Imperfect Information, Optimal Monetary Policy and Informational Consistency*

PAUL LEVINE
UNIVERSITY OF SURREY†

JOSEPH PEARLMAN
CITY UNIVERSITY, LONDON

BO YANG

XI'AN JIAOTONG - LIVERPOOL UNIVERSITY (XJTU), SUZHOU, CHINA
AND
UNIVERSITY OF SURREY

AUGUST 26, 2014

Abstract

This paper examines the implications of imperfect information (II) for optimal monetary policy with a consistent set of informational assumptions for the modeller and the private sector an assumption we term the *informational consistency*. We use an estimated simple NK model from Levine *et al.* (2012), where the assumption of symmetric II significantly improves the fit of the model to US data to assess the welfare costs of II under commitment, discretion and simple Taylor-type rules. Our main results are: first, common to all information sets we find significant welfare gains from commitment only with a zero-lower bound constraint on the interest rate. Second, optimized rules take the form of a price level rule, or something very close across all information cases. Third, the combination of limited information and a lack of commitment can be particularly serious for welfare. At the same time we find that II with lags introduces a 'tying ones hands' effect on the policymaker that may improve welfare under discretion. Finally, the impulse response functions under our most extreme imperfect information assumption (output and inflation observed with a two-quarter delay) exhibit hump-shaped behaviour and the fiscal multiplier is significantly enhanced in this case.

JEL Classification: C11, C52, E12, E32.

Keywords: Imperfect Information, DSGE Model, Optimal Monetary Policy, Bayesian Estimation

* The paper has been presented at a CCBS, Bank of England workshop on "Modelling Monetary Policy" October 17-19; the National Bank of Poland Conference, "DSGE and Beyond", Warsaw, September 29 - 30, 2011; the MONFISPOL final Conference at Goethe University, September 19 - 20, 2011; the CDMA Conference "Expectations in Dynamic Macroeconomic Models" at St Andrews University, August 31 - September 2, 2011; the 17th International Conference on Computing in Economics and Finance, San Francisco, June 29 - July 1, 2011 and the European Monetary Forum, University of York, March 4 - 5, 2011. Comments by participants at these events are gratefully acknowledged, especially those of discussants Andrzej Torój and Martin Ellison at the NBP and MONFISPOL conferences respectively, as are those by seminar participants at Glasgow University, the University of Surrey, the Norges Bank and the University of Kent. We also acknowledge financial support from ESRC project RES-062-23-2451 and from the EU Framework Programme 7 project MONFISPOL.

† Corresponding author; School of Economics, University of Surrey, Guildford, Surrey GU2 7XH UK; tel: +44(0)1483 689928; email: p.levine@surrey.ac.uk

Contents

1 Introduction

The formal estimation of DSGE models by Bayesian methods has now become standard.[1] However, as Levine *et al.* (2007) first pointed out, in the standard approach there is an implicit asymmetric informational assumption that needs to be critically examined: whereas perfect information about current shocks and other macroeconomic variables is available to the economic agents, it is not to the econometricians. By contrast, in Levine *et al.* (2007) and Levine *et al.* (2012) a symmetric information assumption is adopted in which rational expectations amounts to model consistency. This approach can be thought of as the informational counterpart to the "*cognitive consistency principle*" proposed in Evans and Honkapohja (2009) which holds that economic agents should be assumed to be "about as smart as, but no smarter than good economists". The assumption that agents have imperfect and certainly no more information than the econometrician who constructs and estimates the model on behalf of the policymaker, amounts to what we term *informational consistency (IC)*. IC may seem plausible, but does it improve the empirical performance of DSGE models? Drawing upon Levine *et al.* (2012), we show this is indeed is the case for a standard NK model.[2]

The main focus of this paper is on the implications of imperfect information for optimal monetary policy. A sizeable literature now exists on this subject - a by no means exhaustive selection of contributions includes: Cukierman and Meltzer (1986), Pearlman (1992), Svensson and Woodford (2001), Svensson and Woodford (2003), Faust and Svensson (2001), Faust and Svensson (2002), Aoki (2003), Aoki (2006) and (Melecky *et al.* (2008).[3]

Our contribution is to study three policy issues using an estimated DSGE model with IC at both the estimation and policy design stages of the exercise. We compare the Bayesian-estimated NK model under two information assumptions: *perfect information* (PI) where the private sector observes all the macroeconomic state variables in the model (including exogenous shock processes) and *imperfect information* (II) where they only observe the data used by the econometrician and IC applies. A marginal likelihood race shows strong evidence in favour of II. Moreover the impulse response functions for all policy regimes display the hump-shaped behaviour highlighted in the literature.[4]

The three policy questions posed are first, what are the welfare costs associated with the private sector possessing only II of the state variables? Indeed, in a second-best world where the policymaker cannot commit, can withdrawal of information be welfare-improving? Second, what are the implications of II for the gains from commitment (or, equivalently, the costs of discretion) and third, how does II affect the form of welfare-optimized Taylor rules? A novel feature of our analysis, irrespective of the information assumptions, is the consideration of the zero lower bound (ZLB) constraint in the design of interest rate

[1] See Fernandez-Villaverde (2009) for a comprehensive review.

[2] The possibility that imperfect information in NK models improves the empirical fit has also been examined by Collard and Dellas (2004), Collard and Dellas (2006), Collard *et al.* (2009), although an earlier assessment of the effects of imperfect information for an IS-LM model dates back to Minford and Peel (1983)

[3] Imperfect information covers a wide range of informational assumptions: section 5 provides a discussion of the various assumed information sets assumed for both the private sector and policymaker/modeller in these papers.

[4] An interesting feature of the impulse response analysis is a comparison of the fiscal multipliers under PI or II.

rules. Our paper is, we believe, the first to provide an integrated treatment of information, commitment, the simplicity constraint on rules and zero-lower bound aspects of the policy problem, enabling us to explore the important interaction between these aspects.

The rest of the paper is organized as follows. Section 2 describes a fairly standard new Keynesian (NK) model subsequently used for the policy analysis. Section 3 sets out the general solution procedure for solving such a model under II given a particular (and usually sub-optimal) policy rule. Section 4 describes the estimation by Bayesian methods. Section 5 sets out the general framework for calculating optimal policy under different information assumptions. Section 6 turns to numerical solutions of optimal policy in the estimated model first assuming PI for both the private sector and the policymaker. For all information assumptions our policy exercises use the ex post welfare-optimal (Ramsey) as a benchmark against which to assess the policies where either no commitment mechanism is available and the central bank exercises discretion, or where policy conducted in the form of a simple interest rate, Taylor-type commitment rules. Under perfect information, in sub-section 6.1, both sets of agents, the central bank and the private sector observes the full state vector describing the model model dynamics (though the econometrician only uses a sub-set of this data). Sub-section 6.2 then relaxes this assumption by introducing different II sets that correspond to IC adopted at the estimation stage. Section 7 concludes.

2 The Model

We utilize a fairly standard NK model with a Taylor-type interest rate rule, one factor of production (labour) with decreasing returns to scale. The model has external habit in consumption, price indexing and a Taylor interest-rate rule with persistence. These are part of the model, albeit ad hoc in the case of indexing, and therefore are endogenous sources of persistence. Persistent exogenous shocks to demand, technology and the price mark-up classify as exogenous persistence. A key feature of the model and the focus of the paper is a further endogenous source of persistence that arises when agents have imperfect information and learn about the state of the economy using Kalman-filter updating.[5]

The full model in non-linear form is as follows

$$1 = \beta R_t E_t \left[\frac{MU^C_{t+1}}{MU^C_t \Pi_{t+1}} \right] \tag{1}$$

$$\frac{W_t}{P_t} = -\frac{1}{(1-\frac{1}{\eta})} \frac{MU^N_t}{MU^C_t} \tag{2}$$

$$Y_t = F(A_t, N_t, \Delta_t) = \frac{A_t N_t^\alpha}{\Delta_t} \text{ where } \Delta_t \equiv \frac{1}{n} \sum_{j=1}^{n} (P_t(j)/P_t)^{-\zeta} \tag{3}$$

$$MC_t = \frac{W_t}{P_t F_{N,t}} = \frac{W_t \Delta_t}{P_t A_t \alpha N_t^{\alpha-1}} \tag{4}$$

[5]The simplicity of the model facilitates the separate examination of different sources of persistence in the model – see Levine *et al.* (2012).

$$(5)$$

$$\begin{aligned} H_t - \xi\beta E_t[\tilde{\Pi}_{t+1}^{\zeta-1} H_{t+1}] &= Y_t MU_t^C & (6)\\ J_t - \xi\beta E_t[\tilde{\Pi}_{t+1}^{\zeta} J_{t+1}] &= \frac{1}{1-\frac{1}{\zeta}} MC_t MS_t Y_t MU_t^C & (7)\\ 1 &= \xi\tilde{\Pi}_t^{\zeta-1} + (1-\xi)\left(\frac{J_t}{H_t}\right)^{1-\zeta} \quad \text{where } \tilde{\Pi}_t \equiv \frac{\Pi_t}{\Pi_{t-1}^{\gamma}} & (8)\\ \tilde{\Pi}_t &\equiv \frac{\Pi_t}{\Pi_{t-1}^{\gamma}} & (9)\\ Y_t &= C_t + G_t & (10) \end{aligned}$$

Equation (1) is the familiar Euler equation with β the discount factor, R_t the gross nominal interest rate, MU_t^C the marginal utility of consumption and $\Pi \equiv \frac{P_t}{P_{t-1}}$ the gross inflation rate, with P_t the price level. The operator $E_t[\cdot]$ denotes rational expectations conditional upon a general information set (see section 3). In (2) the real wage, $\frac{W_t}{P_t}$ is a mark-up on the marginal rate of substitution between leisure and consumption. MU_t^N is the marginal utility of labour supply N_t. Equation (4) defines the marginal cost. Equations (6) – (8) describe Calvo pricing with $1-\xi$ equal to the probability of a monopolistically competitive firm re-optimizing its price $P_t^0 = \frac{J_t}{H_t}$, indexing by an amount γ with an exogenous mark-up shock MS_t. ζ is the elasticity of substitution of each variety entering into the consumption basket of the representative household.

Equation (3) is the production function, with labour the only variable input into production and the technology shock A_t exogenous. Price dispersion Δ_t, defined in (3), can be shown for large n, the number of firms, to be given by

$$\Delta_t = \xi\tilde{\Pi}_t^{\zeta}\Delta_{t-1} + (1-\xi)\left(\frac{J_t}{H_t}\right)^{-\zeta} \tag{11}$$

Finally (10), where C_t denotes consumption, describes output equilibrium, with an exogenous government spending demand shock G_t. To close the model we assume a current inflation based Taylor-type interest-rule

$$\begin{aligned} \log R_t &= \rho_r \log R_{t-1} + (1-\rho_r)\left(\theta_\pi \log\frac{\Pi_t}{\Pi_{tar,t}} + \log(\frac{1}{\beta}) + \theta_y \log\frac{Y_t}{\bar{Y}_t}\right) + \epsilon_{e,t}\\ \log\frac{\Pi_{tar,t+1}}{\Pi} &= \rho_\pi \log\frac{\Pi_{tar,t}}{\Pi} + \epsilon_{\pi,t+1} & (12) \end{aligned}$$

where $\bar{Y}_t$ is the output trend and $\Pi_{tar,t}$ is a time-varying inflation target following an AR(1) process, (12), and $\epsilon_{e,t}$ is a monetary policy shock.[6] The following form of the single period utility for household r is a non-separable function of consumption and labour effort that is consistent with a balanced growth steady state:

$$U_t = \frac{\left[(C_t(r) - h_C C_{t-1})^{1-\varrho}(1-N_t(r))^{\varrho}\right]^{1-\sigma}}{1-\sigma} \tag{13}$$

[6]Note the Taylor rule feeds back on output relative to its steady state rather than the output gap so we avoid making excessive informational demands on the central bank when implementing this rule.

where $h_C C_{t-1}$ is external habit. In equilibrium $C_t(r) = C_t$ and marginal utilities MU_t^C and MU_t^N are obtained by differentiation:

$$MU_t^C = (1-\varrho)(C_t - h_C C_{t-1})^{(1-\varrho)(1-\sigma)-1}(1-N_t)^{\varrho(1-\sigma)} \quad (14)$$
$$MU_t^N = -(C_t - h_C C_{t-1})^{(1-\varrho)(1-\sigma)}\varrho(1-N_t)^{\varrho(1-\sigma)-1} \quad (15)$$

Shocks $A_t = Ae^{a_t}$, $G_t = Ge^{g_t}$, $\Pi_{tar,t}$ are assumed to follow log-normal AR(1) processes, where A, G denote the non-stochastic balanced growth values or paths of the variables A_t, G_t. Following Smets and Wouters (2007) and others in the literature, we decompose the price mark-up shock into persistent and transient components: $MS_t = MS_{per}e^{msper_t}MS_{tra}e^{\varepsilon_{mstra,t}}$ where $msper_t$ is an AR(1) process and $\varepsilon_{mstra,t}$ is i.i.d., which results in MS_t being an ARMA(1,1) process. We can normalize $A = 1$ and put $MS = MS_{per} = MS_{tra} = 1$ in the steady state. The innovations are assumed to have zero contemporaneous correlation. This completes the model. The equilibrium is described by (1) – (12), the expressions for MU_t^C and MU_t^N, (14) – (15), and processes for the six exogenous shocks in the system: A_t, G_t, $MS_{per,t}$, $MS_{tra,t}$, $\Pi_{tar,t}$ and $\epsilon_{e,t}$.

Bayesian estimation is based on the rational expectations solution of the log-linear model. The conventional approach assumes that the private sector has perfect information of the entire state vector including crucially, all six current shocks. These are extreme information assumptions and exceed the data observations on three data sets output, inflation and the nominal interest rate (Y_t, Π_t and R_t) that we subsequently use to estimate the model. If the private sector can only observe these data series (we refer to this as *symmetric information*) we must turn from a solution under *perfect* information by the private sector (later referred to as *asymmetric information* – *AI* – since the private sector's information set exceeds that of the econometrician) to one under *imperfect information* – *II.*

3 General Solution with Imperfect Information

The model with a particular and not necessarily optimal rule is a special case of the following general setup in non-linear form

$$Z_{t+1} = J(Z_t, E_tZ_t, X_t, E_tX_t) + \nu\sigma_\epsilon\epsilon_{t+1} \quad (16)$$
$$E_tX_{t+1} = K(Z_t, E_tZ_t, X_t, E_tX_t) \quad (17)$$

where Z_t, X_t are $(n-m)\times 1$ and $m \times 1$ vectors of backward and forward-looking variables, respectively, and ϵ_t is a $\ell \times 1$ shock variable, ν is an $(n-m)\times\ell$ matrix and σ_ϵ is a small scalar. Either analytically, or numerically using the methods of Levine and Pearlman (2011), a log-linearized form state-space representation can be obtained as

$$\begin{bmatrix} z_{t+1} \\ E_tx_{t+1} \end{bmatrix} = A^1 \begin{bmatrix} z_t \\ x_t \end{bmatrix} + A^2 \begin{bmatrix} E_tz_t \\ E_tx_t \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ 0 \end{bmatrix} \quad (18)$$

where z_t, x_t are vectors of backward and forward-looking variables, respectively, and u_t is a shock variable. [7] We also define $A^1 = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. In addition we assume that agents all make the same observations at time t, which are given, in non-linear and linear forms respectively, by

$$M_t^{obs} = m(Z_t, E_t Z_t, X_t, E_t X_t) + \mu\sigma_\epsilon \epsilon_t \tag{19}$$

$$m_t = \begin{bmatrix} M_1 & M_2 \end{bmatrix} \begin{bmatrix} z_t \\ x_t \end{bmatrix} + \begin{bmatrix} L_1 & L_2 \end{bmatrix} \begin{bmatrix} E_t z_t \\ E_t x_t \end{bmatrix} + v_t \tag{20}$$

where $\mu\sigma_\epsilon\epsilon_t$ and v_t represents measurement errors. Given the fact that expectations of forward-looking variables depend on the information set, it is hardly surprising that the absence of full information will impact on the path of the system.

In order to simplify the exposition we assume terms in $E_t Z_t$ and $E_t X_t$ do not appear in the set-up so that in the linearized form $A^2 = L = 0$.[8] Full details of the solution for the general setup are provided in PCL.

3.1 Linear Solution Procedure

Now we turn to the solution for (18) and (20). First assume *perfect information*. Following Blanchard and Kahn (1980), it is well-known that there is then a saddle path satisfying:

$$x_t + N z_t = 0 \quad \text{where} \quad \begin{bmatrix} N & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \Lambda^U \begin{bmatrix} N & I \end{bmatrix}$$

where Λ^U has unstable eigenvalues. In the *imperfect information* case, following PCL, we use the Kalman filter updating given by

$$\begin{bmatrix} z_{t,t} \\ x_{t,t} \end{bmatrix} = \begin{bmatrix} z_{t,t-1} \\ x_{t,t-1} \end{bmatrix} + J \left[m_t - \begin{bmatrix} M_1 + L_1 & M_2 + L_2 \end{bmatrix} \begin{bmatrix} z_{t,t-1} \\ x_{t,t-1} \end{bmatrix} \right]$$

where we denote $z_{t,t} \equiv E_t[z_t]$ etc. Thus the best estimator of the state vector at time $t-1$ is updated by multiple J of the error in the predicted value of the measurement. The matrix J is given by

$$J = \begin{bmatrix} PD^T \\ -NPD^T \end{bmatrix} \Gamma^{-1}$$

where $D \equiv M_1 - M_2 A_{22}^{-1} A_{21}$, $M \equiv [M_1 \;\; M_2]$ is partitioned conformably with $\begin{bmatrix} z_t \\ x_t \end{bmatrix}$, $\Gamma \equiv EPD^T + V$ where $E \equiv M_1 + L_1 - (M_2 + L_2)N$, $V = \mathrm{cov}(v_t)$ is the covariance matrix of the measurement errors and P satisfies the Riccati equation (24) below.

Using the Kalman filter, the solution as derived by PCL[9] is given by the following

[7] In Pearlman *et al.* (1986), henceforth PCL, a more general setup allows for shocks to the equations involving expectations.

[8] In fact our model is of this simplified form.

[9] A less general solution procedure for linear models with imperfect information is in Lungu *et al.* (2008)

processes describing the pre-determined and non-predetermined variables z_t and x_t and a process describing the innovations $\tilde{z}_t \equiv z_t - z_{t,t-1}$:

$$\begin{aligned}
\text{Predetermined}: \quad z_{t+1} &= Cz_t + (A-C)\tilde{z}_t + (C-A)PD^T(DPD^T+V)^{-1}(D\tilde{z}_t + v_t) \\
&\quad + u_{t+1} && (21)\\
\text{Non-predetermined}: \quad x_t &= -Nz_t + (N - A_{22}^{-1}A_{21})\tilde{z}_t && (22)\\
\text{Innovations}: \quad \tilde{z}_{t+1} &= A\tilde{z}_t - APD^T(DPD^T+V)^{-1}(D\tilde{z}_t + v_t) + u_{t+1} && (23)
\end{aligned}$$

where

$$C \equiv A_{11} - A_{12}N, \quad A \equiv A_{11} - A_{12}A_{22}^{-1}A_{21}, \quad D \equiv M_1 - M_2A_{22}^{-1}A_{21}$$

and P is the solution of the Riccati equation given by

$$P = APA^T - APD^T(DPD^T+V)^{-1}DPA^T + \Sigma \tag{24}$$

where $\Sigma \equiv \mathrm{cov}(u_t)$ is the covariance matrix of the shocks to the system. The measurement m_t can now be expressed as

$$m_t = Ez_t + (D-E)\tilde{z}_t + v_t - (D-E)PD^T(DPD^T+V)^{-1}(D\tilde{z}_t + v_t) \tag{25}$$

We can see that the solution procedure above is a generalization of the case when there are $n - m$ ($= dim(z)$) measurements at each time t, and when the observations are uncontaminated by noise; provided the matrix D is of full rank and square[10], so that its inverse exists, from (24) we have $P = \Sigma$. This then yields the standard Blanchard-Kahn solution for perfect information

$$z_{t+1} = Cz_t + u_{t+1}\,; \quad x_t = -Nz_t \tag{26}$$

By comparing (26) with (21) we see that the determinacy of the system is independent of the information set. This is an important property that contrasts with the case where private agents use statistical learning to form forward expectations.

3.2 The Filtering and Likelihood Calculations

Now consider the computations of the Bayesian econometrician estimating the model. To evaluate the likelihood for a given set of parameters (prior to multiplying by their prior probabilities), the econometrician takes the equations (21), (23) and (25) as representing the dynamics of the system under imperfect information. In order to reduce the amount of subsequent notation, we now augment the state space so that the measurement errors $\{v_t\}$ are incorporated into the system errors $\{u_t\}$, which entails augmenting the states $\{z_t\}, \{\tilde{z}_t\}$ to incorporate these as well; for convenience we then retain the notation above, but now the covariance matrix $V = 0$. It is a standard result that apart from constants, we can

with an application to a small open economy model, which they also extend to a non-linear version.

[10] An obvious example of this is when $M = [I\ 0]$. Another more useful example is when M_2 is of the same rank r as $A_{22}^{-1}A_{21}$, and linearly independent of M_1, which has rank $n - m - r$.

write the likelihood function as:

$$2\ln L = -\sum \ln\det(\mathrm{cov}(e_t)) - \sum e_t^T(\mathrm{cov}(e_t))^{-1}e_t \tag{27}$$

where the innovations process $e_t \equiv m_t - E_{t-1}m_t$.

In order to obtain $E_{t-1}m_t$, we need to solve the appropriate filtering problem. At first sight, it would seem that the obvious way to do this is to subtract (23) from (21) to obtain

$$z_{t+1,t} = Cz_{t,t-1} + CPD^T(DPD^T)^{-1}D\tilde{z}_t \tag{28}$$

and to substitute for $D\tilde{z}_t$ from the measurement equation now written correspondingly as

$$m_t = Ez_{t,t-1} + EPD^T(DPD^T)^{-1}D\tilde{z}_t \tag{29}$$

However this is incorrect, because whereas these equations do describe the steady state dynamics, they do not generate the covariance matrix of the innovations process, which evolves over time. In order to generate this, we apply the Kalman filter to equations (21), (28) and (29), where we note that the initial covariance matrix of $\tilde{z}_0$ is P and $cov(\tilde{z}_0, z_{0,-1}) = 0$. It is then easy to show by induction that the Kalman filter generates $E_{t-1}\tilde{z}_t = 0$, with corresponding covariance matrix equal to P for all t, and in addition the filtering covariance matrix between $\tilde{z}_t$ and $z_{t,t-1}$ is 0 for all t. Finally, defining $\bar{z}_t = z_{t,t-1}$, the remaining updates from the Kalman filter are given by:

$$\bar{z}_{t+1,t} = C\bar{z}_{t,t-1} + CZ_tE^T(EZ_tE^T)^{-1}e_t \qquad e_t \equiv m_t - E\bar{z}_{t,t-1}$$

$$Z_{t+1} = CZ_tC^T + PD^T(DPD^T)^{-1}DP - CZ_tE^T(EZ_tE^T)^{-1}EZ_tC^T$$

the latter being a time-dependent Riccati equation. The initial value of Z_t is given by

$$Z_0 = H + PD^T(DPD^T)^{-1}DP \qquad \text{where } H = CHC^T + CPD^T(DPD^T)^{-1}DPC^T$$

and $H = cov(z_{0,-1})$. Finally, $cov(e_t) = EZ_tE^T$.

In principle, this cannot be directly extended to the case when there are unit roots, which typically may originate from technology shocks. However Koopman and Durbin (2003) have shown that the initial covariance matrix can be decomposed in the form $H = \kappa H^a + H^b$, as $\kappa \to \infty$, where H^a and H^b can be directly obtained computationally. In practice, one would set higher and higher values for κ until the likelihood converged, which would then permit marginal likelihood comparisons for differing models. This decomposition of the initial covariance matrix is a better computational strategy than the arbitrary approach of setting its diagonals equal to a very large number.

3.3 When Can Perfect Information be Inferred?

We now pose the question: under what conditions do the RE solutions under perfect and imperfect information actually differ? By observing a subset of outcomes can agents actually infer the full state vector, including shocks?

To answer this basic question we first explore the possibility of representing the solution to the model under imperfect information as a VAR.[11] First define
$s_t \equiv \begin{bmatrix} z_t \\ \tilde{z}_t \end{bmatrix}$ and $\epsilon_t \equiv \begin{bmatrix} u_t \\ v_{t-1} \end{bmatrix}$ and

$$m_t = \begin{bmatrix} \tilde{M}_1 & \tilde{M}_2 \end{bmatrix} \begin{bmatrix} s_t \\ x_t \end{bmatrix} + v_t \tag{30}$$

Then the solution set out in the previous section can be written as

$$s_{t+1} = \tilde{A}s_t + \tilde{B}\epsilon_{t+1} \tag{31}$$
$$x_t = -\tilde{N}s_t \tag{32}$$

where $\tilde{A}$, $\tilde{B}$ and $\tilde{N}$ are functions of A, B, C, N, P, D, U and V. Hence

$$m_{t+1} = (\tilde{M}_1 - \tilde{M}_2\tilde{N})(\tilde{A}s_t + \tilde{B}\epsilon_{t+1}) + v_{t+1} \equiv \tilde{C}s_t + \tilde{D}\epsilon_{t+1}$$

Suppose that *the number of shocks=the number of observed variables.* With at least one shock this can only be true if there is no measurement error; so we also put $v_t = 0$. With this assumption D is square. Suppose first that it is invertible. Then we can write

$$\epsilon_{t+1} = \tilde{D}^{-1}(m_{t+1} - \tilde{C}s_t)$$

Substituting into (31) we then have

$$[I - (\tilde{A} - \tilde{B}\tilde{D}^{-1}\tilde{C})L]s_{t+1} = \tilde{B}\tilde{D}^{-1}m_{t+1}$$

Iterating we arrive at

$$s_t = \sum_{j=0}^{\infty}[\tilde{A} - \tilde{B}\tilde{D}^{-1}\tilde{C}]^j\tilde{B}\tilde{D}^{-1}m_{t-j} \tag{33}$$
$$m_{t+1} = \tilde{C}\sum_{j=0}^{\infty}[\tilde{A} - \tilde{B}\tilde{D}^{-1}\tilde{C}]^j\tilde{B}\tilde{D}^{-1}w_{t-j} + \tilde{D}\epsilon_{t+1} \tag{34}$$

Then provided matrix $[\tilde{A} - \tilde{B}\tilde{D}^{-1}\tilde{C}]$ has stable eigenvalues, the summations converge.[12] Then (34) is an *infinite* VAR representation of the solution to our DSGE model. Furthermore, from (33), observations on the history of m_t imply that s_t is observed. This is consistent with our full information RE assumption. Thus we have the result that *if agents observe* m_t *without measurement error and if the number of shocks = the number of observations, then by observing the latter agents can infer the full state vector if* $\tilde{D}$ *is invertible. Imperfect information is equivalent to complete information in this special case.*

[11] This section essentially generalizes Fernandez-Villaverde *et al.* (2007) to the case of imperfect information.

[12] This is an innocuous requirement - see the online Appendix of Levine *et al.* (2012).

Under what conditions would $\tilde{D}$ be singular? An obvious case pursued later in the optimal policy exercises is under imperfect information where some variables are observed only with one or two lags. Then the current shocks cannot influence these observed variables so some of rows (two in this case) are zero, meaning $\tilde{D}$ is not invertible. In our model then, both these sufficient conditions for imperfect information collapsing to the perfect information case do not hold, so we can expect differences between the two cases.[13]

4 Bayesian Estimation

The Bayesian approach combines the prior distributions for the individual parameters with the likelihood function, evaluated using the Kalman filter, to form the posterior density. The likelihood does not admit an analytical solution, so the posterior density is computed through the use of the Monte-Carlo Markov Chain sampling methods. The linearized model is estimated using the Dynare software (Juillard (2003)), which has been extended by the paper's authors to allow for imperfect information on the part of the private sector.

4.1 Data and Priors

To estimate the system, we use three macro-economic observables at quarterly frequency for the US: real GDP, the GDP deflator and the nominal interest rate. Since the variables in the DSGE model are measured as deviations from the trend, the time series for GDP is de-trended and those for inflation and the nominal interest rate are demeaned. Following Smets and Wouters (2003), for GDP we use a linear trend.[14] Real variables in the model are now measured in proportional (virtually identical to logarithmic) deviations from linear trends, in percentage points, while inflation (the GDP deflator) and the nominal interest rate are de-trended by the same linear trend in inflation and converted to quarterly rates. The estimation results are based on a sample from 1981:1 to 2006:4.

The values of priors are taken from Levin *et al.* (2006) and Smets and Wouters (2007). Table 7 in Appendix B provides an overview of the priors used for each model variant described below. In general, inverse gamma distributions are used as priors when non-negativity constraints are necessary, and beta distributions for fractions or probabilities. Normal distributions are used when more informative priors seem to be necessary. We use the same prior means as in previous studies and allow for larger standard deviations, i.e. less informative priors. For the parameters γ, h_C and ξ we center the prior density in the middle of the unit interval. The priors related to the process for the price mark-up shock are taken from Smets and Wouters (2007). One structural parameter, $\beta = 0.99$, is kept fixed in the estimation procedure. A consumption-output ratio $c_y \equiv \frac{C}{Y} = 0.6$ is imposed in the steady state. Given c_y and h_C, the parameter ϱ is calibrated to target hours worked in

[13] In fact many NK DSGE models do have the property that the number of shocks equal the number of observables, and the latter are current values without lags - for example Smets and Wouters (2003).

[14] In Levine *et al.* (2012) a more comprehensive empirical exercise is carried out that includes second moment comparisons, and identification and robustness checks. Estimations were run using a linear-quadratic trend obtaining virtually identical parameter estimates, with the ordering of data densities under II and AI assumptions remaining unchanged.

the steady state at $N = 0.4$.[15]

4.2 Estimation Results

We examine two information sets: first we make the assumption that private agents are better informed than the econometricians and have perfect information (PI) – the standard information assumption in the estimation literature). Then we examine a symmetric information set for both econometrician and private agents: imperfect Information with observable sets $I_t = [y_t, \pi_t, r_t]$. Table 8 in Appendix B reports the parameter estimates using Bayesian methods. It summarizes posterior means of the studied parameters and 90% uncertainty bands for the two information sets, PI and II, as well as the posterior model odds. Overall, the parameter estimates are plausible, and are generally similar to those of Levin *et al.* (2006) and Smets and Wouters (2007).

It is interesting to note that the parameter estimates are fairly consistent across the information assumptions despite the fact that these alternatives lead to a better model fit based on the corresponding posterior marginal likelihood. Focusing on the parameters characterizing the degree of price stickiness and the existence of real rigidities, we find that the price indexation parameter, γ, is estimated to be smaller than assumed in the prior distribution (in line with those reported by Smets and Wouters (2007)). The estimate of γ imply that inflation is intrinsically not very persistent. The posterior mean estimates for the Calvo price-setting parameter, ξ, imply an average price contract duration of about 7 quarters (compared with the prior of 2 quarters). This is rather high, but is consistent with findings in much of the literature including Smets and Wouters (2007).[16]. The external habit parameter is estimated to be around $60-90\%$ of past consumption, which is consistent with other estimates reported in the literature.

In Table 1 we report the posterior marginal likelihood from the estimation which is computed using the Geweke (1999) modified harmonic-mean estimator. This can be interpreted as maximum log-likelihood values, penalized for the model dimensionality, and adjusted for the effect of the prior distribution (Chang *et al.* (2002)). Whichever model variant has the highest marginal likelihood attains the best relative model fit.

Information set	PI	II with $I_t = [y_t,\, \pi_t,\, r_t]$
Log ML	-105.84	-102.36

Table 1: Log Marginal Likelihood Values Across Information Sets

In order to compare models we calculate the relative model posterior probabilities as follows. Let $p_i(\theta|m_i)$ represent the *prior* distribution of the parameter vector $\theta \in \Theta$ for some model $m_i \in M$ and let $L(y|\theta, m_i)$ denote the likelihood function for the observed data $y \in Y$ conditional on the model and the parameter vector. Then by Bayes' rule the joint posterior distribution of θ for model m_i combines the likelihood function with the prior

[15] A full discussion of the choice of priors is provided in Levine *et al.* (2012).

[16] Modifying the model to have Kimball preferences (Kimball (1995)) enables a flat estimated Philips curve to be made consistent with shorter contracts - see Smets and Wouters (2007)

distribution:

$$p_i(\theta|y, m_i) \propto L(y|\theta, m_i)\, p_i(\theta|m_i)$$

Bayesian inference provides a framework for comparing alternative and potentially misspecified models based on their *marginal likelihood.* For a given model $m_i \in M$ and common data set, the latter is obtained by integrating out the vector θ,

$$L(y|m_i) = \int_{\Theta} L(y|\theta, m_i)\, p(\theta|m_i)\, d\theta$$

where $p_i(\theta|m_i)$ is the prior density for model m_i, and $L(y|m_i)$ is the data density for model m_i given parameter vector θ. For m_i and m_j, the *Bayes Factor* is then the ratio of their posterior model probabilities when the prior odds ratio, $\frac{p(m_i)}{p(m_j)}$, is set to unity:

$$BF_{i,j} \equiv \frac{p(m_i|y)}{p(m_j|y)} = \frac{L(y|m_i)p(m_i)}{L(y|m_j)p(m_j)} = \frac{L(y|m_i)}{L(y|m_j)} = \frac{\exp(LL(y|m_i))}{\exp(LL(y|m_j))} = \exp(LL(y|m_i) - LL(y|m_j))$$

in terms of the log-marginal likelihoods (LL). According to Jeffries (1996), a BF of 3-10 is "slight evidence" in favour of model i over j. This corresponds to a LL difference in the range [ln 3, ln 10]= [1.10,2.30]. A BF of 10-100 or a LL range of [2.30, 4.61] is "strong to very strong" evidence; a BF over 100 (LL over 4.61) is "decisive" evidence.

Table 1 now reveals that information set II outperforms information set PI by a Bayes factor of 3.48 which is "strong to very strong" evidence of II over PI. This is a striking result; although informational consistency is intuitively appealing, there is no inevitability that models that assume this will perform better in LL terms than the traditional assumption of PI. The evidence in favour of II confirms the significant persistence effect seen in the analytic models of Collard and Dellas (2006) and Levine *et al.* (2012).[17]

5 The General Set-Up and Optimal Policy Problem

This section describes the general set-up that applies irrespective of the informational assumptions. Removing the estimated rule (12), for a given set of observed policy instruments w_t we now consider a linearized model in a general state-space form:

$$\begin{bmatrix} \mathsf{z}_{t+1} \\ E_t\mathsf{x}_{t+1} \end{bmatrix} = A^1 \begin{bmatrix} \mathsf{z}_t \\ \mathsf{x}_t \end{bmatrix} + A^2 \begin{bmatrix} E_t\mathsf{z}_t \\ E_t\mathsf{x}_t \end{bmatrix} + B\mathsf{w}_t + \begin{bmatrix} \mathsf{u}_{t+1} \\ 0 \end{bmatrix} \tag{35}$$

where $\mathsf{z}_t, \mathsf{x}_t$ are vectors of backward and forward-looking variables, respectively, w_t is a vector of policy variables, and u_t is an i.i.d. zero mean shock variable with covariance matrix Σ_u; as before the more general setup in PCL allows for shocks to the equations involving expectations. In addition for the imperfect information case, we assume that

[17] A limitation of the likelihood race methodology is that the assessment of model fit is only relative to its other rivals with different restrictions. The outperforming model in the space of competing models may still be poor (potentially misspecified) in capturing the important dynamics in the data. To further evaluate the absolute performance of one particular model (or information assumption) against data, it is necessary to compare the model's implied characteristics with those of the actual data and with a benchmark DSGE-VAR model. See Levine *et al.* (2012)

agents all make the same observations at time t, which are still given by (20).

Define target variables s_t by

$$\mathsf{s}_t = J\mathsf{y}_t + H\mathsf{w}_t \tag{36}$$

Then the policy-maker's loss function at time t by

$$\Omega_t = \frac{1}{2}E_t\left[\sum_{\tau=0}^{\infty}\beta^t(\mathsf{s}_{t+\tau}^T Q_1 \mathsf{s}_{t+\tau} + \mathsf{w}_{t+\tau}^T Q_2 \mathsf{w}_{t+\tau})\right] \tag{37}$$

where Q_1 and Q_2 are symmetric and non-negative definite and $\beta \in (0,1)$ is a discount factor. This could be an ad hoc loss function or a large distortions approximation to the household's utility as described in Levine *et al.* (2008a) and summarized in Appendix E. Substituting (36) into (37) results in the following form of the loss function used subsequently in the paper

$$\Omega_t = \frac{1}{2}E_t\left[\sum_{i=0}^{\infty}\beta^t(\mathsf{y}_{t+\tau}^T Q \mathsf{y}_{t+\tau} + 2\mathsf{y}_{t+\tau}^T U \mathsf{w}_{t+\tau} + \mathsf{w}_{t+\tau}^T R \mathsf{w}_{t+\tau})\right] \tag{38}$$

where $Q = J^T Q_1 M$, $U = J^T Q_1 H$ and $R = Q_2 + H^T Q_1 H$.

For the literature described in the introduction, rational expectations are formed assuming the following information sets:

1. For what we refer to as 'perfect information' (see Svensson and Woodford (2001), Aoki (2003), Aoki (2006) and standard Bayesian estimation):
   $I_t^{ps} = \{\mathsf{z}_\tau, \mathsf{x}_\tau\},\ \tau \leq t; A^1, A^2, B, \Sigma_u, [Q, U, R, \beta]$ or the monetary rule for the private sector and
   $I_t^{pol} = \{\mathsf{m}_\tau\},\ \tau \leq t; A^1, A^2, B, M, L, \Sigma_u, \Sigma_v, [Q, U, R, \beta]$ and the monetary rule for the policymaker/modeller.

2. For 'imperfect information' (see Pearlman (1992), Svensson and Woodford (2003) and for Bayesian estimation with IC in this paper):
   $I_t^{ps} = I^{pol} = \{\mathsf{m}_\tau\},\ \tau \leq t$ given by (20); $A^1, A^2, B, M, L, \Sigma_u, \Sigma_v, [Q, U, R, \beta]$ and the monetary rule.

3. For an alternative category of asymmetric imperfect information (see Cukierman and Meltzer (1986), Faust and Svensson (2001), Faust and Svensson (2002)) and (Melecky *et al.* (2008)):
   $I_t^{pol} = \{\mathsf{m}_\tau\},\ \tau \leq t; A^1, A^2, B, M, L, \Sigma_u, \Sigma_v, [Q, U, R, \beta]$ and the monetary rule for the policymaker sector and
   $I_t^{pol} \supset I_t^{ps} = \{\mathsf{m}_\tau\},\ \tau \leq t; A^1, A^2, B, M, L, \Sigma_u, \Sigma_v$ for the private sector.

In the rest of the paper we confine ourselves to information set 1 for perfect information (PI) and information set 2 for imperfect information (II). Information set 1 is incompatible with IC. Information set 3 is however compatible and is needed to address the issue of optimal ambiguity. However this interesting case is beyond the scope of this paper.

The welfare losses for information sets 1 and 2 can be summarised in the theorem below, and proved in Appendix D:

Theorem: The expected welfare for each of the regimes is given by

$$W^J = \mathsf{z}_{0,0}^T S^J \mathsf{z}_{0,0} + \frac{\lambda}{1-\lambda}\mathrm{tr}\Big(S^J P D^T (DPD^T + V)^{-1} DP\Big) + \frac{1}{1-\lambda}\mathrm{tr}\Big(Q_{11} - Q_{12}A_{22}^{-1}A_{21} - A_{21}^T A_{22}^{-T} Q_{21} + A_{21}^T A_{22}^{-T} Q_{22} A_{22}^{-1} A_{21}\Big)\bar{P} \tag{39}$$

where J =OPT, TCT, SIM refer to the optimal, time-consistent and optimized simple rules respectively; the second term is the expected value of the first three terms of (D.8) under each of the rules, and the final term is independent of the policy rule, and is the expected value of the final term of (D.8), utilising (D.2). Also note that from the perfect information case in Appendix C:

$$S^{OPT} = S_{11} - S_{12}S_{22}^{-1}S_{21} \tag{40}$$

- S_{ij} are the partitions of S, the Riccati matrix used to calculate the welfare loss under optimal policy with commitment.
- S^{TCT} is used to calculate the welfare loss in the time consistent solution algorithm.
- $S^{SIM} = V^{LYA}$ is calculated from the Lyapunov equation used to calculate the welfare under the optimized simple rule.

In the special case of perfect information on all z_t, $M_1 = I$, L, v_t, V are all zero, so that $D = I$. It follows that $\bar{P} = 0$ and the last term in (39) disappears. Moreover $P = \Sigma$, $\mathsf{z}_{0,0} = \mathsf{z}_0$ and (39) reduces to the welfare loss expressions obtained in Appendix C. Thus the effect of imperfect information is to introduce a new term into the welfare loss that depends only on the model filter, but is independent of policy and to modify the first policy-dependent term by an effect that depends on the solution P to the Riccati equation associated with the Kalman Filter.

6 Optimal Monetary Policy in the Estimated NK Model

This section sets out numerical results for optimal policy under commitment, optimal discretionary (or time consistent) policy and for an optimized simple Taylor rule. The model is the estimated form of the best-fitting one, namely that under II with observables $I_t = [y_t, \pi_t, r_t]$. For the first set of results we ignore ZLB considerations. The questions we pose are first, what are the welfare costs associated with the private sector possessing only imperfect information of the state variables? Second, what are the implications of imperfect information for the gains from commitment? To assess these we compare the welfare outcomes under commitment and discretion. Third, how does imperfect information affect the form of optimized Taylor rules and the costs of simplicity, and finally what are the impulse responses to shocks under different information assumptions and policy regimes?

With one preferred estimated model in place, to address these questions we now examine the following forms of imperfect information (II) for the private sector and policymaker.[18]

[18]Note for $j > 0$ informational consistency still holds, in that the econometrician using historical data has more information than the private sector at the time it forms rational expectations.

In log-linearized form[19]

$$\textbf{Information Set II}: \quad m_t \;=\; \begin{bmatrix} y_{t-j} \\ \pi_{t-j} \\ r_t \end{bmatrix}; \quad j = 0, 1, 2 \tag{41}$$

This contrasts with the information set under perfect information (PI) which consists of all the state variables including the shock processes a_t, g_t, etc.

We considered simple inflation targeting rules that respond to both inflation and output:

$$r_t = \rho_r r_{t-1} + \theta_\pi \pi_t + \theta_y y_t$$

for PI and for the II information set with $j \geq 0$, one of the two forms:

$$r_t = \rho_r r_{t-1} + \theta_\pi E_t \pi_t + \theta_y E_t y_t \quad (\textbf{Form A})$$

$$r_t = \rho_r r_{t-1} + \theta_\pi \pi_{t-j} + \theta_y y_{t-j} \quad (\textbf{Form B})$$

Thus for form A the rule responds to the best estimate of inflation and output given observations of m_t. For form B the response is to direct observations available to both the private sector and the policymaker at the time the interest rate is set. Of course for PI and II with $j = 0$ forms A and B are identical.

With this choice of Taylor rule the case where $\rho_r = 1$ and $\alpha_y = 0$ is of particular interest as this then corresponds to a *price-level* rule. There has been a recent interest in the case for price-level rather than inflation stability. Gaspar *et al.* (2010) provide an excellent review of this literature. The basic difference between the two regimes in that under an inflation targeting mark-up shock leads to a commitment to use the interest rate to accommodate an increase in the inflation rate falling back to its steady state. By contrast a price-level rule commits to an inflation rate below its steady state after the same initial rise. Under inflation targeting one lets bygones be bygones allowing the price level to drift to a permanently different price-level path whereas price-level targeting restores the price level to its steady state path. The latter can lower inflation variance and be welfare enhancing because forward-looking price-setters anticipates that a current increase in the general price level will be undone giving them an incentive to moderate the current adjustment of its own price. In our results we will see whether price-level targeting is indeed welfare optimal across different information assumptions.

6.1 Optimal Policy without Zero Lower Bound Considerations

Results are presented for a loss function that is formally a quadratic approximation about the steady state of the Lagrangian, and which represents the true approximation about the fully optimal solution appropriate for a distorted steady state. This welfare-based loss function has been obtained numerically using the procedure set out in Appendix E.

Table 2 sets out the stochastic inter-temporal welfare loss for our three policy regimes

[19] Strictly speaking, we use proportional deviations from steady state, so that lower case variables are defined as $x_t = \frac{X_t - X}{X}$. r_t and π_t are proportional deviations of *gross* rates.

under PI and II. Consumption equivalent losses relative to the optimal policy under PI are shown in brackets.[20] We immediately observe that the welfare losses associated with either an inability to commit or from simplicity are small under all information sets, but grow in significance as information available at time t declines. At most, welfare costs are $c_e = 0.02\%$.

Information	Information Set	Optimal	Time Cons	Simple Rule A	Simple Rule B
Perfect	Full state vector	6.78 (0)	7.70 (0.01)	6.81 (0.005)	6.81 (0.005)
Imperfect	$I_t = [y_t, \pi_t, r_t]$	7.50 (0.004)	8.46 (0.01)	7.52 (0.004)	7.52 (0.004)
Imperfect	$I_t = [y_{t-1}, \pi_{t-1}, r_t]$	8.63 (0.01)	9.22 (0.01)	8.64 (0.01)	8.70 (0.01)
Imperfect	$I_t = [y_{t-2}, \pi_{t-2}, r_t]$	9.41 (0.01)	9.96 (0.02)	9.42 (0.01)	9.54 (0.02)

Table 2: Welfare Loss and Information without ZLB Considerations

Simple rules are able to closely replicate the welfare outcome under the fully optimal solution. Table 3 shows that for PI and II but no lags in information, this is achieved with a first-difference interest rate rule ($\rho_r = 1$) and no significant feedback from output ($\alpha_y \simeq 0$), *a price level rule* in other words. In fact α_y is slightly negative indicating that monetary policy accommodates an increase in output above the steady state, rather than 'leaning against the wind', but the effect is very small. For information set II with one or two lags, the form of the rule is close to a price level rule. For simple rule B which responds only to directly observed data on inflation and output, interest rates respond less to inflation as the information lag increases. This is intuitive: policy responds less to dated information and less than that for estimates of the target variables (form A), a result broadly in accordance with the Brainard principle (Brainard (1967)).

Information	Information Set	Simple Rule A $[\rho_r,\ \theta_\pi,\ \theta_y]$	Simple Rule B $[\rho_r,\ \theta_\pi\ \theta_y]$
Perfect	Full state vector	[1.000, 2.162, −0.014	[1.000, 2.162, −0.014]
Imperfect	$I_t = [y_t, \pi_t, r_t]$	[1.000, 2.439, −0.026]	[1.000, 2.439, −0.026]
Imperfect	$I_t = [y_{t-1}, \pi_{t-1}, r_t]$	[0.951, 2.235, −0.025]	[0.914, 0.729, −0.008]
Imperfect	$I_t = [y_{t-2}, \pi_{t-2}, r_t]$	[0.962, 2.291, −0.015]	[0.862, 0.706, −0.013]

Table 3: Optimized Coefficients in Simple Rules without ZLB Considerations

[20]To derive the welfare in terms of a consumption equivalent percentage increase ($c_e \equiv \frac{\Delta C}{C} \times 10^2$), expanding $U(X_t, 1 - N_t)$ as a Taylor series, a $\Delta U = U_C \Delta C = CMU^C c_e \times 10^{-2}$. Losses X reported in the Table are of the order of variances expressed as percentages and have been scaled by $1 - \beta$. Thus $X \times 10^{-4} = \Delta U$ and hence $c_e = \frac{X \times 10{-}2}{CMU^C}$. For the steady state of this model, $CMU^C = 1.77$. It follow that a welfare loss difference of $X = 100$ gives a consumption equivalent percentage difference of $c_e = 0.566\%$.

Figure 1: IRFs with Technology Shock

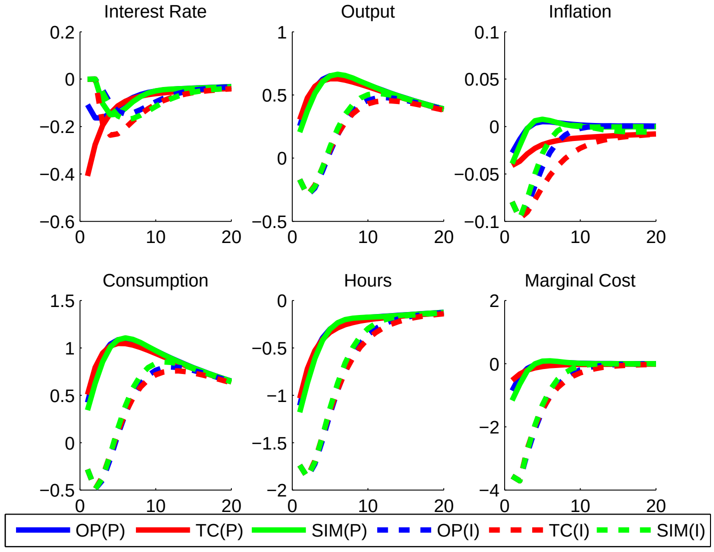


Figure 2: IRFs with Government Spending Shock

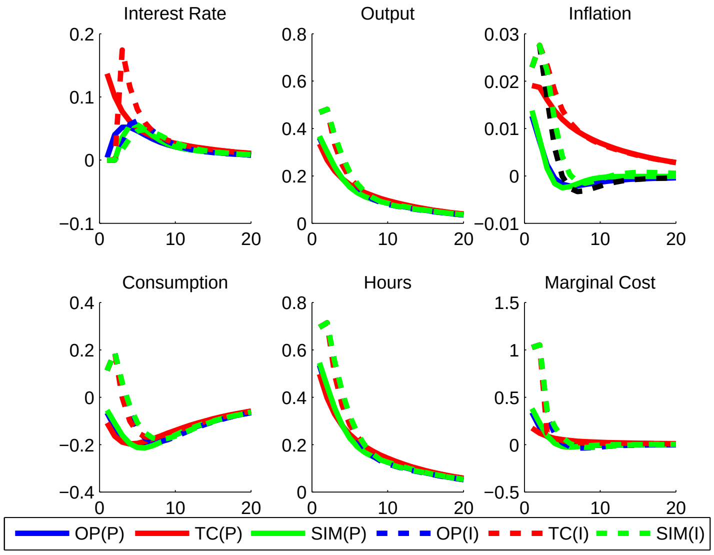
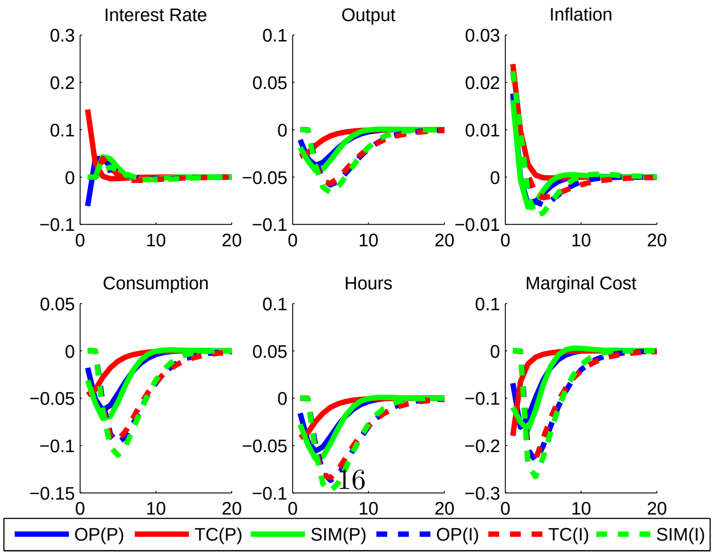


Figure 3: IRFs with Persistent Mark-up Shock



To gain further insights into our results we compare the impulse response functions for the NK model under the optimal, time-consistent and optimized simple rules.[21] We consider two information assumptions: PI and II with $j = 2$. Figures $1 - 3$ display the impulse responses to three shocks, technology, technology and the persistent component of the mark-up shock.

Under PI we see the familiar responses in a NK model. For a technology shock output immediately rises and, inflation falls. The optimal policy is to raise the interest rate a little initially to contain inflation, but then to commit to a sharp monetary relaxation before gradually returning to the steady state. Both consumption and leisure rise (the latter a familiar result in the NK literature) and hours fall. The productivity shocks results in a fall in the marginal cost, which is why inflation falls in the first place. The $\sqcup$-shaped interest rate path is time-inconsistent. Only an increasing interest rate path after the initial fall will be time-consistent; regime TC sees this happening with a larger drop in both the interest rate and inflation. Real variables - output, hours and consumption differ little between OP and TC for all shocks which explains the small welfare differences for all shocks combined.

Under II with two lags the interest rate only responds when information is received. At the micro-level firms respond to the shock but with the delayed drop in the nominal interest rate consumers save more and consume less, demand falls and output initially falls as well. The main impact of the productivity shock is now a larger and more prolonged fall in inflation because of the delay in the interest rate response. There is also a sharp fall in the real wage adding to the fall in the marginal cost. With II we see endogenous persistence arising from the rational learning of the private sector about the unobserved shock using Kalman updating. Output, inflation, consumption, hours and marginal cost all exhibit *hump-shaped* responses, a feature stressed in the II literature (see, for example, Collard *et al.* (2009) and Levine *et al.* (2012) among others cited in the introduction).

The mark-up shock is similar to the technology shock but with opposite effects; only the qualitative response of hours differ. The government spending shock however provides more interesting results. Under PI an increase in demand acts as a fiscal stimulus - in fact with $\frac{G}{Y} = 0.4$ in the steady state the impact multiplier is over unity in our estimated model and almost identical across all policy regimes. [22] Inflation also rises which elicits an interest rate rise, again for all regimes. The increase in government spending is financed by non-distortionary tax; in anticipation of this households save more and consume less. The real wage and therefore marginal costs rise, the marginal utility of consumption rises and there is a switch away from leisure (hours increase). Under II there is a delayed upward response of the interest rate to the inflation response. The demand increase is therefore greater, the fiscal multiplier reaches almost 2 on impact and the real wage, marginal cost and inflation increase by more. Now both leisure and consumption increase on impact and the crowding out of consumption is delayed for around 5 quarters.

To summarize, although the welfare effects of II are modest in consumption equivalent terms we see significant differences in impulse responses with II bringing about hump-shaped reactions to shocks. However Table 4 indicates the aggressive nature of these rules leads

[21] Only the simple rule of type A is shown - type B is very similar.

[22] $\frac{\Delta Y_t}{\Delta G_t} = \frac{Y_t}{G_t} \times$ irf, but note that 'government spending' consists of all non-consumption demand in our model.

Information	Information Set	Optimal	Time Cons	Simple Rule A	Simple Rule B
Perfect	Full state vector	0.235	0.669	0.134	0.134
Imperfect	$I_t = [y_t, \pi_t, r_t]$	0.200	0.729	0.165	0.166
Imperfect	$I_t = [y_{t-1}, \pi_{t-1}, r_t]$	0.117	0.364	0.118	0.121
Imperfect	$I_t = [y_{t-2}, \pi_{t-2}, r_t]$	0.118	0.366	0.116	0.123

Table 4: Interest Rate Variances

to high interest rate variances resulting in a ZLB problem for all the rules and information sets. From Table 4 with our zero-inflation steady state and nominal interest rate of 1% per quarter, optimal policy variances between 0.118 and 0.235 of a normally distributed variable imply a probability per quarter of hitting the ZLB in the range $[0.004, 0.04]$. Probabilities for the optimized simple rules are within this range whilst for the time consistent policy these rise to a range $[0.05, 0.11]$. At the upper end of these ranges the ZLB would be hit almost once every two years. In the next section we address this issue.

6.2 Imposing an Interest Rate Zero Lower Bound Constraint

In the absence of a lower bound constraint on the nominal interest rate the policymaker's optimization problem is to minimize Ω_0 given by (38) subject to (35) and (36) and given z_0. If the variances of shocks are sufficiently large, this will lead to a large nominal interest rate variability and the possibility of the nominal interest rate becoming negative.

We can impose a lower bound effect on the nominal interest rate by modifying the discounted quadratic loss criterion as follows.[23] Consider first the ZLB constraint on the nominal on the nominal interest rate. Rather than requiring that $R_t \geq 0$ for *any* realization of shocks, we impose the constraint that the mean rate should at least k standard deviation above the ZLB. For analytical convenience we use discounted averages.

Define $\bar{R} \equiv E_0\left[(1-\beta)\sum_{t=0}^{\infty}\beta^t R_t\right]$ to be the discounted future average of the nominal interest rate path $\{R_t\}$. Our 'approximate form' of the ZLB constraint is a requirement that $\bar{R}$ is at least k_r standard deviations above the zero lower bound; i.e., using discounted averages that

$$\bar{R} \geq k\sqrt{\overline{(R_t - \bar{R})^2}} = k\sqrt{\overline{R_t^2} - (\bar{R})^2} \tag{42}$$

Squaring both sides of (42) we arrive at

$$E_0\left[(1-\beta)\sum_{t=0}^{\infty}\beta^t R_t^2\right] \quad \leq \quad K\left[E_0\left[(1-\beta)\sum_{t=0}^{\infty}\beta^t R_t\right]\right]^2 \tag{43}$$

where $K = 1 + k^{-2} > 1$

We now maximize $\sum_{t=0}^{\infty}\beta^t[U(X_{t-1}, W_t)$ subject to the additional constraint (43) alongside the other dynamic constraints in the Ramsey problem. Using the Kuhn-Tucker theorem this results in an additional term $w_r\left(\overline{R^2} - K(\bar{R})^2\right)$ in the Lagrangian to incorporate this

[23] This follow the treatment of the ZLB in Woodford (2003) and Levine *et al.* (2008b)

extra constraint, where $w_r > 0$ is a Lagrangian multiplier. From the first order conditions for this modified problem this is equivalent to adding terms $E_0(1-\beta)\sum_{t=0}^{\infty}\beta^t w_r(R_t^2 - 2K\bar{R}R_t)$ where $\bar{R} > 0$ is evaluated at the constrained optimum. It follows that the effect of the extra constraint is to follow the same optimization as before, except that the single period loss function terms of in log-linearized variables is replaced with

$$L_t = \mathrm{y}_t^T Q \mathrm{y}_t + w_r(r_t - r^*)^2 \tag{44}$$

where $r^* = (K-1)\bar{R} > 0$ is a nominal interest rate target for the constrained problem.

In what follows, we linearize around a zero-inflation steady state. With a ZLB constraint, the policymaker's optimization problem is now to choose an unconditional distribution for r_t, shifted to the right by an amount r^*, about a new *positive* steady-state inflation rate, such that the probability of the interest rate hitting the lower bound is extremely low. This is implemented by choosing the weight w_r *for each of our policy rules* so that $z_0(p)\sigma_r < R^*$ where $z_0(p)$ is the critical value of a standard normally distributed variable Z such that prob $(Z \leq z_0) = p$, $R^* = (1+\pi^*)R + \pi^*$ is the steady state nominal interest rate, R is the shifted steady state real interest rate, $\sigma_r^2 = \mathrm{var}(R)$ is the unconditional variance and π^* is the new steady state positive net inflation rate. Given σ_r the steady state positive inflation rate that will ensure $R_t \geq 0$ with probability $1-p$ is given by

$$\pi^* = \max\left[\frac{z_0(p)\sigma_r - R + 1}{R} \times 100, 0\right] \tag{45}$$

In our linear-quadratic framework we can write the intertemporal expected welfare loss at time $t = 0$ as the sum of stochastic and deterministic components, $\Omega_0 = \tilde{\Omega}_0 + \bar{\Omega}_0$. By increasing w_r we can lower σ_r thereby decreasing π^* and reducing the deterministic component, but at the expense of increasing the stochastic component of the welfare loss. By exploiting this trade-off, we then arrive at the optimal policy that, in the vicinity of the steady state, imposes a ZLB constraint, $r_t \geq 0$ with probability $1-p$. Figure $4-6$ shows this solution to the problem for all three policy regimes and PI with $p = 0.0025$; ie., a very stringent ZLB requirement that the probability of hitting the zero lower bound is only once every 400 quarters or 100 years.

Note that in our LQ framework, the zero interest rate bound is very occasionally hit; then the interest rate is allowed to become negative, possibly using a scheme proposed by Gesell (1934) and Keynes (1936). Our approach to the ZLB constraint (following Woodford (2003))[24] in effect replaces it with a nominal interest rate variability constraint which ensures the ZLB is hardly ever hit. By contrast the work of a number of authors including Adam and Billi (2007), Coenen and Wieland (2003), Eggertsson and Woodford (2003) and Eggertsson (2006) study optimal monetary policy with commitment in the face of a non-linear constraint $R_t \geq 0$ which allows for frequent episodes of liquidity traps in the form of $R_t = 0$.

[24] As in Levine *et al.* (2008b), we generalize the treatment of Woodford however by allowing the steady-state inflation rate to rise. Our policy prescription has recently been described as a "dual mandate" in which a central bank committed to a long-run inflation objective sufficiently high to avoid the ZLB constraint as well as a Taylor-type policy stabilization rule about such a rate - see Blanchard *et al.* (2010) and Gavin and Keen (2011).

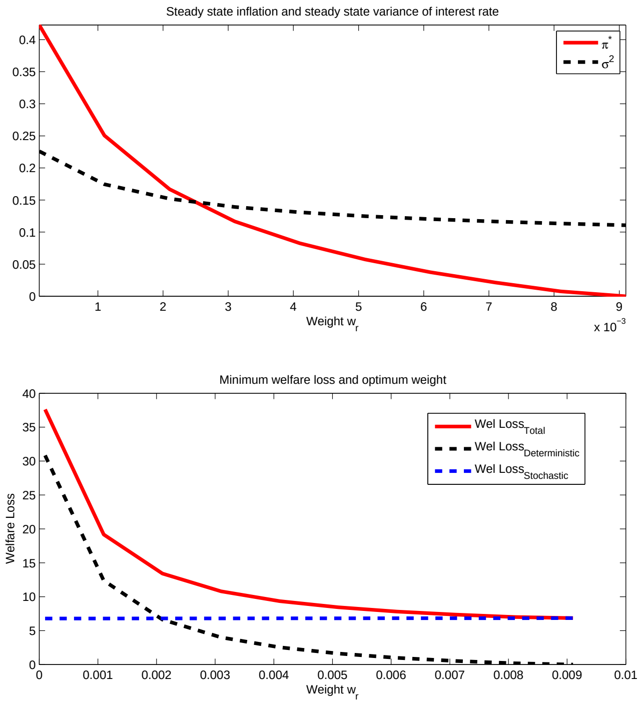


Figure 4: Imposition of ZLB for Optimal Policy and Perfect Information

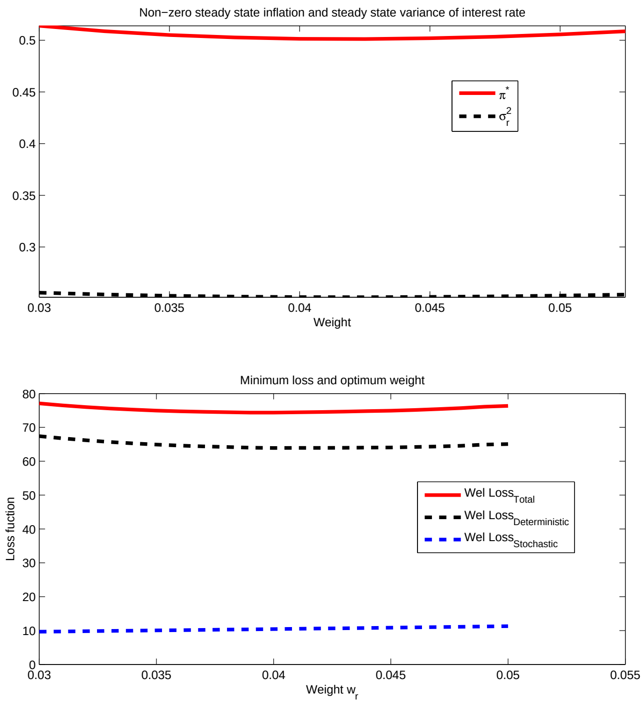


Figure 5: Imposition of ZLB for Time-Consistent Policy and Perfect Information

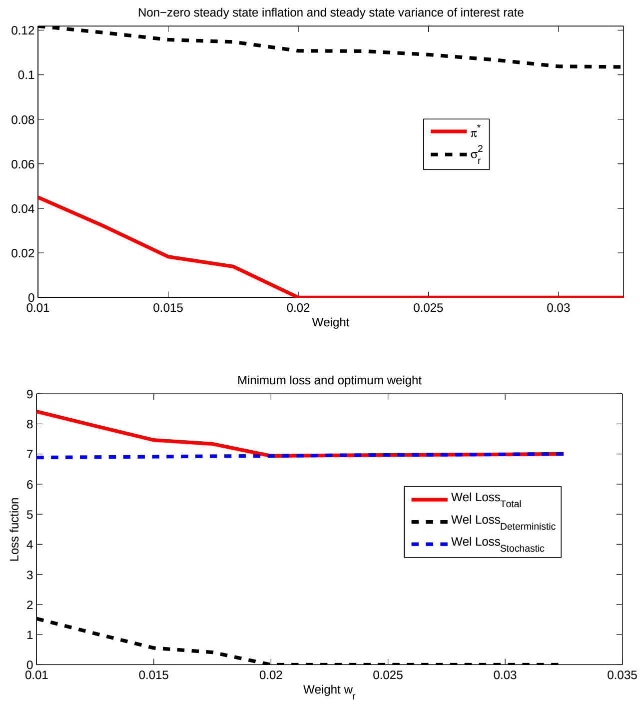


FIGURE 6: IMPOSITION OF ZLB FOR THE OPTIMIZED SIMPLE RULE AND PERFECT INFORMATION

Table 5 shows that introducing the ZLB constraint significantly changes the relative welfare performance of commitment, simple rules and the withdrawal of information. Now there are substantial gains from commitment of over $0.39 - 0.50\%$ consumption equivalent. Simple rules are still able to mimic their optimal counterpart. The form of the optimized simple rules is now a difference rule that is very close to a price level rule for all cases. Again the response to positive output deviations is slightly negative, offsetting the contractionary response to inflation. We also see a far less aggressive response of monetary policy to inflation that lowers the variance of the interest rate and prevents the ZLB problem seen previously.

The reason why the discretionary policy performs so badly with a ZLB constraint is that under discretion the policymaker lacks the leverage over private sector behaviour that is possible under commitment from say temporary loosening (or tightening) of monetary policy with promises to reverse this in the future. This in turn greatly inhibits the ability to reduce the unconditional variance of the nominal interest rate when it is penalized by an increasing size of the weight w_r. Consequently to achieve a low probability of hitting the ZLB one needs a larger shift of the nominal interest rate distribution to the right. Whereas under commitment $\pi^* = 0$, under discretion this rises to $\pi^* = 0.57 - 0.67\%$ or around 2.5% per year. Our ZLB constraint then results in a *long-run inflationary bias* in addition to the familiar *stabilization bias* highlighted by Currie and Levine (1993), Clarida *et al.* (1999) and others.

These results of imposing the ZLB are fairly uniform across all three information sets.

Information	Information Set	Optimal	Time Consis	Sim Rule A	Sim Rule B
Perfect (Wel Loss)	Full state vector	6.84 (0.003)	75.8 (0.39)	6.88 (0.006)	6.88 (0.006)
Perfect (Weight w_r)	Full state vector	0.009	0.032	0.10	0.10
Perfect (Inflation π^*)	Full state vector	0.00	0.62	0.00	0.00
Imperfect (Wel Loss)	$I_t = [y_t, \pi_t, r_t]$	7.61 (0.005)	95.8 (0.50)	7.71 (0.005)	7.71 (0.005)
Imperfect ((Weight w_r)	$I_t = [y_t, \pi_t, r_t]$	0.018	0.0375	0.03	0.03
Imperfect (Inflation π^*)	$I_t = [y_t, \pi_t, r_t]$	0.00	0.67	0.00	0.00
Imperfect (Wel Loss)	$I_t = [y_{t-1}, \pi_{t-1}, r_t]$	8.64 (0.01)	71.8 (0.39)	8.64 (0.01)	8.73 (0.01)
Imperfect (Weight w_r)	$I_t = [y_{t-1}, \pi_{t-1}, r_t]$	0.002	0.275	0.003	0.0003
Imperfect (Inflation π^*)	$I_t = [y_{t-1}, \pi_{t-1}, r_t]$	0.00	0.58	0.00	0.00
Imperfect (Wel Loss)	$I_t = [y_{t-2}, \pi_{t-2}, r_t]$	9.43 (0.02)	69.1 (0.35)	9.43 (0.02)	9.52 (0.02)
Imperfect (Weight w_r)	$I_t = [y_{t-2}, \pi_{t-2}, r_t]$	0.003	0.030	0.002	0.005
Imperfect (Inflation π^*)	$I_t = [y_{t-2}, \pi_{t-2}, r_t]$	0.00	0.57	0.00	0.00

Table 5: Welfare Costs per period of Imperfect Information *with* ZLB Considerations. Consumption Equivalent Losses (%) in brackets. Prob of hitting ZLB=0.0025.

Information	Information Set	Simple Rule A $[\rho_r,\ \theta_\pi,\ \theta_y]$	Simple Rule B $[\rho_r,\ \theta_\pi,\ \theta_y]$
Perfect	Full state vector	[1.00, 0.417, −0.006]	[1.00, 0.417, −0.006]
Imperfect	$I_t = [y_t, \pi_t, r_t]$	[1.00, 0.397, −0.017]	[1.00, 0.397, −0.017]
Imperfect	$I_t = [y_{t-1}, \pi_{t-1}, r_t]$	[1.00, 0.370, −0.009]	[1.00, 0.256, −0.020]
Imperfect	$I_t = [y_{t-2}, \pi_{t-2}, r_t]$	[1.00, 0.335, −0.010]	[1.00, 0.170, −0.015]

Table 6: Optimized Coefficients in Simple Rules *with* ZLB Considerations

What then are the particular implications of II then? There are two results to highlight. First under commitment with both optimal policy and optimized rules, the welfare consequences of limiting information to lagged output and inflation is similar to before without ZLB considerations. But the combination of II and a lack of commitment can have particularly severe welfare implications. It should be noted that without commitment we are in a world of second-best and the withdrawal of information is not automatically welfare-reducing as it actually could *improve* the welfare outcome by the "tying one's hands" of the policymaker to respond to current information. However the delay in the response imposed by II could go the other way and in our estimated model this is precisely what happens as one proceeds from PI to II with no lags in available information. But then with such lags the tying one's hands effect dominates and the welfare loss from an inability to commit falls from $c_e = 0.5\%$ at its peak with no lags to $c_e = 0.35\%$ with two lags.

Finally in Figures 7 – 9 we examine the impulse responses with the ZLB constraint.

These are now about a non-zero inflation steady state for the time-consistent case, but apart from this feature they are similar to those obtained before. The most marked differences is the noticeable divergence between the OP and SIM regimes that we expect from the larger welfare difference reported in Table 5 for simple rule A and lag 2 ($c_e = 0.02\%$ with the ZLB compared with $c_e = 0.01\%$ without).

7 Conclusions

We believe this to be the first paper to examine optimal policy in an estimated DSGE NK model where informational consistency is applied at both the estimation and policy stages. Our main results can be summarized as follows. First, common to all information sets only with a ZLB constraint do we see substantial welfare gains from commitment. Second, optimized rules take the form of a price level rule, or something very close across all information cases. Third, the combination of limited information and a lack of commitment can be particulary serious for welfare. At the same time we find that II with lags introduces a 'tying ones hands' effect on the policymaker that improves welfare under discretion. Finally, the impulse response functions under our most extreme imperfect information assumption (output and inflation observed with a two-quarter delay) exhibit hump-shaped behaviour and the fiscal multiplier is significantly enhanced in this case.

There are a number of potential areas for future research. Our model is very basic with low costs of business cycle fluctuations in the absence of ZLB considerations. If anything we underestimate the costs of imperfect information and the importance of the ZLB. It seems therefore worthwhile to revisit the issues raised in the context of a richer DSGE model that includes capital, sticky wages, search-match labour market and financial frictions. A second avenue for research would be to extend the work to allow the policymaker to have *more* information than the private sector. This satisfies informational consistency and would allow the proper examination of the benefits or otherwise of transparency. A third research direction is the address the same the policy questions using other ways of modelling information limitations associated with the 'rational inattention' and 'sticky information' literatures (see, for example, Sims (2005), Adam (2007), Luo and Young (2009), Reis (2009) and Mackowiak and Wiederholt (2009)). The basic idea for the latter is that only a fraction of agents can update their information each period and, for the former, that agents process information subject to a constraint placing an upper bound on the information flow. Finally, we assume rational (model consistent) expectations. It would be of interest to combine some aspects of learning (for example about the policy rule) alongside model consistent expectations with II, as in Ellison and Pearlman (2011).

References

Adam, K. (2007). Optimal Monetary Policy with Imperfect Common Knowledge . *Journal of Monetary Economics*, **54**(2), 267–301.

Adam, K. and Billi, R. M. (2007). Discretionary Monetary Policy and the Zero Lower Bound on Nominal Interest Rates. *Journal of Monetary Economics*. Forthcoming.

Figure 7: IRFs with Technology Shock and ZLB

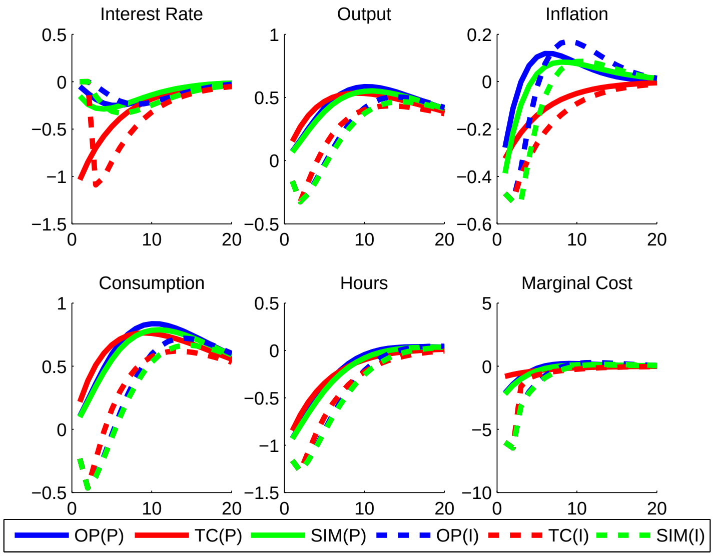


Figure 8: IRFs with Government Spending Shock and ZLB

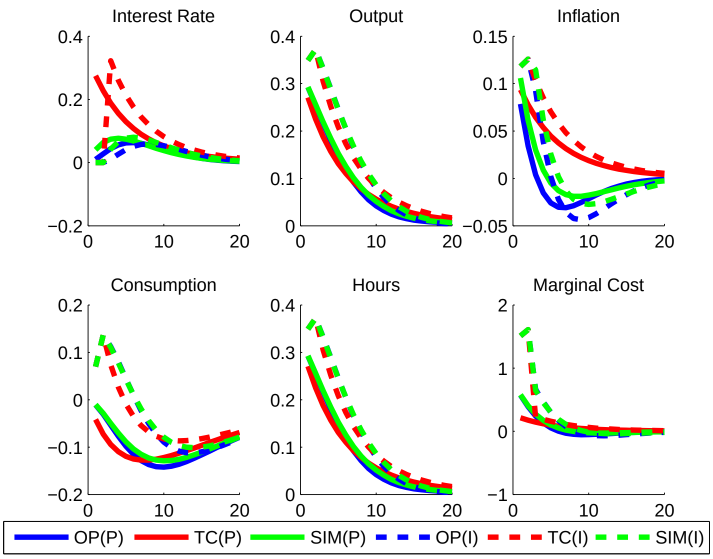
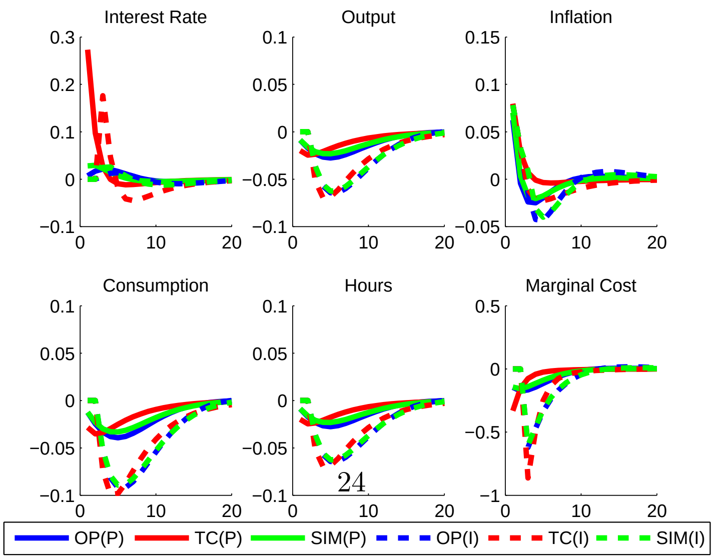


Figure 9: IRFs with Persistent Mark-up Shock and ZLB



Aoki, K. (2003). On the optimal monetary policy response to noisy indicators. *Journal of Monetary Economics*, **113**(3), 501 – 523.

Aoki, K. (2006). Optimal commitment policy under noisy information. *Journal of Economic Dynamics and Control*, **30**(1), 81 – 109.

Blanchard, O. and Kahn, C. (1980). The Solution of Linear Difference Models under Rational Expectations. *Econometrica*, **48**, 1305–1313.

Blanchard, O., Giovanni, D., and Mauro, P. (2010). Rethinking Macroeconomic Policy. IMF Staff Position Note, SPN/10/03 .

Brainard, W. (1967). Uncertainty and the effectiveness of policy. *American Economic Review*, **47**(2), 411–425.

Chang, Y., Gomes, J. F., and Schorfheide, F. (2002). Learning-by-Doing as a propagation Mechanism. *American Economic Review*, **92**(5), 1498–1520.

Clarida, R., Galí, J., and Gertler, M. (1999). The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature*, **37**(4), 1661–1707.

Coenen, G. and Wieland, V. (2003). The Zero-Interest Rate Bound and the Role of the Exchange Rate for Monetary Policy in Japan. *Journal of Monetary Economics*, **50**, 1071–1101.

Collard, F. and Dellas, H. (2004). The New Keynesian Model with Imperfect Information and Learning. *mimeo, CNRS-GREMAQ*.

Collard, F. and Dellas, H. (2006). Misperceived Money and Inflation Dynamics. *mimeo, CNRS-GREMAQ*.

Collard, F., Dellas, H., and Smets, F. (2009). Imperfect Information and the Business Cycle. *Journal of Monetary Economics*. Forthcoming.

Cukierman, A. and Meltzer, A. H. (1986). A theory of ambiguity, credibility and inflation under discretion and asymmetric information. *Econometrica*, **54**, 1099–1128.

Currie, D. and Levine, P. L. (1993). *Rules, Reputation and Macroeconomic Policy Coordination*. Cambridge University Press.

Eggertsson, G. (2006). The Deflation Bias and Committing to Being Irresponsible. *Journal of Money, Credit and Banking*, **36**(2), 283–322.

Eggertsson, G. and Woodford, M. (2003). The Zero Interest-Rate Bound and Optimal Monetary Policy. *Brooking Papers on Economic Activity*, **1**, 139–211.

Ellison, M. and Pearlman, J. (2011). Saddlepath learning. *Journal of Economic Theory*, **146**(4), 1500–1519.

Evans, G. W. and Honkapohja, S. (2009). Learning and Macroeconomics. *Annual Review of Economics*, **1**, 421–449.

Faust, J. and Svensson, L. (2001). Transparency and credibility: monetary policy with unobservable goals. *International Economic Review*, **42**, 369–397.

Faust, J. and Svensson, L. (2002). The equilibrium degree of transparency and control in monetary policy. *Journal of Money, Credit and Banking*, **34**(2), 520–539.

Fernandez-Villaverde, J. (2009). The Econometrics of DSGE Models. CEPR Discussion Paper No. 7159.

Fernandez-Villaverde, J., Rubio-Ramirez, J., Sargent, T., and Watson, M. W. (2007). ABC (and Ds) of Understanding VARs. *American Economic Review*, **97**(3), 1021–1026.

Gaspar, V., Smets, F., and Vestin, D. (2010). Is Time Ripe for Price Level Path Stability? In P. L. Siklos, M. T. Bohl, and M. E. Wohar, editors, *Challenges in central banking: the current institutional environment and forces affecting monetary policy*. Cambridge University Press.

Gavin, W. T. and Keen, B. D. (2011). The Zero Lower Bound and the Dual Mandate. Mimeo. Presented to the CEF 2011 Conference in San Francisco .

Gesell, S. (1934). *The Natural Economic Order*. Free-Economy Publishing Co., Phlip Pye, San Antonio.

Geweke, J. (1999). Computational Experiments and Reality. University of Minnesota and Federal Reserve Bank of Minneapolis.

Jeffries, H. (1996). *Theory of Probability*. Oxford: Clarendon Press. Third Edition.

Juillard, M. (2003). DYNARE: A Program for Solving Rational Expectations Models. CEPREMAP.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. Macmillan, New York.

Kimball, M. (1995). The Quantitative Analytics of the Basic Neomonetarist Model. *Journal of Monetary Economics*, **27**(4), 1241 – 1277. Part 2.

Koopman, S. J. and Durbin, J. (2003). Filtering and smoothing of state vector for diffuse state-space models. *Journal of Time Series Analysis*, **24**(1), 85–98.

Levin, A., Onatski, A., Williams, J. C., and Williams, N. (2006). Monetary Policy Under Uncertainty in Micro-Founded Macroeconomic Models. in M. Gertler and K. Rogoff (eds.), NBER Macroeconomics Annual, 2005, pp 229–387 .

Levine, P., Pearlman, J., and Perendia, G. (2007). Estimating DSGE Models under Partial Information. Department of Economics Discussion Papers 1607, Department of Economics, University of Surrey .

Levine, P., Pearlman, J., and Pierse, R. (2008a). Linear-Quadratic Approximation, Efficiency and Target-Implementability. *Journal of Economic Dynamics and Control*, **32**, 3315–3349.

Levine, P., McAdam, P., and Pearlman, J. (2008b). Quantifying and Sustaining Welfare Gains from Monetary Commitment. *Journal of Monetary Economics*, **55**(7), 1253–1276.

Levine, P., Pearlman, J., Perendia, G., and Yang, B. (2012). Endogenous Persistence in an Estimated DSGE Model under Imperfect Information. *Economic Journal*, **122**(565), 1287 – 1312.

Levine, P. L. and Pearlman, J. G. (2011). Computation of LQ Approximations to Optimal Policy Problems in Different Information Settings under Zero Lower Bound Constraints. Dynare Discussion Paper 10.

Lungu, L., Matthews, K., and Minford, A. (2008). Partial Current Information and Signal Extraction in a Rational Expectations Macroeconomic Model: A Computational Solution. *Economic Modelling*, **25**(2), 255–273.

Luo, Y. and Young, E. R. (2009). Rational Inattention and Aggregate Fluctuations. *The B.E. Journal of Macroeconomics*, **9**(1). Article 14.

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

Magill, M. (1977). A Local Analysis of Capital Accumulation under Uncertainty. *Journal of Economic Theory*, **15**(2), 211–219.

Melecky, M., Rodriguez Palenzuela, D., and Soderstrom, U. (2008). Inflation Target Transparency and the Macroeconomy. MPRA Paper No. 10545.

Minford, A. and Peel, D. (1983). Some Implications of Partial Information Sets in Macroeeconomic Models Embodying Rational Expectations. *Manchester School*, **51**, 235–249.

Pearlman, J. G. (1992). Reputational and Non-Reputational Policies with Partial Information. *Journal of Economic Dynamics and Control*, **16**, 339–357.

Pearlman, J. G., Currie, D., and Levine, P. (1986). Rational Expectations Models with Private Information. *Economic Modelling*, **3**(2), 90–105.

Reis, R. (2009). A Sticky Information General Equilibrium Model for Policy Analysis. NBER WP No. 14732.

Sims, C. (2005). Rational Inattention: A Research Agenda. Deutche Bundesbank, W.P. no. 34/2005.

Smets, F. and Wouters, R. (2003). An estimated Stochastic Dynamic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association*, **1(5)**, 1123–1175.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, **97**(3), 586–606.

Svensson, L. E. O. and Woodford, M. (2001). Indicator variables for Optimal Policy. *Journal of Monetary Economics*, **50**(3), 691–720.

Svensson, L. E. O. and Woodford, M. (2003). Indicator variables for Optimal Policy under Asymmetric Information. *Journal of Economic Dynamics and Control*, **28**(4), 661–680.

Woodford, M. (2003). *Foundations of a Theory of Monetary Policy*. Princeton University Press.

ONLINE APPENDICES

A Linearization of Model

The log-linearization[25] of the model about the non-stochastic steady state zero-growth[26], zero-inflation is given by

$$y_t = c_y c_t + (1-c_y)g_t \quad \text{where } c_y = \frac{C}{Y} \tag{A.1}$$

$$E_t mu^C_{t+1} = mu^C_t - (r_t - E_t\pi_{t+1}) \tag{A.2}$$

$$\pi_t = \frac{\beta}{1+\beta\gamma}E_t\pi_{t+1} + \frac{\gamma}{1+\beta\gamma}\pi_{t-1} + \frac{(1-\beta\xi)(1-\xi)}{(1+\beta\gamma)\xi}(mc_t + ms_t) \tag{A.3}$$

where marginal utilities, mu^C_t, mu^N_t, and marginal costs, mc_t, and output, y_t are defined by

$$mu^C_t = \frac{(1-\varrho)(1-\sigma)-1}{1-h_C}(c_t - h_C c_{t-1}) - \frac{\varrho(1-\sigma)N}{1-N}n_t \tag{A.4}$$

$$mu^N_t = \frac{1}{1-h_C}(c_t - h_C c_{t-1}) + \frac{N}{1-N}n_t + mu^C_t \tag{A.5}$$

$$w_t - p_t = mu^N_t - mu^C_t \tag{A.6}$$

$$mc_t = w_t - p_t - a_t + (1-\alpha)n_t \tag{A.7}$$

$$y_t = a_t + \alpha n_t \tag{A.8}$$

Equations (A.1) and (A.2) constitute the micro-founded 'IS Curve' and demand side for the model, given the monetary instrument. According to (A.2) solved forward in time, the marginal utility of consumption is the sum of all future expected real interest rates. (A.3) is the 'NK Philips Curve', the supply side of our model. In the absence of indexing it says that the inflation rate is the discounted sum of all future expected marginal costs. Note that price dispersion, Δ_t, disappears up to a first order approximation and therefore does not enter the linear dynamics. Finally, shock processes and the Taylor rule are given by

$$\begin{aligned}
g_{t+1} &= \rho_g g_t + \epsilon_{g,t+1} \\
a_{t+1} &= \rho_a a_t + \epsilon_{a,t+1} \\
msper_{t+1} &= \rho_{ms} msper_t + \epsilon_{msper,t+1} \\
ms_t &= msper_t + \epsilon_{mstra,t} \\
\pi_{tar,t+1} &= \rho_a \pi_{tar,t} + \epsilon_{tar,t+1} \\
r_t &= \rho_r r_{t-1} + (1-\rho_r)\theta(E_t\pi_{t+1} - \rho_{tar}\pi_{tar,t}) + \epsilon_{e,t}
\end{aligned}$$

$\epsilon_{e,t}$, $\epsilon_{a,t}$, $\epsilon_{g,t}$, $\epsilon_{msper,t}$, $\epsilon_{mstra,t}$ and $\epsilon_{tar,t}$ are i.i.d. with mean zero and variances $\sigma^2_{\epsilon_e}$, $\sigma^2_{\epsilon_a}$, $\sigma^2_{\epsilon_g}$, $\sigma^2_{\epsilon_{msper}}$, $\sigma^2_{\epsilon_{mstra}}$ and $\sigma^2_{\epsilon_{tra}}$ respectively.

Calibration

ϱ to target $N = 0.4$ given $\frac{C}{Y} = 0.6$ and the estimate of h_C, so $\frac{G}{Y} = 0.4$ with G_t including exogenous investment.

[25] Lower case variables are defined as $x_t = \log \frac{X_t}{X}$. r_t and π_t are log-deviations of *gross* rates. The validity of this log-linear procedure for general information sets is discussed in the online Appendix of Levine *et al.* (2012).

[26] With growth we simply replace β and h_C with $\beta_g \equiv \beta(1+g)^{(1-\varrho)(1-\sigma)-1}$ and $h_{Cg} = \frac{h_C}{1+g}$.

B Priors and Posterior Estimates

Parameter	Notation	Prior distribution		
		Density	Mean	
Risk aversion	σ	Normal	1.50	0.375
Price indexation	γ	Beta	0.50	0.15
Calvo prices	ξ	Beta	0.50	0.10
Consumption habit formation	h_C	Beta	0.50	0.10
Labour Share	α	Beta	0.70	0.10
Interest rate rule				
Inflation	θ_π	Normal	1.50	0.25
Output	θ_y	Normal	0.125	0.05
Interest rate smoothing	ρ_r	Beta	0.80	0.10
AR(1) coefficient				
Technology	ρ_a	Beta	0.85	0.10
Government spending	ρ_g	Beta	0.85	0.10
Price mark-up	ρ_{ms}	Beta	0.50	0.20
Inflation target	ρ_{tar}	Beta	0.85	0.10
Standard deviation of AR(1) innovations				
Technology	$sd(\epsilon_a)$	Inv. gamma	0.40	2.00
Government spending	$sd(\epsilon_g)$	Inv. gamma	1.50	2.00
Price mark-up	$sd(\epsilon_{ms})$	Inv. gamma	0.10	2.00
Inflation target	$sd(\epsilon_{tar})$	Inv. gamma	0.10	10.00
Standard deviation of I.I.D. shocks				
Mark-up process	$sd(\epsilon_m)$	Inv. gamma	0.10	2.00
Monetary policy	$sd(\epsilon_e)$	Inv. gamma	0.10	2.00

Table 7: Prior Distributions

Parameter	Information PI	Information II
σ	2.22 [1.66:2.79]	2.15[1.89:2.70]
γ	0.26 [0.08:0.43]	0.18 [0.08:0.31]
ξ	0.86 [0.80:0.93]	0.87 [0.84:0.91]
h_C	0.77 [0.63:0.91]	0.66 [0.54:0.84]
α	0.70 [0.56:0.85]	0.67 [0.58:0.80]
Interest rate rule		
θ_π	1.79 [1.40:2.18]	2.03 [1.65:2.26]
θ_y	0.15 [0.09:0.22]	0.13 [0.10:0.19]
ρ_r	0.65 [0.53:0.78]	0.63 [0.58:0.74]
AR(1) coefficient		
ρ_a	0.96 [0.93:0.99]	0.96 [0.94:0.98]
ρ_g	0.92 [0.88:0.95]	0.91 [0.89:0.94]
ρ_{ms}	0.27 [0.04:0.49]	0.16 [0.03:0.40]
ρ_{targ}	0.72 [0.55:0.91]	0.85 [0.71:0.92]
SD of AR(1) innovations		
$sd(\epsilon_a)$	0.43 [0.27:0.60]	0.51 [0.35:0.61]
$sd(\epsilon_g)$	1.89 [1.63:2.14]	1.99 [1.74:2.13]
$sd(\epsilon_{ms})$	0.05 [0.02:0.08]	0.05 [0.03:0.06]
$sd(\epsilon_{targ})$	0.28 [0.03:0.50]	0.11 [0.04:0.22]
SD of I.I.D. shocks		
$sd(\epsilon_m)$	0.10 [0.06:0.13]	0.09 [0.05:0.11]
$sd(\epsilon_e)$	0.12 [0.04:0.18]	0.18 [0.15:0.21]
Price contract length		
$\frac{1}{1-\xi}$	7.30	7.42
Log Marginal Likelihood (LL) and posterior model odd		
LL	-105.84	–102.36
Prob.	0.037	0.963

Table 8: Bayesian Posterior Distributions$^\diamond$

$^\diamond$ Notes: we report posterior means and 90% probability intervals (in parentheses) based on the output of the Metropolis-Hastings Algorithm. Sample range: 1981:I to 2006:IV.

C Optimal Policy Under Perfect Information

Under perfect information, $\begin{bmatrix} E_t\mathsf{z}_t \\ E_t\mathsf{x}_t \end{bmatrix} = \begin{bmatrix} \mathsf{z}_t \\ \mathsf{x}_t \end{bmatrix}$. Let $A \equiv A^1 + A^2$ and first consider the purely deterministic problem with a model then in state-space form:

$$\begin{bmatrix} \mathsf{z}_{t+1} \\ \mathsf{x}^e_{t+1,t} \end{bmatrix} = A \begin{bmatrix} \mathsf{z}_t \\ \mathsf{x}_t \end{bmatrix} + B\mathsf{w}_t \tag{C.1}$$

where z_t is an $(n-m)\times 1$ vector of predetermined variables including non-stationary processed, z_0 is given, w_t is a vector of policy variables, x_t is an $m \times 1$ vector of non-predetermined variables and $\mathsf{x}^e_{t+1,t}$ denotes rational (model consistent) expectations of x_{t+1} formed at time t. Then $\mathsf{x}^e_{t+1,t} = \mathsf{x}_{t+1}$

and letting $\mathsf{y}_t^T = \left[\mathsf{z}_t^T \ \mathsf{x}_t^T\right]$ (C.1) becomes

$$\mathsf{y}_{t+1} = A\mathsf{y}_t + B\mathsf{w}_t \tag{C.2}$$

The procedures for evaluating the three policy rules are outlined in the rest of this section (or Currie and Levine (1993) for a more detailed treatment).

C.1 The Optimal Policy with Commitment

Consider the policy-maker's *ex-ante* optimal policy at $t = 0$. This is found by minimizing Ω_0 given by (38) subject to (C.2) and (36) and given z_0. We proceed by defining the Hamiltonian

$$\mathcal{H}_t(y_t, y_{t+1}, \mu_{t+1}) = \frac{1}{2}\beta^t(\mathsf{y}_t^T Q\mathsf{y}_t + 2\mathsf{y}_t^T U\mathsf{w}_t + \mathsf{w}_t^T R\mathsf{w}_t) + \mu_{t+1}(A\mathsf{y}_t + B\mathsf{w}_t - \mathsf{y}_{t+1}) \tag{C.3}$$

where μ_t is a row vector of costate variables. By standard Lagrange multiplier theory we minimize

$$\mathcal{L}_0(y_0, y_1, \ldots, w_0, w_1, \ldots, \mu_1, \mu_2, \ldots) = \sum_{t=0}^{\infty} \mathcal{H}_t \tag{C.4}$$

with respect to the arguments of L_0 (except z_0 which is given). Then at the optimum, $\mathcal{L}_0 = \Omega_0$.

Redefining a new costate column vector $\mathsf{p}_t = \beta^{-t}\mu_t^T$, the first-order conditions lead to

$$\mathsf{w}_t = -R^{-1}(\beta B^T \mathsf{p}_{t+1} + U^T \mathsf{y}_t) \tag{C.5}$$

$$\beta A^T \mathsf{p}_{t+1} - \mathsf{p}_t = -(Q\mathsf{y}_t + U\mathsf{w}_t) \tag{C.6}$$

Substituting (C.5) into (C.6), we arrive at the following system under control

$$\begin{bmatrix} I & \beta BR^{-1}B^T \\ 0 & \beta(A^T - UR^{-1}B^T) \end{bmatrix} \begin{bmatrix} \mathsf{y}_{t+1} \\ \mathsf{p}_{t+1} \end{bmatrix} = \begin{bmatrix} A - BR^{-1}U^T & 0 \\ -(Q - UR^{-1}U^T & I \end{bmatrix} \begin{bmatrix} \mathsf{y}_t \\ \mathsf{p}_t \end{bmatrix} \tag{C.7}$$

To complete the solution we require $2n$ boundary conditions for (C.7). Specifying z_0 gives us $n - m$ of these conditions. The remaining condition is the 'transversality condition'

$$\lim_{t\to\infty} \mu_t^T = \lim_{t\to\infty} \beta^t \mathsf{p}_t = 0 \tag{C.8}$$

and the initial condition

$$\mathsf{p}_{20} = 0 \tag{C.9}$$

where $\mathsf{p}_t^T = \left[\mathsf{p}_{1t}^T \, \mathsf{p}_{2t}^T\right]$ is partitioned so that p_{1t} is of dimension $(n - m) \times 1$. Equation (36), (C.5), (C.7) together with the $2n$ boundary conditions constitute the system under optimal control.

Solving the system under control leads to the following rule

$$\mathsf{w}_t = -F \begin{bmatrix} I & 0 \\ -N_{21} & -N_{22} \end{bmatrix} \begin{bmatrix} \mathsf{z}_t \\ \mathsf{p}_{2t} \end{bmatrix} \equiv D \begin{bmatrix} \mathsf{z}_t \\ \mathsf{p}_{2t} \end{bmatrix} = -F \begin{bmatrix} \mathsf{z}_t \\ \mathsf{x}_{2t} \end{bmatrix} \tag{C.10}$$

where

$$\begin{bmatrix} \mathsf{z}_{t+1} \\ \mathsf{p}_{2t+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ S_{21} & S_{22} \end{bmatrix} G \begin{bmatrix} I & 0 \\ -N_{21} & -N_{22} \end{bmatrix} \begin{bmatrix} \mathsf{z}_t \\ \mathsf{p}_{2t} \end{bmatrix} \equiv H \begin{bmatrix} \mathsf{z}_t \\ \mathsf{p}_{2t} \end{bmatrix} \tag{C.11}$$

$$N = \begin{bmatrix} S_{11} - S_{12}S_{22}^{-1}S_{21} & S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21} & S_{22}^{-1} \end{bmatrix} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \tag{C.12}$$

$$\mathsf{x}_t = -\left[\begin{array}{cc} N_{21} & N_{22} \end{array}\right]\left[\begin{array}{c} \mathsf{z}_t \\ \mathsf{p}_{2t} \end{array}\right] \tag{C.13}$$

where $F = -(R + B^T SB)^{-1}(B^T S^{OPT} A + U^T), G = A - BF$ and

$$S = \left[\begin{array}{cc} S_{11} & S_{12} \\ S_{21} & S_{22} \end{array}\right] \tag{C.14}$$

partitioned so that S_{11} is $(n-m)\times(n-m)$ and S_{22} is $m \times m$ is the solution to the steady-state Riccati equation

$$S = Q - UF - F^T U^T + F^T RF + \beta(A - BF)^T S(A - BF) \tag{C.15}$$

The welfare loss for the optimal policy (OPT) at time t is

$$\Omega_t^{OPT} = -\frac{1}{2}(\mathrm{tr}(N_{11}Z_t) + \mathrm{tr}(N_{22}\mathsf{p}_{2t}\mathsf{p}_{2t}^T)) \tag{C.16}$$

where $Z_t = \mathsf{z}_t\mathsf{z}_t^T$. To achieve optimality the policy-maker sets $\mathsf{p}_{20} = 0$ at time $t = 0$.[27] At time $t > 0$ there exists a gain from reneging by resetting $\mathsf{p}_{2t} = 0$. It can be shown that $N_{11} < 0$ and $N_{22} < 0$.[28], so the incentive to renege exists at all points along the trajectory of the optimal policy. This is the time-inconsistency problem.

C.2 The Dynamic Programming Discretionary Policy

To evaluate the discretionary (time-consistent) policy we rewrite the welfare loss Ω_t given by (38) as

$$\Omega_t = \frac{1}{2}[\mathsf{y}_t^T Q\mathsf{y}_t + 2\mathsf{y}_t^T U\mathsf{w}_t + \mathsf{w}_t^T R\mathsf{w}_t + \beta\Omega_{t+1}] \tag{C.17}$$

The dynamic programming solution then seeks a stationary solution of the form $\mathsf{w}_t = -\mathsf{F}\mathsf{z}_t$ in which Ω_t is minimized at time t subject to (1) in the knowledge that a similar procedure will be used to minimize Ω_{t+1} at time $t+1$.

Suppose that the policy-maker at time t expects a private-sector response from $t+1$ onwards, determined by subsequent re-optimization, of the form

$$\mathsf{x}_{t+\tau} = -N_{t+1}\mathsf{z}_{t+\tau},\ \tau \geq 1 \tag{C.18}$$

The loss at time t for the *ex ante* optimal policy was from (C.16) found to be a quadratic function of x_t and p_{2t}. We have seen that the inclusion of p_{2t} was the source of the time inconsistency in that case. We therefore seek a lower-order controller

$$\mathsf{w}_t = -F\mathsf{z}_t \tag{C.19}$$

with the welfare loss in z_t only. We then write $\Omega_{t+1} = \frac{1}{2}\mathsf{z}_{t+1}^T S_{t+1}^{TCT}\mathsf{z}_{t+1}$ in (C.17). This leads to the following iterative process for F_t

$$\mathsf{w}_t = -F_t\mathsf{z}_t \tag{C.20}$$

[27] Noting from (C.13) that for the optimal policy we have $\mathsf{x}_t = -N_{21}\mathsf{z}_t - N_{22}\mathsf{p}_{2t}$, the optimal policy "from a timeless perspective" proposed by Woodford (2003) replaces the initial condition for optimality $p_{20} = 0$ with $J\mathsf{x}_0 = -N_{21}\mathsf{z}_0 - N_{22}\mathsf{p}_{20}$ where J is some $1 \times m$ matrix. Typically in New Keynesian models the particular choice of condition is $\pi_0 = 0$ thus avoiding any once-and-for-all initial surprise inflation. This initial condition applies only at $t = 0$ and only affects the deterministic component of policy and not the stochastic, stabilization component.

[28] See Currie and Levine (1993), chapter 5.

where

$$
\begin{aligned}
F_t &= (\overline{R}_t + \lambda \overline{B}_t^T S_{t+1}^{TCT} \overline{B}_t)^{-1} (\overline{U}_t^T + \beta \overline{B}_t^T S_{t+1}^{TCT} \overline{A}_t) \\
\overline{R}_t &= R + K_t^T Q_{22} K_t + U^{2T} K_t + K_t^T U^2 \\
K_t &= -(A_{22} + N_{t+1} A_{12})^{-1} (N_{t+1} B^1 + B^2) \\
\overline{B}_t &= B^1 + A_{12} K_t \\
\overline{U}_t &= U^1 + Q_{12} K_t + J_t^T U^2 + J_t^T Q_{22} J_t \\
\overline{J}_t &= -(A_{22} + N_{t+1} A_{12})^{-1} (N_{t+1} A_{11} + A_{12})
\end{aligned}
$$

$$
\begin{aligned}
\overline{A}_t &= A_{11} + A_{12} J_t \\
S_t^{TCT} &= \overline{Q}_t - \overline{U}_t F_t - F_t^T \overline{U}^T + \overline{F}_t^T \overline{R}_t F_t + \beta (\overline{A}_t - \overline{B}_t F_t)^T S_{t+1}^{TCT} (\overline{A}_t - \overline{B}_t \overline{F}_t) \\
\overline{Q}_t &= Q_{11} + J_t^T Q_{21} + Q_{12} J_t + J_t^T Q_{22} J_t \\
N_t &= -J_t + K_t F_t
\end{aligned}
$$

where $B = \begin{bmatrix} B^1 \\ B^2 \end{bmatrix}$, $U = \begin{bmatrix} U^1 \\ U^2 \end{bmatrix}$, $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, and Q similarly are partitioned conformably with the predetermined and non-predetermined components of the state vector.

The sequence above describes an iterative process for F_t, N_t, and S_t^{TCT} starting with some initial values for N_t and S_t^{TCT}. If the process converges to stationary values, F, N and S say, then the time-consistent feedback rule is $\mathsf{w}_t = -F\mathsf{z}_t$ with loss at time t given by

$$
\Omega_t^{TCT} = \frac{1}{2} \mathsf{z}_t^T S^{TCT} \mathsf{z}_t = \frac{1}{2} \mathrm{tr}(S^{TCT} Z_t) \tag{C.21}
$$

C.3 Optimized Simple Rules

We now consider simple sub-optimal rules of the form

$$
\mathsf{w}_t = D\mathsf{y}_t = D \begin{bmatrix} \mathsf{z}_t \\ \mathsf{x}_t \end{bmatrix} \tag{C.22}
$$

where D is constrained to be sparse in some specified way. Rule (C.22) can be quite general. By augmenting the state vector in an appropriate way it can represent a PID (proportional-integral-derivative)controller.

Substituting (C.22) into (38) gives

$$
\Omega_t = \frac{1}{2} \sum_{i=0}^{\infty} \beta^t \mathsf{y}_{t+i}^T P_{t+i} \mathsf{y}_{t+i} \tag{C.23}
$$

where $P = Q + UD + D^T U^T + D^T RD$. The system under control (C.1), with w_t given by (C.22), has a rational expectations solution with $\mathsf{x}_t = -N\mathsf{z}_t$ where $N = N(D)$. Hence

$$
\mathsf{y}_t^T P \mathsf{y}_t = \mathsf{z}_t^T T \mathsf{z}_t \tag{C.24}
$$

where $T = P_{11} - N^T P_{21} - P_{12} N + N^T P_{22} N$, P is partitioned as for S in (C.14) onwards and

$$
\mathsf{z}_{t+1} = (G_{11} - G_{12} N) \mathsf{z}_t \tag{C.25}
$$

where $G = A + BD$ is partitioned as for P. Solving (C.25) we have

$$\mathsf{z}_t = (G_{11} - G_{12}N)^t \mathsf{z}_0 \tag{C.26}$$

Hence from (C.27), (C.24) and (C.26) we may write at time t

$$\Omega_t^{SIM} = \frac{1}{2} z_t^T V z_t = \frac{1}{2} \mathrm{tr}(V Z_t) \tag{C.27}$$

where $Z_t = z_t z_t^T$ and V^{LYA} satisfies the *Lyapunov* equation

$$V^{LYA} = T + H^T V^{LYA} H \tag{C.28}$$

where $H = G_{11} - G_{12}N$. At time $t = 0$ the optimized simple rule is then found by minimizing Ω_0 given by (C.27) with respect to the non-zero elements of D given z_0 using a standard numerical technique. An important feature of the result is that unlike the previous solution the optimal value of D, D^* say, is not independent of z_0. That is to say

$$D^* = D^*(z_0)$$

C.4 The Stochastic Case

Consider the stochastic generalization of (C.1)

$$\begin{bmatrix} \mathsf{z}_{t+1} \\ \mathsf{x}^e_{t+1,t} \end{bmatrix} = A \begin{bmatrix} \mathsf{z}_t \\ \mathsf{x}_t \end{bmatrix} + B\mathsf{w}_t + \begin{bmatrix} \mathsf{u}_t \\ 0 \end{bmatrix} \tag{C.29}$$

where u_t is an $n \times 1$ vector of white noise disturbances independently distributed with $\mathrm{cov}(\mathsf{u}_t) = \Sigma$. Then, it can be shown that certainty equivalence applies to all the policy rules apart from the simple rules (see Currie and Levine (1993)). The expected loss at time t is as before with quadratic terms of the form $\mathsf{z}_t^T X \mathsf{z}_t = \mathrm{tr}(X\mathsf{z}_t, Z_t^T)$ replaced with

$$E_t \left(\mathrm{tr} \left[X \left(z_t z_t^T + \sum_{i=1}^{\infty} \beta^t \mathsf{u}_{t+i} \mathsf{u}_{t+i}^T \right) \right] \right) = \mathrm{tr} \left[X \left(z_t^T z_t + \frac{\lambda}{1-\lambda} \Sigma \right) \right] \tag{C.30}$$

where E_t is the expectations operator with expectations formed at time t.

Thus for the optimal policy with commitment (C.16) becomes in the stochastic case

$$\Omega_t^{OPT} = -\frac{1}{2} \mathrm{tr} \left(N_{11} \left(Z_t + \frac{\beta}{1-\beta} \Sigma \right) + N_{22} \mathsf{p}_{2t} \mathsf{p}_{2t}^T \right) \tag{C.31}$$

For the time-consistent policy (C.21) becomes

$$\Omega_t^{TCT} = -\frac{1}{2} \mathrm{tr} \left(S \left(Z_t + \frac{\beta}{1-\beta} \Sigma \right) \right) \tag{C.32}$$

and for the simple rule, generalizing (C.27)

$$\Omega_t^{SIM} = -\frac{1}{2} \mathrm{tr} \left(V^{LYA} \left(Z_t + \frac{\beta}{1-\beta} \Sigma \right) \right) \tag{C.33}$$

The optimized simple rule is found at time $t = 0$ by minimizing Ω_0^{SIM} given by (C.33). Now we find that

$$D^* = D^* \left(\mathsf{z}_0 \mathsf{z}_0^T + \frac{\beta}{1-\beta} \Sigma \right) \tag{C.34}$$

or, in other words, the optimized rule depends both on the initial displacement z_0 and on the covariance matrix of disturbances Σ.

A very important feature of optimized simple rules is that unlike their optimal commitment or optimal discretionary counterparts they are *not certainty equivalent*. In fact if the rule is designed at time $t = 0$ then $D^* = f^* \left(Z_0 + \frac{\beta}{1-\beta} \Sigma \right)$ and so depends on the displacement z_0 at time $t = 0$ and on the covariance matrix of innovations $\Sigma = \text{cov}(\epsilon_t)$. From non-certainty equivalence it follows that if the simple rule were to be re-designed at ant time $t > 0$, since the re-optimized D^* will then depend on Z_t the new rule will differ from that at $t = 0$. This feature is true in models with or without rational forward-looking behaviour and it implies that *simple rules are time-inconsistent even in non-RE models.*

D Optimal Policy Under Imperfect Information

The proof of the theorem generalizes Pearlman (1992) slightly, in that it allows for cross-product terms in states and instruments i.e. matrix $U \neq 0$, and provides a more elegant proof. Pearlman (1992) shows that optimal policy is certainty equivalent in the sense that all the rules under imperfect information correspond to those under perfect information, but with $z_{t,t}$ and $x_{t,t}$ replacing z_t, x_t. In particular, for the fully optimal rule p_{2t} then depends only on past values $\{z_{s,s},\ x_{s,s} :\ s < t\}$, so that $p_{2t} = p_{2t,t} = p_{2t,t-1}$. The updating equation for $z_{t,t}$ is then derived as follows:

$$x_{t,t} + N_{21} z_{t,t} + N_{22} p_{2t} = 0 \qquad x_t - x_{t,t} = A_{22}^{-1} A_{21} (z_t - z_{t,t}) \tag{D.1}$$

where $N_{22} = 0$ for TCT and SIM, N_{21}, N_{22} were defined for OPT in (C.13) and N_{21} is dependent on which rule is in place; the second equation is obtained by taking time-t expectations of the equation involving $E_t x_{t+1}$ and subtracting from the original:

$$0 = A_{12}(\mathsf{z}_t - \mathsf{z}_{t,t}) + A_{22}(\mathsf{x}_t - \mathsf{x}_{t,t}) \tag{D.2}$$

After taking expectations of each of these at $t-1$, it then follows that we can write

$$m_t - m_{t,t-1} = D(z_t - z_{t,t-1}) + v_t + (E - D)(z_{t,t} - z_{t,t-1}) \tag{D.3}$$

using the definitions of D and E in Section 3.1. Now assume that

$$z_{t,t} - z_{t,t-1} = J_1 (D(z_t - z_{t,t-1}) + v_t) \tag{D.4}$$

which will be verified shortly. It then follows that

$$m_t - m_{t,t-1} = (I + (E - D) J_1)(D(z_t - z_{t,t-1}) + v_t) \tag{D.5}$$

and hence the updated value $z_{t,t}$ using the measurement m_t is given by

$$\begin{aligned} z_{t,t} - z_{t,t-1} &= PD^T (DPD^T + V)^{-1} (I + (E - D) J_1)^{-1} (m_t - m_{t,t-1}) \\ &= PD^T (DPD^T + V)^{-1} (D(z_t - z_{t,t-1}) + v_t) \end{aligned} \tag{D.6}$$

where the second equality is obtained by substituting from (D.5); hence $J_1 = PD^T (DPD^T + V)^{-1}$. Finally Pearlman (1992) shows that $E[(z_t - z_{t,t}) z_{s,s}] = 0,\ s \leq t$. This enables us to rewrite the welfare loss in the form of (D.8), and to obtain its value in (39) using (D.9), where P is the solution of the Riccati equation

$$P = APA^T - APD^T(DPD^T + V)^{-1}DPA^T + \Sigma \tag{D.7}$$

where $A = A_{11} - A_{12}A_{22}^{-1}A_{21}$. Note that this Riccati equation *is independent of policy.* We may then write the expected utility as

$$\frac{1}{2}E_t\Big[\sum_{i=0}^{\infty}\beta^t(\mathsf{y}_{t+\tau,t+\tau}^TQ\mathsf{y}_{t+\tau,t+\tau} \;+\; 2\mathsf{y}_{t+\tau,t+\tau}^TU\mathsf{w}_{t+\tau} \;+\; \mathsf{w}_{t+\tau}^TR\mathsf{w}_{t+\tau} \\ +\; (\mathsf{y}_{t+\tau} - \mathsf{y}_{t+\tau,t+\tau})^TQ(\mathsf{y}_{t+\tau} - \mathsf{y}_{t+\tau,t+\tau}))\Big] \tag{D.8}$$

where we note that $\mathsf{w}_{t+\tau}$ is dependent only on current and past $\mathsf{y}_{t+s,t+s}$. The discounted expected sum of the last term of (D.8) corresponds to the last term of (39) and is independent of policy, while the other terms are minimized subject to the *expected* dynamics

$$\begin{bmatrix} \mathsf{z}_{t+1,t+1} \\ E_t\mathsf{x}_{t+1,t+1} \end{bmatrix} = (A^1 + A^2)\begin{bmatrix} \mathsf{z}_{t,t} \\ \mathsf{x}_{t,t} \end{bmatrix} + B\mathsf{w}_t + \begin{bmatrix} \mathsf{z}_{t+1,t+1} - \mathsf{z}_{t+1,t} \\ 0 \end{bmatrix} \tag{D.9}$$

allowing the problem to be solved using the techniques of Appendix C. We note by the chain rule that $E_t\mathsf{x}_{t+1,t+1} \equiv E_t[E_{t+1}x_{t+1}] = E_tx_{t+1}$, and that $\mathrm{cov}(\mathsf{z}_{t+1,t+1} - \mathsf{z}_{t+1,t}) = PD^T(DPD^T + V)^{-1}DP$ and $\mathrm{cov}(\mathsf{z}_{t+1} - \mathsf{z}_{t+1,t+1}) = P - PD^T(DPD^T + V)^{-1}DP \equiv \bar{P}$. This implies that the the other terms of the welfare loss are as given in (39).

Furthermore, as in Pearlman (1992) we can show that certainty equivalence holds for both the fully optimal and the time consistent solutions (but not for optimized simple rules).

E The Hamiltonian Quadratic Approximation of Welfare

Consider the following general deterministic optimization problem

$$\max \sum_{t=0}^{\infty}\beta^tU(X_{t-1}, W_t) \;\; s.t. \;\; X_t = f(X_{t-1}, W_t) \tag{E.1}$$

where X_{t-1} is vector of state variables and W_{t-1} a vector of instruments.[29] There are given initial and the usual tranversality conditions. For our purposes, we consider this as including models with forward-looking expectations, so that the optimal solution to the latter setup is the pre-commitment solution. Suppose the solution converges to a steady state X, W as $t \to \infty$ for the states X_t and the policies W_t. Define $x_t = X_t - X$ and $w_t = W_t - W$ as representing the first-order approximation to absolute deviations of states and policies from their steady states.[30]

[29] An alternative representation of the problem is $U(X_t, W_t)$ and $E_t[X_{t+1}] = f(X_t, W_t)$ where X_t includes forward-looking non-predetermined variables and $E_t[X_{t+1}] = X_{t+1}$ for the deterministic problem where perfect foresight applies. Whichever one uses, it is easy to switch from one to the other by a simple re-definition. Note that Magill (1977) adopted a continuous-time model without forward-looking variables. As we demonstrate in Levine *et al.* (2008b), although the inclusion of forward-looking variables significantly alters the nature of the optimization problem, these changes only affect the boundary conditions and the second-order conditions, but not the steady state of the optimum which is all we require for LQ approximation.

[30] Alternatively $x_t = (X_t - X)/X$ and $w_t = (W_t - W)/W$, depending on the nature of the economic variable. Then the Theorem follows in a similar way with an appropriate adjustment to the Jacobian Matrix.

The Lagrangian for the problem is defined as,

$$\sum_{t=0}^{\infty} \beta^t [U(X_{t-1}, W_t) - \lambda_t^T (X_t - f(X_{t-1}, W_t))] \tag{E.2}$$

so that a necessary condition for the solution to (E.1) is that the Lagrangian is stationary at all $\{X_s\}$, $\{W_s\}$ i.e.

$$U_W + \lambda_t^T f_W = 0 \qquad U_X - \frac{1}{\beta}\lambda_{t-1}^T + \lambda_t^T f_X = 0 \tag{E.3}$$

Assume a steady state λ for the Lagrange multipliers exists as well. Now define the Hamiltonian $H_t = U(X_{t-1}, W_t) + \lambda^T f(X_{t-1}, W_t)$. The following is the discrete time version of Magill (1977):

Theorem: If a steady state solution (X, W, λ) to the optimization problem (E.1) exists, then any perturbation (x_t, w_t) about this steady state can be expressed as the solution to

$$\max \frac{1}{2}\sum_{t=0}^{\infty} \beta^t \left[\begin{array}{cc} x_{t-1} & w_t \end{array} \right] \left[\begin{array}{cc} H_{XX} & H_{XW} \\ H_{WX} & H_{WW} \end{array} \right] \left[\begin{array}{c} x_{t-1} \\ w_t \end{array} \right] \quad s.t. \; x_t = f_X x_{t-1} + f_W w_t \tag{E.4}$$

where H_{XX}, etc denote second-order derivatives evaluated at (X, W). This can be directly extended to the case incorporating disturbances.

Thus our general procedure is as follows:

1. Set out the deterministic non-linear problem for the Ramsey Problem, to maximize the representative agents' utility subject to non-linear dynamic constraints.
2. Write down the Lagrangian for the problem.
3. Calculate the first order conditions. We do not require the initial conditions for an optimum since we ultimately only need the steady-state of the Ramsey problem.
4. Calculate the steady state of the first-order conditions. The terminal condition implied by this procedure is such that the system converges to this steady state.
5. Calculate a second-order Taylor series approximation, about the steady state, of the Hamiltonian associated with the Lagrangian in 2.
6. Calculate a first-order Taylor series approximation, about the steady state, of the first-order conditions and the original constraints.
7. Use 4. to eliminate the steady-state Lagrangian multipliers in 5. By appropriate elimination both the Hamiltonian and the constraints can be expressed in minimal form. This then gives us the accurate LQ approximation of the original non-linear optimization problem in the form of a minimal linear state-space representation of the constraints and a quadratic form of the utility expressed in terms of the states.