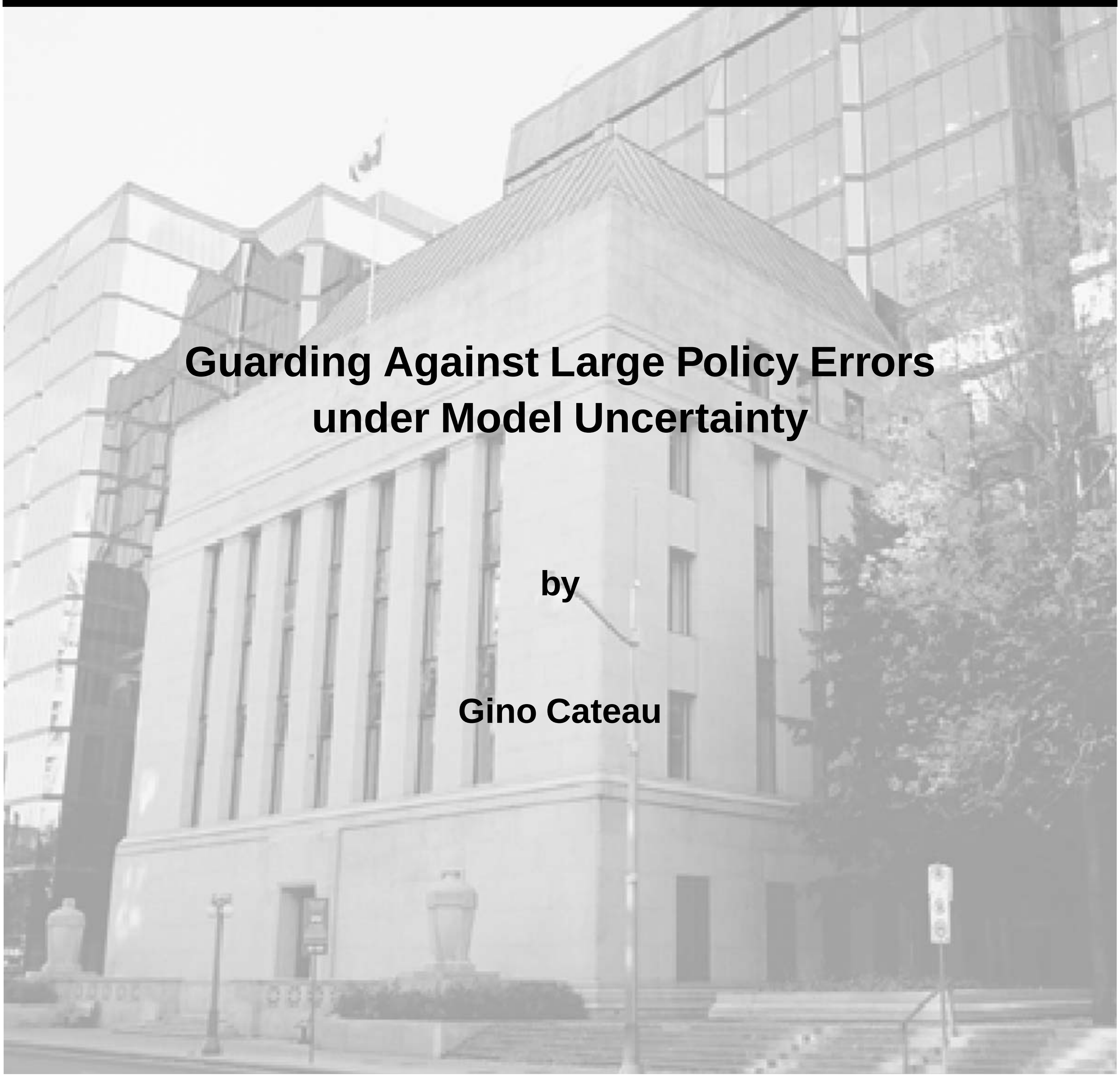
Bank of Canada Banque du Canada

Working Paper 2006-13 / Document de travail 2006-13



Guarding Against Large Policy Errors under Model Uncertainty

by

Gino Cateau

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2006-13

April 2006

Guarding Against Large Policy Errors under Model Uncertainty

by

Gino Cateau

Research Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
gcateau@bankofcanada.ca

Contents

Acknowledgements iv
Abstract/Résumé v

1. Introduction 1

2. Robust Control 3
   - 2.1 The intuition behind robust control 3
   - 2.2 Robust control à la Hansen-Sargent 4
   - 2.3 The sense in which robust control yields robust decision rules 6
   - 2.4 Choosing the degree of robustness, θ: detection-error probabilities 9

3. Implementing Robust Control 13
   - 3.1 The models 13
   - 3.2 Detection probabilities 15
   - 3.3 Effects of concerns for robustness on monetary policy 17

4. Multiple-Model Approaches 20
   - 4.1 Bayesian approach 21
   - 4.2 Worst-case approach 23
   - 4.3 Trade-off between average performance and robustness 24
   - 4.4 Trade-off between the CEE and FM model 31

5. Conclusion 33

References 36

Appendix A: Robustness in Forward-Looking Models 38

Appendix B: An Alternative Algorithm for Solving Forward-Looking Linear Quadratic Robust Control 48

Appendix C: The Kalman Filter 52

Appendix D: Impulse Responses from a VAR 54

Acknowledgements

I thank Don Coletti and Eva Ortega for encouragement in starting this project. I especially thank Amanda Armstrong and Wendy Chan for their excellent research assistance. The paper would have been far less complete without their help.

Abstract

How can policy-makers avoid large policy errors when they are uncertain about the true model of the economy? The author discusses some recent approaches that can be used for that purpose under two alternative scenarios: (i) the policy-maker has *one reference model* for choosing policy but cannot take a stand as to how that model is misspecified, and (ii) the policy-maker, being uncertain about the economy's true structure, entertains *multiple distinct models* of the economy. The author shows how these approaches can be implemented in practice using as benchmark models simplified versions of Fuhrer and Moore (1995) and Christiano, Eichenbaum, and Evans (2005).

JEL classification: E5, E58, D8, D81
Bank classification: Uncertainty and monetary policy

Résumé

Comment les autorités peuvent-elles éviter de grosses erreurs dans le choix de leur politique monétaire lorsqu'elles ne savent pas quel modèle de l'économie est le plus indiqué? L'auteur expose quelques-unes des méthodes avancées récemment à cette fin en distinguant deux cas de figure : i) les décideurs se réfèrent à *un seul modèle*, mais ignorent lesquels de ses éléments ont été mal spécifiés; ii) les décideurs retiennent *plusieurs modèles différents* parce qu'ils ne connaissent pas la véritable structure de l'économie. En s'appuyant sur des versions simplifiées des modèles proposés par Fuhrer et Moore (1995) et par Christiano, Eichenbaum et Evans (2005), l'auteur montre comment les méthodes considérées peuvent être mises concrètement à profit.

Classification JEL : E5, E58, D8, D81
Classification de la Banque : Incertitude et politique monétaire

1. Introduction

Suppose that, as a policy-maker at the Bank of Canada, you are presented with Figure 1. The solid line shows how inflation and the output gap respond to an unanticipated 1 per cent increase in the nominal interest rate according to a simplified version of the Fuhrer and Moore (1995) model (hereafter, FM model) estimated for Canada. The dotted line shows the corresponding impulse responses according to a simplified version of the Christiano, Eichenbaum, and Evans (2005) model (hereafter, CEE model)(see Dennis 2004), calibrated for the Canadian economy.

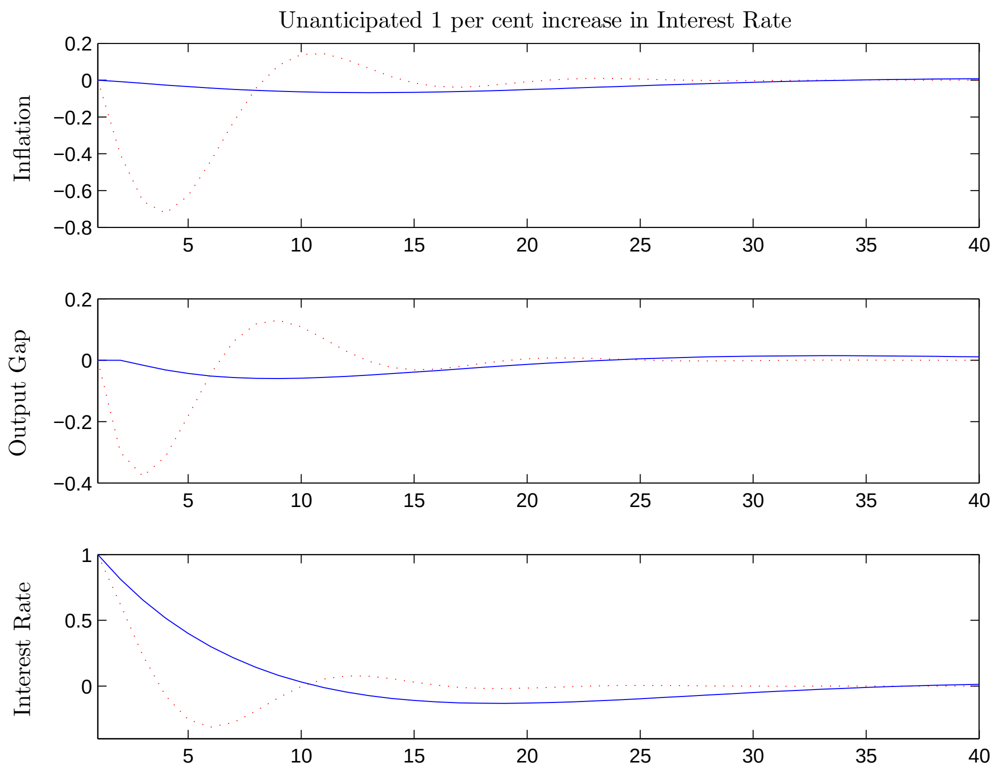


Figure 1: Different models imply differences in the policy transmission mechanism

Now suppose, for the sake of argument, that you are uncertain about which model is the most relevant for the Canadian economy, but you need to set monetary policy by appropriately choosing the nominal interest rate. Figure 1 illustrates that the decision problem in the face of this uncertainty is likely to be non-trivial. On the one hand, both models seem to be consistent with economic theory: an unanticipated increase in the interest rate has downward effects on inflation and the output gap. However, the models generate markedly different predictions as to how policy affects the policy-relevant variables, in terms of the magnitude of effects, the duration of the effects, and the dynamic path of those variables after the change in policy. And because they have different implications about the policy transmission mechanism, the models will, in general, lead to different policy implications. Indeed, a policy-maker concerned with controlling inflation, who believes that the Canadian economy is well represented by the CEE model, will presumably consider that increasing the interest rate is an effective way of combatting inflation; a 1 per cent increase in the interest rate reduces inflation by 0.72 per cent after 4 quarters, and inflation is stabilized after about 16 quarters. If a policy-maker, however, views the FM model as the reference model for Canada, then that policy-maker will presumably consider that the interest rate is not a very effective instrument for controlling inflation; a 1 per cent increase in the interest rate reduces inflation by 0.07 per cent after 14 quarters!

The root of the difficulty in choosing policy in the above example lies in the fact that the policy-maker entertains two models of the economy that predict fairly different effects of policy. In practice, it is not unusual to find different departments in central banks using different models to help their policy analysis. At the Bank of Canada, the Research Department uses TOTEM (Binette, Murchison, Perrier, and Rennison 2004) and policy analysis, while the Monetary and Financial Analysis Department is currently developing a model that emphasizes household sector financial frictions in the economy (Gammoudi and Mendes 2005). But even if the policy-maker uses one reference model for policy analysis, it does not mean that model uncertainty is not a concern. Indeed, any model is a simplification of a more complex reality. So, what if the reference model is misspecified? What if the model is built around an economic paradigm that is further than assumed from the economic reality? Or what if it ignores economic relationships that are in fact relevant?

In this paper, we discuss some recent approaches to dealing with model uncertainty where the policy-maker has one reference model for policy-making or has multiple reference models. We present the theoretical ideas behind each approach, use simple examples to illustrate what each approach does, and discuss when a particular approach may be preferable. But we do not limit ourselves to a theoretical discussion. We show how to sensibly determine the various parameters that a theoretical discussion takes for granted that are nonetheless essential ingredients in practical applications. We also work through the implementation using as a benchmark the simplified CEE and FM models (described in section 3.1).

The rest of this paper is organized as follows. Section 2 discusses the theory behind robust control and other concepts that are necessary for implementing robust control in practice. Section 3 implements robust control in the CEE and FM models. Section 4 describes various multiple-model approaches (Bayesian, worst-case, and trade-off (Cateau 2005)), and implements the methods when the policy-maker is uncertain between the CEE and FM models. Section 5 offers some conclusions.

2. Robust Control

2.1 The intuition behind robust control

Suppose, to start, that the policy-maker has one model of the economy. Suppose further that the policy-maker considers it a good *approximating* model of the economy, but feels that it may deviate from some unknown true model in possibly important ways. How should the policy-maker make decisions in such a situation? Robust control is designed to work well in this scenario.

The main feature of robust control is that it formally allows the decision-maker to recognize that data may not be generated by the approximating model but by an unknown member of a set of models *near* the approximating model. Robust control provides a way for the decision-maker to make decisions that would perform well over the set of nearby models. Therefore, robust control aims to yield decisions that would work reasonably well even if the approximating model does not coincide with the true data-generating model, as opposed to decisions that would be the best if they do coincide but possibly very bad if they do not.

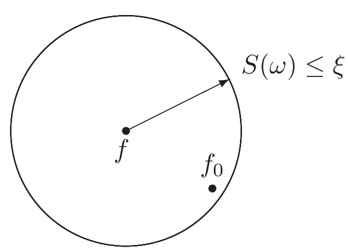


Figure 2: Robust decision making

In Figure 2, the decision-maker specifies a model, f, but suspects that the data are actually generated by a nearby model, f_0, which cannot be specified. In robust control, the decision-maker confronts the model uncertainty by first recognizing that there are specification errors (say, ω) and then seeking a decision rule that will work well, not only for the approximating model but for a set of models in the neighbourhood of f. To express the idea that f is a good approximating model, the neighbourhood of f for which the decision-maker wants a decision rule that works well is restricted to the set of models for which the size of the specification errors (say, $S(\omega)$) is bounded by a certain value, ξ. Thus, ξ represents how much the decision-maker believes that f is a good approximating model. A larger ξ means that the decision-maker believes the true model is further apart and therefore considers a wider set of models around f, and vice versa. Finally, to ensure that the decisions made perform well even when $f \neq f_0$, robust control instructs the decision-maker to make decisions according to the worst-case model in the set.

2.2 Robust control à la Hansen-Sargent

Since Hansen and Sargent (2004), the literature on robust control has expanded considerably, and there are now different ways of implementing robust control versions of decision-making problems. The discussion below follows the approach of Hansen and Sargent (2004) because it can be more easily implemented than the other approaches suggested in the literature (Giannoni 2002, Onatski and Williams 2003), is non-parametric, and is general.

Let X_t be a vector of state variables and U_t be the vector of controls to be chosen at time t, and let the policy-maker's model take the form of the linear transition law,

$$X_{t+1} = AX_t + BU_t + C\check{\epsilon}_{t+1}, \tag{1}$$

where $\{\check{\epsilon}_t\}$ is an identically, independently distributed (i.i.d.) shock process with mean 0 and identity covariance matrix. Suppose that the policy-maker regards model (1) as approximating another model that he cannot specify. How should the notion that model (1) is misspecified be represented? The i.i.d. process $\{\check{\epsilon}_t\}$, by definition, can represent only a very limited class of approximation errors and, in particular, it cannot represent misspecified dynamics, since it does not influence the conditional mean of the state. To represent dynamic misspecification, Hansen and Sargent (2004) suggest surrounding model (1) by a set of models of the form

$$X_{t+1} = AX_t + BU_t + C(\epsilon_{t+1} + \omega_{t+1}), \tag{2}$$

where $\{\epsilon_t\}$ is another i.i.d. shock process with mean 0 and identity covariance matrix, ω_{t+1}

is a vector process that can feed back on the history of X_t in a possibly non-linear way,

$$\omega_{t+1} = g_t(X_t, X_{t-1}, ...),$$

and g_t is a sequence of measurable functions. In what sense does augmenting model (1) to (2) allow for dynamic misspecification? When model (2) generates the data, the errors $\{\check{\epsilon}_t\}$ of model (1) have conditional mean ω_{t+1}, rather than 0. Thus, the idea that the approximating model is misspecified is captured by allowing the conditional mean of the shock process of model (2) that actually generates the data to feed back on the history of the state.

To express the idea that model (1) is a *good* approximation when model (2) generates the data, the misspecification errors must not be unbounded. Hansen and Sargent (2004) restrain the approximation errors by

$$E_0 \sum_{t=0}^{\infty} \beta^{t+1} \omega_{t+1}' \omega_{t+1} \leq \eta_0, \tag{3}$$

where E_t denotes mathematical expectation conditioned on $X^t = (X_t, X_{t-1}, X_{t-2}, ...)$ calculated with model (2).

The policy-maker believes that the data are generated by model (2) with some unknown process, ω_{t+1}, satisfying (3). The policy-maker's distrust of the approximating model leads the policy-maker to want decision rules that perform well over a set of models (2) satisfying (3). The robust decision rule is obtained by solving

$$\min_{\{U_t\}} \max_{\{\omega_{t+1}\}} E_0 \sum_{t=0}^{\infty} \beta^t \{X_t' Q X_t + U_t' R U_t\} \tag{4}$$

subject to (2), given (3) and X_0. Therefore, robust control involves switching from a typical minimization problem (the policy-maker minimizes the loss function) to an appropriately specified min-max problem.[1] The policy function that the policy-maker ultimately chooses is best understood as the equilibrium outcome of a two-player game: the policy-maker chooses the best possible policy, given that a fictitious evil agent whose purpose is to hurt the policy-maker as much as possible chooses the worst model from the possible set of models. The

[1] In Appendix A, I work out the solution to the above robust control problem. For generality, I assume that X_t includes both backward-looking state variables (i.e., states inherited from the past) and forward-looking variables (i.e., state variables not inherited from the past but that need to be determined in equilibrium). I work out the mechanics of the robust solution under two cases: (i) the policy-maker chooses the fully optimal policy rule (Ramsey solution), and (ii) the policy-maker chooses policy according to a simple rule.

evil agent is a metaphor for the policy-maker being cautious. The policy-maker's cautious behaviour implies that the policy-maker wants a policy function (i.e., a U_t) that would perform well should the worst possible model generate the data.

The robust control problem described above is set up as a *constraint game*. The constraint game directly constrains the distortions that the evil agent can make to the approximating model through the constraint (3). In practical applications, it is often more convenient to set up the robust control problem as a *multiplier game*:

$$\min_{\{U_t\}} \max_{\{\omega_{t+1}\}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ X_t' Q X_t + U_t' R U_t - \theta \beta \omega_{t+1}' \omega_{t+1} \right\}, \tag{5}$$

subject to (1), given X_0. Thus, the multiplier game *penalizes* the distortions that the evil agent can make. Hansen and Sargent (2004) show that the two formulations are equivalent under conditions that allow the Lagrange multiplier theorem to apply. Then θ can be interpreted as a Lagrange multiplier on the constraint measuring the size of the set of models. It is related to η_0 in (3). Hansen and Sargent (2004), in fact, show that there is a positive relationship between η_0 and θ^{-1}. Because $\theta = \infty$ corresponds to $\eta_0 = 0$, θ corresponds to the case where the set of models collapses to the approximating model; there is no desire for robustness.[2] On the other hand, lowering θ and thereby increasing η_0 increases the preference for robustness.

2.3 The sense in which robust control yields robust decision rules

To determine how robust control yields robust decisions, let us analyze a simple static robust control problem that can be solved using pen and paper. Let U, π, π_e be the unemployment rate, the inflation rate, and the public's expected rate of inflation, respectively, and suppose that the policy-maker's model is:

$$U = U^* - \gamma(\pi - \pi_e) + \hat{\epsilon}, \tag{6}$$

[2]Equivalently consider setting θ to ∞ in (5). Since the objective is to minimize loss, $\theta = \infty$ sets the loss function to $-\infty$ irrespective of the value that ω_{t+1} takes. Therefore, $\theta = \infty$ corresponds to the case where there is no role for the evil agent; the agent can do nothing to harm the policy-maker. This corresponds to the case where there is no desire for robustness. Lowering θ, however, reduces the shadow price of achieving some robustness. Hansen and Sargent (2004) show that there is a lower bound on θ below which θ cannot be set. For values of θ below that bound, the evil agent is penalized so little for distorting the approximating model that the agent can find a distortion that sends the loss function of the policy-maker to $+\infty$.

where $\gamma > 0$ and $\hat{\epsilon}$ is $N(0,1)$. Thus, U^* is the natural rate of unemployment that prevails on average if $\pi = \pi_e$. The policy-maker sets π, the public sets π_e, and nature draws $\hat{\epsilon}$. Suppose that the policy-maker believes that the model may be misspecified and, as a result, suppose that the policy-maker views (6) as an approximation because of a suspicion that U may actually be governed by

$$U = U^* - \gamma(\pi - \pi_e) + \epsilon + w, \tag{7}$$

where ϵ is another random variable distributed as $N(0,1)$ and w is an unknown distortion to the mean. Thus, the policy-maker suspects that the natural rate of unemployment might be $U^* + w$ for some unknown w. The policy-maker, however, knowing that (6) is a *good* approximating model, knows that the distortion is not too big. To capture that notion, suppose that the policy-maker considers distortions that fall in the set:

$$w^2 \leq \xi. \tag{8}$$

Therefore, the policy-maker, although unable to determine the distortion to the mean, knows that the squared deviation is bounded by ξ.

Assume that the policy-maker sets inflation, π, to minimize the loss function,

$$E(U^2 + \pi^2).$$

In robust control, given that the policy-maker believes that (7) generates U, the policy-maker takes π_e as given and chooses π so that:

$$\min_{\pi} \max_{w} E(U^2 + \pi^2)$$
$$\text{subject to (7), (8).}$$

In other words, the policy-maker chooses π such that it works well for the worst possible distortion, w, that falls in the set given by (8). The above robust control problem can be solved by writing the Lagrangian

$$L(\pi, w, \theta) = (U^* - \gamma(\pi - \pi_e) + w)^2 + \pi^2 - \theta w^2, \tag{9}$$

where θ can be seen as the Lagrange multiplier[3] for constraint (8), and ϵ and the expectation operator are omitted without loss of generality, because of certainty equivalence. The first-

[3]As mentioned in section 2.2, there is a connection between θ and ξ. For now, we take this as given, but in section 2.4 we describe how to pin down a reasonable θ using *detection probabilities*.

order conditions are:

$$\pi \quad : \quad \pi = \gamma(U^* - \gamma(\pi - \pi_e) + w), \tag{10}$$
$$w \quad : \quad \theta w = U^* - \gamma(\pi - \pi_e) + w. \tag{11}$$

Using (10) and (11), we can solve for the optimal π and w. These yield

$$\pi(\theta^{-1}) = \left(\frac{\gamma}{(1+\gamma^2) - \theta^{-1}}\right)(U^* + \gamma\pi_e), \tag{12}$$
$$w(\theta^{-1}) = \left(\frac{\theta^{-1}}{(1+\gamma^2) - \theta^{-1}}\right)(U^* + \gamma\pi_e). \tag{13}$$

The solutions (12) and (13) for π and w, respectively, are in terms of the inverse of the multiplier. Define $\sigma = \theta^{-1}$ and consider Figure 3.

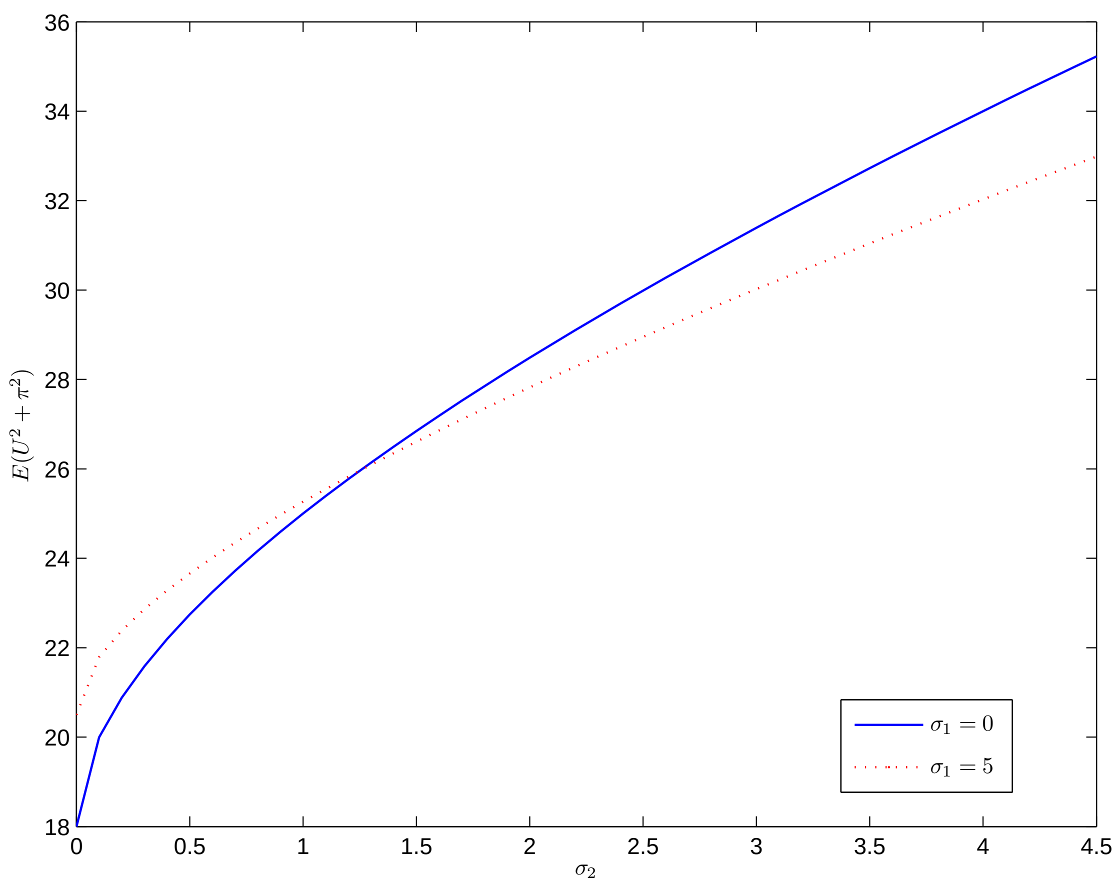


Figure 3: The sense in which robust control yields robustness

Figure 3 illustrates the sense in which robust control yields a robust decision for π. We assume that $U^* = 4$, $\gamma = 1$, and take $\pi_e = 2$ as given. The graph plots the value of $E(U^2 + \pi^2)$ associated with setting $\pi = \pi(\sigma_1)$ when $w = w(\sigma_2)$. The idea, therefore, is to check how costly it is to make mistakes: inflation is set according to a certain assumption on the size of the specification error, which may turn out to be wrong. σ_2 is varied between 0 and 4.5, while σ_1 is allowed to take two values: 0, and 5. The solid line is the value of the loss criterion for $\sigma_1 = 0$, while the dotted line refers to $\sigma_1 = 5$. Notice how the two curves cross: the $\sigma_1 = 0$ curve minimizes loss when there is no specification error (i.e., $\sigma_2 = 0$, which implies $w = 0$), but its performance deteriorates rapidly as the specification error increases along the σ_2 axis. The robust rule, on the other hand, sacrifices performance when there is no distortion to the mean. But its performance deteriorates less rapidly as the specification error increases.

2.4 Choosing the degree of robustness, θ: detection-error probabilities

So far, we have taken θ as given; thus, the policy-maker knows the size of the set of models surrounding the approximating model to consider in the robust decision-making problem. But how can the policy-maker determine the size of the set of surrounding models in practice? This question is particularly important in a linear quadratic set-up, because the evil agent's constraint always binds.[4] Therefore, a policy-maker who gets ready for the worst chooses decision rules for policy tailored on a model lying on the boundary of the set of models from which the evil agent can choose. The decision rules in a linear quadratic robust control problem will hence depend on θ. It is therefore crucial to make a sensible choice for θ.

As a guide to choosing θ, Hansen and Sargent (2004) suggest a *detection-error-probability* approach based on the idea that if it is difficult to statistically distinguish between the models in the set, in a data sample of finite size, then there is a possibility of making the wrong choice of models when determining the true data-generating model. Hansen and Sargent's detection-error-probability approach disciplines the choice of θ by linking it to the probability of making the wrong choice of model. Essentially, this is how the approach works: (i) the decision-maker takes an agnostic position on whether the true data-generating process is given by the approximating model or the worst-case model, (ii) he computes the detection-error probability, i.e., probability of making the wrong choice between the two models on

[4]Since first-order conditions are linear, the most the evil agent can hurt the policy-maker is by choosing a model that deviates the most from the approximating model. Thus, the best choice the evil agent can make is to choose a model on the boundary of the set of models.

the basis of in-sample fit in a data-sample of finite size, (iii) he chooses θ on the basis of the detection-error probability that he wants to achieve. In the sections below, we give a more formal definition of the detection-error probabilities; we describe how to compute them and how they can be related to θ.

2.4.1 Detection-error probabilities and the degree of robustness

Consider the approximating model (1) and the distorted model (2). For a given decision rule $U_t = FX_t$ and a worst-case distortion $\omega_{t+1} = K(\theta)X_t$ when the degree of robustness is θ, define $A_0 = A + BF$ and $\hat{A} = A + BF + CK(\theta)$. The approximating model can then be depicted as

$$X_{t+1} = A_0 X_t + C\check{\epsilon}_{t+1} \tag{14}$$

and the distorted model can be represented as

$$X_{t+1} = \hat{A} X_t + C\epsilon_{t+1}. \tag{15}$$

Now assume that $\check{\epsilon}_{t+1}$ and ϵ_{t+1} are both Gaussion vector process with mean 0 and identity covariance matrices. Detection error probabilities are calculated from likelihood ratios. Thus, consider our two alternative models. Model A is the approximating model (14), and model B is the distorted model (15) associated with the context-specific worst-case shock implied by θ. Consider a fixed sample of observations on the state x_t, $t = 0, ..., T-1$. Let L_{ij} be the likelihood of that sample for model j assuming that model i generates the data. Define the log-likelihood ratio

$$r_i \equiv \log \frac{L_{ii}}{L_{ij}}, \tag{16}$$

where $j \neq i$ and $i = A, B$. When model i generates the data, r_i should be positive. But in a sample of finite size, we may mistakenly conclude that model j generates the data if L_{ij} turns out to be greater than L_{ii}. So supposing that we replicated data a large number of times and computed the likelihood ratio r_i for each sample, the probability of mistakenly concluding that model j generates the data when in fact model i is the true data-generating model is

$$p_i = \text{freq}(r_i \leq 0). \tag{17}$$

Thus, p_i is the frequency of negative log-likelihood ratios r_i when model i is true. Hansen and Sargent (2004) call the probability of a detection error

$$p(\theta) = \frac{1}{2}(p_A + p_B). \tag{18}$$

The probability of a detection error is therefore the average of two kinds of mistakes: (i) concluding that model B generates the data when it is in fact model A and (ii) concluding that model A generates data when it is in fact model B. Here, θ is the robustness parameter used to generate a particular distorted model B. In a given context, Hansen and Sargent (2004) propose to choose θ by setting $p(\theta)$ to a reasonable number and then inverting $p(\theta)$. In other words, if the decision-maker wanted to achieve a detection error probability of 10 per cent, he would pick the θ, say $\hat{\theta}$, that yields $p(\hat{\theta}) = 0.1$.

2.4.2 Computing detection-error probabilities

We now derive formulae for L_{ii} and L_{jj} and provide a recipe for computing detection-error probabilities under our assumption that $\check{\epsilon}_{t+1}$ and ϵ_{t+1} are Gaussian processes with mean 0 and identity covariance matrix. First notice that we can relate the innovations under the approximating model and worst-case model by

$$\begin{aligned} \check{\epsilon}_{t+1} &= \omega_{t+1} + \epsilon_{t+1} \\ &= K(\theta)X_t + \epsilon_{t+1}. \end{aligned}$$

Now suppose that the approximating model is the true data-generating model. Defining X_t^A as data generated from the approximating model and $\omega_{t+1}^A = K(\theta)X_t^A$ as the worst-case distortion, the relationship between the innovations when the approximating model generates data is

$$\check{\epsilon}_{t+1} = \omega_{t+1}^A + \epsilon_{t+1}.$$

The log-likelihood under the approximating model is

$$\log L_{AA} = -\frac{1}{T}\sum_{t=0}^{T-1}\left\{\log\sqrt{2\pi} + \frac{1}{2}(\check{\epsilon}_{t+1}\cdot\check{\epsilon}_{t+1})\right\}. \tag{19}$$

The log-likelihood for the distorted model, given that the approximating model (14) is the data-generating process, is

$$\log L_{AB} = -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(\epsilon_{t+1}\cdot\epsilon_{t+1})\} \tag{20}$$

$$= -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(\check{\epsilon}_{t+1} - w_{t+1}^{A})'(\check{\epsilon}_{t+1} - w_{t+1}^{A})\}. \tag{21}$$

Hence, assuming that the approximating model is the data-generating process, the likelihood ratio r_A is

$$r_A \equiv \log L_{AA} - \log L_{AB}, \tag{22}$$

$$= \frac{1}{T}\sum_{t=0}^{T-1}\{\frac{1}{2}w_{t+1}^{A}{}'w_{t+1}^{A} - w_{t+1}^{A}{}'\check{\epsilon}_{t+1}\}. \tag{23}$$

Suppose, on the other hand, that the distorted model (15) generates the data. Defining $\omega_{t+1}^{B} = K(\theta)X_t^B$ as the worst-case distortion where X_t^B is generated from the distorted model, we can similarly show that

$$r_B \equiv \log L_{BB} - \log L_{BA}, \tag{24}$$

$$= \frac{1}{T}\sum_{t=0}^{T-1}\{\frac{1}{2}w_{t+1}^{B}{}'w_{t+1}^{B} + w_{t+1}^{B}{}'\epsilon_{t+1}\}. \tag{25}$$

Now that we know how to compute the likelihood ratios for a given sample of size T, we compute the detection-error probability as follows:

(i) For a given θ, compute F and $K(\theta)$ given the distorted model (2) and the loss function of the policy-maker.

(ii) For a given θ, use F and $K(\theta)$ to pin down matrices A_0 and $\hat{A}$ for the approximating model (14) and distorted model (15), respectively. Generate a sample j, $\{X_{j,t}^{A} : t = 0, 1, ..., T\}$ according to the approximating model and a sample $\{X_{j,t}^{B} : t = 0, 1, ..., T\}$ according to the distorted model. Compute the likelihood ratios r_A and r_B for sample j according to (23) and (25) respectively.

(iii) Repeat steps 2 and 3 above for samples $j = 1, ..., J$. Compute $p_i = \text{freq}(r_i < 0), i = A, B$ and compute the detection-error probability $p(\theta) = \frac{1}{2}(p_A + p_B)$.

(iv) Repeat steps 1, 2, and 3 for different θs. This will give us a profile linking $p(\theta)$ and θ.

(v) Supposing that the decision-maker wants to achieve a detection-error probability of p^*, determine the degree of robustness that will yield a detection-error probability of p^* as $p(\theta^*) = p^*$.

3. Implementing Robust Control

We now consider a policy-maker whose problem is to set monetary policy according to the following simple rule

$$i_t = \rho_i i_{t-1} + \rho_\pi \pi_t + \rho_x x_t. \tag{26}$$

The policy-maker has a well-defined reference model for carrying policy analysis but is, however, concerned that the reference model may depart from the economic reality in some unknown but potentially important way. As a result, the policy-maker uses robust control to allow for those unknown deviations. In the analysis that follows, we will apply robust control in turn to two candidate reference models: a simplified version of Fuhrer and Moore (1995) and a simplified version of Christiano, Eichenbaum, and Evans (2005). Our objective is to compute the theoretical objects defined in section 2 and provide a practical example of how robust control affects policy choice under different models.

3.1 The models

The first model we consider is a simplified version of Fuhrer and Moore (1995) which we estimate for Canada over the period 1962Q1-2005Q1 by full information maximum likelihood.[5] The version of Fuhrer and Moore (1995) that we estimate assumes a world where agents negotiate nominal-wage contracts that remain in effect for three quarters. Inflation is defined as

$$\pi_t = 4(p_t - p_{t-1}) \tag{27}$$

and the price index, p_t, is defined as a moving average of current and past nominal contract prices, w_t:

$$p_t = \sum_{i=0}^{3} f_i w_{t-i}, \tag{28}$$

[5] We use the AIM algorithm to implement this. See Anderson and Moore (1985).

where the weight, f_i, is given by $f_i = 1/3 + (1-i)s$. The real contract price index, v_t, is defined as the weighted average of current and past real contract prices, $w_t - p_t$:

$$v_t = \sum_{i=0}^{2} f_i(w_{t-i} - p_{t-i}). \tag{29}$$

Agents determine the current real contract price as a function of the real contract prices that are expected to prevail over the duration of the contract, adjusted for excess demand conditions and an identically, independently distributed (i.i.d.) shock,

$$w_t - p_t = \sum_{i=0}^{2} f_i E_t(v_{t+i} + c_x\, x_{t+i}) + \epsilon_{w,t}. \tag{30}$$

The aggregate demand relation makes the output gap a function of its own lags and the *ex ante long-term* real interest rate, ρ_t:

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} - a_\rho \rho_{t-1} + \epsilon_{x,t}. \tag{31}$$

ρ_t is defined, according to the pure expectations' hypothesis, as a weighted average of the future short-term real interest rate,

$$\rho_t = d\, E_t \rho_{t+1} + (1-d)(i_t - E_t \pi_{t+1}). \tag{32}$$

We set d to 40/41, or an average bond-holding period of 40 quarters (see Fuhrer and Moore 1995). Therefore, in the FM model, the short-term nominal interest rate i_t - the policy-maker's policy instrument - influences the output gap only through its effect on the long-term interest rate.

The second model we consider follows Dennis (2003) to derive a simplified version of Christiano, Eichenbaum, and Evans (2005). The model assumes that a fixed proportion of firms $1 - \xi_p$ re-optimize their price every period (Calvo 1983). The remaining proportion of firms, ξ_p, indexes their price change to last period's inflation rate following Christiano, Eichenbaum, and Evans (2005). The first-order condition for optimal price-setting, combined with price-indexation by non-optimizing firms and log-linearized around a non-stochastic steady state, yields the transition equation for inflation:

$$\pi_t = \frac{1}{1+\beta}\pi_{t-1} + \frac{\beta}{1+\beta}E_t\pi_{t+1} + \frac{(1-\beta\xi_p)(1-\xi_p)}{(1+\beta)\xi_p}(x_t + u_t) \tag{33}$$

where β is the discount rate and u_t is an i.i.d. shock.[6] Dennis (2003) assumes household preferences to exhibit habit formation. The log-linearized first-order condition for the household consumption decision yields the transition equation for the output gap:

$$x_t = \frac{\gamma}{1+\gamma} x_{t-1} + \frac{1}{1+\gamma} E_t x_{t+11} - \frac{1-\gamma}{\sigma(1+\gamma)} E_t(i_t - \pi_{t+1} - g_t) \tag{34}$$

where γ is the parameter indexing the degree of habit formation, σ is the parameter determining the curvature of the household's utility function with respect to consumption relative to habit, and g_t is an i.i.d. shock. Following Murchison, Rennison, and Zhu (2004), we set $\beta = 0.99$, $\gamma = 0.85$, $\sigma = 1/0.92$, and $\xi_p = 0.5$. We then calibrate the standard errors of u_t and g_t so that the long-run variances of inflation, output gap, and interest rate match those estimated from a VAR (see Appendix D).

The policy-maker is assumed to set i_t according to the simple rule (26) to minimize a weighted average of the squared deviations of inflation, output gap, and the change in the interest rate:

$$E_0 \sum_{t=0}^{\infty} \left\{ \pi_t^2 + \omega x_t^2 + \nu \Delta i_t^2 \right\}. \tag{35}$$

Both the FM and CEE models can be put in state space form to fit equation (1) once we determine the forward-looking and predetermined state variables. Once the state vector is determined, we can similarly write (35) in the quadratic form given in (4). The solution methods described in Appendixes A and B can therefore be applied. To implement robust control, it is necessary to choose θ; i.e., the size of the set of models surrounding the reference model. In the next section, we compute detection probabilities to discipline the choice of θ.

3.2 Detection probabilities

We follow the steps in section 2.4.2 to compute detection probabilities for each model and make a suitable choice of θ, which dictates the size of the set of models the policy-maker considers when developing policy. Figure 4 plots the detection-error probabilities for distinguishing the FM model from the worst-case models associated with various choices of θ^{-1}. Notice that when $\theta^{-1} = 0$, the detection-error probability is 0.5. This is what it should be since, when $\theta^{-1} = 0$, the approximating and the worst-case model are the same. When θ^{-1} increases, we admit more models in the set of models surrounding the approximating model. As a result, the probability of distinguishing the approximating model from the

[6]The i.i.d. shock appears because the f.o.c. relates inflation to the firm's marginal cost, mc_t, which is assumed to be related to the output gap by $mc_t = x_t + u_t$.

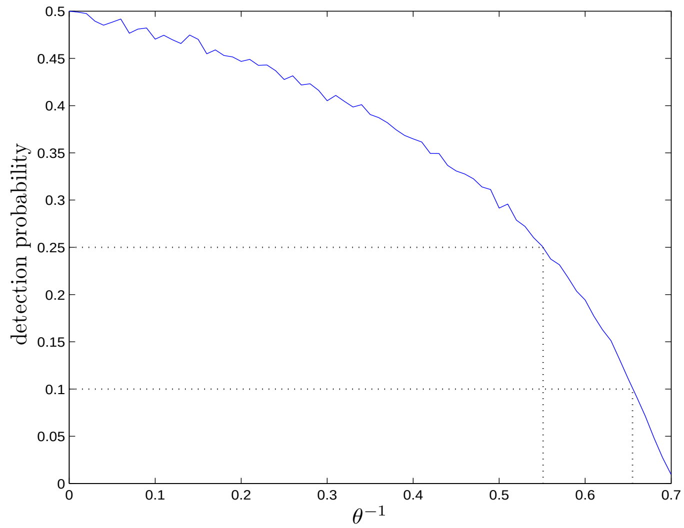


Figure 4: Detection-error probabilities for the FM model

worst-case model declines. Hansen and Sargent (2004) recommend choosing values of θ^{-1} that correspond to detection-error probabilities from 10 to 25 per cent, on a case-by-case basis. For the FM model, this range suggest values of θ^{-1} from 0.66 to 0.55.

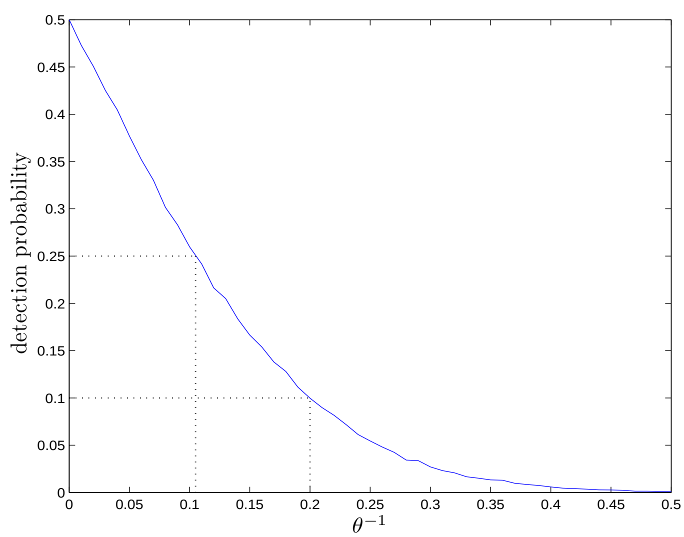


Figure 5: Detection-error probabilities for the CEE model

Figure 5 similarly plots the detection-error probabilities for the CEE model. The values of θ^{-1} corresponding to detection-error probabilities of 10 and 25 per cent are 0.2 and 0.11, respectively.

3.3 Effects of concerns for robustness on monetary policy

Table 1: FM Model: how coefficients of the simple rule vary with θ^{-1}

θ^{-1}	ρ_i	ρ_π	ρ_x
0	0.8629	0.9594	0.5316
0.1000	0.8524	1.0238	0.4962
0.2000	0.8390	1.1059	0.4581
0.3000	0.8222	1.2114	0.4161
0.4000	0.7993	1.3605	0.3696
0.5000	0.7666	1.5883	0.3162

Table 1 shows how the coefficients of the rule change as we increase θ^{-1} from 0 to 0.5 when the reference model is the FM model. Recall that when $\theta^{-1} = 0$, there is no concern for robustness. The optimal rule in that case requires the policy-maker to put a weight of 0.86 on interest rate inertia and weights of 0.96 and 0.53 on contemporaneous inflation and the output gap, respectively. But as θ^{-1} increases, optimal policy changes in three ways: first, the policy-maker gives less and less importance to interest rate inertia; second, the policy-maker responds with a smaller weight to contemporaneous output gap; and third, the policy-maker responds more aggressively to contemporaneous inflation. Why does that happen? Figure 6 shows how the dynamic responses of inflation, the output gap, and interest rate vary with respect to the two shocks in the FM model: $\epsilon_{w,t}$, the real contract price shock, and $\epsilon_{x,t}$, the output gap shock. We see that in contrast to the approximating model, the worst-case model when $\theta^{-1} = 0.5$ makes the effect of a real contract price shock and output gap shock on inflation and output more persistent. Indeed, in the worst-case model, a contract price shock increases inflation more importantly than the approximating model, and its effect takes 25 quarters to die out (relative to 17 quarters in the approximating model). On the other hand, a real contract price shock yields a more pronounced decline in the output gap, and the effect dies out after more than 30 quarters (relative to 28 in the approximating model). Similarly, the worst-case model also makes the effects of an output gap shock on inflation and the output gap bigger and more persistent. By virtue of those more important and persistent positive effects on inflation and more important persistent declines in the output gap, the optimal policy rule in the worst-case model is to respond more aggressively to contemporaneous inflation and less aggressively to the contemporaneous output gap. Note that although the policy-maker chooses a lower degree of inertia, the worst-case model prescribes more important positive effects on the interest rate. The worst-case model also makes the effects of an output gap shock bigger and more persistent.

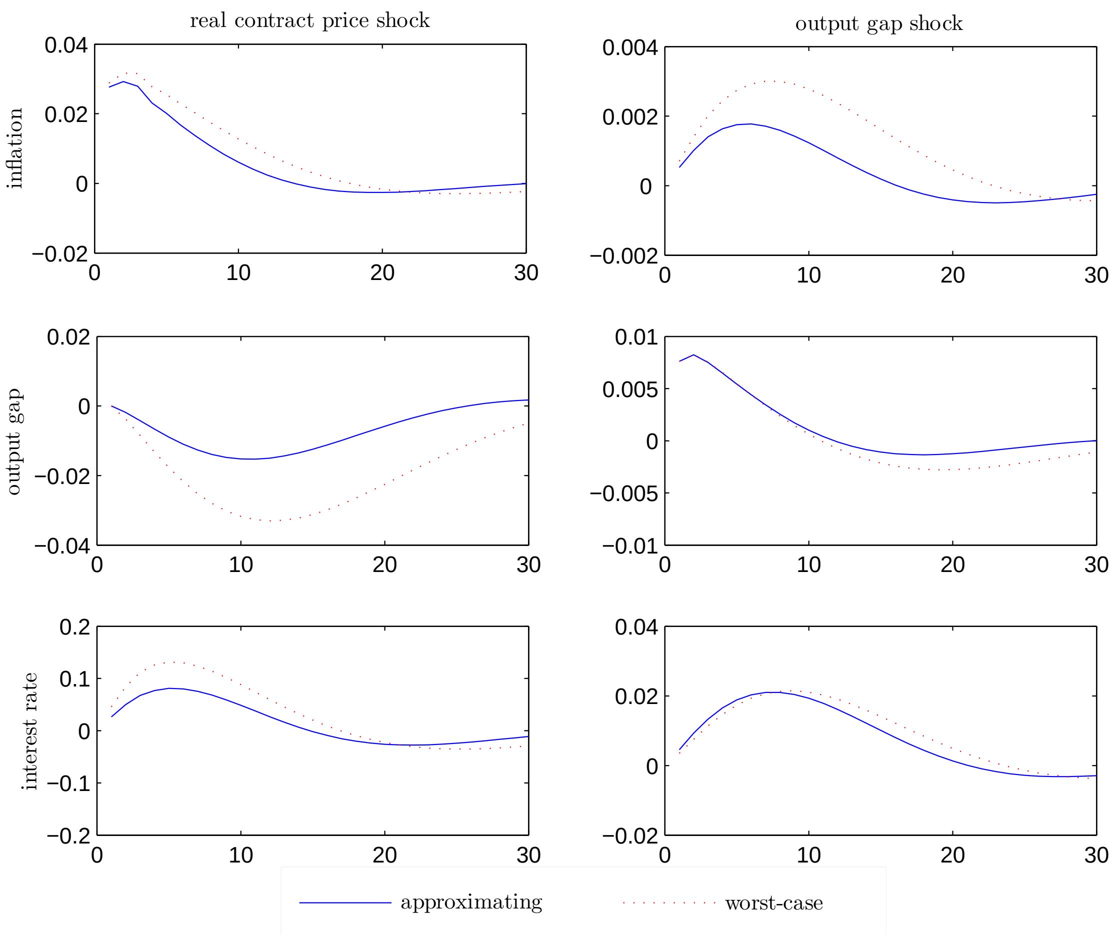


Figure 6: FM model: Approximating model vs worst-case model ($\theta = 0.5$)

Table 2 displays the coefficients of the policy rule for various choices of θ^{-1} for the Dennis model. When there is no concern for robustness, the optimal policy rule requires a high degree of inertia in the interest rate (0.95), a relatively high weight to contemporaneous inflation (1.89), and a moderate response to the output gap (0.66). When we increase θ^{-1}, we first notice that for $\theta^{-1} = 0.22$ (the size of θ^{-1}) corresponding to a detection-error probability of 25 per cent, the optimal policy rule remains virtually unchanged. Even for very high values of θ^{-1}, which according to the detection-error probability criterion is too high, we end up with policy rules that are quantitatively quite close to the $\theta^{-1} = 0$ rule. Figure 7 confirms why this is the case. The dotted lines are the impulse responses of the worst-case model when $\theta^{-1} = 10$. We see that even for a high value of θ^{-1}, the worst-case model predicts that inflation and the output gap respond to the output gap shock and inflation shock with almost the same magnitude and the same persistence as the approximating model. As a result, the optimal policy rule that works best in the worst-case model also works well in

Table 2: How coefficients of the simple rule vary with θ^{-1}

θ^{-1}	ρ_i	ρ_π	ρ_x
0	0.9502	1.8859	0.6607
0.22	0.9501	1.8862	0.6604
1.00	0.9498	1.8870	0.6596
5.00	0.9482	1.8912	0.6552
10.00	0.9461	1.8966	0.6497
30.00	0.9367	1.9191	0.6262
50.00	0.9255	1.9432	0.6008
70.00	0.9122	1.9691	0.5729
100.00	0.8878	2.0115	0.5250

the approximating model.

The above result illustrates one stand on which the robust control approach to dealing with model uncertainty has been criticized. Robust control assumes that the policy-maker's approximating model is a good reference model. The reference model posits the structural features of the economy and robust control constructs a set of models in the neighbourhood of that model for decision making. A sensible degree of robustness (a sensible θ^{-1}), however, may yield a set of models that do not differ much from the approximating model. While this may be suitable when the policy-maker is confident that the reference model captures the main features of the economy, it may not as suitable when the policy-maker has more than one reference model that behaves differently.

From a practical point of view, it is in fact not unusual for policy-makers in central banks to be willing to consider predictions from different models that emphasize different economic paradigms to reduce the risk of policy errors (Engert and Selody 1998). Indeed, central banks are often composed of different departments, each of which uses a particular model to inform their policy judgment. At the Bank of Canada, for instance, although QPM has been the main model used for policy analysis over the past decade, the MFA Department has used the M1-VECM model - a model based on the paradigm that there is a unique cointegrating relationship between M1, real GDP, the consumer price index, and the interest rate - to produce the Blue Book. Perhaps more importantly, significant resources are currently being devoted to developing new DSGE models for policy analysis. MFA, for instance, is currently developing a model that will emphasize household sector financial frictions in the economy (Gammoudi and Mendes 2005). The Research Department is currently developing TOTEM

(Binette, Murchison, Perrier, and Rennison 2004) - a new model that will be used for both economic projections and policy analysis. Since these models will be non-nested models, it is probably not far-fetched to imagine that they will have different predictions along different economic dimensions.[7] So, how should the Bank deal with model uncertainty when it does not have one but two or more reference models that it considers relevant for policy making? In the next section, we present various approaches for making decisions when the policy-maker considers more than one model for policy-making.

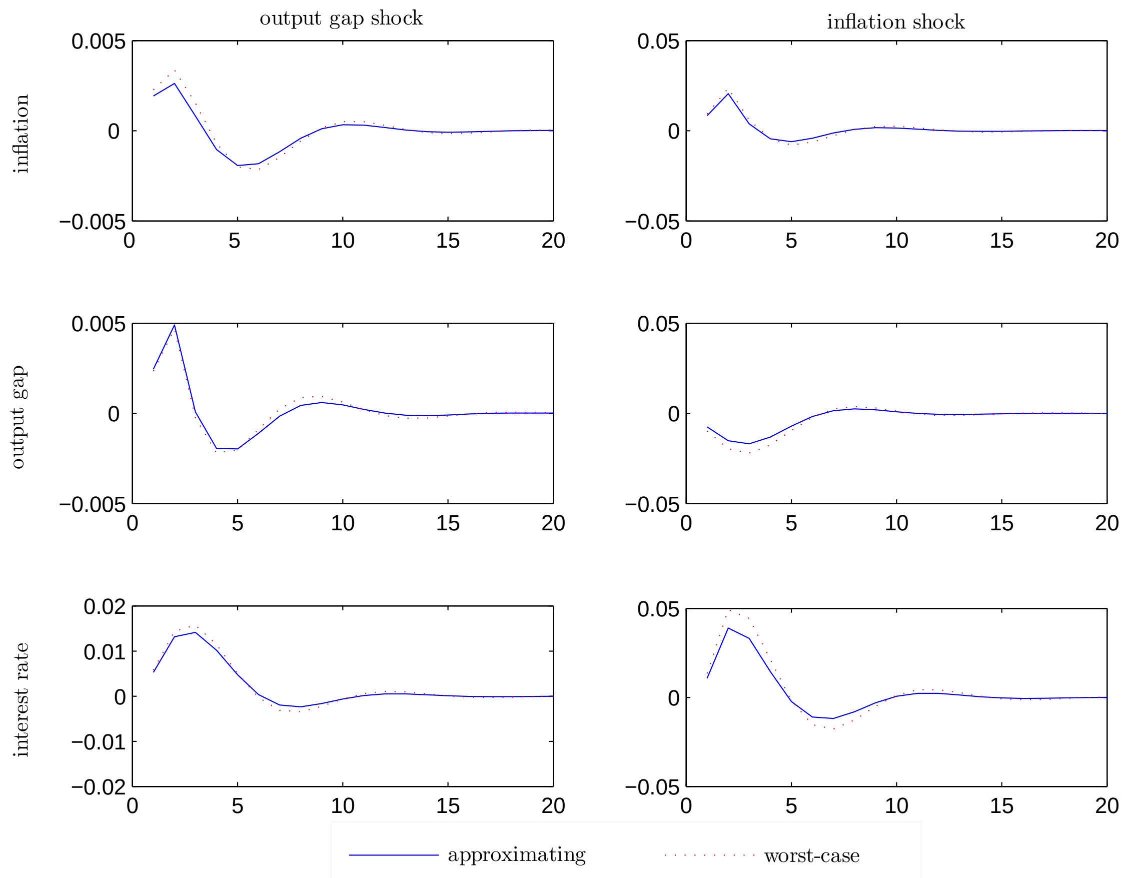


Figure 7: CEE model: Approximating model vs worst-case model ($\theta = 10$)

4. Multiple-Model Approaches

Suppose that the true data-generating model is G, but that the policy-maker does not know it. Suppose also that, being faced with competing theories suggesting different models, the

[7]The models cited in this paragraph are not an exhaustive list of the models that can be used for policy analysis at the Bank of Canada. Ortega and Rebei (2005), for instance, can analyze optimal policy in the context of a multi-sector small open-economy model estimated for the Canadian economy.

policy-maker finds it difficult to settle on a particular model. As a result, suppose that the policy-maker considers $\{G_k, k = 1, ..., n\}$ as the set of *distinct* plausible models, albeit with varying degrees of belief. Assume that π_k is the weight that the policy-maker attaches to the relevance of model k such that $\sum_k \pi_k = 1$.

Moreover, suppose that the policy-maker is given a set of feasible rules,

$$\{K(\gamma), \gamma \in \Gamma\}, \tag{36}$$

and a loss function,

$$v_k(\gamma) = V(G_k, K(\gamma)),$$

which measures the loss of applying the rule $K(\gamma)$ when the model is G_k. The policy-maker's objective is to find *one* rule, say $K(\hat{\gamma})$, that minimizes the loss given the policy-maker's inability to decide between the competing models.

4.1 Bayesian approach

One possible method of designing a rule that works reasonably well is to minimize the Bayesian criterion function with respect to γ:

$$av(\gamma) = \sum_{k=1}^{n} \pi_k v_k(\gamma). \tag{37}$$

What the Bayesian criterion says is that in the face of various plausible models, the policy-maker, believing in each model, should simply try to do well on average. This approach has the advantage that the least plausible models are given the least weight in the decision process. Therefore, it is consistent with the idea that a rule that does not perform very well in a particular model is permissible if that particular model is not very plausible; what matters most is that the rule performs reasonably well in the models that are more likely to be relevant. A disadvantage, however, is that there is no notion that the agent may care about something more than the average performance. In particular, there is no notion that the policy-maker may be unwilling to accept rules that yield very bad performances in some models even though they may perform well on average. Therefore, there is no notion that the agent may, per se, want to achieve some robustness.

To see how this can be pertinent, consider Figure 8. Figure 8 considers a policy-maker who seeks to control the variability of inflation by changing the policy instrument, the

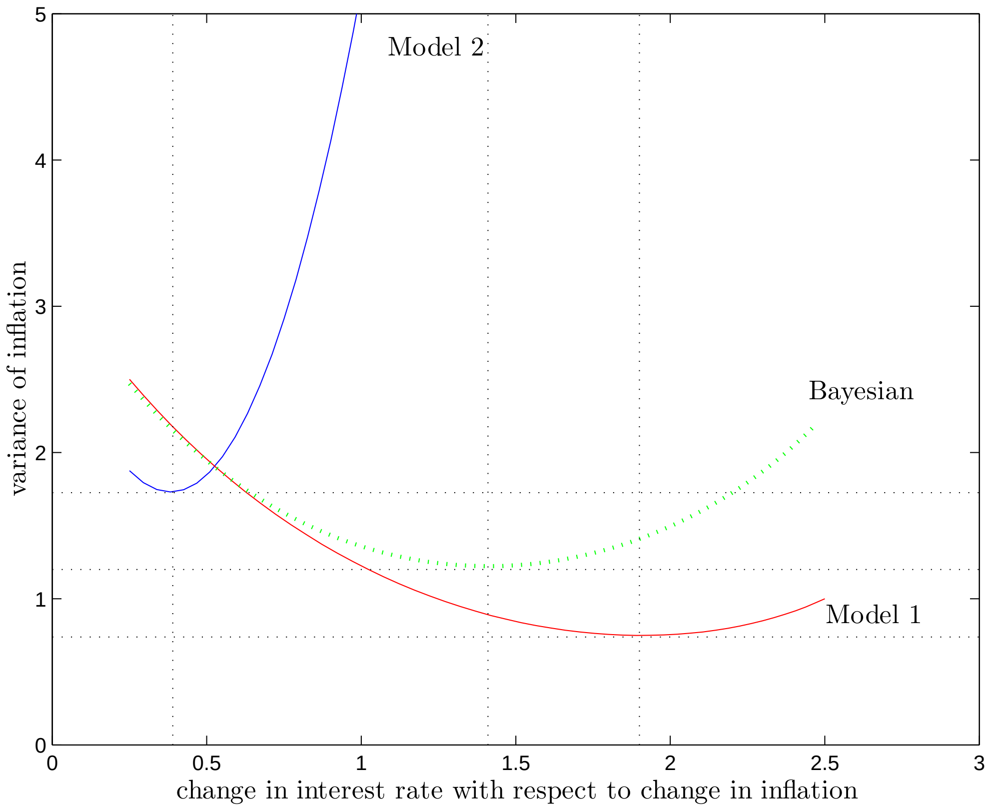


Figure 8: The Bayesian approach

interest rate, with respect to changes in inflation. The policy-maker has two competing reference models for policy-making. Model 1 requires that the policy-maker change the interest rate by 1.9 per cent with respect to a 1 per cent change in inflation, while Model 2 requires that the policy-maker change the interest rate by 0.4 per cent, thus less aggressively. Notice, however, that Model 2 is much more sensitive to policy changes than Model 1: small increases in the interest rate above 0.4 per cent lead to relatively large increases in the variance of inflation. In a situation like this, one can ask: should a policy-maker care only about average performance, or should they also care about preventing those policy choices that can lead to extreme performances in Model 2? If the policy-maker values robustness, then the dotted curve illustrates that the Bayesian approach may not be appropriate. The dotted curve corresponds to a case where the policy-maker assigns a weight of 0.9 to Model 1. The optimal Bayesian policy in that case is to change the interest rate by 1.4 per cent in response to a 1 per cent change in inflation. This performs well on average given the beliefs of the policy-maker, but it leads to very variable inflation in Model 2.

4.2 Worst-case approach

An alternative approach that can yield robustness in the context of distinct models is the worst-case approach. In the worst-case approach, the policy-maker chooses policy according to the following criterion:

$$wc(\gamma) = \max \{v_1(\gamma), v_2(\gamma), ..., v_n(\gamma)\}. \tag{38}$$

What the criterion above entails can again be illustrated by way of an example. Consider

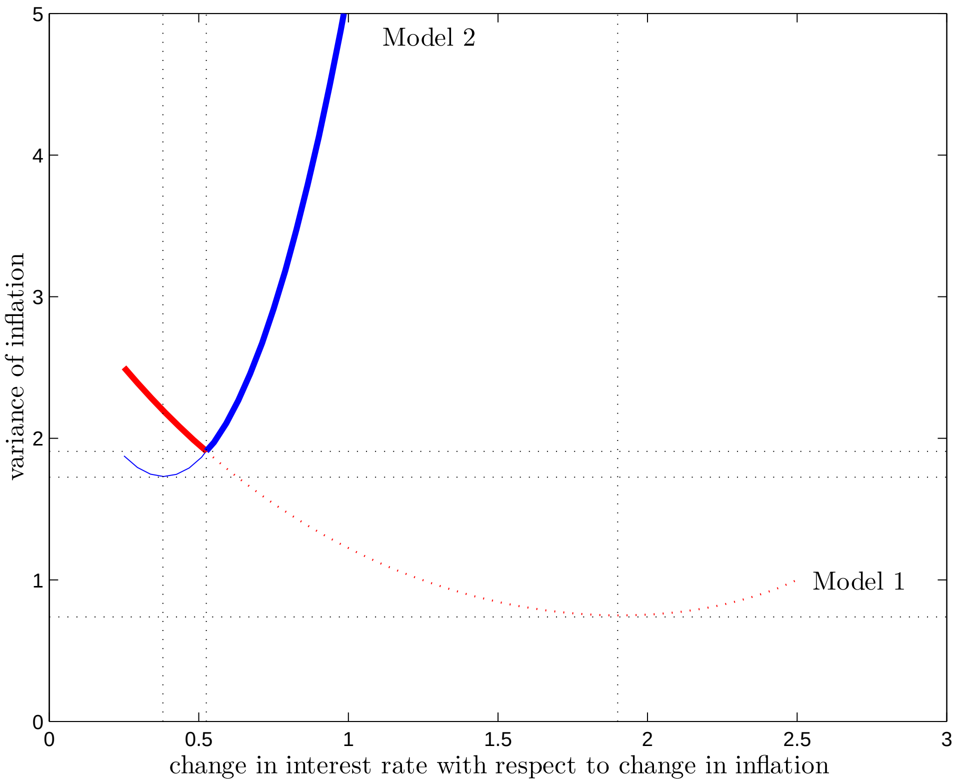


Figure 9: The worst-case approach

again the policy-maker with Model 1 and Model 2 as competing reference models for the economy. With the worst-case approach, the policy-maker's objective is to ensure that the policy decision rule works reasonably well no matter which of the two models is true. To do that, the policy-maker contemplates the policy choices and determines which model fares the worst under each of these choices. Therefore, for the example in Figure 9, the policy-maker determines that, for changes in the interest rate of less than 0.55 per cent, the highest variability of inflation that can arise is given by Model 1 (thicker section of the dotted curve) and conversely, for changes in the interest rate of more than 0.55 per cent, the highest variability of inflation that can arise is given by Model 2 (thicker section of the solid curve). Therefore, the policy choice that would minimize the variability of inflation, irrespective of Model 1 or Model 2 being the true model, is a change in the interest rate of 0.55 per cent.

The main advantage of the worst-case approach is that it does not require the decision-maker to have beliefs about the plausibility of each model for decision making. Although there are techniques for assigning weights to competing models, it may not always be easy for the policy-maker to formulate priorities over a set of models. Therefore, a nice feature of the worst-case approach is that it can be applied even when the decision-maker cannot determine a reliable priority over a set of models. This property, however, turns out to be a drawback when the policy-maker does have information that can help discriminate between models. In that case, by neglecting information that may be relevant for assessing the plausibility of each model, the worst-case approach can lead to corner solutions (as in our example) that are much more restrictive than necessary, especially if the model leading to the corner solution is not very plausible.

4.3 Trade-off between average performance and robustness

The worst-case approach can lead to robust decision rules that may be too conservative, whereas the Bayesian approach yields decision rules that may not be sufficiently robust although they perform well on average. An approach that balances the desire for good average performance and a concern for robustness is suggested in Cateau (2005). Cateau (2005) adapts the decision theory literature (see Klibanoff, Marinacci, and Mukerji 2002; Segal 1990; and Ergin and Gul 2004) to suggest the criterion:

$$h(\gamma) = \sum_{k=1}^{n} \pi_k \phi(\ v_k(\gamma)\ ). \tag{39}$$

The main feature of criterion (39) is that: (i) it incorporates the beliefs that the decision-maker may have about each model, as in the Bayesian approach, and (ii) it allows the decision-maker to distinguish between the *within-model risk* (the risk which arises naturally because the model is stochastic) and the *across-model risk* (the risk which is associated with the multiplicity of models). The decision-maker allows for the across-model risk by evaluating the performance of a model not only by its loss (e.g. how much variability of inflation it leads to) but by a *transform* of that loss. That transform, as motivated by Klibanoff, Marinacci, and Mukerji (2002), reflects the attitude of the decision-maker towards the across-model risk; basically, its curvature will represent the degree of aversion towards the across-model risk.

The decision-maker's care about the across-model risk leads them to decide to achieve a balance between average performance and robustness. Indeed, Cateau (2005) shows that his framework nests both the Bayesian approach and the worst-case approach as special cases: the Bayesian approach is the special case where the decision-maker is neutral to the across-model risk (the decision-maker's degree of aversion towards the across-model risk is 0), and the worst-case approach is the special case where the decision-maker's degree of aversion towards the across-model risk is infinite. Therefore, the degree of aversion towards the across-model risk which reflects the attitude of the decision-maker towards model uncertainty determines the extent to which the decision-maker wants to trade-off average performance for robustness.

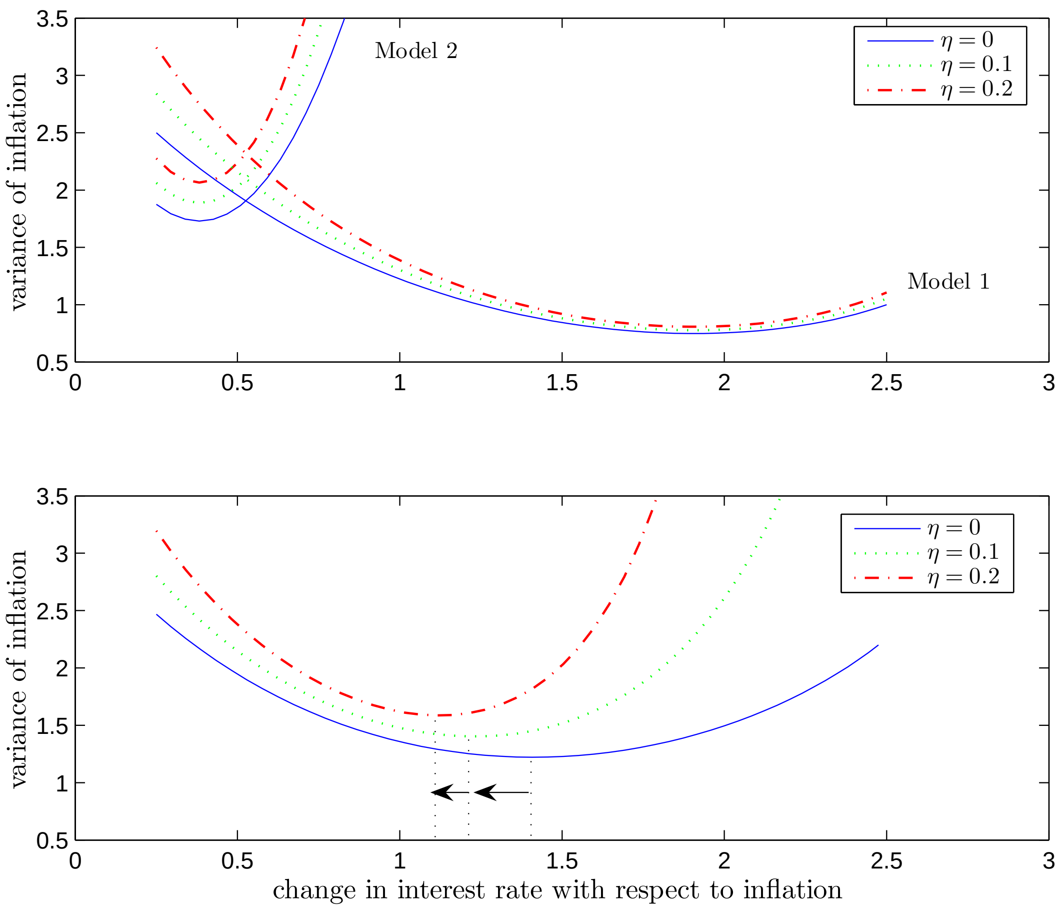


Figure 10: Aversion to across-model risk

Figure 10 shows how accounting for the across-model risk helps the policy-maker to balance performance and robustness. I assume that the policy-maker re-evaluates the performance of a particular model by transforming the variability of inflation it leads to, say x, according to a function, $\phi(x)$, which exhibits constant absolute across-model risk aversion,[8] η. When $\eta = 0$, the policy-maker is neutral to the across-model risk and therefore adopts a Bayesian approach for dealing with indecision between Model 1 and Model 2. The solid curves in the upper panel, as before, show how much variability of inflation, Model 1 and Model 2, lead to under each policy choice. As in Figure 8, when the policy-maker assigns a weight of 0.9 to Model 1, the optimal policy choice is to change the interest rate by 1.4 per cent with respect to a 1 per cent change in inflation (the solid curve in the lower panel). When the policy-maker is averse to the across-model risk, the aversion forces the policy-maker to add a risk premium to the cost of each model under each policy choice. The upper panel shows that increasing the degree of aversion to the across-model risk shifts the loss profile of Model 2 upwards in a quantitatively more important way than that of Model 1. As a result, even after giving a weight of 0.9 to Model 1, the more averse the policy-maker becomes, the more the policy-maker prefers less aggressive policy to avoid bad performances in Model 2 (the dotted and dashed curves reach their minimum at smaller changes in the interest rate). It is in this sense that accounting for the across-model risk induces the policy-maker to search for more robust policy.

4.3.1 Implementing the trade-off approach

The trade-off approach described above assumes that (i) the policy-maker can assign weights to the models in the decision set, and (ii), the policy-maker knows their own degree of aversion towards the across-model risk. To implement the approach, it is therefore important to determine those parameters. Section 4.3.1 shows how one can use Bayes law to calculate model weights from the data starting from a certain prior on the set of models. Section 4.3.1 shows how one can proceed to give economic meaning to the degree of aversion towards the across-model risk.

Determining weights for models

Suppose that model i implies the transition equation[9]

$$x^i_{t+1} = a x^i_t + c^i \epsilon^i_{t+1}, \tag{40}$$

[8] Figure 10 assumes that $\phi(x) = \frac{\exp(\eta x) - 1}{\eta}$ such that η denotes the degree of aversion towards the across-model risk.

[9] Think of this transition equation as resulting from the policy-maker's linear-quadratic control problem if the model is i.

where x_t^i is a vector of state variables that may be completely or only partially observed, and ϵ_{t+1}^i is a vector of Gaussian random variables with mean 0 and identity covariance matrix, and let

$$y_{t+1} = h^i x_{t+1}^i \tag{41}$$

be the vector of observable state variables.

Now suppose that $\pi_{i,0} = P(m_i)$ is the initial probability that the policy-maker assigns to model i as time 0. The policy-maker can revise that probability at time t after observing $y^t = (y_0, y_1, ..., y_t)$ using Bayes law to derive:

$$\begin{aligned} \pi_{i,t} &= P(m_i|y^t) \\ &= \frac{P(y^t|m_i)P(m_i)}{\sum_j P(y^t|m_j)P(m_j)} \\ &\propto P(y^t|m_i)P(m_i) \\ &= \alpha_{i,t}. \end{aligned} \tag{42}$$

Therefore, the posterior probability, $\pi_{i,t}$, for model i is proportional to the marginalized likelihood of the data according to model i weighted by the initial prior for model i. To develop a recursion for $\pi_{i,t}$, it is in fact easier to first develop a recursion for $\alpha_{i,t}$ and then normalize $\alpha_{i,t}$ to relate it to $\pi_{i,t}$. Consider the ratio

$$\frac{\alpha_{i,t+1}}{\alpha_{i,t}} = \frac{P(y^{t+1}|m_i)}{P(y^t|m_i)}. \tag{43}$$

Under the state space form (40) for model i and the assumption of normality for ϵ_{t+1}^i, we can use the Kalman filter to construct the marginalized likelihood in the numerator and denominator. In fact, if ϵ_{t+1}^i is independent across time, since

$$P(y^{t+1}|m_i) = \prod_{\tau=0}^{t} P(y_{\tau+1}|y^\tau, m_i), \tag{44}$$

we can further simplify (43) to

$$\frac{\alpha_{i,t+1}}{\alpha_{i,t}} = P(y_{t+1}|y^t, m_i). \tag{45}$$

Therefore, $\alpha_{i,t+1}$ gets updated every period according to the contribution that the time $t+1$ piece of data, y_{t+1}, makes to the marginalized likelihood. With normality, given data y^t,

y_{t+1} is normally distributed with mean $y_{t+1}|y^t$ and variance Δ_{t+1}.[10] So, given a starting value $\alpha_{i,0}$ at time 0, we can use (45) to update $\alpha_{i,t}$ every period. We can then calculate the posterior probability for each model by normalizing $\alpha_{i,t}$ as follows:

$$\pi_{i,t} = \frac{\alpha_{i,t}}{\sum_j \alpha_{j,t}}. \tag{46}$$

Figure 11 uses quarterly data on inflation and the output gap from 2000Q1 to update

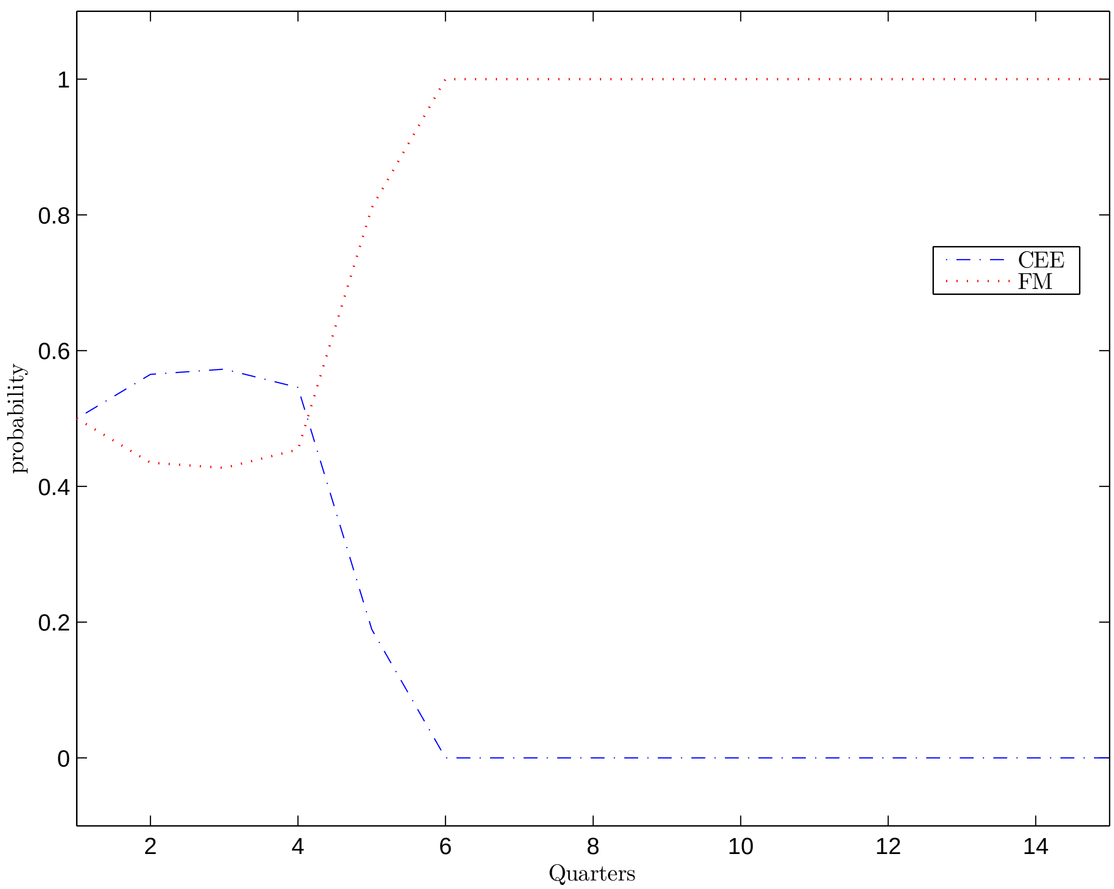


Figure 11: Updating probability weights of model

weights to the CEE and FM model. Initially, the models are assigned equal weights, but as new data arrived, the weights to each model are recalculated using equations (45) and (46). The weight to the FM model initially declines, but we see that after 6 quarters of data have been accumulated, the statistical evidence in favour of the FM model causes its weight to converge to 1.

[10] Appendix C explains how we can use the Kalman filter to construct $y_{t+1}|y^t$ and Δ_{t+1}.

Determining the degree of aversion towards the across-model risk

To pin down the degree of across-model risk aversion, we proceed in analogy to risk theory. In risk theory, to interpret the size of the risk-aversion parameter, it is related to the risk premium that the decision maker would be prepared to pay to be indifferent between accepting a small risk or obtaining a non-random amount for sure.

So suppose that the policy-maker wants to achieve a loss of v^*. However, the policy-maker faces N models that suggest losses v_i, with probability π_i, $i = 1, ..., N$. In an analogy to risk theory, we can ask: how much of a premium, δ, would the policy-maker be ready to pay to be indifferent between achieving v^* for sure or else face model uncertainty and, hence, N possible different losses with probability π_i. That is,

$$\phi(v^* + \delta) = \sum_{i=1}^{N} \pi_i \phi(v_i). \tag{47}$$

Now suppose that $\phi(x)$ is the exponential function given by $\phi(x) = \frac{e^{\eta x}-1}{\eta}$ such that $\eta = \frac{\phi''(x)}{\phi'(x)}$ is the degree of aversion to the across-model risk. Substituting for ϕ into (47) yields

$$\sum_{i=1}^{N} \pi_i e^{\eta(v_i - v^* - \delta)} = 1. \tag{48}$$

Equation (48) denotes the relationship between η and δ. Given the set of probabilities attached to the models, the loss, v^*, and the premium, δ, that the policy-maker is ready to pay to achieve v^*, we can solve numerically for η as a function of these parameters. Two caveats are in order: first, there need not be a unique η that solves (48). This follows since the left-hand side of (48) need not be monotonic in η. Second, if we take the total derivative of η with respect to δ, we obtain:

$$\frac{d\eta}{d\rho} = \frac{\eta}{\sum_{i=1}^{N} \pi_i e^{\eta(v_i - v^* - \delta)}(v_i - v^* - \delta)}. \tag{49}$$

Therefore, $\frac{d\eta}{d\rho}$ can in general be positive or negative depending on the π_i's, v^*, η and δ. These concerns appear in the risk literature and it should not come as a surprise that they appear here as well. Indeed, just as it is known that agents do not necessarily react to a bet on \$10 as they would to a bet on \$1 million, the policy-maker need not react in the same way in situations where models exhibit small or big losses. Equation (48) is nevertheless useful if we want to characterize how the degree of aversion to across-model risk behaves when the risk in question can potentially be large.

Conversely, we may be interested to know how η behaves with respect to slight across-model risk. Rewrite (47) as

$$\phi(v^* + \delta) = E\phi(\tilde{v}) \tag{50}$$

where v^* is the non-stochastic loss that the policy-maker wants to achieve, E denotes expectation with respect to the prior over models, and $\tilde{v}$ is the random variable representing the loss of each model. Taking a first-order approximation of the left-hand side of (50) around v^*, we obtain

$$\phi(v^* + \delta) = \phi(v^*) + \phi'(v^*)\delta. \tag{51}$$

Taking a second-order approximation of the right-hand side of (50) around v^*, we obtain

$$E\phi(\tilde{v}) = \phi(v^*) + \phi'(v^*)E(\tilde{v} - v^*) + \frac{1}{2}\phi''(v^*)E(\tilde{v} - v^*)^2. \tag{52}$$

Equating (51) and (52) and noting that $\frac{\phi''(v^*)}{\phi'(v^*)} = \eta$, we obtain

$$\delta = \eta\frac{1}{2}E(\tilde{v} - v^*)^2 + E\tilde{v} - v^*, \tag{53}$$

such that

$$\eta = \frac{\delta + v^* - E\tilde{v}}{\frac{1}{2}E(\tilde{v} - v^*)^2}. \tag{54}$$

Therefore, in the neighbourhood of a given non-stochastic v^*, $\frac{d\eta}{d\rho} > 0$. Hence, the larger the premium that the policy-maker is willing to pay to achieve v^*, the more the policy-maker is averse to the across-model risk.

Equation (54) tells us how to calculate the local degree of across-model risk aversion, while (48) tells us how to obtain a measure of the degree of aversion in the large. Both equations yield a relationship of η as a function of δ. So (54) and (48) can help us determine a sensible η provided that we can sensibly pin down δ, the premium that the policy-maker is willing to pay in the face of model uncertainty. This is not without difficulty; however, it should be far easier to determine such a premium than to directly determine the degree of across-model risk aversion. Why this is the case can be illustrated by the following example.

Suppose that the policy-maker considers two models of the economy and cares only about controlling inflation. Prior to choosing monetary policy, suppose that Model 1 predicts that a particular policy choice will yield a squared deviation of inflation from target of 0.1 per cent, while Model 2 predicts that the same policy choice will yield a squared deviation of 10 per cent. Suppose further that the policy-maker's objective is to achieve a squared deviation of 3 per cent. To deduce the premium that the policy-maker would be willing to pay to eliminate model uncertainty, we can ask the following question: Given that the squared deviation of inflation from target can be either 0.1 or 10 per cent tomorrow, what is the maximum squared deviation you would tolerate today if we could guarantee that tomorrow we will have a single model that tells us how to achieve the 3 per cent squared deviation? Suppose that the policy-maker would tolerate a squared deviation of 5 per cent. Then the premium that the policy-maker would be ready to pay to eliminate model uncertainty and achieve 3 per cent squared deviation for sure is 5 - 3 = 2 per cent. Once we know the premium that the policy-maker is willing to pay, we can then determine the policy maker's degree of across-model risk aversion from (54) or (48).

A different strategy for implementing the trade-off approach to work is to assume a particular degree of aversion and work backwards from (53) to determine the across-model risk premium that corresponds to the assumed degree of aversion. When repeated for various degrees of aversion, that exercise will lead to a profile linking each degree of aversion to the corresponding premium that the policy-maker would tolerate to eliminate model uncertainty. That profile is likely to be more useful from a policy-making perspective because, instead of being asked to quote a particular across-model risk premium, the policy-maker can directly see the effect of increasing or decreasing the across-model degree of risk aversion on the premium.

4.4 Trade-off between the CEE and FM model

We now implement the trade-off approach for the policy-maker that views the CEE model and the FM model as two plausible models of the Canadian economy. We assume that the policy-maker values both models equally and assigns equal weight to them. The policy-maker's across-model risk aversion is captured by the function $\phi(x) = \frac{e^{\eta x}-1}{\eta}$. The policy-maker's decision problem is to choose policy according to the simple rule (26) given the two models considered plausible and across-model risk aversion. Table 3 displays how the coefficients of the policy rule vary with η. The last two columns display the losses that the optimal policy rule lead to in each model. When $\eta = 0$, the policy-maker uses the Bayesian criterion to choose policy. The optimal policy rule arrived at leads to a loss of 0.12 in the CEE model, while it leads to a loss of 0.49 in the FM model. Since the policy

Table 3: Simple rule coefficients, the degree of across-model risk aversion and model losses

η	ρ_i	ρ_π	ρ_x	loss CEE	loss FM
0	0.785	1.337	0.496	0.1138	0.4932
0.5	0.794	1.292	0.502	0.1144	0.4926
1	0.803	1.250	0.507	0.1152	0.4920
1.5	0.811	1.210	0.511	0.1159	0.4916
2	0.818	1.173	0.514	0.1167	0.4912
3	0.831	1.112	0.517	0.1181	0.4906
5	0.847	1.031	0.520	0.1203	0.4901
7	0.855	0.989	0.520	0.1217	0.4899
10	0.861	0.961	0.520	0.1226	0.4899
15	0.863	0.950	0.520	0.1230	0.4899
20	0.864	0.948	0.520	0.1231	0.4899
25	0.864	0.947	0.520	0.1231	0.4899
30	0.864	0.947	0.520	0.1231	0.4899

rule is more restrictive for the FM model than for the CEE model, increasing the degree of aversion makes the policy-maker more worried about improving the performance of the rule in the FM model. As a result, as we increase η, the policy-maker chooses rules that exhibit more inertia, a slightly higher contemporaneous response to the output gap but a lower contemporaneous response to inflation. These rules perform better in the FM model but worse in the CEE model. It is in this sense that increasing η makes the policy-maker more conservative and gradually convinces the policy-maker to act on a worst-case scenario.

Now that we know the effect of different degrees of across-model risk aversion on the optimal policy rule, it is important to interpret their economic significance. We do that by computing the implied premium that each degree of aversion leads to. We use (54) and (48) to compute the local premium (δ_l) and the global premium (δ_g), respectively. Table 4 assumes that $v^* = \tilde{E(v)}$. Therefore Table 4 computes the premium that the policy-maker would be willing to pay ex ante (before resolving model uncertainty) to be certain of achieving the average of the model losses ex post (after resolving model uncertainty). First we see that δ_g, the global premium, increases with η for the range of values considered. Therefore, in our case, increasing the degree of across-model risk aversion implies that the policy-maker is willing to pay a higher premium ex ante to resolve the model uncertainty. From (54), the same holds for the local premium δ_l. Second, we notice that δ_l and δ_g are almost the same for small degrees of aversion. But at higher degrees of aversion ($\eta >= 10$), the local premium grows much faster than the global premium. Therefore, the local premium

Table 4: Implied premiums and the degree of across-model risk aversion

η	average loss	δ_l	δ_g
0.5	0.3035	0.009	0.009
1	0.3036	0.018	0.018
1.5	0.3037	0.027	0.026
2	0.3039	0.035	0.034
3	0.3043	0.052	0.050
5	0.3052	0.086	0.076
7	0.3058	0.119	0.096
10	0.3062	0.169	0.117
15	0.3065	0.252	0.138
20	0.3065	0.336	0.149
25	0.3065	0.420	0.156
30	0.3065	0.504	0.160

is less reliable at higher degrees of aversion. This is not surprising. Higher η's imply that $\phi(x)$, the function that the policy-maker uses to re-evaluate models because of across-model risk aversion has much more curvature to it. Therefore, local linear approximations are less accurate.

How can Table 4 be used for policy analysis? Recall that the loss function in the calculations above is a weighted average of the squared deviations of inflation, output gap, and the change in the interest rate. Column 2 in the table shows the average of the model losses which we assume the policy-maker would be happy to achieve. By adding the average loss and δ_g (or δ_l, which is less accurate but easier to compute), we obtain the maximum loss that the policy-maker would tolerate to resolve model uncertainty ex ante. So once we have the profile linking η to the maximum tolerable loss, the policy-maker can select the value of η, say η^* that yields a reasonable maximum tolerable loss. We can then compute the optimal policy rule that η^* leads to.

5. Conclusion

In this paper, we have discussed some recent methods for dealing with model uncertainty. We presented the theoretical ideas behind each method, illustrated what each method does through simple examples but also discussed why and when a particular method may be more appropriate than the other. But we did not limit ourselves to a theoretical discussion. We also showed how to compute the various parameters that a theoretical discussion takes for granted and we have also worked through the implementation using as benchmark two models of the Canadian economy. Now we summarize the key points of our discussion and also suggest some ways, which we believe can be implemented in the short and medium term, in which our analysis can be used to help the Bank of Canada deal with model uncertainty.

Dealing with Model Uncertainty:

- Robust control is most useful when the policy-maker has one good model of the economy; by definition, it is designed to choose decision rules that work well in a neighbourhood of a particular model.

- When the policy-maker has competing reference models of the economy, the Bayesian approach, worst-case model approach, and the trade-off approach of Cateau (2005) are likely to perform better than robust control since they take into account that the policy-maker may, in fact, use models that are arbitrarily *far* from each other.

- The critical distinction between the Bayesian approach, worst-case approach, and Cateau's approach is that they make different assumptions about the attitude of the policy-maker towards model uncertainty. The Bayesian approach assumes that the policy-maker is neutral to the across-model risk, only caring about average performance; the worst-case model approach assumes that the policy-maker is infinitely averse to the across-model risk, only caring about how robust the policy choice is; Cateau (2005) allows the aversion of the policy-maker to vary between zero (Bayesian) and infinity (worst-case) - the policy-maker's degree of aversion determines how much the policy-maker trades off average performance for robustness.

Some Proposals:

- Robust control for TOTEM. TOTEM will be the main model used by Bank staff for policy exercises. Depending on how reasonable it is to consider TOTEM as **the** reference model for policy-making at the Bank of Canada, robust control will provide policy rules that are robust to misspecification within a certain distance of TOTEM. To put this in practice, we will need to specify the size of the set of models surrounding TOTEM for which we want robust rules (i.e., ξ in Figure 2). We show in section 2.4 how we can compute that parameter.

- Robust control for sensitivity analysis. At the FAD meetings, representatives of departments at the Bank are often asked not only for their perspective on the current and future state of the Canadian economy but also about the confidence of their assessment. Robust control can be used for sensitivity analysis relative to misspecification. If the department, for instance, uses a model for predicting how inflation would respond to a cost-push shock, the department can also calculate how inflation would respond if its model is subject to misspecification. That response can be calculated for various degrees of misspecification to get an idea of how robust the prediction is.

- Optimal policy rule with competing models. With the development of TOTEM and MFA's financial frictions model, for example, the Bank will soon have different models providing different pictures of the Canadian economy that can be used for policy analysis. What kind of policy rules should the Bank use when there is model uncertainty? Côté, Kuszczak, Lam, Liu, and St-Amant (2002) analyze the performance and robustness of some simple policy rules in twelve models of the Canadian economy. While their analysis shows that Taylor-type rules, interest rate smoothing rules, and open economy rules are not robust (at least in the set of models that they consider), their approach is not useful to answer the question asked above. Indeed, they consider simple policy rules *optimized for a particular model* and evaluate how it performs in the other models. A more useful way to derive a policy rule that works well across various models is to start with a general type of policy rule and evaluate how it performs relative to a criterion that *involves all the models in the decision set of the policy-maker*. In fact, if we use the criterion that Cateau (2005) suggests, we can seek to answer a more interesting question: what kind of policy rules should the bank use to achieve a certain trade-off between average performance and robustness when it faces model uncertainty? Of course, the trade-off will depend on the policy-maker's aversion to model uncertainty, so it will be important to estimate the policy-maker's degree of aversion. We suggest two approaches in section 4.3.1 to determine the degree of across-model risk aversion.

References

Anderson, G. and G. Moore. 1985. "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models." *Economic Letters* 17(3): 247–52.

Binette, A., S. Murchison, P. Perrier, and A. Rennison. 2004. "An Introduction To TOTEM." Work in Progress, Bank of Canada.

Calvo, G. 1983. "Staggered Prices in a Utility Maximizing Framework." *Journal of Monetary Economics* 12: 383–98.

Cateau, G. 2005. "Monetary Policy Under Model and Data-Parameter Uncertainty." Bank of Canada Working Paper No. 2005-6.

Christiano, L., M. Eichenbaum, and C. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113(1): 1–45.

Côté, D., J. Kuszczak, J.P. Lam, Y. Liu, and P. St-Amant. 2002. *The Performance and Robustness of Simple Monetary Policy Rules in Models of the Canadian Economy.* Technical Report No. 92. Ottawa: Bank of Canada.

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

Dennis, R. 2003. "The Policy Preference of the U.S. Federal Reserve." Federal Reserve Bank of San Francisco Working Paper No. 2001-08.

———. 2004. "New Keynesian Optimal-Policy Models: An Empirical Assessment." Federal Reserve Bank of San Francisco Working Paper No. 2003-16.

Engert, W. and J. Selody. 1998. "Uncertainty and Multiple Paradigms of the Transmission Mechanism." Bank of Canada Working Paper No. 98-7.

Ergin, H. and F. Gul. 2004. "A Subjective Theory of Compound Lotteries." M.I.T. Working Paper.

Fuhrer, J.C. and G.R. Moore. 1995. "Monetary Policy Trade-Offs and the Correlation between Nominal Interest Rates and Real Output." *American Economic Review* 85(1): 219–39.

Gammoudi, M. and R. Mendes. 2005. "Household Sector Financial Frictions in a Small Open Economy." Work in Progress, Bank of Canada.

Giannoni, M.P. 2002. "Does Model Uncertainty Justify Caution? Robust Optimal Monetary Policy in a Forward-Looking Model." *Macroeconomic Dynamics* 6(1): 111–44.

Hansen, L.P. and T.J. Sargent. 2004. "Robust Control and Model Uncertainty in Macroeconomics." University of Chicago and New York University Manuscript.

Klibanoff, P., M. Marinacci, and S. Mukerji. 2002. *A Smooth Model of Decision Making under Ambiguity*. Technical report. International Centre for Economic Research.

Murchison, S., A. Rennison, and Z. Zhu. 2004. "A Structural Small Open-Economy Model for Canada." Bank of Canada Working Paper No. 2004-4.

Onatski, A. and N. Williams. 2003. "Modeling Model Uncertainty." *Journal of the European Economic Association* 1(5): 1087–1122.

Ortega, E. and N. Rebei. 2005. "Optimal Monetary Policy with a Two Sector Small Open Economy Model Estimated for Canada." Work in Progress, Bank of Canada.

Segal, U. 1990. "Two-Stage Lotteries without the Reduction Axiom." *Econometrica* 58(2): 349–77.

Appendix A: Robustness in Forward-Looking Models

Forward-looking models differ from backward-looking ones (also known as the linear regulator problem in the literature) in that part of the state is not inherited from the past but are jump variables; i.e., variables that need to adjust for a solution to a stabilizing solution to exist. To solve a forward-looking problem, however, knowing how to solve a backward-looking problem is still very useful because many of the objects that appear in the solution to the backward-looking problem are the same ones that we need to construct the solution to the forward-looking problem. So below, we introduce the forward-looking problem and in section A.1, solve the problem just as if it was a linear regulator problem. In section A.1.2, however, we show how to obtain the solution to the forward-looking problem.

A.1 The Ramsey problem

Let $X_t = \begin{bmatrix} Y_t \\ Z_t \end{bmatrix}$ be the state vector at time t. The Y_t are natural state variables in the sense that they are inherited from the past. The Z_t are forward-looking state variables (such as those coming from the Euler equations of the private sector in a Ramsey problem, for example) that need to be determined by the model at time t. Suppose that the policy-maker's approximating model is given by

$$X_{t+1} = AX_t + BU_t + C\epsilon_{t+1} \tag{A1}$$

where the U_t are the policy-maker's control and ϵ_t is an i.i.d. shock process with mean 0 and identity covariance matrix and let the loss function be given by

$$E_0 \sum_{t=0}^{\infty} \beta^t \left\{ X_t' Q X_t + U_t' R U_t \right\}. \tag{A2}$$

As before, without a concern for robustness, the objective of the policy-maker is to minimize (A2) subject to (A1) and given Y_0. [1] The policy-maker, however, doubts the model. As a result, the policy-maker considers the approximating model to be a good approximation of the data-generating model that falls in the set of models given by

$$X_{t+1} = AX_t + BU_t + C(\epsilon_{t+1} + \omega_{t+1}), \tag{A3}$$

[1] Notice, that since the Z_t are forward-looking variables and not inherited from the past, there are no initial conditions Z_0.

where ϵ_t is an i.i.d. shock process with mean 0 and identity covariance matrix. Robust policy choices are then obtained by solving

$$\min_{\{U_t\}} \max_{\{\omega_{t+1}\}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ X_t' Q X_t + U_t' R U_t - \theta \beta \omega_{t+1}' \omega_{t+1} \right\} \tag{A4}$$

subject to (A3) and given Y_0.

The Lagrangian associated to the *extremization*[2] problem above is

$$L = \sum_{t=0}^{\infty} \beta^t \left\{ X_t' Q X_t + U_t' R U_t - \theta \beta \omega_{t+1}' \omega_{t+1} + 2\beta \mu_{t+1}' (A X_t + B U_t + C \omega_{t+1} - X_{t+1}) \right\}. \tag{A5}$$

Notice, that we have dropped the expectation sign and ϵ_t from the problem above. This is because, a version of certainty equivalence continues to apply owing to the linear quadratic nature of the robust control problem. The first-order conditions with respect to U_t, X_t, and ω_{t+1} are given by:

$$\begin{aligned} 0 &= R U_t + \beta B' \mu_{t+1} \\ \mu_t &= Q X_t + \beta A' \mu_{t+1} \\ 0 &= \beta \theta \omega_{t+1} - \beta C' \mu_{t+1}. \end{aligned}$$

Solving the f.o.c.'s for U_t and ω_{t+1} and substituting in (A3) yields

$$X_{t+1} = A X_t - \beta (B R^{-1} B' - \beta^{-1} \theta^{-1} C C') \mu_{t+1}. \tag{A6}$$

Now, define $\tilde{B} = [B \ C]$ and $\tilde{R} = \begin{bmatrix} R & 0 \\ 0 & -\beta \theta I_\omega \end{bmatrix}$. We can rewrite (A6) as

$$X_{t+1} = A X_t - \beta \tilde{B} \tilde{R}^{-1} \tilde{B}' \mu_{t+1}. \tag{A7}$$

Collecting (A7) and (A6), we have the system of difference equations:

$$\begin{bmatrix} I & \beta \tilde{B} \tilde{R}^{-1} \tilde{B}' \\ 0 & \beta A' \end{bmatrix} \begin{bmatrix} X_{t+1} \\ \mu_{t+1} \end{bmatrix} = \begin{bmatrix} A & 0 \\ -Q & I \end{bmatrix} \begin{bmatrix} X_t \\ \mu_t \end{bmatrix}, \tag{A8}$$

[2]Extremization is the term used for the optimization problem where we both minimize and maximize the objective function with respect to some arguments.

or

$$L \begin{bmatrix} X_{t+1} \\ \mu_{t+1} \end{bmatrix} = N \begin{bmatrix} X_t \\ \mu_t \end{bmatrix}. \tag{A9}$$

Solving the model therefore boils down to finding a stabilizing solution to (A9) i.e., one which satisfies

$$\sum_{t=0}^{\infty} \beta^t X_t' X_t < +\infty.$$

The stabilizing solution is attained for a P such that $\mu_0 = PX_0$ where P solves the Riccati equation connected to the system of difference equations (A9). The solution $\mu_0 = PX_0$ replicates itself over time in the sense that

$$\mu_t = PX_t. \tag{A10}$$

A.1.1 Riccati equation

We mention above that finding a solution to (A9) involves finding a P which solves the Riccati equation connected to (A9). What is the Riccati equation and how do we find P? In practice, we can find P by using one of the *invariant subspace methods* documented in Hansen and Sargent (2004). This involves locating the stable invariant subspace of the *matrix pencil* $\lambda L - N$ of system (A9). We do this by first taking the matrices L and N and computing the generalized Schur decomposition of the pencil $\lambda L - N$. We can then construct P from the matrix of right Schur eigenvectors after carefully reordering the eigenvectors in terms of the stable and unstable eigenvalues.

Another approach for computing P is to solve the Riccati equation connected to (A9). The Riccati equation is more easily derived by writing down the Bellman equation of the dynamic problem. The Bellman equation associated to the dynamic problem (A4) is

$$V(X) = \min_U \max_\omega \{X'QX + U'RU - \theta\beta\omega'\omega + \beta V(X^*)\}, \tag{A11}$$

subject to

$$X^* = AX + BU + Cw. \tag{A12}$$

Since the objective is quadratic, and the constraint linear, we conjecture that the value function is also quadratic[3]:

$$V(X) = X'PX. \tag{A13}$$

Consider the inner maximization with respect to ω. Substituing (A13) for $V(X^*)$ and collecting terms in ω, the value of the inner maximization is:

$$(AX+BU)'D(P)(AX+BU) = \max_{\omega}\{-\theta\omega'\omega+(AX+BU+Cw)'P(AX+BU+Cw)\}. \tag{A14}$$

The first-order condition for ω yields

$$\omega = \theta^{-1}C'PX^*, \tag{A15}$$

or

$$\omega = (\theta I_{\omega} - C'PC)^{-1}C'P(AX+BU). \tag{A16}$$

Substituting (A15) in (A12), we get

$$X^* = (I - \theta^{-1}CC'P)(AX+BU). \tag{A17}$$

Using (A17) and (A15) and some algebraic manipulations, we can solve for $D(P)$. The result is

$$\begin{aligned} D(P) &= (I - \theta^{-1}PCC')^{-1}P \\ &= P + \theta^{-1}PC(I - \theta^{-1}C'PC)^{-1}C'P. \end{aligned} \tag{A18}$$

Having solved for the inner maximization, we can now solve the outer minimization problem. The problem reduces to

$$X'PX = \min_{U}\{X'QX + U'RU + \beta(AX+BU)'D(P)(AX+BU)\}. \tag{A19}$$

Equation (A19) illustrates how a concern for robustness modifies the problem of the policy-maker. Basically, it modifies the Bellman equation of the policy-maker by distorting the continuation value of the value function according to D. The f.o.c. for U implies that

$$U = -(R + \beta B'D(P)B)^{-1}B'D(P)AX \tag{A20}$$

[3]We rely on certainty equivalence to solve the non-stochastic version of the robust control problem. The same P solves both the stochastic and non-stochastic problem. The stochastic problem, however, adds a scalar to the value function i.e., $V(X) = X'PX + p$. We can easily show that $p = \frac{\beta}{1-\beta}\text{trace}(CC'P)$.

So, defining $F = -(R+\beta B'D(P)B)^{-1}B'D(P)A$ and substituting $U = FX$ in (A19), we get a recursion for P:

$$\begin{aligned} P &= Q + F'RF + \beta(A+BF)'D(P)(A+BF) \\ &= Q + \beta A'D(P)A - \beta^2 A'D(P)B(R+\beta B'D(P)B)^{-1}B'D(P)A. \end{aligned} \tag{A21}$$

Denote $T(v)$ as the operator

$$T(v) = Q + \beta A'vA - \beta^2 A'vB(R+\beta B'vB)^{-1}B'vA. \tag{A22}$$

Those familiar with solving dynamic linear quadratic control problems will recognize that the operator above appears in typical dynamic linear quadratic problems where there is no concern for robustness. Indeed, we solve for the value function by finding the fixed point of $T(v) = v$, the so-called Riccati equation. The usefulness of defining the operator above is that it illustrates how a desire for robustness affects the ordinary Bellman equation. A concern for robustness modifies the ordinary Riccati equation by finding the fixed point of $T(D(P)) = P$ with D given by (A18) rather than $T(P) = P$.

The P that solves the Riccati equation (A21) is the same P that will result from using the invariant subspace methods (again, see Hansen and Sargent 2004) to solve the system (A9). These different solution methods determine different solution algorithms, and the one we pick usually depends on the degree of accuracy or efficiency we require. Whichever method we use, once we obtain P we obtain the solutions $U = FX$ and $\omega = KX$ where

$$F = -(R+\beta B'D(P)B)^{-1}B'D(P)A, \tag{A23}$$

$$K = (\theta I_\omega - C'PC)^{-1}C'P(A+BF). \tag{A24}$$

A.1.2 Constructing the solution to the forward-looking problem

In a typical linear regulator problem, X_t is a state vector inherited from the past at time t, meaning that at time 0, X_0 is given. Here, however, $X_t = \begin{bmatrix} Y_t \\ Z_t \end{bmatrix}$ and only Y_t are inherited from the past. Z_t, on the other hand, are *jump* state variables that need to adjust for a solution to exist. Suppose that the dimension of Y_t and Z_t are n_y and n_z, respectively. Consider the system of difference equations given by (A9). Since $\mu_t = \begin{bmatrix} \mu_t^y \\ \mu_t^z \end{bmatrix}$ are the co-state variables associated with the state variables X_t, the system is of order $2(n_y+n_z)$. Therefore, for a solution to exist, we need $2(n_y + n_z)$ *boundary conditions*. Since the dynamic problem is of infinite horizon, the terminal states are unspecified. We obtain $n_y + n_z$ boundary conditions by imposing the transversality conditions

$$\lim_{t\to\infty} \beta^t \mu_t = 0.$$

At time 0, Y_0 is given. This gives us n_y initial conditions. However, because the Z_t are forward-looking variables, at time 0, we cannot take Z_0 as given. We obtain the missing n_z boundary conditions by imposing initial conditions for the co-state variables associated with Z_t; i.e., μ_t^z. But now, what initial conditions do we impose for μ_0^z? Following Currie and Levine (1993), it can be shown that the value function is decreasing in μ_t^z. Therefore, it is optimal to set $\mu_0^z = 0$.[4]

Recall that the solution to (A9) is

$$\mu_t = PX_t,$$

which we rewrite as

$$\begin{bmatrix} \mu_t^y \\ \mu_t^z \end{bmatrix} = \begin{bmatrix} P_{yy} & P_{yz} \\ P_{zy} & P_{zz} \end{bmatrix} \begin{bmatrix} Y_t \\ Z_t \end{bmatrix}. \tag{A25}$$

Using (A25) to solve for Z_t in terms of μ_t^z, we get

$$Z_t = -P_{zz}^{-1} P_{zy} Y_t + P_{zz}^{-1} \mu_t^z. \tag{A26}$$

The objective is to find the transition equation for the backward-looking state vector $\begin{bmatrix} Y_t \\ \mu_t^z \end{bmatrix}$ Since $\mu_{t+1}^z = P_{zy} Y_{t+1} + P_{zz} Z_{t+1}$,

$$\begin{bmatrix} Y_{t+1} \\ \mu_{t+1}^z \end{bmatrix} = \begin{bmatrix} I & 0 \\ P_{zy} & P_{zz} \end{bmatrix} \begin{bmatrix} Y_{t+1} \\ Z_{t+1} \end{bmatrix}. \tag{A27}$$

Now, since $U_t = FX_t$ and $\omega_{t+1} = KX_t$ with F and K given by (A23) and (A24), respectively,

[4] This pinpoints why there is a time-consistency problem in forward-looking problems. The policy-maker's loss function is decreasing in μ_t^z. At time 0, since μ_0^z must be non-negative, it is optimal to set it equal to 0. But from (A9) and $\mu_t = PX_t$, starting from $\mu_0^z = 0$, μ_t^z is non-zero forever after. Therefore, there is an incentive for the policy-maker to reset $\mu_\tau^z = 0$ for any $\tau > 0$. This is the time-consistency problem.

the evolution of the state vector X_t is

$$\begin{bmatrix} Y_{t+1} \\ Z_{t+1} \end{bmatrix} = (A + BF + CK) \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} + C\epsilon_{t+1}. \tag{A28}$$

Therefore,

$$\begin{aligned} \begin{bmatrix} Y_{t+1} \\ \mu^z_{t+1} \end{bmatrix} &= \begin{bmatrix} I & 0 \\ P_{zy} & P_{zz} \end{bmatrix} (A + BF + CK) \begin{bmatrix} I & 0 \\ -P_{zz}^{-1}P_{zy} & P_{zz}^{-}1 \end{bmatrix} \begin{bmatrix} Y_t \\ \mu^z_t \end{bmatrix} \\ &\quad + \begin{bmatrix} I & 0 \\ P_{zy} & P_{zz} \end{bmatrix} C\epsilon_{t+1}. \end{aligned} \tag{A29}$$

We can also express U_t and ω_{t+1} in terms of $\begin{bmatrix} Y_t \\ \mu^z_t \end{bmatrix}$:

$$\begin{bmatrix} U_t \\ \omega_{t+1} \end{bmatrix} = \begin{bmatrix} F \\ K \end{bmatrix} \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} F \\ K \end{bmatrix} \begin{bmatrix} I & 0 \\ -P_{zz}^{-1}P_{zy} & P_{zz}^{-}1 \end{bmatrix} \begin{bmatrix} Y_t \\ \mu^z_t \end{bmatrix}. \tag{A30}$$

Finally, for the stochastic problem, we can show that the value function in terms of the predetermined variables is

$$\begin{bmatrix} Y_0 \\ \mu^z_0 \end{bmatrix}' \tilde{P} \begin{bmatrix} Y_0 \\ \mu^z_0 \end{bmatrix} + \tilde{p}, \tag{A31}$$

where

$$\tilde{P} = \begin{bmatrix} I & 0 \\ P_{zy} & P_{zz} \end{bmatrix}' P \begin{bmatrix} I & 0 \\ P_{zy} & P_{zz} \end{bmatrix} \tag{A32}$$

and

$$\tilde{p} = \mathrm{trace}(\tilde{P}CC'). \tag{A33}$$

A.2 A simple rule

Suppose that instead of pursuing the Ramsey solution, the policy-maker chooses to set policy according to a simple rule:

$$U_t = FX_t. \tag{A34}$$

Hence, the policy-maker's problem is to minimize the loss function (A2) subject to the simple rule (A34) and the set of models (A3). The robust control problem with a simple rule is found by solving

$$\min_{\{F\}} \max_{\{\omega_{t+1}\}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ X_t' Q X_t + U_t' R U_t - \theta \beta \omega_{t+1}' \omega_{t+1} \right\}. \tag{A35}$$

For a given F, we can focus on determining the worst-case perturbation by solving the associated non-stochastic problem. The Lagrangian for the inner maximization becomes

$$L = \sum_{t=0}^{\infty} \beta^t \left\{ X_t' Q^* X_t - \theta \beta \omega_{t+1}' \omega_{t+1} + 2\beta \mu_{t+1}' (A^* X_t + C \omega_{t+1} - X_{t+1}) \right\}, \tag{A36}$$

where $Q^* = Q + F'RF$ and $A^* = A + BF$. The first order conditions with respect to ω_{t+1}, μ_t, and X_t are given by:

$$\omega_{t+1} = \theta^{-1} C' \mu_{t+1} \tag{A37}$$

$$\mu_t = Q^* X_t + \beta A^{*\prime} \mu_{t+1} \tag{A38}$$

$$X_{t+1} = A^* X_t + C \omega_{t+1}. \tag{A39}$$

By substituting the f.o.c. for ω_{t+1} into that for X_{t+1}, we get

$$X_{t+1} = A^* X_t + \theta^{-1} C' C \mu_{t+1}. \tag{A40}$$

Finally, collecting (A40) and the f.o.c. for μ_t yields the recursive system in X_{t+1} and μ_{t+1}:

$$\begin{bmatrix} I & -\theta CC' \\ 0 & \beta A^{*\prime} \end{bmatrix} \begin{bmatrix} X_{t+1} \\ \mu_{t+1} \end{bmatrix} = \begin{bmatrix} A^* & 0 \\ -Q^* & I \end{bmatrix} \begin{bmatrix} X_t \\ \mu_t \end{bmatrix}. \tag{A41}$$

The stabilizing solution is attained by finding a P that solves the Riccati equation connected to the system of difference equations (A41). Since $\mu_t = PX_t$ is true in all periods, substituting this condition into (A40) yields

$$X_{t+1} = \left[I - \theta^{-1} CC'P \right]^{-1} A^* X_t. \tag{A42}$$

From the f.o.c. for μ_t and (A42), we get

$$P = Q^* + \beta A^{*\prime} P \left[I - \theta^{-1} CC'P \right]^{-1} A^*. \tag{A43}$$

As before, it is instructive to derive the Riccati equation via the Bellman equation associated with the dynamic problem

$$V(X) = \min_F \max_\omega \{X'Q^*X - \theta\beta\omega'\omega + \beta V(X^*)\}, \tag{A44}$$

subject to

$$X^* = A^*X + C\omega. \tag{A45}$$

For a given F, we conjecture a quadratic value function $V(X) = X'PX$. The first-order condition with respect to ω for the inner maximization is

$$\omega = \theta^{-1}C'PX^*. \tag{A46}$$

Using the above in (A45) gives us

$$X^* = (I - \theta^{-1}CC'P)^{-1}A^*X. \tag{A47}$$

Restricting our consideration to the inner maximization with respect to ω, we find the continuation value

$$\begin{aligned}(A^*X)'D(P)(A^*X) &= \max_\omega\{-\theta\omega'\omega + X^{*\prime}PX^*\} && \text{(A48)}\\ &= -\theta X^{*\prime}PC\theta^{-1}\theta^{-1}C'PX^* + \beta X^{*\prime}PX^* && \text{(A49)}\\ &= X^{*\prime}\left[-\theta^{-1}PCC'P + P\right]X^* && \text{(A50)}\\ &= (A^*X)'(I - \theta^{-1}CC'P)^{-1}P(A^*X), && \text{(A51)}\end{aligned}$$

and, hence,

$$D(P) = (I - \theta^{-1}CC'P)^{-1}P. \tag{A52}$$

Given our previous conjecture on the form of the value function, the Bellman equation can be restated in terms of X, X^*, and $D(P)$:

$$X'PX = X'Q^*X + \beta(A^*X)'D(P)(A^*X). \tag{A53}$$

By substituting (A52) into (A53), the solution for P is

$$P = Q^* + \beta A^{*\prime}(I - \theta^{-1}CC'P)^{-1}PA^*. \tag{A54}$$

Having determined the worst-case pertubation for a given simple rule F, we can then find the rule that will minimize the loss function of the policy-maker. For the problem at hand, the loss function is the value function expressed in terms of predetermined variables:

$$L(F) = \begin{bmatrix} Y_0 \\ \mu_0^z \end{bmatrix}' \tilde{P} \begin{bmatrix} Y_0 \\ \mu_0^z \end{bmatrix} + \tilde{p}, \tag{A55}$$

where

$$\tilde{P} = \begin{bmatrix} I & 0 \\ P_{zy} & P_{zz} \end{bmatrix}' P \begin{bmatrix} I & 0 \\ P_{zy} & P_{zz} \end{bmatrix}, \tag{A56}$$

and

$$\tilde{p} = \text{trace}(\tilde{P}CC'), \tag{A57}$$

and P and its submatrices solve (A54).

Appendix B: An Alternative Algorithm for Solving Forward-Looking Linear Quadratic Robust Control

The solution methods presented above proceed by first finding a P that stabilizes the system given by full state and co-state of the economy (that is, including backward- and forward-looking variables and their associated co-states). We then use P to characterize the law of motion of the predetermined state variables, which is the solution we are ultimately after. An alternative algorithm is to directly find the solution of a system composed of only predetermined state variables. The next section outlines how this can be done. The algorithm was more efficient in solving the FM model for our parametrization of the model. To derive the alternative algorithm, we specify the robust control problem as follows:

$$\min_{U_t} \max_{\omega_{t+1}} \sum_{t=0}^{\infty} \beta^t \big\{ X_t' Q X_t + U_t' R U_t - \theta \beta \omega_{t+1}' \omega_{t+1} \big\}, \tag{B1}$$

subject to our approximating model, written in the form of a difference quation with one lead and one lag:

$$H_1 X_{t-1} + H_2 X_t + H_3 X_{t+1} + B U_t + C \omega_{t+1} = 0. \tag{B2}$$

The Lagrangian for this problem is

$$\begin{aligned} L \;=\; & \sum_{t=0}^{\infty} \beta^t \big\{ X_t' Q X_t + U_t' R U_t - \theta \beta \omega_{t+1}' \omega_{t+1} \big\} \\ & + 2\lambda_t \big\{ H_1 X_{t-1} + H_2 X_t + H_3 X_{t+1} + B U_t + C \omega_{t+1} \big\}. \end{aligned} \tag{B3}$$

The first-order conditions are

$$U_t \;:\; \beta^t R U_t + B' \lambda_t = 0 \tag{B4}$$

$$\omega_{t+1} \;:\; -\theta \beta \omega_{t+1} \beta^t + C' \lambda_t = 0 \tag{B5}$$

$$X_t \;:\; \beta^t Q X_t + H_1' \lambda_{t+1} + H_2' \lambda_t + H_3' \lambda_{t-1} = 0. \tag{B6}$$

Defining $\mu_t = \frac{\lambda_t}{\beta^t}$, we can rewrite the first-order conditions as

$$U_t = -R^{-1} B' \mu_t \tag{B7}$$

$$\omega_{t+1} = \theta^{-1} \beta^{-1} C' \mu_t \tag{B8}$$

$$Q X_t + H_1' \beta \mu_{t+1} + H_2' \mu_t + H_3' \beta^{-1} \mu_{t-1} = 0. \tag{B9}$$

By substituting the f.o.c.'s for U_t and ω_{t+1} into the constraint (B2), we obtain

$$H_1 X_{t-1} + H_2 X_t + H_3 X_{t+1} + BR^{-1}B'\mu_t + \theta^{-1}\beta^{-1}CC'\mu_t. \tag{B10}$$

From (B10) and the f.o.c. for X_t, we can construct a system of difference equations in X_t and μ_t :

$$\begin{bmatrix} H_1 & 0 \\ 0 & H_3' B^{-1} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ \mu_{t-1} \end{bmatrix} + \begin{bmatrix} H_2 & (\theta^{-1}\beta^{-1}CC' - BR^{-1}B') \\ Q & H_2' \end{bmatrix} \begin{bmatrix} X_t \\ \mu_t \end{bmatrix} + \begin{bmatrix} H_3 & 0 \\ 0 & H_1'\beta \end{bmatrix} \begin{bmatrix} X_{t+1} \\ \mu_{t+1} \end{bmatrix} = 0, \tag{B11}$$

which can be rewritten as

$$A_1 \begin{bmatrix} X_{t-1} \\ \mu_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} X_t \\ \mu_t \end{bmatrix} + A_3 \begin{bmatrix} X_{t+1} \\ \mu_{t+1} \end{bmatrix} = 0. \tag{B12}$$

We are searching for a solution for $\begin{bmatrix} X_t \\ \mu_t \end{bmatrix}$ in terms of what is predetermined at time t. At time t, X_{t-1} is inherited from the past; μ_{t-1} is the discounted $t-1$ shadow price of X_t and thus known at t (at time 0 it is optimal to initialize $\mu_{-1} = 0$). So we are looking for a solution characterized by

$$Y_t = NY_{t-1}, \tag{B13}$$

where $Y_s = \begin{bmatrix} X_s \\ \mu_s \end{bmatrix}$. For the system

$$A_1 Y_{t-1} + A_2 Y_t + A_3 Y_{t+1} = 0, \tag{B14}$$

let $Y_t = N_t Y_{t-1}$. Substituing this into (B14), the solution becomes

$$N_t = -(A_2 + A_3 N_{t+1})^{-1} A_1. \tag{B15}$$

To find a N that satisfies (B13), we start with a guess for N_{t+1} and iterate on (B15) until convergence.

Once we have solved the system of difference equations and obtained

$$Y_t = NY_{t-1}, \tag{B16}$$

$$\begin{bmatrix} X_t \\ \mu_t \end{bmatrix} = N \begin{bmatrix} X_{t-1} \\ \mu_{t-1} \end{bmatrix}, \tag{B17}$$

we need to back out decision rules for U_t and ω_{t+1}. We obtain from the respective first-order conditions for U_t and ω_{t+1}. From (B17),

$$\mu_t = \begin{bmatrix} 0 & I \end{bmatrix} N \begin{bmatrix} X_{t-1} \\ \mu_{t-1} \end{bmatrix}. \tag{B18}$$

Hence,

$$U_t = F \begin{bmatrix} X_{t-1} \\ \mu_{t-1} \end{bmatrix}, \tag{B19}$$

where $F = -R^{-1}B' \begin{bmatrix} 0 & I \end{bmatrix} N$, and

$$\omega_{t+1} = K \begin{bmatrix} X_{t-1} \\ \mu_{t-1} \end{bmatrix}, \tag{B20}$$

where $K = \theta^{-1}\beta^{-1}C' \begin{bmatrix} 0 & I \end{bmatrix} N$.

B.1 Dynamics in a stochastic system

In this section we consider the problem recast as a stochastic system. Beginning with

$$H_1X_{t-1} + H_2X_t + H_3X_{t+1} + BU_t + C\omega_{t+1} + C\epsilon_{t+1} = 0, \tag{B21}$$

and performing similar substitutions and manipulations, we obtain the difference system

$$A_1 \begin{bmatrix} X_{t-1} \\ \mu_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} X_t \\ \mu_t \end{bmatrix} + A_3 \begin{bmatrix} X_{t+1} \\ \mu_{t+1} \end{bmatrix} + C\epsilon_{t+1} = 0. \tag{B22}$$

Using the fact that $Y_{t+1} = NY_t$, we set

$$A_1Y_{t-1} + (A_2 + A_3N)Y_t + C\epsilon_{t+1} = 0, \tag{B23}$$

and therefore

$$Y_t = -(A_2 + A_3 N)^{-1} A_1 Y_{t-1} - (A_2 + A_3 N)^{-1} C \epsilon_{t+1} \tag{B24}$$

$$= N Y_{t-1} + C_N \epsilon_{t+1}. \tag{B25}$$

A_1, A_2, and A_3 are exactly as in the non-stochastic case because of certainty equivalence. If we take expectations at time t, $\epsilon_{t+1} = 0$ and the last term drops out. Therefore, our solutions to the non-stochastic and stochastic problems are equivalent.

Appendix C: The Kalman Filter

Suppose that the state of the economy evolves according to

$$X_{t+1} = AX_t + C\epsilon_{t+1}, \tag{C1}$$

but assume that we observe only a subset of the state variables, possibly with some measurement error

$$\begin{aligned} Y_{t+1} &= \tilde{h}X_{t+1} + \tilde{D}\epsilon_{t+1} \\ &= \tilde{h}AX_t + (\tilde{h}C + \tilde{D})\epsilon_{t+1} && \text{(C2)} \\ &= hX_t + D\epsilon_{t+1}. && \text{(C3)} \end{aligned}$$

Let $y^t = [y_t, y_{t-1}, y_{t-2}, ..., y_0]$ be information available at time t. The best prediction of the state at $t+1$ given information at time t is:

$$X_{t+1|y^t} = AX_{t|y^t}, \tag{C4}$$

and the forecast error for X_{t+1} is:

$$X_{t+1} - X_{t+1|y^t} = A(X_t - X_{t|y^t}) + C\epsilon_{t+1}.$$

The mean squared error of our forecast is:

$$\begin{aligned} \Omega_{t+1|y^t} &= E[(X_{t+1} - X_{t+1|y^t})(X_{t+1} - X_{t+1|y^t})'] && \text{(C5)} \\ &= A\Omega_{t|y^t}A' + CC'. && \text{(C6)} \end{aligned}$$

Conversely, the best prediction of Y_{t+1} at time t is:

$$Y_{t+1|y^t} = hX_{t|y^t}. \tag{C7}$$

The forecast error is

$$v_{t+1} = Y_{t+1} - Y_{t+1|y^t} = h(X_t - X_{t|y^t}) + D\epsilon_{t+1}.$$

The MSE of v_{t+1} is

$$\begin{aligned}\Delta_{t+1|y^t} &= E(v_{t+1}v'_{t+1}|y^t) && \text{(C8)}\\ &= E[(h(X_t - X_{t|y^t}) + D\epsilon_{t+1})(h(X_t - X_{t|y^t}) + D\epsilon_{t+1})'|y^t] && \text{(C9)}\\ &= h\Omega_{t|y^t}h' + DD'. && \text{(C10)}\end{aligned}$$

Now, note that from linear projection theory,

$$X_{t+1} = X_{t+1|y^t} + \kappa_{t+1}v_{t+1} + \xi_{t+1}. \qquad \text{(C11)}$$

We are interested in deriving κ_{t+1}. Take expectation given y^t to obtain,

$$\begin{aligned}\kappa_{t+1} &= E[(X_{t+1} - X_{t+1|y^t})v'_{t+1}]E(v_{t+1}v'_{t+1}|y^t)^{-1} && \text{(C12)}\\ &= = [A\Omega_{t|y^t}h' + CD']\Delta^{-1}_{t+1}. && \text{(C13)}\end{aligned}$$

Given κ_{t+1} and (C11), we can then develop a recursion for updating our estimates of the state and its mean square error. We do this by taking expectations of (C11) given y^{t+1}. We obtain

$$X_{t+1|y^{t+1}} = X_{t+1|y^t} + \kappa_{t+1}vt + 1. \qquad \text{(C14)}$$

Using (C11) and (C14), we can similarly obtain

$$\Omega_{t+1|y^{t+1}} = \Omega_{t+1|y^t} - \kappa_{t+1}\Delta_{t+1}\kappa'_{t+1}. \qquad \text{(C15)}$$

(C14) and (C15) are the updating equations of the Kalman filter. They tell us how to update our estimates of the state and the associated MSE once a new piece of data arrives.

The Kalman filter is useful in constructing the likelihood function. Recall that,

$$Y_{t+1} = Y_{t+1|y^t} + v_{t+1}.$$

So the conditional distribution of $Y_{t+1|y^t}$ is $N(Y_{t+1|y^t}, \Delta_{t+1})$. Therefore, we can use the Kalman filter to write down the contribution of each piece of data to the likelihood function.

Appendix D: Impulse Responses from a VAR

To estimate the historical impulse responses of inflation, interest rates, and the output gap to an interest rate shock, output gap shock, and inflation shock, we estimate a restricted VAR using data from 1980Q1 to 2005Q1. Our VAR is ordered as follows: inflation, consumption growth, log of real investment, output gap, log of nominal foreign exchange rate, and the 90-day commercial paper rate. We also include the following exogenous variables: U.S. output gap, U.S. inflation, log of world commodity prices, and the federal funds rate. Finally, we include a dummy variable that takes the value 1 from 1991Q1 to 2005Q1 to indicate the period of explicit inflation targeting in Canada.

We use historical data from the Bank of Canada's Quarterly Projection Model database. Real variables are based on the GDP deflator. Quarter-over-quarter inflation and quarter-over-quarter consumption growth are calculated as the first-difference of the log of GDP deflator and log of real consumption, respectively. We calculate the output-gap variables as deviations of the log of real GDP from a linear quadratic trend. Finally, the foreign exchange rate used is the log of the trade-weighted G6 nominal exchange rate. To illustrate, an appreciation of the Canadian dollar relative to a basket of foreign currencies translates to a decrease in the nominal exchange rate.

We estimate our restricted VAR by forcing some coefficients to be zero and excluding non-significant variables from the regression. Table D1 presents the equations we estimate and the lag order of the included variables. A '0' indicates that the variable enters the equation contemporaneously, a '1' indicates the first lag of the variable, and so on. Our methodology and results are comparable to those in Murchison, Rennison, and Zhu (2004). Figure D1 shows the impulse responses of inflation, the output gap, and interest rates to an interest rate shock, output gap shock, and inflation shock.

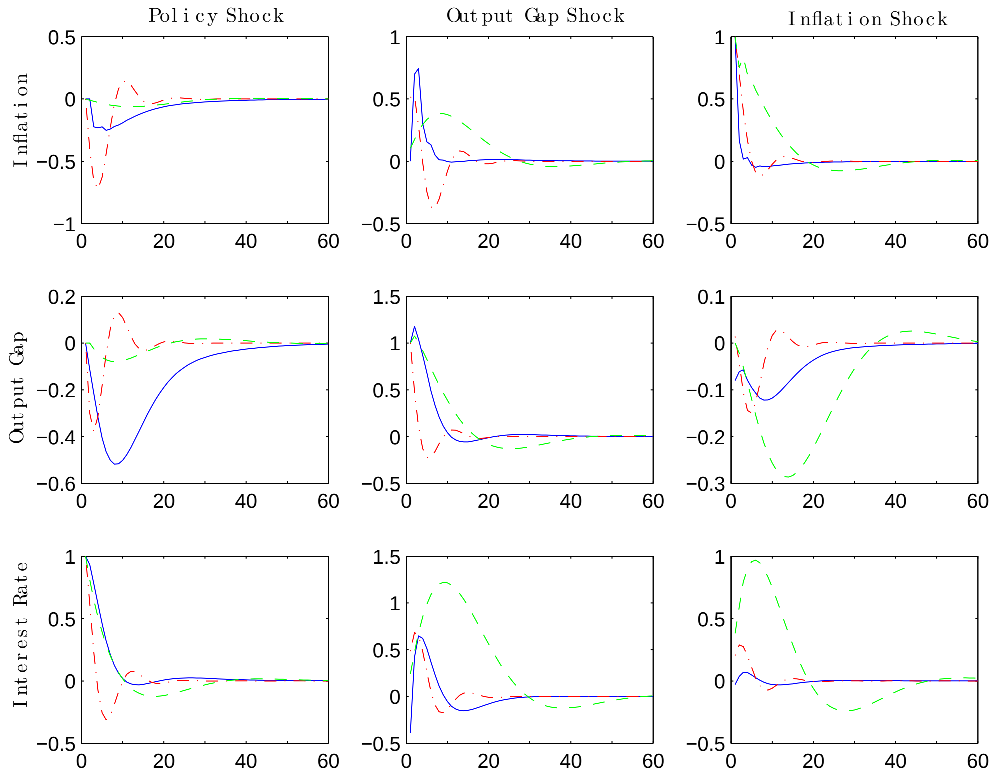


Figure D1: Historical impulse responses vs model-based impulse responses

Table D1: Benchmark VAR

		Equations					
		inf	grcons	lrinv	ogap	lforexn	r1n
	inf	1	1	1		1	1
	grcons	1	1 to 2		1		
	lrinv			1 to 2			
	ogap	1 to 2	1	1	1 to 2		1 to 2
Variables	lforexn		1	2	1	1 to 2	
	r1n		1	1	1	1 to 2	1
	ogapUS	0	0 to 1	0 to 1	0 to 1		
	infUS	0		0	0		
	lpcom	0 to 1	0			0 to 1	0
	ffr		0 to 1			1	
	dummy		0		0		

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages, and are available on the Bank's website (see bottom of page). ***Les documents de travail sont généralement publiés dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue. On peut les consulter dans le site Web de la Banque du Canada, dont l'adresse est indiquée au bas de la page.***

2006

2006-12	The Welfare Implications of Inflation versus Price-Level Targeting in a Two-Sector, Small Open Economy	E. Ortega and N. Rebei
2006-11	The Federal Reserve's Dual Mandate: A Time-Varying Monetary Policy Priority Index for the United States	R. Lalonde and N. Parent
2006-10	An Evaluation of Core Inflation Measures	J. Armour
2006-9	Monetary Policy in an Estimated DSGE Model with a Financial Accelerator	I. Christensen and A. Dib
2006-8	A Structural Error-Correction Model of Best Prices and Depths in the Foreign Exchange Limit Order Market	I. Lo and S.G. Sapp
2006-7	Ownership Concentration and Competition in Banking Markets	A. Lai and R. Solomon
2006-6	Regime Shifts in the Indicator Properties of Narrow Money in Canada	T. Chan, R. Djoudad, and J. Loi
2006-5	Are Currency Crises Low-State Equilibria? An Empirical, Three-Interest-Rate Model	C.M. Cornell and R.H. Solomon
2006-4	Forecasting Canadian Time Series with the New Keynesian Model	A. Dib, M. Gammoudi, and K. Moran
2006-3	Money and Credit Factors	P.D. Gilbert and E. Meijer
2006-2	Structural Change in Covariance and Exchange Rate Pass-Through: The Case of Canada	L. Khalaf and M. Kichian
2006-1	The Institutional and Political Determinants of Fiscal Adjustment	R. Lavigne

2005

2005-45	An Evaluation of MLE in a Model of the Nonlinear Continuous-Time Short-Term Interest Rate	I. Lo
2005-44	Forecasting Core Inflation in Canada: Should We Forecast the Aggregate or the Components?	F. Demers and A. De Champlain
2005-43	The 1975–78 Anti-Inflation Program in Retrospect	J. Sargent

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
Telephone: 1 877 782-8248
(toll free in North America)
Email: publications@bankofcanada.ca
Website: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Téléphone : 1 877 782-8248 (sans frais en Amérique du Nord)
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca