In the model and the data, consumption responds positively on impact, reaches its peak after 4 years and persistently remains positive. Finally, model-generated and empirical impulse responses of relative prices have a large and negative impact response which dies out in about 6 years.

[47] Estimated parameter values are strongly robust to various alternative specifications. Using different empirical results, changing the initial calibration, the number of parameters matched or the methodology do not qualitative change the estimated value of γ close to 0.6.

Figure 20: Empirical VAR versus model responses to an ICT shock.



The black solid lines show the empirical responses to an identified ICT shock in Specification 76. The dotted red lines show the model's responses to a shock to the ICT-specific technological component ζ_t. The horizontal axes refer to forecast horizon and the units of the vertical axes are percentage deviations.

3.4.3 Interpretation of the Results

Although both the model and impulse-response matching exercise are simple and preliminary, we believe that they are already able to rationalize the empirical results obtained in Section 3.2 and thus convey the main message of this paper. Formally, the question we tried to answer is why a contemporaneous technological shock specific to the ICT sector is able to trigger increases in future total factor productivity even when it is too small to affect TFP today. In other words, why is an ICT-investment shock able to increase the productivity of ICT-good users with lags?

We interpret ICT capital as being the general-purpose technology of the U.S. economy over the last three decades. Specifically, when an ICT technology shock hits the economy, both sectors invest more on ICT capital goods since they are now cheaper to purchase and easier to produce. Now, in order to link current higher investment in ICT goods to future TFP, we assume that the general-purpose nature of ICT goods manifests itself in the form of a spillover on both production functions which is not internalized by the agents.[48] Since the spillover comes from the overall stock of ICT capital, we interpret it as the diffusion of ICT capital.

This assumption is legitimate from a mathematical standpoint for two main reasons. First of all, the spillover enters the estimated Solow residual since, being an externality, it is a non-renumerated input. Additionally, since the spillover arises from a state variable, it leads to a dynamics where the benefits of an ICT technology shock show up in the productivity of ICT-capital users with some lags because the accumulation of ICT goods is a slow process. Thus, the spillover effect only comes into play only gradually once enough ICT capital is being accumulated. The slow-moving nature of the state variable capital also means that this effect tends to be persistent, which also helps us match the empirical responses we see in the data.[49]

More importantly, there is clear economic intuition underlying our assumption of the spillover. Since the purpose of ICT goods is to enhance communication, the quality and speed of information acquisition and sharing is also determined by the overall diffusion of ICT goods. As a simple example different from the one in the Introduction, assume that in 1995 a scholar decides to purchase a computer with an Internet subscription in order to speed up the acquisition of information. She is very satisfied with this investment and does not regret the amount paid. However, at the same time, a former colleague who is now working in a different university decides to buy a computer with an Internet subscription as well. Once she learns of this development, the initial researcher decides to henceforth use her computer to exchange research material with her former colleague. They eventually decide to start a new project together, since being in two different universities is no longer a constraint for their joint research. None of the two

[48] This externality is governed by the parameter γ which has been estimated via impulse-response matching to be robustly larger than zero.

[49] We could slow down the dynamics of ICT capital further, for example by assuming investment adjustment costs. Obviously, this would improve the fit of the model since it would allow the externality effect to play out even slower. Our key point here, however, is to show how the spillover is necessary to match the main features of the empirical responses. Without it, the model has no chance to capture the large lagged effect of ICT shocks on TFP.

scholars paid for this additional benefit when they purchased a computer, since they bought it for their personal use and did not expect the other to simultaneously acquire one also. What is actually happening here is that the research production of both researchers has been boosted by the positive externality that also the other colleague is endowed with the same technology. This externality is what we understand under the general-purpose nature of ICT goods and what our model captures with the assumption of the spillover.

The impulse response matching exercise of Section 3.4.2 shows that a two-sector model featuring ICT capital and hard capital can only successfully match the empirical responses to an ICT shock if indeed the spillover effect is in action. Thus the main take-away of this paper is that the interpretation of ICT as a general-purpose technology of the US between 1989-2017 is a good explanation of medium-run TFP growth dynamics. This relies on understanding the general-purpose nature of ICT goods as the extent to which their diffusion in the economy matters.

3.5 Conclusion

In this paper, we look at the effect of ICT shocks on medium-run TFP from both an empirical and a theoretical perspective. In the first part of the paper, we investigate the effect of technological shocks specific to the ICT sector on future TFP. Using Structural VAR techniques on quarterly aggregate U.S. data from 1989, we impose two identifying restrictions to correctly identify ICT shocks: (i) they must be orthogonal to current productivity and (ii) maximize the impact response of ICT investment. In response to an ICT shock, ICT investment increases on impact and stays significant for more than 10 years; relative prices persistently decline as an additional support for the reliability of our identification strategy; and TFP rises after three quarters and remains significant for more than 10-years. In addition, variance decomposition analysis suggests that the quantitative impact of ICT shocks on TFP is economically significant since they explain about one third of the overall TFP fluctuations over a 10-year horizon.

In order to check whether our identification assumptions may confound our shock with a news shock which contemporaneously boosts ICT investment, we provide a series of different identification strategies and show that they deliver analogous results. Our main robustness check is a sequential identification of news shocks first and then ICT shocks. Encouragingly, neither this, nor any other robustness check affects our initial estimates. We view this as a confirmation of the causality relation which goes from current ICT investment to future TFP.

In the second part of the paper, we dig deeper into the nature of the link between current ICT investment and future TFP. We rationalize our empirical results by analyzing a two-sector DSGE model à la Greenwood et al. (1997) augmented by two main assumptions: (i) the sector with the faster technological improvement produces specifically ICT capital, and (ii) both production functions are augmented with a spillover related to the overall level of ICT capital which is not internalized by the agents. In particular, this last assumption points out our interpretation of the general-purpose nature of ICT capital: the effectiveness of ICT capital is enhanced by its diffusion, and the cost of ICT goods does not reflect this indirect benefit.

We use both empirical and theoretical results to estimate main parameters of the model via impulse-response matching to an ICT technology shock. Results show that the parameter which governs the size of the spillover effect is positive and large, suggesting that our interpretation of the general-purpose nature of ICT capital is confirmed by the data. Thus, our investigation suggests not only that ICT is an important driver of medium-run TFP, but specifically that what matters for the evolution of TFP is whether ICT can act as a general-purpose technology. In particular, large externalities from using ICT goods can explain why a sector comprising a tiny fraction of both output and investment can play an important role for future productivity developments.

A Monetary Policy and Anchored Expectations - An Endogenous Gain Learning Model

A.1 Robustness of rolling regression results

I investigate the following alternative specifications. As a first check, I replace the forecast error with realized CPI inflation. Thus Equation (1) becomes

$$\Delta\bar{\pi}_t^i = \beta_0 + \beta_1^w \pi_t + \epsilon_t^i. \tag{A.1}$$

I then consider the original specification of Equation (1), adding realized CPI inflation to control for inflation levels.[50] This results in the following regression:

$$\Delta\bar{\pi}_t^i = \beta_0^w + \beta_1^w f_{t|t-1}^i + \beta_2^w \pi_t + \epsilon_t^i. \tag{A.2}$$

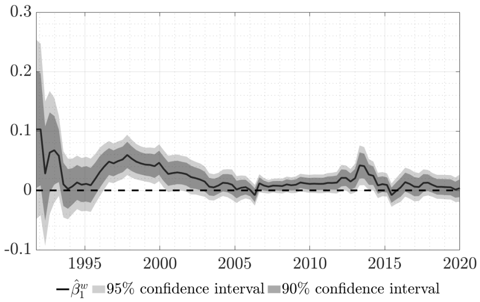
In the last alternative specification, I replace both the ten and the one-year ahead CPI inflation expectation with PCE core. The regression equation thus looks identical to Equation (1) but cleans out the effects of movements in oil prices:

$$\Delta\bar{\pi}_t^i = \beta_0^w + \beta_1^w f_{t|t-1}^i + \epsilon_t^i. \tag{A.3}$$

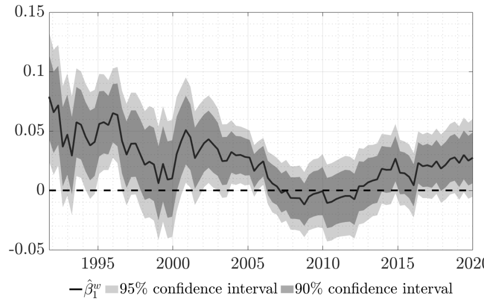
The three alternative time series of the estimate $\hat{\beta}_1^w$ are depicted in Fig. 21. Qualitatively, the same conclusion obtains as in the specification in the main text: the coefficients are time-varying in size and in statistical significance. In particular, large sensitivity in the 1990s gives way to an insignificant coefficient in the early 2000s, to

[50] In a separate specification, I have similarly added uncertainty as a control, but it is so similar that I suppress it here.

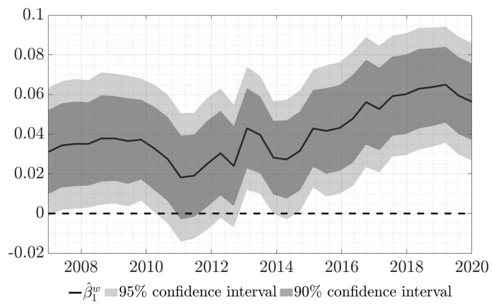
Figure 21: Time series of responsiveness of long-run inflation expectations to inflation



(a) Specification (A.1)



(b) Specification (A.2)



(c) Specification (A.3)

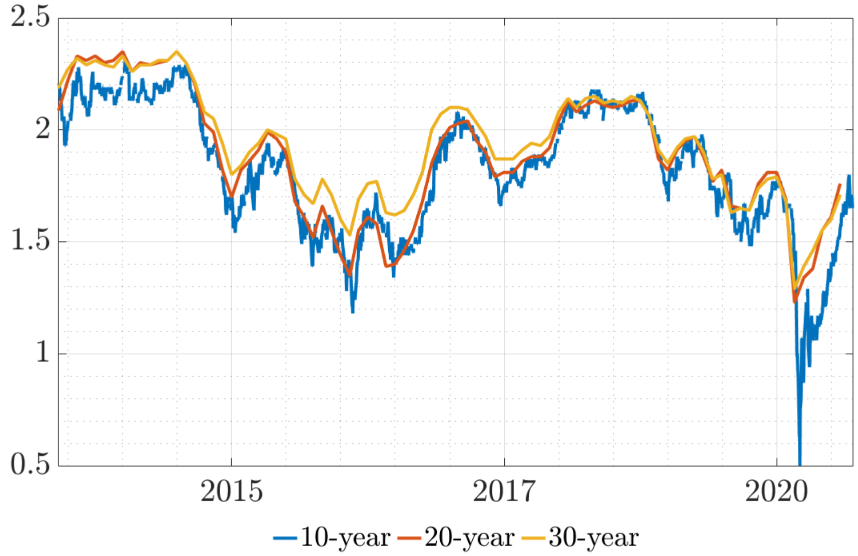
drift back up shortly after 2010. The data is much shorter for the PCE core specification, but also in this case one observes the upward drift in the estimated coefficient starting around 2010.

A.2 Alternative measures of long-run inflation expectations

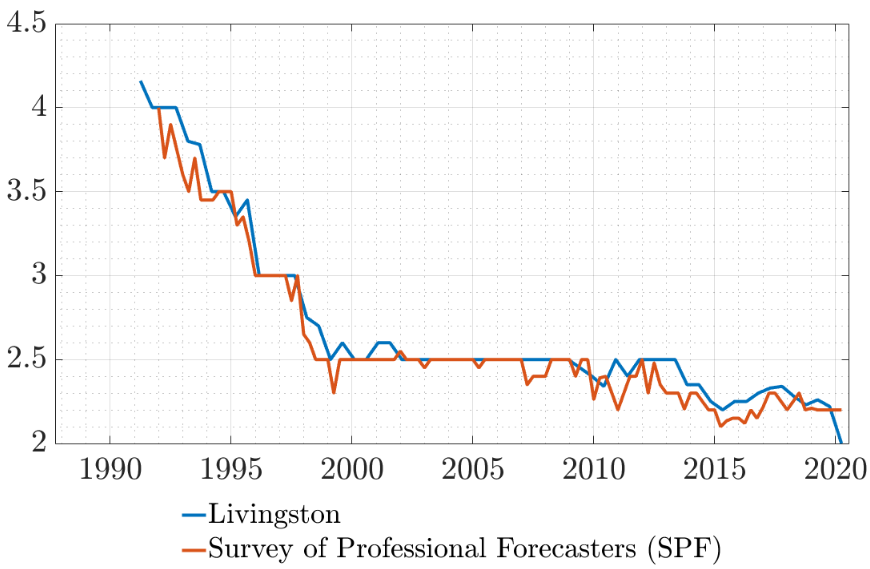
Measuring long-run expectations is challenging. In particular, two dimensions of expectations are hard to line up between models and data: the identity of the economic agent forming expectations, and the forecast horizon. To start with the former, one would ideally wish to measure the expectations of the private sector, firms and households, since their expectations correspond to the ones in economic models. As for the forecast horizon, no single horizon is an exact counterpart to the concept of expectations of average inflation. Arguably, the longer the forecast horizon, the closer the measured expectation is to expectations of the average.

These considerations can lead the researcher to utilize breakeven inflation, that

Figure 22: Alternative long-run inflation expectations measures



(a) Market-based inflation expectations from TIPS, various horizons



(b) 10-year ahead inflation expectations of firms (Livingston) and professional forecasters (SPF)

All inflation rates are annualized percentages. The series are at the following frequency: daily (10-year TIPS), monthly (20- and 30-year TIPS), quarterly (SPF), twice per year (Livingston).

is, the difference between Treasury yields and Treasury Inflation Protected Securities (TIPS) (top panel of Fig. 22). There are two main advantages of constructing inflation expectation measures from TIPS. First, since this measure is based on trades that happened in the Treasury market, it pertains to agents who are active in a market where inflation expectations matter. Such expectations are therefore a good proxy for the private sector's expectations in economic models. Second, the fact that breakeven inflation is not elicited from surveys allows the econometrician to bypass many of the challenges that survey data involve. It has been widely documented that survey participants may have poor understanding of the economic concept elicited, may misunderstand the survey questions or be unduly influenced by the wording. Most troubling is the fact that there is no way the econometrician could control for the noise thus introduced in survey data.

Breakeven inflation measures are not a panacea, however. The illiquidity of TIPS markets introduces a distortion into the measured expectations in the form of a time-varying liquidity premium (Appendix A.3 corrects for this by filtering out an approximate liquidity premium series from the 10-year expectation series.) One could also contend that many firms and the majority of households are not active on the TIPS market, and therefore that breakeven inflation is not representative of the economy-wide expectation. For this reason, I turn to alternative long-run inflation measures.

The forecast horizon is a binding constraint: consumer expectations data have a maximal forecast horizon of three (New York Fed Survey of Consumer Expectations (SCE)) or five years (University of Michigan Survey of Consumers). To have at least a forecast horizon of ten years, I thus resort to the Livingston Survey of the Philadelphia Fed and the Survey of Professional Forecasters (SPF). Another advantage of the SPF is that comprehensive individual-level data is available, a feature which I exploit for the regressions in the Introduction and in Appendix A.1.

Fig. 22 plots the breakeven inflation series for a horizon of 10, 20 and 30 years (top panel) alongside Livingston and SPF (bottom panel). The breakeven inflation measures are clearly and consistently below the Livingston and SPF measures. Part of this may be driven by liquidity premia, which I control for in Appendix A.3. At the same time, both the Livingston and the SPF expectations may potentially suffer from representativeness issues. The fact that the two expectations series are so closely aligned indicates that the Livingston survey may only be reaching a subset of firms.[51] This subset seems to be highly correlated with the set of professional forecasters filling out the SPF. This is a concern because it suggests that the elicited expectations may not be representative of the economy-wide firm expectations. Moreover, if professional forecasters respond to the survey by running a (by assumption stationary) econometric model, then the survey responses may not even accurately reflect their held beliefs.

But all these caveats notwithstanding, there is a single element that is consistent across all of these measures of long-run inflation expectations: following 2010, they all shift downward. In other words, they all show responsiveness to the fact that the Federal Reserve has undershot the 2% inflation target. Given that the Livingston and SPF measures are initially above the 2% target, one may argue that expectations are only adjusting following the 2012 announcement of an official target on the part of the Fed.

However, the 2% target was already understood prior to the 2012 announcement. Moreover, the Fed's target is specified in terms of PCE inflation, while the Livingston and SPF forecasts pertain to CPI inflation. The fact that the latter tends to be higher than the former also helps to explain why expectations of CPI inflation are above the Fed's PCE inflation target.

A.3 Filtering out liquidity risk from TIPS

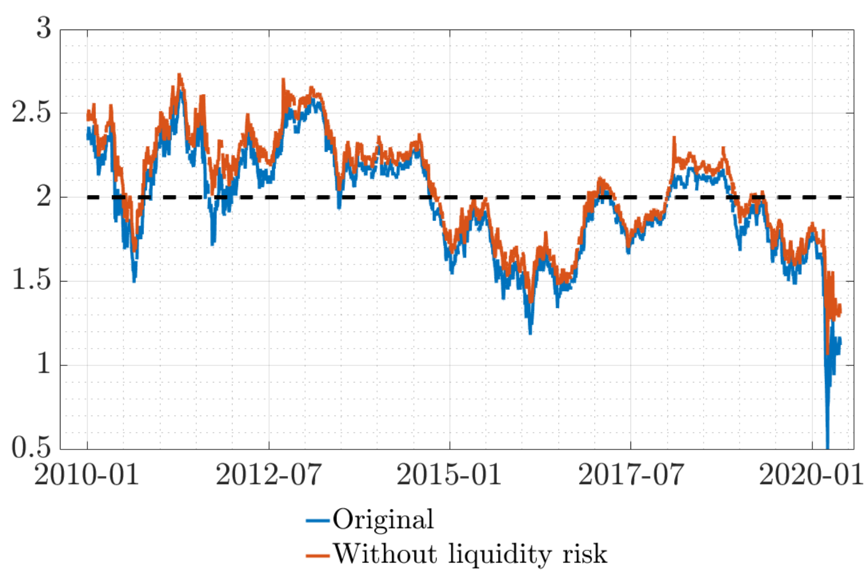
Some caution is needed when inferring inflation expectations from TIPS. The underlying idea is that the difference between real (indexed to inflation) and nominal yields should be a good metric for the market's expectations of inflation on average for the duration of the particular maturity. But since TIPS markets face liquidity issues, especially for

[51] The category codes of individual survey participates in the Livingston survey seems to underscore this. According to the documentation, the surveyed businesses fall into the following categories: academic institution, commercial banking, consulting, Federal Reserve, government, industry trade group, insurance company, investment banking, labor, non-financial business and other.

seasoned securities, the TIPS yield also incorporates a liquidity premium. The positive bias in the TIPS yield thus leads to a negative bias in expected inflation.

To gauge the presence of liquidity risk in TIPS, I rely on Andreasen et al. (2018)'s estimation of the liquidity premium. Since their series only covers the period between July 11, 1997 - Dec 27, 2013, I make use of the fact that they demonstrate a high correlation between liquidity risk and uncertainty. In particular, a regression of their average TIPS liquidity premium measure on the VIX index and controls yields an estimated coefficient of 0.85 (significant at the 1 percent level), with a constant of -5.21. Thus I can use the VIX to back out a fitted Andreasen et al. (2018) estimate of the TIPS liquidity premium after 2013. Doing so, I subtract this fitted liquidity premium series from the TIPS yields, allowing me to construct an estimate of breakeven inflation corrected for the liquidity risk bias. Fig. 23 presents the original breakeven inflation series, along with the bias-corrected estimate.

Figure 23: Market-based inflation expectations, 10 year, average, %



Breakeven inflation, constructed as the difference between the yields of 10-year Treasuries and 10-year TIPS (blue line), difference between 10-year Treasury and 10-year TIPS, the latter cleaned from liquidity risk (red line).

In line with Andreasen et al. (2018)'s findings, I obtain a negative bias in the inflation expectations series throughout the sample. This bias averages -0.0966 pp, which is sizable, but not as significant as one might have suspected. Also in analogy with Andreasen et al. (2018)'s results, I find that liquidity issues in the TIPS market pose a bigger problem in recessions, when the market worries more about future TIPS becoming illiquid. In particular once the Covid-19-shock raises worries at the end of the sample, the estimated liquidity premium hits its highest value of 0.6508 pp. Even this large upward correction in breakeven inflation does not change the overall picture, however. The conclusion that long-run inflation expectations trend downward in the second half of the decade remains solid.

A.4 Compact model notation

The A matrices are given by

$$A_a = \begin{pmatrix} g_{\pi a} \\ g_{xa} \\ \psi_\pi g_{\pi a} + \psi_x g_{xa} \end{pmatrix}, \quad A_b = \begin{pmatrix} g_{\pi b} \\ g_{xb} \\ \psi_\pi g_{\pi b} + \psi_x g_{xb} \end{pmatrix}, \quad A_s = \begin{pmatrix} g_{\pi s} \\ g_{xs} \\ \psi_\pi g_{\pi s} + \psi_x g_{xs} + \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \end{pmatrix}, \tag{A.4}$$

$$g_{\pi a} = (1 - \frac{\kappa\sigma\psi_\pi}{w}) \left[(1-\alpha)\beta, \kappa\alpha\beta, 0\right], \tag{A.5}$$

$$g_{xa} = \frac{-\sigma\psi_\pi}{w} \left[(1-\alpha)\beta, \kappa\alpha\beta, 0\right], \tag{A.6}$$

$$g_{\pi b} = \frac{\kappa}{w} \left[\sigma(1-\beta\psi_\pi), (1-\beta-\beta\sigma\psi_x, 0\right], \tag{A.7}$$

$$g_{xb} = \frac{1}{w} \left[\sigma(1-\beta\psi_\pi), (1-\beta-\beta\sigma\psi_x, 0\right], \tag{A.8}$$

$$g_{\pi s} = (1 - \frac{\kappa\sigma\psi_\pi}{w}) \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} (I_3 - \alpha\beta h)^{-1} - \frac{\kappa\sigma}{w} \begin{bmatrix} -1 & 1 & 0 \end{bmatrix} (I_3 - \beta h)^{-1}, \tag{A.9}$$

$$g_{xs} = \frac{-\sigma\psi_\pi}{w} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} (I_3 - \alpha\beta h)^{-1} - \frac{\sigma}{w} \begin{bmatrix} -1 & 1 & 0 \end{bmatrix} (I_3 - \beta h)^{-1}, \tag{A.10}$$

$$w = 1 + \sigma\psi_x + \kappa\sigma\psi_\pi. \tag{A.11}$$

The matrices of the state transition equation (15) are

$$h \equiv \begin{pmatrix} \rho_r & 0 & 0 \\ 0 & \rho_i & 0 \\ 0 & 0 & \rho_u \end{pmatrix}, \quad \epsilon_t \equiv \begin{pmatrix} \varepsilon_t^r \\ \varepsilon_t^i \\ \varepsilon_t^u \end{pmatrix}, \quad \text{and} \quad \eta \equiv \begin{pmatrix} \sigma_r & 0 & 0 \\ 0 & \sigma_i & 0 \\ 0 & 0 & \sigma_u \end{pmatrix}. \tag{A.12}$$

Note that this is the formulation for the case where a Taylor rule is in effect and is known by the private sector. It is straightforward to remove any of these two assumptions.

A.5 The observation matrix for learning

Instead of the matrix g in the rational expectations observation equation

$$z_t = gs_t, \tag{A.13}$$

agents in the anchoring model use the estimated matrix g^l

$$g^l_{t-1} = \begin{bmatrix} F_{t-1} & G_{t-1} \end{bmatrix}, \tag{A.14}$$

with

$$F_{t-1} = \left(A_a \frac{1}{1-\alpha\beta} + A_b \frac{1}{1-\beta}\right) a_{t-1}, \tag{A.15}$$

$$G_{t-1} = A_a b_{t-1} \left(I_3 - \alpha\beta h\right)^{-1} + A_b b_{t-1} \left(I_3 - \beta h\right)^{-1} + A_s. \tag{A.16}$$

A.6 Alternative specifications for the anchoring function

The general law of motion for the gain in the main text is given by Equation (23), reproduced here for convenience:

$$k_t = \mathbf{g}(k_{t-1}, f_{t|t-1}). \tag{A.17}$$

The baseline specification of the anchoring function $\mathbf{g}$ (Equation (31) in the main text) is

$$\mathbf{g} = \sum_i \gamma_i b_i(f_{t|t-1}), \tag{A.18}$$

where $b_i(f_{t|t-1})$ is the piecewise linear basis at node i and $\hat{\gamma}$ are the coefficients estimated in Section 1.4.

To my knowledge, there are only two other papers that consider an endogenous gain as a model for anchored expectations. The first one, more related to this paper, is Carvalho et al. (2019). In their model, the anchoring function is a discrete choice function as follows. Let θ_t be a criterion to be defined. Then, for a threshold value $\tilde{\theta}$, the gain evolves according to

$$k_t = \begin{cases} (k_{t-1}+1)^{-1} & \text{if} \quad \theta_t < \tilde{\theta}, \\ \bar{g} & \text{otherwise.} \end{cases} \tag{A.19}$$

In other words, agents choose a decreasing gain when the criterion θ_t is lower than the threshold $\tilde{\theta}$; otherwise they choose a constant gain. The criterion employed by Carvalho et al. (2019) is computed as the absolute difference between subjective and model-consistent expectations, scaled by the variance of shocks:

$$\theta_t = \max|\Sigma^{-1}(\phi_{t-1} - \begin{bmatrix} F_{t-1} & G_{t-1} \end{bmatrix})|, \tag{A.20}$$

where Σ is the VC matrix of shocks, ϕ_{t-1} is the estimated matrix and $[F, G]$ is the ALM (see Appendix A.5).

As a robustness check, Carvalho et al. (2019) also compute an alternative criterion.[52]

Let ω_t denote agents' time t estimate of the forecast error variance and θ_t be a statistic

[52] Note that for both criteria, I present the matrix generalizations of the scalar versions considered by Carvalho et al. (2019).

evaluated by agents in every period as

$$\omega_t = \omega_{t-1} + \tilde{\kappa} k_{t-1}(f_{t|t-1} f'_{t|t-1} - \omega_{t-1}), \tag{A.21}$$

$$\theta_t = \theta_{t-1} + \tilde{\kappa} k_{t-1}(f'_{t|t-1} \omega_t^{-1} f_{t|t-1} - \theta_{t-1}), \tag{A.22}$$

where $\tilde{\kappa}$ is a parameter that allows agents to scale the gain compared to the previous estimation and $f_{t|t-1}$ is the most recent forecast error, realized at time t. Indeed, this is a multivariate time series version of the squared CUSUM test.[53]

It is interesting to compare the two discrete anchoring functions with my smooth specification. On the one hand, Carvalho et al. (2019)'s preferred specification requires the private sector to evaluate model-consistent expectations, which runs counter to the maintained informational assumptions. It is more consistent with the present model, then, to assume that firms and households employ a statistical test of structural change, as is the case with the CUSUM-based and smooth functions. These are therefore more appealing on conceptual grounds.

On the other hand, simulation of the model using the different anchoring specifications reveals that Carvalho et al. (2019)'s preferred functional form leads to the opposite comparative statics of anchoring with respect to monetary policy aggressiveness as the smooth or the CUSUM-based specifications. In particular, policy that is more aggressive on inflation (a higher ψ_π in the Taylor rule) leads to more anchoring in a model with the smooth or the CUSUM-inspired criterion. If one uses Carvalho et al. (2019)'s criterion, the same comparative static involves *less* anchoring. This comes from the fact that Carvalho et al. (2019)'s criterion endows the private sector with capabilities to disentangle volatility due to the learning mechanism from that owing to exogenous disturbances. Thus agents in the Carvalho et al. (2019) model are able to make more advanced inferences about the performance of their forecasting rule and understand that a higher ψ_π causes more learning-induced volatility. This is however not possible

[53]See Brown et al. (1975) and Lütkepohl (2013) for details.

for agents who process data in real time without knowledge of the model. Therefore my smooth and the discrete CUSUM-inspired specifications are preferable both on conceptual and quantitative grounds.

There are two main reasons I opt for a smooth anchoring function instead of the discrete CUSUM criterion. First, the analytical analysis of this paper requires derivatives of the anchoring function to exist. In such a case, the smooth specification is necessary. Second, a smooth anchoring function can capture varying degrees of unanchoring, which is a feature of the data I document in the Introduction.

The second paper with an anchoring function is Gobbi et al. (2019). In their model, which is a three-equation New Keynesian model, firms and households entertain the possibility that the model may switch from a "normal" regime to a liquidity trap regime that the authors name the "new normal." Expectations are a probability-weighted average of the regime-specific expectations. The concept of unanchoring in the model is when p, the probability of the liquidity trap regime, rises significantly. The function governing the evolution of p, which Gobbi et al. (2019) refer to as the deanchoring function (DA), is the analogy of my anchoring function. The authors use the following logistic specification for the DA function:

$$p = h(y_{t-1}) = A + \frac{BCe^{-Dy_{t-1}}}{(Ce^{-Dy_{t-1}} + 1)^2}, \tag{A.23}$$

where y_{t-1} denotes the output gap and A, B, C and D are parameters.

A.7 Estimation procedure

The estimation of Section 1.4 is a simulated method of moments (SMM) exercise. As elaborated in the main text, I target the autocovariances of CPI inflation, the output gap, the federal funds rate and the 12-months ahead inflation forecasts coming from the Survey of Professional Forecasters. For the autocovariances, I consider lags $0, \dots, 4$. The target moment vector, Ω, is the vectorized autocovariance matrices for the lags considered, 80×1.

Let Θ denote the set of parameters in the New Keynesian model that I calibrate. Then, for each proposed coefficient vector γ, the estimation procedure consists of simulating the model conditional on γ, Θ and N different sequences of disturbances, then computing model-implied moments for each simulation, and lastly choosing γ such that the squared distance between the data- and model-implied mean moments is minimized. Thus

$$\hat{\gamma} = \left(\Omega^{data} - \frac{1}{N}\sum_{n=1}^{N}\Omega^{model}(\gamma, \Theta, \{e_t^n\}_{t=1}^{T})\right)' W^{-1}\left(\Omega^{data} - \frac{1}{N}\sum_{n=1}^{N}\Omega^{model}(\gamma, \Theta, \{e_t^n\}_{t=1}^{T})\right), \tag{A.24}$$

where the observed data is of length $T = 151$ quarters. Here $\{e_t^n\}_{t=1}^{T}$ is a sequence of disturbances of the same length as the data; note that I use a cross-section of N such sequences and take average moments across the cross-section to wash out the effects of particular disturbances. Experimentation with the number N led me to choose $N = 1000$, as estimates no longer change upon selecting larger N.

Before computing moments, I filter both the observed and model-generated data using the Baxter and King (1999) filter, with thresholds at 6 and 32 quarters and truncation at 12 lags, the recommended values of the authors. I then compute the moments by fitting a reduced-form VAR to the filtered series and using the estimated coefficients to back out autocovariances. Because there are four observables to three structural shocks and occasionally low volatility in the expectation series, I estimate the VAR coefficients by ridge regression with a tuning parameter of 0.001. This is to ensure that the VAR coefficients are estimated with a lower standard error, so that estimated variances of the moments are more accurate. As the weighting matrix of the quadratic form in the moments, I use the inverse of the estimated variances of the target moments, W^{-1}, computed from 10000 bootstrapped samples.

To improve identification, I also impose restrictions on the estimates. First, I require that the γ-coefficients be convex, that is, that larger forecast errors in absolute value be associated with higher gains. Second, since forecast errors close to zero render the size of the gain irrelevant (cf. the learning equation (21)), I impose that the coefficient associated with a zero forecast error should be zero. Both restrictions are implemented with weights penalizing the loss function, and the weights are selected by experimentation.

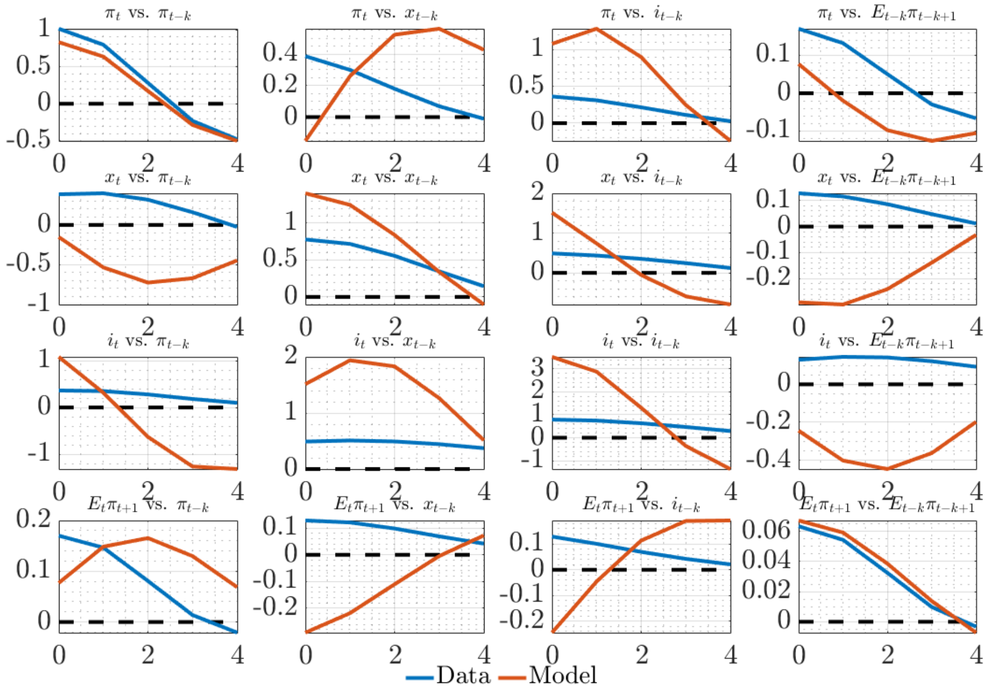
Both additional assumptions reflect properties that the anchoring function should reasonably exhibit. The convexity assumption captures the very notion that larger forecast errors in absolute value suggest bigger changes to the forecasting procedure are necessary. This is thus a very natural requirement. As for the zero gain for zero forecast error assumption, the idea here is to supply the estimation with information where it is lacking. Since the updating of learning coefficients corresponds to gain times forecast error, as Equation (21) recalls, a zero forecast error supplies no information for the value of the gain. To impose a zero value here also seems natural, given that since forecast errors switch sign at zero, one would expect the zero forecast error point to be an inflection point in the anchoring function. By the same token, Gobbi et al. (2019) also impose a related restriction when they require that their deanchoring function should yield a zero value at the zero input. Lastly, the objective function does not deteriorate upon imposing either assumption, suggesting that they are not at odds with the data.

Lastly, I supply 100 different initial points and select the estimate involving the lowest value for the loss (A.24). Fig. 24 presents the autocovariances of the observed variables for the estimated coefficients so obtained.

A.8 The policy problem in the simplified baseline model

Denote by $\mathbf{g}_{i,t} \in (0,1),\ i = \pi, \bar{\pi}$, the potentially time-varying derivatives of the anchoring function $\mathbf{g}$. In this simplified setting, $\bar{\pi}_t = e_1 a_t$, the estimated constant for the inflation process. As in the main text, e_i is a selector vector, selecting row i of the subsequent

Figure 24: Autocovariogram



Estimates for $N = 1000$

matrix. I also use the notation $b_i \equiv e_i b$. The planner chooses $\{\pi_t, x_t, f_{a,t}, f_{b,t}, \bar{\pi}_t, k_t\}_{t=t_0}^{\infty}$ to minimize

$$\mathcal{L} = \mathbb{E}_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} \Big\{ (\pi_t^2 + \lambda_x x_t^2) \tag{A.25}$$

$$+ \varphi_{1,t} \Big(\pi_t - \kappa x_t - (1-\alpha)\beta f_{a,t} - \kappa\alpha\beta b_2 (I_3 - \alpha\beta h)^{-1} s_t - e_3 (I_3 - \alpha\beta h)^{-1} s_t \Big) \tag{A.26}$$

$$+ \varphi_{2,t} \Big(x_t + \sigma i_t - \sigma f_{b,t} - (1-\beta) b_2 (I_3 - \beta h)^{-1} s_t + \sigma\beta b_3 (I_3 - \beta h)^{-1} s_t - \sigma e_1 (I_3 - \beta h)^{-1} s_t) \Big) \tag{A.27}$$

$$+ \varphi_{3,t} \Big(f_{a,t} - \frac{1}{1-\alpha\beta} \bar{\pi}_{t-1} - b_1 (I_3 - \alpha\beta h)^{-1} s_t \Big) \tag{A.28}$$

$$+ \varphi_{4,t} \Big(f_{b,t} - \frac{1}{1-\beta} \bar{\pi}_{t-1} - b_1 (I_3 - \beta h)^{-1} s_t \Big) \tag{A.29}$$

$$+ \varphi_{5,t} \Big(\bar{\pi}_t - \bar{\pi}_{t-1} - k_t (\pi_t - (\bar{\pi}_{t-1} + b_1 s_{t-1})) \Big) \tag{A.30}$$

$$+ \varphi_{6,t} \Big(k_t - \mathbf{g}(\pi_t - \bar{\pi}_{t-1} - b_1 s_{t-1}) \Big) \Big\}, \tag{A.31}$$

where ρ_i are Lagrange-multipliers on the constraints. After a little bit of simplifying, the first-order conditions boil down to the following three equations:

$$2\pi_t + 2\frac{\lambda_x}{\kappa} x_t - \varphi_{5,t} k_t - \varphi_{6,t} \mathbf{g}_{\pi,t} = 0, \tag{A.32}$$

$$-\frac{2(1-\alpha)\beta}{1-\alpha\beta} \frac{\lambda_x}{\kappa} \mathbb{E}_t\, x_{t+1} + \varphi_{5,t} - (1-k_t)\, \mathbb{E}_t\, \varphi_{5,t+1} + \mathbf{g}_{\bar{\pi},t}\, \mathbb{E}_t\, \varphi_{6,t+1} = 0, \tag{A.33}$$

$$\varphi_{6,t} = (\pi_t - \bar{\pi}_{t-1} - b_1 s_{t-1}) \varphi_{5,t}. \tag{A.34}$$

Note that Equation (A.32) is the analogue of Gaspar et al. (2010)'s Equation (22) (or, equivalently, of Molnár and Santoro (2014)'s (16)), except that there is an additional multiplier, φ_6. This multiplier reflects the fact that in addition to the constraint coming from the expectation process itself, with shadow value φ_5, learning involves the gain equation as a constraint as well. One can also clearly read off Proposition 2: when the learning process has converged such that neither expectations nor the gain process are constraints ($\varphi_5 = \varphi_6 = 0$), the discretionary inflation-output gap tradeoff familiar from

Clarida et al. (1999) obtains. Combining the above three equations and solving for $\varphi_{5,t}$, using the notation that $\prod_0^0 \equiv 1$, one obtains the target criterion (34).

The system of first-order conditions (38)-(39) and model equations for this simplified system also reveal how the endogenous gain introduces nonlinearity to the equation system. In particular, notice how in equations (38)-(39), the gain k_t shows up multiplicatively with the Lagrange multiplier, $\varphi_{2,t}$. In fact, the origin of the problem is the recursive least squares learning equation for the learning coefficient $\bar{\pi}_t$

$$\bar{\pi}_t = \bar{\pi}_{t-1} + k_t f_{t|t-1} \tag{A.35}$$

where the first interaction terms between the gain and other endogenous variables show up. This results in an equation system of nonlinear difference equations that does not admit an analytical solution.

Considering equation (A.35) is instructive to see how it is indeed the endogeneity of the gain that causes these troubles. Were we to specify a constant gain setup, k_t would merely equal the constant $\bar{g}$ and the anchoring function $\mathbf{g}(\cdot)$ would trivially reduce to zero as well. In such a case, all interaction terms would reduce to multiplication between endogenous variables and parameters; linearity would be restored and a solution for the optimal time paths of endogenous variables would be obtainable. Similarly, a decreasing gain specification would also be manageable since for all t, the gain would simply be given by t^{-1}, and the anchoring function would also be deterministic and exogenous.

A.9 Relaxing Assumption 2

Consider the general anchoring mechanism of Equation (23):

$$k_t = \mathbf{g}(k_t, f_{t|t-1}). \tag{A.36}$$

With this assumption, the FOCs of the Ramsey problem are

$$2\pi_t + 2\frac{\lambda_x}{\kappa}x_t - k_t\varphi_{5,t} - \mathbf{g}_{\pi,t}\varphi_{6,t} = 0, \tag{A.37}$$

$$c\,\mathbb{E}_t\, x_{t+1} + \varphi_{5,t} - (1-k_t)\,\mathbb{E}_t\,\varphi_{5,t+1} + \mathbf{g}_{\bar{\pi},t}\,\mathbb{E}_t\,\varphi_{6,t+1} = 0, \tag{A.38}$$

$$\varphi_{6,t} + \mathbb{E}_t\,\varphi_{6,t+1} = f_{t|t-1}\varphi_{5,t}, \tag{A.39}$$

where the red multiplier is the new element vis-à-vis the case where the anchoring function is specified in levels ($k_t = \mathbf{g}(f_{t|t-1})$), and I am using the shorthand notation

$$c = -\frac{2(1-\alpha)\beta}{1-\alpha\beta}\frac{\lambda_x}{\kappa}, \tag{A.40}$$

$$f_{t|t-1} = \pi_t - \bar{\pi}_{t-1} - b_1 s_{t-1}. \tag{A.41}$$

One can simplify this three-equation-system to:

$$\varphi_{6,t} = -cf_{t|t-1}\,\mathbb{E}_t\,x_{t+1} + \mathbb{E}_t\left(1 + \frac{f_{t|t-1}}{f_{t+1|t}}(1-k_{t+1}) - f_{t|t-1}\mathbf{g}_{\bar{\pi},t}\right)\varphi_{6,t+1} - \mathbb{E}_t\,\frac{f_{t|t-1}}{f_{t+1|t}}(1-k_{t+1})\varphi_{6,t+2}, \tag{A.42}$$

$$0 = 2\pi_t + 2\frac{\lambda_x}{\kappa}x_t - \left(\frac{k_t}{f_{t|t-1}} + \mathbf{g}_{\pi,t}\right)\varphi_{6,t} + \frac{k_t}{f_{t|t-1}}\,\mathbb{E}_t\,\varphi_{6,t+1}. \tag{A.43}$$

Thus a central bank that follows the target criterion has to compute $\varphi_{6,t}$ as the solution to (A.43), and then evaluate (A.42) as a target criterion. The solution to (A.43) is given by:

$$\varphi_{6,t} = -2\,\mathbb{E}_t\sum_{i=0}^{\infty}(\pi_{t+i} + \frac{\lambda_x}{\kappa}x_{t+i})\prod_{j=0}^{i-1}\frac{\frac{k_{t+j}}{f_{t+j|t}}}{\frac{k_{t+j}}{f_{t+j|t}} + \mathbf{g}_{\pi,t+j}}. \tag{A.44}$$

The interpretation of (A.44) is that the anchoring constraint is not binding ($\varphi_{6,t} = 0$) if the central bank always hits the target ($\pi_{t+i} + \frac{\lambda_x}{\kappa}x_{t+i} = 0,\ \forall i$); or expectations are always anchored ($k_{t+j} = 0,\ \forall j$).

A.10 Parameterized expectations algorithm (PEA)

The objective of the parameterized expectations algorithm is to solve for the sequence of interest rates that solves the model equations including the target criterion, representing the first-order condition of the Ramsey problem. For convenience, I list the model equations:

$$x_t = -\sigma i_t + \begin{bmatrix} \sigma & 1-\beta & -\sigma\beta \end{bmatrix} f_{b,t} + \sigma \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} (I_3 - \beta h)^{-1} s_t, \tag{A.45}$$

$$\pi_t = \kappa x_t + \begin{bmatrix} (1-\alpha)\beta & \kappa\alpha\beta & 0 \end{bmatrix} f_{a,t} + \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} (I_3 - \alpha\beta h)^{-1} s_t, \tag{A.46}$$

$$f_{a,t} = \frac{1}{1-\alpha\beta}\bar{\pi}_{t-1} + b(I_3 - \alpha\beta h)^{-1} s_t, \tag{A.47}$$

$$f_{b,t} = \frac{1}{1-\beta}\bar{\pi}_{t-1} + b(I_3 - \beta h)^{-1} s_t, \tag{A.48}$$

$$f_{t|t-1} = \pi_t - (\bar{\pi}_{t-1} + b_1 s_{t-1}), \tag{A.49}$$

$$k_t = \sum_i \gamma_i b_i(f_{t|t-1}), \tag{A.50}$$

$$\bar{\pi}_t = \bar{\pi}_{t-1} + k_t f_{t|t-1}, \tag{A.51}$$

$$\pi_t = -\frac{\lambda_x}{\kappa}\left\{ x_t - \frac{(1-\alpha)\beta}{1-\alpha\beta}\left(k_t + f_{t|t-1}\mathbf{g}_{\pi,t}\right)\left(\mathbb{E}_t \sum_{i=1}^{\infty} x_{t+i} \prod_{j=0}^{i-1}(1 - k_{t+1+j} - f_{t+1+j|t+j}\mathbf{g}_{\bar{\beta},\mathbf{t+j}})\right)\right\}. \tag{A.52}$$

Denote the expectation on the right hand side of (A.52) as E_t. The idea of the PEA is to approximate this expectation and to solve model equations given the approximation $\hat{E}_t$. The algorithm is as follows:[54]

Objective: Obtain the sequence $\{i_t\}_{t=1}^T$ that solves Equations (A.45) - (A.52) for a history of exogenous shocks $\{s_t\}_{t=1}^T$ of length T.

1. Conjecture an initial expectation $\hat{E}_t = \beta^0 s(X_t)$.

   The expectation is approximated as a projection on a basis, $s(X_t)$, where β^0 are initial projection coefficients, and $X_t = (k_t, \bar{\pi}_{t-1}, r_t^n, u_t)$ is the state vector. I use

[54] For a thorough treatment of the PEA, see Christiano and Fisher (2000).

a monomial basis consisting of the first, second and third powers of X_t.

2. Solve model equations given conjectured $\hat{E}_t$ for a given sequence of shocks $\{s_t\}_{t=1}^T$.

   Compute residuals to the model equations (A.45) - (A.52) given $\{s_t\}_{t=1}^T$ and $\{\hat{E}_t\}_{t=1}^T$. Obtain a sequence $\{i_t\}_{t=1}^T$ that sets the residuals to zero. The output of this step is $\{v_t\}_{t=1}^T$, the simulated history of endogenous variables (Christiano and Fisher (2000) refer to this as a "synthetic time series").

3. Compute realized analogues of $\{E_t\}_{t=1}^T$ given $\{v_t\}_{t=1}^T$.

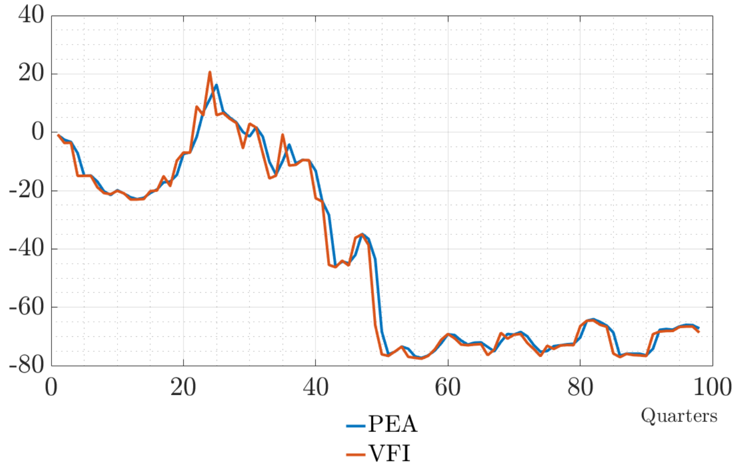
4. Update β regressing the synthetic E_t on $s(X_t)$.

   The coefficient update is $\beta^{i+1} = (s(X_t)'s(X_t))^{-1}s(X_t)'E_t$. Then iterate until convergence by evaluating at every step $||\beta^i - \beta^{i+1}||$.

A.11 Parametric value function iteration

This is an alternative approach I implement as a robustness check to the PEA. The objective is thus the same: to obtain the interest rate sequence that solves the model equations. The general value function iteration (VFI) approach is fairly standard, for which reason I refer to the Judd (1998) textbook for details. Specific to my application is that the state vector is five-dimensional, $X_t = (\bar{\pi}_{t-1}, r_t^n, u_t, r_{t-1}^n, u_{t-1})$, and that I approximate the value function using a cubic spline. Thus the output of the algorithm is a cubic spline approximation of the value function and a policy function for each node on the grid of states. Next, I interpolate the policy function using a cubic spline as well. As a last step I pass the state vector from the PEA simulation, obtaining an interest rate sequence conditional on the history of states. Fig. 25 shows the resulting interest rate sequence, obtained through the two approaches, conditional on a simulated sequence for the exogenous states.

Figure 25: Policy function for a particular history of states, PEA against VFI



A.12 Oscillatory dynamics in adaptive learning models

Here I present an illustration for why adaptive learning models produce oscillatory impulse responses if the gain is high enough. Consider a stylized adaptive learning model in two equations:

$$\pi_t = \beta e_t + u_t, \tag{A.53}$$

$$e_t = e_{t-1} + k(\pi_t - e_{t-1}). \tag{A.54}$$

The reader can recognize in (A.53) a simplified Phillips curve in which I am abstracting from output gaps to keep the presentation as clear as possible. Like in the simple model of Section 1.5.1 in the main text, e_t represents the one-period inflation expectation $\hat{\mathbb{E}}_t \pi_{t+1}$. (A.54), then, represents the simplest possible recursive updating of the expectations e_t. My notation of the gain as k indicates a constant gain specification, but the intuition remains unchanged for decreasing (or endogenous) gains.

Combining the two equations allows one to solve for the time series of expectations

$$e_t = \frac{1-k}{1-k\beta} e_{t-1} + \frac{k}{1-k\beta} u_t, \tag{A.55}$$

which, for β close but smaller than 1, is a near-unit-root process. (In fact, if the gain were to go to zero, this would be a unit root process.) Defining the forecast error as $f_{t|t-1} \equiv \pi_t - e_{t-1}$, one obtains

$$f_{t|t-1} = -\frac{1-\beta}{1-k\beta} e_{t-1} + \frac{1}{1-k\beta} u_t. \tag{A.56}$$

Equation (A.56) shows that in this simple model, the forecast error loads on a near-unit-root process with a coefficient that is negative and less than one in absolute value. Damped oscillations are the result.

Note that even if the gain would converge to zero, the coefficient on e_{t-1} would be negative and less than one in absolute value. Thus even for decreasing gain learning, one obtains oscillations, but the lower the gain, the more damped the oscillations become. This corroborates my findings in the impulse responses of Fig. 5. But importantly, the opposite extreme, when $k \to 1$, results in a coefficient of exactly -1, giving perpetual oscillations. This clearly illustrates how the oscillatory behavior of impulse responses comes from the oscillations in the forecast error that obtain when the gain is sufficiently large.

B Talking in Time - Dynamic Central Bank Communication

B.1 A two-signals formulation of the static model

Consider the same environment as in Section (2.2). Instead of signal (48), the CB can now choose to send two separate signals about the two states.

$$s_1 = \theta_1 + \varepsilon_1 \qquad \varepsilon_1 \sim \mathbb{N}(0, \sigma_1^2) \tag{B.1}$$

$$s_2 = \theta_2 + \varepsilon_2 \qquad \varepsilon_2 \sim \mathbb{N}(0, \sigma_2^2) \tag{B.2}$$

The inference of the PS is given by

$$I = \phi_1 s_1 + \phi_2 s_2,$$

where

$$\phi_1 = \frac{\rho\sigma_2}{-\rho^2 + \sigma_1 + \sigma_1\sigma_2 + \sigma_2 + 1} \tag{B.3}$$

$$\phi_2 = \frac{-\rho^2 + \sigma_1 + 1}{-\rho^2 + \sigma_1 + \sigma_1\sigma_2 + \sigma_2 + 1} \tag{B.4}$$

$$\tag{B.5}$$

Taking the squares of the CB loss yields, plugging in $I = \phi_1 s_1 + \phi_2 s_2$, the expressions for s_1, s_2 and ϕ_1, ϕ_2, and subsequently integrating the resulting expressions over Θ, the support of $\theta_i,\ i = 1, 2$, one obtains the expected loss function:

$$\mathcal{V} = \frac{\left(b^2 - 2b\rho + 1\right)\left(\rho^2 - \sigma_1 - 1\right) - \sigma_2\left(b^2\sigma_1 + (b - \rho)^2\right)}{\rho^2 - \left(\sigma_1 + 1\right)\left(\sigma_2 + 1\right)} \tag{B.6}$$

As in the main text, I observe the following properties when integrating:

$$\int_{\Theta} \theta_i \, d\theta_i = \mathbb{E}[\theta_i] = 0 \quad \text{for} \quad i = 1, 2 \quad \text{and} \quad \theta_i \in \Theta \tag{B.7}$$

$$\int_{\Theta} \theta_i^2 \, d\theta_i = \mathbf{Var}[\theta_i] = 1 \tag{B.8}$$

$$\int_{\Theta} \int_{\Theta} \theta_i \theta_j \, d\theta_i \, d\theta_j = \mathbf{Cov}[\theta_i, \theta_j] = \rho \quad \text{for} \quad i, j = 1, 2 \tag{B.9}$$

To simplify, suppose the communication about θ_2 is already given, that is, normalize $\sigma_2^2 = 1$. Then, the central bank's optimization problem can be stated as:

$$\sigma_1^2 = \arg\min \mathcal{V}, \tag{B.10}$$

with an associated first order condition

$$-\frac{\rho \left(2b \left(\rho^2 - 2\right) + \rho\right)}{\left(\rho^2 - 2 \left(\sigma_1 + 1\right)\right)^2} = 0, \tag{B.11}$$

which clearly has no solution. So I go on to investigate the behavior of the loss function to see whether we can find a corner solution. This is generalized in the Proposition below.

Proposition 10. *The solution* $(\sigma_1^2, \sigma_2^2)^*$ *to the central bank's communication problem, for an arbitrary value* $\hat{\sigma}_2^2 > 0$*, takes the form*

$$\begin{cases} (\sigma_1^2, \sigma_2^2) \to (\infty, \infty) & \textit{if} \quad b\rho < \frac{1}{2} \\ (\sigma_1^2, \sigma_2^2) = (0, \sigma_2^2) & \textit{else, with} \quad \sigma_2^2 < \hat{\sigma}_2^2 \end{cases} \tag{B.12}$$

Note that if $b\rho > \frac{1}{2}$*, then* $(\sigma_1^2, \sigma_2^2) = (0, 0)$ *is strictly better than* $(\sigma_1^2, \sigma_2^2) \to (\infty, \infty)$*, and if* $b\rho = \frac{1}{2}$*, then the CB is indifferent between* (∞, ∞) *and (0,0).*

Corollary 10.1. *Complete silence is optimal when neither* b *nor* ρ *are big enough, that is when i) the central bank does not want the PS' action to be sizable relatively to* θ_1

ii) the comovement between the states is not strong enough for the PS's action to be aligned with the central bank's preferences.

Corollary 10.2. *Full disclosure is better than complete silence when either condition i) or ii) from Corollary 10.1 is violated. But in this case there exists a threshold* $\bar{\sigma}_1^2$ *such that for higher precision* $\sigma_1^2 < \bar{\sigma}_1^2$ *partial disclosure of the form* $(\sigma_1^2, \sigma_2^2) = (0, \hat{\sigma}_2^2)$ *makes the CB strictly better off.*

Why is this the optimal thing for the CB? Because when full disclosure achieves a lower loss than silence, the states are either sufficiently close to each other or the CB forcibly wishes the PS's action to track θ_1, so it can communicate unambiguously about θ_1, and get the PS to respond to this information in their efforts to back out θ_2.

Corollary 10.3. *When the CB is indifferent between full and no disclosure, partial disclosure always makes it strictly better off.*

B.2 The Kalman filter for the dynamic model

Denoting $\theta_t^{'} := \theta_{t+1}$, set up the state-space system as:

$$x_{t+1} = hx_t + \eta\epsilon_{t+1} \quad \text{state equation,} \tag{B.13}$$

$$y_t = gx_t + v_t \quad \text{observation equation,} \tag{B.14}$$

with the observation vector given by $y_t = s_t$, and the state vector x_t and transition matrices are

$$x_t = \begin{bmatrix} \theta_t^{'} \\ \theta_t \end{bmatrix}, \quad h = \begin{bmatrix} \rho & 0 \\ 1 & 0 \end{bmatrix}, \quad g = \begin{bmatrix} \frac{1}{\psi} & 1 \end{bmatrix}, \quad \eta = \begin{bmatrix} \sigma_\varepsilon & 0 \\ 0 & 0 \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}, \quad Q = \eta\eta' = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad R = \sigma_v^2.$$

The optimal forecast of the state vector at $t+1$ is:

$$x_{t+1|t} = hx_{t|t-1} + h\kappa(y_t - gx\, x_{t|t-1}), \tag{B.15}$$

where $\kappa = Pg'\Omega^{-1}$ and the steady-state forecast-error-variance matrix P solves

$$hPh' - hPg'\Omega^{-1}gPh' + Q = P, \tag{B.16}$$

$$\Omega = gPg' + R. \tag{B.17}$$

Given the analytical solution to the 2×2 matrix $P = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_4 \end{bmatrix}$, suppressed here, m_i $i = 1, \ldots, 4$ are

$$m_1 = \rho - \kappa_1(\frac{\rho}{\psi} + 1), \tag{B.18}$$

$$m_2 = m_4 = \kappa_1, \tag{B.19}$$

$$m_3 = \frac{\kappa_1}{\psi}, \tag{B.20}$$

where κ_1 is the first element of the 2×1 Kalman gain and is given by

$$\kappa_1 = \frac{\rho p_4 + \frac{1}{\psi} p_1}{p_4 + \frac{1}{\psi^2} p_1 + 2\frac{\rho}{\psi} p_4 + \sigma_v^2}. \tag{B.21}$$

For simplicity, I refer to κ_1 as κ in the main text.

B.3 Expected loss functions

The expected loss function in the static model is given by

$$\mathcal{V}^{static} = \frac{b^2\left(\psi\left(2\rho + \psi\sigma_v^2 + \psi\right) + 1\right) - 2b(\rho + \psi)(\rho\psi + 1) + (\rho\psi + 1)^2}{\psi\left(2\rho + \psi\sigma_v^2 + \psi\right) + 1}, \tag{B.22}$$

with an associated first order condition of

$$\frac{2\left(b\left(\rho^2 + \psi^2\left(\rho^2\left(\sigma_v^2 - 1\right) + \sigma_v^2 + 1\right) + 2\rho\psi\sigma_v^2 - 1\right) + \psi(\rho\psi + 1)\left(\rho^2 - \sigma_v^2 - 1\right)\right)}{\left(\psi\left(2\rho + \psi\sigma_v^2 + \psi\right) + 1\right)^2} = 0. \tag{B.23}$$

In the dynamic model, the expected loss function is

$$\mathcal{V}^{dynamic} = \frac{1}{1-\beta}\Big\{b^2 +$$
$$\frac{\rho^2\psi^2+\rho^2+4\rho\psi+\rho^2\psi^2\left(-\sigma_v^2\right)-\sqrt{((\rho+1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi-1)^2)((\rho-1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi+1)^2)}+\psi^2\sigma_v^2+\psi^2+1}{2(\rho+\psi)^2}$$
$$-\frac{1}{\psi(\rho+\psi)^2\left(\rho^2+\rho\psi\sigma_v^2-1\right)}b\Big(2\rho^3\psi\left(\psi^2-1\right)+2\rho^5\psi+\rho^4\left(\psi^2\left(\sigma_v^2+3\right)+1\right)$$
$$+2\rho\psi^3\left(2\sigma_v^2-1\right)+\psi^4\sigma_v^4+\psi^4\sigma_v^2+2\psi^2\sigma_v^2-\psi^2+1-\rho^2\left(\psi^4\sigma_v^2\left(\sigma_v^2-1\right)+\psi^2\left(\sigma_v^2+2\right)+2\right)$$
$$-\sqrt{((\rho+1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi-1)^2)((\rho-1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi+1)^2)}$$
$$+\rho^2\sqrt{((\rho+1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi-1)^2)((\rho-1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi+1)^2)}$$
$$-\psi^2\sigma_v^2\sqrt{((\rho+1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi-1)^2)((\rho-1)^2\psi^2\sigma_v^2-(\rho^2-1)(\psi+1)^2)}\Big)\Big\},$$

and the first order condition is suppressed.

B.4 Optimal precision

The first order condition in static model is

$$-\frac{2\psi^2(\rho\psi+1)\sigma_v(-2b(\rho+\psi)+\rho\psi+1)}{\left(\psi\left(2\rho+\psi\sigma_v^2+\psi\right)+1\right)^2}=0, \tag{B.24}$$

which is solved by $\sigma_v = 0$. That this is either a minimum or a maximum can be verified by considering the second order condition, which is given by

$$\frac{2\psi^2(\rho\psi+1)(2b(\rho+\psi)-\rho\psi-1)\left(\psi\left(2\rho-3\psi\sigma_v^2+\psi\right)+1\right)}{\left(\psi\left(2\rho+\psi\sigma_v^2+\psi\right)+1\right)^3}. \tag{B.25}$$

The first and second order conditions in the dynamic model are too large and are thus suppressed.

B.5 Prior and posterior variances in the two models

Using the definitions of prior and posterior variances from the main text, in the static model these are given by

$$\pi(\theta_2) = 1, \tag{B.26}$$

$$p(\theta_2) = 1 - Var(\theta_{2|s}), \tag{B.27}$$

$$\pi(\theta_1) = 1, \tag{B.28}$$

$$p(\theta_1) = 1 - Var(\theta_{1|s}). \tag{B.29}$$

In the dynamic model, instead, they are

$$\pi(\theta_{t+1}) = 1 - \rho^2 Var(\theta_{t+1|t}), \tag{B.30}$$

$$p(\theta_{t+1}) = 1 - Var(\theta_{t+1|t}), \tag{B.31}$$

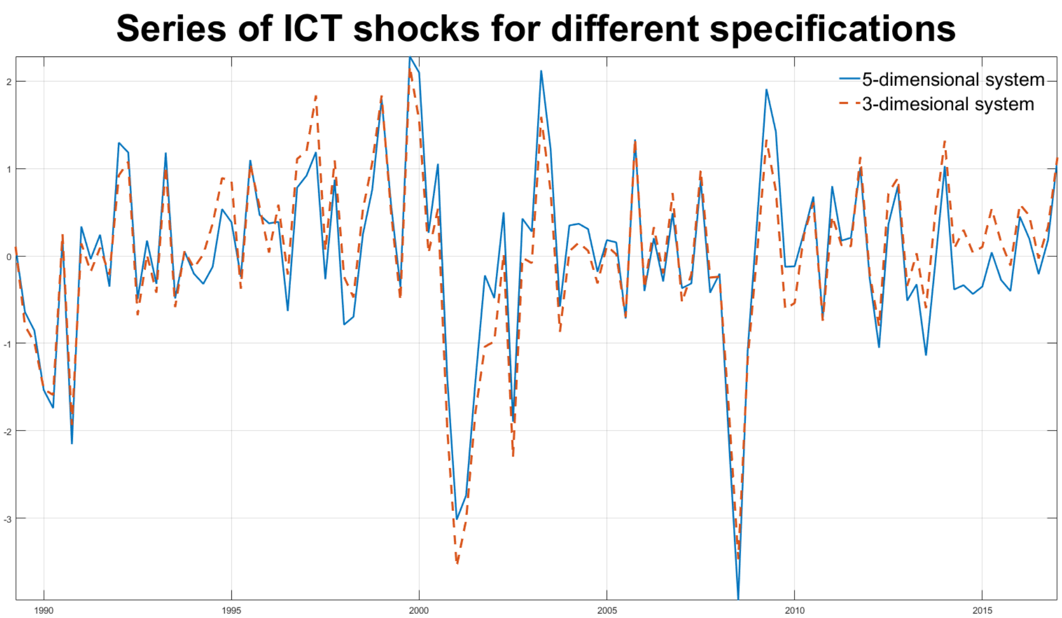
$$\pi(\theta_t) = 1 - Var(\theta_{t|t-1}) = p(\theta_{t+1}), \tag{B.32}$$

$$p(\theta_t) = 1 - Var(\theta_{t|t}). \tag{B.33}$$

C ICT and Future Productivity - Evidence and Theory of a General-Purpose Technology

C.1 Removing Forward-Looking Variables

Figure 26: ICT shock series using the empirical strategy presented in 3.2.1.2.



Blue solid line represents shock series for the 5-dimension system 76 presented in 3.2.1. Red dotted line represents shock series for the 3-dimensional system presented in C.1.

C.2 Proof of Proposition 9

Notice that in this proof we use the same assumptions and the same notation in the 2-sector model presented in Section 3.3.

First order condition with respect to labor input for firm j is

$$w_t = (1 - a - b) p_t \eta_t \theta^c (k_t^i)^\gamma \left(\frac{k_t^c(j)}{l_t(j)} \right)^a \left(\frac{k_t^i(j)}{l_t(j)} \right)^b, \tag{C.1}$$

and first order condition with respect to labor input for firm q is

$$w_t = (1-a-b)\eta_t \theta^c (k_t^i)^\gamma \left(\frac{k_t^c(q)}{l_t(q)}\right)^a \left(\frac{k_t^i(q)}{l_t(q)}\right)^b, \tag{C.2}$$

Since both production functions display identical input shares then in equilibrium input ratios must be the same over the two sectors, which is,

$$\frac{k_t^c(j)}{l_t(j)} = \frac{k_t^c(q)}{l_t(q)}, \tag{C.3}$$

$$\frac{k_t^i(j)}{l_t(j)} = \frac{k_t^i(q)}{l_t(q)}, \tag{C.4}$$

$$\frac{k_t^i(j)}{k_t^c(j)} = \frac{k_t^i(q)}{k_t^c(q)}.$$

Now, dividing C.1 over C.2, we can use C.3 and C.4 to obtain

$$1 = \frac{1}{p_t}\frac{\theta_t^c}{\theta_t^i} \quad \Rightarrow \quad p_t = \frac{\theta_t^c}{\theta_t^i}. \tag{C.5}$$

From Equation C.5, it is clear that the equilibrium path of p_t is fully independent on the past, current, and future values of neutral technology, η_t. □

References

Andreasen, M. M., J. H. Christensen, and S. Riddell (2018). The TIPS liquidity premium. Federal Reserve Bank of San Francisco.

Angeletos, G.-M., Z. Huo, and K. A. Sastry (2020, June). Imperfect macroeconomic expectations: Evidence and theory. Working Paper 27308, National Bureau of Economic Research.

Angeletos, G.-M. and A. Pavan (2007). Efficient use of information and social value of information. *Econometrica 75*(4), 1103–1142.

Angeletos, G.-M. and A. Pavan (2009). Policy with dispersed information. *Journal of the European Economic Association 7*(1), 11–60.

Anufriev, M. and C. Hommes (2012). Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics 4*(4), 35–64.

Anzoategui, D., D. Comin, M. Gertler, and J. Martinez (2016). Endogenous technology adoption and r&d as sources of business cycle persistence. Technical report, National Bureau of Economic Research.

Ball, L. (1994). Credible disinflation with staggered price-setting. *American Economic Review 84*(1), 282–289.

Barro, R. J. and D. B. Gordon (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics 12*(1), 101–121.

Barsky, R. B., S. Basu, and K. Lee (2015). Whither news shocks? *NBER Macroeconomics Annual 29*(1), 225–264.

Barsky, R. B. and E. R. Sims (2011). News shocks and business cycles. *Journal of Monetary Economics 58*(3), 273–289.

Basu, S. and J. Fernald (2007). Information and communications technology as a general-purpose technology: Evidence from us industry data. *German Economic Review 8*(2), 146–173.

Basu, S., J. Fernald, J. Fisher, and M. Kimball (2010). Sector-specific technical change.

Basu, S., J. G. Fernald, N. Oulton, and S. Srinivasan (2003). The case of the missing productivity growth, or does information technology explain why productivity accelerated in the united states but not in the united kingdom? *NBER macroeconomics annual 18*, 9–63.

Baxter, M. and R. G. King (1999). Measuring business cycles: approximate band-pass filters for economic time series. *Review of Economics and Statistics 81*(4), 575–593.

Beaudry, P. and F. Portier (2006). Stock prices, news, and economic fluctuations. *American Economic Review 96*(4), 1293–1307.

Bianchi, F., H. Kung, and G. Morales (2014). Growth, slowdowns, and recoveries. Technical report, National Bureau of Economic Research.

Bils, M. and P. J. Klenow (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy 112*(5), 947–985.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2017). Are ideas getting harder to find? Technical report, National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *The Journal of Finance 73*(1), 199–227.

Branch, W. A. and G. W. Evans (2006). A simple recursive forecasting model. *Economics Letters 91*(2), 158–166.

Branch, W. A. and G. W. Evans (2011, Jun). Monetary policy and heterogeneous expectations. *Economic Theory 47*(2), 365–393.

Bresnahan, T. and M. Trajtenberg (1992). General purpose technologies engines of growth? National Bureau of Economic Research Working Paper No. 4148.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological) 37*(2), 149–163.

Bullard, J. and K. Mitra (2002). Learning about monetary policy rules. *Journal of Monetary Economics 49*(6), 1105–1129.

Byrne, D. M., J. G. Fernald, and M. B. Reinsdorf (2016). Does the United States have a productivity slowdown or a measurement problem? *Brookings Papers on Economic Activity 2016*(1), 109–182.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics 12*(3), 383–398.

Campbell, J. R., C. L. Evans, J. D. M. Fisher, and A. Justiniano (2012). Macroeconomic effects of Federal Reserve forward guidance [with comments and discussion]. *Brookings Papers on Economic Activity*, 1–80.

Carvalho, C., S. Eusepi, E. Moench, and B. Preston (2019). Anchored inflation expectations. *Available at SSRN 3018198*.

Cette, G., J. Fernald, and B. Mojon (2016). The pre-great recession slowdown in productivity. *European Economic Review 88*, 3–20.

Chahrour, R. (2014, July). Public communication and information acquisition. *American Economic Journal: Macroeconomics 6*(3), 73–101.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2000). Sticky price models of the business cycle: can the contract multiplier solve the persistence problem? *Econometrica 68*(5), 1151–1179.

Cho, I.-K. and A. Matsui (1995). Induction and the Ramsey policy. *Journal of Economic Dynamics and Control 19*(5-7), 1113–1140.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*(1), 1–45.

Christiano, L. J. and J. D. Fisher (2000). Algorithms for solving dynamic models with occasionally binding constraints. *Journal of Economic Dynamics and Control 24*(8), 1179 – 1232.

Clarida, R., J. Gali, and M. Gertler (1999). The science of monetary policy: a new Keynesian perspective. *Journal of Economic Literature 37*(4), 1661–1707.

Cogley, T. and J. M. Nason (1993). Impulse dynamics and propagation mechanisms in a real business cycle model. *Economics Letters 43*(1), 77–81.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–78.

Comin, D. and M. Gertler (2006). Medium-term business cycles. *American Economic Review 96*(3), 523–551.

Dombeck, B. (2017). The effects of news shocks and bounded rationality on macroeconomic volatility.

Duffie, D. and K. J. Singleton (1990). Simulated moments estimation of Markov models of asset prices. Technical report, National Bureau of Economic Research.

Eichenbaum, M., N. Jaimovich, and S. Rebelo (2011). Reference prices, costs, and nominal rigidities. *American Economic Review 101*(1), 234–62.

Eusepi, S., M. Giannoni, and B. Preston (2018). On the limits of monetary policy.

Eusepi, S. and B. Preston (2011). Expectations, learning, and business cycle fluctuations. *American Economic Review 101*(6), 2844–72.

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

Evans, G. W., S. Honkapohja, T. J. Sargent, and N. Williams (2013). *Bayesian model averaging, learning and model selection*, Chapter 6, pp. 99–119. Oxford University Press Oxford.

Evans, G. W. and B. McGough (2009). Learning to optimize. Technical report, Citeseer.

Fisher, J. D. (2006). The dynamic effects of neutral and investment-specific technology shocks. *Journal of Political Economy 114*(3), 413–451.

Forni, M. and L. Gambetti (2014). Sufficient information in structural vars. *Journal of Monetary Economics 66*, 124–136.

Frankel, A. and N. Kartik (2018). What kind of central bank competence? *Theoretical Economics 13*(2), 697–727.

Gaballo, G. (2016, January). Rational inattention to news: The perils of forward guidance. *American Economic Journal: Macroeconomics 8*(1), 42–97.

Gaspar, V., F. Smets, and D. Vestin (2010). Inflation expectations, adaptive learning and optimal monetary policy. In *Handbook of Monetary Economics*, Volume 3, pp. 1055–1095. Elsevier.

Gobbi, L., R. Mazzocchi, and R. Tamborini (2019). Monetary policy, de-anchoring of inflation expectations, and the “new normal”. *Journal of Macroeconomics 61*, 103070.

Goldstein, I. and Y. Leitner (2018). Stress tests and information disclosure. *Journal of Economic Theory*.

Goodfriend, M. (1993). Interest rate policy and the inflation scare problem: 1979-1992. *FRB Richmond Economic Quarterly 79*(1), 1–23.

Gordon, R. J. (2012). Is US economic growth over? faltering innovation confronts the six headwinds. Technical report, National Bureau of Economic Research.

Greenwood, J., Z. Hercowitz, and P. Krusell (1997). Long-run implications of investment-specific technological change. *The American Economic Review*, 342–362.

Greenwood, J., Z. Hercowitz, and P. Krusell (2000). The role of investment-specific technological change in the business cycle. *European Economic Review 44*(1), 91–115.

Grossman, G. M. and E. Helpman (1991). Quality ladders in the theory of growth. *The Review of Economic Studies 58*(1), 43–61.

Guerron, P. A. and Q. R. Jinnai (2015). Financial frictions, trends, and the great recession. Technical report, Hitotsubashi Institute for Advanced Study, Hitotsubashi University.

Hamilton, J. D. (1994). *Time series analysis*, Volume 2. Princeton university press Princeton.

Hansen, S. and M. McMahon (2016). First impressions matter: Signalling as a source of policy dynamics. *Review of Economic Studies 83*(4), 1645–1672.

Hebden, J., E. P. Herbst, J. Tang, G. Topa, and F. Winkler (2020). How robust are makeup strategies to key alternative assumptions?

Hellwig, C. and L. Veldkamp (2009). Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies 76*(1), 223–251.

Inostroza, N. and A. Pavan (2017). Persuasion in global games with application to stress testing.

Ireland, P. N. (2000). Expectations, credibility, and time-consistent monetary policy. *Macroeconomic Dynamics 4*(4), 448–466.

Jafari Samimi, A. and Y. E. Roshan (2012). The impact of ICT shocks on business cycle some evidence from iran. *Iranian Economic Review 16*(31), 123–145.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies 76*(1), 283–317.

Judd, K. L. (1998). *Numerical Methods in Economics*. Number 0262100711 in MIT Press Books. The MIT Press.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *The American Economic Review 101*(6), 2590–2615.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics 80*(2), 218–230.

Klenow, P. J. and O. Kryvtsov (2008). State-dependent or time-dependent pricing: Does it matter for recent US inflation? *The Quarterly Journal of Economics 123*(3), 863–904.

Klenow, P. J. and B. A. Malin (2010). Microeconomic evidence on price-setting. In *Handbook of Monetary Economics*, Volume 3, pp. 231–284. Elsevier.

Kreps, D. M. (1998). Anticipated utility and dynamic choice. *Econometric Society Monographs 29*, 242–274.

Kydland, F. E. and E. C. Prescott (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy 85*(3), 473–491.

Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370.

Lee, B.-S. and B. F. Ingram (1991). Simulation estimation of time-series models. *Journal of Econometrics 47*(2-3), 197–205.

Long Jr, J. B. and C. I. Plosser (1983). Real business cycles. *Journal of Political Economy 91*(1), 39–69.

Lubik, T. A. and C. Matthes (2016). Indeterminacy and learning: An analysis of monetary policy in the Great Inflation. *Journal of Monetary Economics 82*, 85 – 106.

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

Lütkepohl, H. (2013). *Introduction to multiple time series analysis*. Springer Science & Business Media.

Mackowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *American Economic Review 99*(3), 769–803.

Mankiw, N. G. and R. Reis (2002). Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. *The Quarterly Journal of Economics 117*(4), 1295–1328.

Marcet, A. and J. P. Nicolini (2003). Recurrent hyperinflations and learning. *American Economic Review 93*(5), 1476–1498.

Mele, A., K. Molnár, and S. Santoro (2019). On the perils of stabilizing prices when agents are learning. *Journal of Monetary Economics*.

Milani, F. (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics 54*(7), 2065–2082.

Milani, F. (2014). Learning and time-varying macroeconomic volatility. *Journal of Economic Dynamics and Control 47*, 94–114.

Mokyr, J. (2014). Secular stagnation? not in your life. *Secular Stagnation: Facts, Causes and Cures 83*.

Molnár, K. and S. Santoro (2014). Optimal monetary policy when agents are learning. *European Economic Review 66*, 39–62.

Moran, P. and A. Queralto (2017). Innovation, productivity, and monetary policy. *Journal of Monetary Economics*.

Morris, S. and H. S. Shin (2002, December). Social value of public information. *American Economic Review 92*(5), 1521–1534.

Moscarini, G. (2007). Competence implies credibility. *The American Economic Review 97*(1), 37–63.

Nakamura, E. and J. Steinsson (2008). Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics 123*(4), 1415–1464.

Oliner, S. D. and D. E. Sichel (2000). The resurgence of growth in the late 1990s: is information technology the story? *Journal of Economic Perspectives 14*(4), 3–22.

Orphanides, A. and J. Williams (2004). Imperfect knowledge, inflation expectations, and monetary policy. In *The inflation-targeting debate*, pp. 201–246. University of Chicago Press.

Oulton, N. (2007). Investment-specific technological change and growth accounting. *Journal of Monetary Economics 54*(4), 1290–1299.

Oulton, N. (2012). Long term implications of the ICT revolution: Applying the lessons of growth theory and growth accounting. *Economic Modelling 29*(5), 1722–1736.

Preston, B. (2005). Learning about monetary policy rules when long-horizon expectations matter. MPRA paper, University Library of Munich, Germany.

Reis, R. (2011). When should policymakers make announcements? 2011 Meeting Papers 122, Society for Economic Dynamics.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy 98*(5, Part 2), S71–S102.

Rotemberg, J. J. and M. Woodford (1997). An optimization-based econometric framework for the evaluation of monetary policy. *NBER Macroeconomics Annual 12*, 297–346.

Sargent, T. J. (1999). *The Conquest of American Inflation*. Princeton University Press.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Sims, E. R. (2012). News, non-invertibility, and structural vars. In *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*, pp. 81–135. Emerald Group Publishing Limited.

Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, S63–S84.

Stiroh, K. J. (2002). Information technology and the us productivity revival: what do the industry data say? *American Economic Review 92*(5), 1559–1576.

Svensson, L. E. (1999). Inflation targeting as a monetary policy rule. *Journal of Monetary Economics 43*(3), 607–654.

Svensson, L. E. O. (2006, March). Social value of public information: Comment: Morris and Shin (2002) is actually pro-transparency, not con. *American Economic Review 96*(1), 448–452.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester conference series on public policy*, Volume 39, pp. 195–214. Elsevier.

Whelan, K. (2003). A two-sector approach to modeling US NIPA data. *Journal of Money, Credit and Banking*, 627–656.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

Woodford, M. (2019). Monetary policy analysis when planning horizons are finite. *NBER Macroeconomics Annual 33*(1), 1–50.