Bank of Canada Banque du Canada

Working Paper 2002-13 / Document de travail 2002-13

Towards a More Complete Debt Strategy Simulation Framework

by

David Jamieson Bolder

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2002-13

May 2002

Towards a More Complete Debt Strategy Simulation Framework

by

David Jamieson Bolder

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
dbolder@bankofcanada.ca

Contents

Acknowledgements

I would like to especially thank, without implicating, Toni Gravelle and Kevin Moran of the Bank of Canada as well as Mark Reesor of the University of Western Ontario. I also want to thank Jeremy Graveline of Stanford University. I would also like to express my appreciation to Anders Holmlund, Lars Hörngren, and Pål Bergström of the Swedish National Debt Office (Riksgälds Kontoret) as well as to Lars Risbjerg and Xiaoming Chen of the Danish National Bank. These individuals are, to my mind, leaders in the area of sovereign debt management and have acted as the inspiration for many of the ideas in this text. Any errors, omissions, or inconsistencies are, of course, my sole responsibility.

Abstract

An effective technique governments use to evaluate the desirability of different financing strategies involves stochastic simulation. This approach requires the postulation of the future dynamics of key macroeconomic variables and the use of those variables in the construction of a debt charge distribution for each individual financing strategy. Summary measures of the resulting debt charge distributions permit comparison of the alternative financing strategies. To defensibly generate a debt charge distribution for a given financing strategy, however, one must have a good model to simulate the future dynamics of a set of key macroeconomic variables into the future. This paper suggests a reduced-form approach to describe the key elements of the stochastic model used to analyze the Government of Canada's debt strategy problem. To this end, a simple algorithm is proposed for the simultaneous simulation of the business cycle, the government's financial position, and the term structure of interest rates. The approach uses the constant parameter hidden-Markov model introduced by Hamilton (1989) in combination with the class of affine term-structure models.

JEL classification: C0, C5, G0

Bank classification: Interest rates; Econometric and statistical methods; Debt management

Résumé

La simulation stochastique est une technique utilisée par les gouvernements pour évaluer l'utilité de diverses stratégies de financement. Pour pouvoir l'effectuer, il faut commencer par formuler un ensemble d'hypothèses sur le comportement dynamique qu'auront certaines variables macroéconomiques clés, puis effectuer, à l'aide de ces variables, une distribution du service de la dette pour chacune des stratégies envisagées. Les statistiques sommaires ainsi produites de ces distributions permettent de comparer les résultats que donne chacune des stratégies de financement, ce qui illustre la nécessité de disposer d'un bon modèle de simulation du comportement dynamique des variables concernées. Dans le présent document, nous proposons d'utiliser une représentation de forme réduite pour décrire les principaux éléments du modèle stochastique servant à analyser le problème de la stratégie de la dette du gouvernement canadien. À cette fin, nous proposons d'utiliser un algorithme simple pour simuler simultanément le cycle économique, la position financière du gouvernement et la structure des taux d'intérêt. Dans le cadre de cette démarche, nous utilisons, conjointement avec des modèles affines relatifs à la structure des taux d'intérêt, le modèle à paramètre constant latent de Markov introduit par Hamilton (1989).

Classification JEL : C0, C5, G0

Classification de la Banque : Taux d'intérêt; Méthodes économétriques et statistiques; Gestion de la dette

1 Introduction

Governments typically borrow funds in domestic and/or foreign capital markets to finance any excess of government expenditures over revenues as well as to refinance maturing debt issued during previous periods. Every government engaged in this practice must determine what is the best way to borrow these funds—whether through short-term treasury bills, longer-term coupon bonds, or a combination of these strategies. This is often called the *debt strategy* problem. To resolve this problem, the government must consider a wide number of alternative financing strategies. The ultimate decision hinges upon the cost and risk characteristics of the portfolios arising from these financing strategies.

In recent years, a number of sovereigns have suggested the use of a stochastic simulation framework to investigate this problem.[1] Conceptually, this is straightforward. One merely selects a financing strategy and then applies it under a variety of stochastically generated future scenarios. In this way, one constructs a distribution of *debt charges* for a given financing strategy. Debt charges are the interest costs associated with financing a given government debt portfolio. Comparison of financing strategies is then reduced to making comparisons between their respective debt charge distributions. In other words, differences between financing strategies can be identified through examination of summary measures of their generated debt charge distributions.

To this end, one can use a number of alternative cost and risk measures to distinguish between the desirability of different financing strategies. For example, one could compare the *expected* debt charges of one financing strategy relative to another. Conversely, one might wish to look at all financing strategies that maintain debt charge *variance* below a predetermined level over a given time interval.[2] In each case, however, the computation of these measures requires one to actually generate the debt charge distribution. Indeed, the greater the confidence one places in the generated debt charge distribution, the more confidently one can distinguish between financing strategies.

This paper is *not* directly concerned with examining and comparing different financing strategies. Instead, it focuses on the stochastic simulation algorithm that is a critical subcomponent of the debt strategy problem. In particular, to defensibly generate a debt charge distribution for a given financing strategy, one must reasonably simulate the future dynamics of a set of key macroeconomic variables. The stochastic processes employed in this debt strategy simulation framework are, therefore, analyzed in substantial detail. We are convinced that if the stochastic processes can be selected and structured in a reasonable and defensible manner, the subsequent analysis will be more meaningful.

[1] Excellent references include the Danish National Bank (1998a,b), Hörngren (1999), Holmlund (1999), and Bergström and Holmlund (2000).

[2] In addition, the Danish National Bank (1998a) suggested the idea of using a *cost-at-risk* analysis, which is very similar in spirit to *value-at-risk*.

This is a rather broad problem. We need to determine the macroeconomic variables of the greatest importance, decide on the appropriate model to describe their future realizations, and estimate or calibrate any model parameters. To assist in this effort, we define certain guidelines for developing of a framework that generates the random processes that describe the evolution of our key macroeconomic variables. Specifically, we require that:

- the stochastic processes employed capture the general empirical properties of the individual random macroeconomic variables we wish to model;
- the model capture any important co-movements between those individual random variables;
- the model be conceptually simple, with a minimum of parameters, and have sufficient *flexibility* so as to lend itself to sensitivity analysis and stress-testing;
- the computational time required to generate future sample paths not be prohibitively expensive.

The final guideline is important because, to construct our debt charge distributions for a number of different financing strategies, we must consider literally hundreds of thousands of future outcomes. A lengthy algorithm for the generation of our stochastic processes can render the computation of these debt charge distributions infeasible. Keeping these guidelines in mind, we can consider the basic macroeconomic ingredients required in our stochastic simulation framework.

One obvious macroeconomic quantity required is the term structure of interest rates. Future interest rates *determine* the cost of future borrowing. Clearly, future term structure outcomes are not known in advance and thus the term structure must be modelled explicitly. In an earlier paper (Bolder 2001) the author considers the applicability of employing the affine class of term structure models in the debt strategy problem. This is a reasonable place to start—and, indeed, these models will form the foundation for our term structure analysis—but more work is required because the term structure of interest rates is not the *only* source of future uncertainty. A second key random variable is the financial position of the government. Simply put, the future state of the government's finances *determines* how much the government needs to borrow in future periods.[3]

The government's financial position is a function of its receipts and its *expenditures*. While the government makes detailed plans with respect to those receipts and expenditures, there still remains a non-trivial amount of residual future uncertainty. For instance, government tax receipts are related in an important manner to the state of the economy. Tax receipts tend to fall in recessionary periods while they tend to rise during periods of strong economic growth. Government expenditures exhibit a similar pattern. Government

[3]Technically, the borrowing requirement in any given future period depends upon the state of the government's finances *and* maturing debt. The maturing debt, however, is an entirely deterministic function of the composition of the government's debt portfolio and thus need not be modelled as a stochastic process.

spending programs, which constitute the bulk of the government's expenditures, also exhibit this business-cyclical pattern.[4] The government's financial position, therefore, depends on the general *macroeconomic* conditions that prevail in the economy during that period.

We also know that the term structure of interest rates is *not* independent of the macroeconomy. In particular, we empirically observe a rather steep term structure preceding periods of economic expansion and a flat, or inverted, term structure prior to recessionary periods. In short, the steepness of the term structure of interest rates actually serves as a good *leading* indicator of economic activity.

Based on this reasoning, we expect to observe an empirical relationship between the surplus position of the government, the evolution of the term structure of interest rates, and the general macroeconomy. Indeed, we believe that any reasonable stochastic simulation framework for the debt strategy problem must incorporate this fact. In reality, however, this is a two-way relationship. That is, the government's financial position and the term structure of interest rates influence the state of the macroeconomy and vice versa. One of the key simplifying assumptions of this work is that the relationship is one-directional. Simply put, the macroeconomic state is assumed to influence the evolution of the term structure and the government's financial position, but these variables do *not* impact the macroeconomic state.[5] Conceptually, therefore, we assume that the state of the macroeconomy acts as a one-way *link* between the government's financial position and the term structure of interest rates. The primary objective of this paper is, conditional on this key assumption, to describe a simple, reduced-form model that *jointly* captures the random evolution of these three fundamental macroeconomic quantities.

Our suggested approach is described in the following three sections. Section 2 highlights the challenges associated with constructing the joint model. The goal is to illustrate what the joint model must accomplish to be a useful component of our simulation framework. Section 3 describes Hamilton's (1989) hidden-Markov model. This incredibly convenient approach is used to describe the evolution of the business cycle. In fact, it forms the *backbone* of our joint model. Section 4 describes in detail our reduced-form model of the business cycle, the term structure of interest rates, and the government's financial position.

2 The Problem

Our goal is not to generate ex-post measures of the state of the economy, precisely forecast future interest rate outcomes, nor construct a full-blown model of the economy that is entirely consistent with macroeconomic theory. It is to find a model with sufficient flexibility to permit the simulation of future economic business cycles in a manner that is broadly consistent with past behaviour. At the same time, we require arbitrage-

[4] This is particularly evident with automatic stabilizers such as social welfare spending.

[5] This is clearly a simplification of reality made to facilitate the analysis. We hope to be in a position to relax this assumption in future work.

free term structures and a reasonable model of future government financial positions.[6] More specifically, we desire a simple model that describes the fundamental stylized facts observed in interest rates, the business cycle, and the government's financial position. In this section, we will discuss the challenges of modelling the government's financial position and the term structure on an individual basis. Then we will consider how one might attempt to model their evolution in a joint fashion.

First, we consider how we could model the government's financial position. We could use a stochastic process to describe the evolution of this macroeconomic variable. A reasonable suggestion might be the mean-reverting Ornstein-Uhlenbeck process. We might, therefore, define the government's financial position as a stochastic process $\{F(t), t \in [0, T]\}$ with the following form,

$$dF(t) = \alpha(\beta - F(t))dt + \xi dW(t), \tag{1}$$

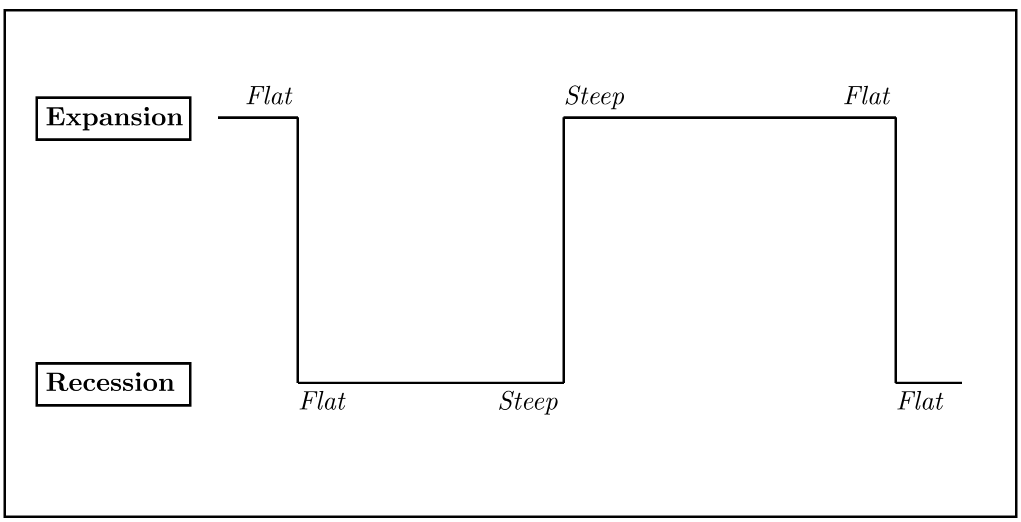
where $\{W(t), t \in [0, T]\}$ is a standard, scalar Wiener process. This is a mean-reverting process with a long-term mean denoted by β and a mean-reversion parameter of α. A bit of reflection suggests that this is not a bad start. A healthy government financial position will typically vary around zero, because a government that incurs a deficit in a given period can alter its discretionary spending to help move its financial position into positive territory in subsequent periods. On the opposite side of the spectrum, strong spending pressures from various political groups imply that a large surplus position will also not likely persist over time. Finally, random shocks can hit the economy that might move the government into either a deficit or surplus position. In this model, the variability of the government's financial position is governed by the Wiener process, $W(t)$, and the volatility parameter, ξ.

This is not the whole story. Unfortunately, the stand-alone stochastic process for the government's financial position suggested in equation (1) says nothing about the origin of these random shocks. We argue that the underlying state of the macroeconomy is an important determinant in the random surprises to the government's financial position. Why is this the case? There are a number of ways in which the business cycle can matter in this respect. First, *automatic stabilizers* such as unemployment insurance payments and welfare spending imply that government expenditure will automatically increase as the economy enters into a recession. Moreover, as output falls we would expect to see an associated decrease in tax revenues. Falling revenue and increasing expenditure would indicate a smaller surplus or a possible deficit position. Conversely, the more robust tax revenues and lower social welfare spending that typically accompany an expansionary period would create a larger surplus position for the government. These pressures are not likely to be extreme, because the government will (and does) use discretionary spending to control, as much as possible, its financial position. Nevertheless, the previous logic implies that there is an important role for the state of macroeconomy in determining a government's financial position.

[6] If an interest rate model is not appropriately constructed, it can lead to negative forward interest rates and hence arbitrage opportunities. We avoid this problem by considering models that, by construction, are free of arbitrage.

We next consider how we could model the term structure of interest rates. Fortunately, previous work has been done on applying the class of affine term-structure models to the debt strategy problem. Note, however, that the term structure of interest rates also depends on the state of the macroeconomy. Substantial evidence suggests that the relationship between the term structure of interest rates and the business cycle is fairly involved. Clinton (1994), Atta-Mensah and Tkacz (1998), and Cozier and Tkacz (1994) provide both empirical and theoretical arguments that describe how the term structure is a leading indicator of business cycles. Specifically, the steepness of the term structure—as measured by the differential between short-term and long-term interest rates—provides information regarding the future economic state. A flat or inverted term structure (i.e., a small or negative differential) is typically followed by a slowdown in economic activity. Increases in output, conversely, generally follow a steep term-structure environment. Figure 1 provides a stylized view of this relationship.

Figure 1: **Interest-Rate Differentials and Economic Cycles**: This figure represents a stylized view of the relationship between the interest-rate differential in long- and short-term interest rates and the business cycle. Observe that preceding an economic slowdown this differential is small (i.e., the term structure is flat), while the differential is large (i.e., the term structure is steep) before economic expansion.



The class of affine term-structure models that we employ in our modelling efforts says nothing about the general macroeconomic conditions in the economy. How might we then hope to incorporate the business cycle into our term-structure model? One possible approach would be to assume that the parameterization of any term-structure model is different depending on the state of the economy. For example, we might assume that the parameter set, θ_0, describes the evolution of the term structure of interest rates under a recessionary period. Similarly, we could let an alternative parameter set, θ_1, summarize expansionary term-structure dynamics. While this approach offers conceptual simplicity and permits a relatively straightforward approach to parameter estimation, it is logically flawed. In particular, observe in Figure 1 that the periodicities of the business cycle and the interest rate cycle do *not* coincide; that is, the flattest term structure—though related to the possibility of recession—will not be temporally aligned with the trough of the business cycle. Solving this problem is one of the principal challenges in developing our stochastic simulation framework.

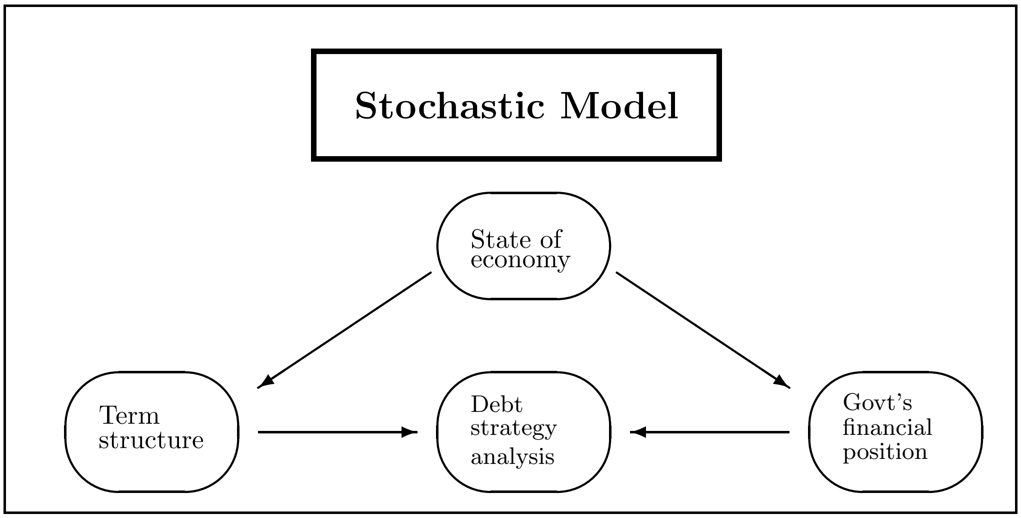
The previous logic suggests that, if we wish to jointly describe the government's financial position and the term structure of interest rates, we should incorporate the business cycle. Indeed, in our analysis, the business cycle will serve as a one-directional *bridge* between the government's financial position and the term structure of interest rates.[7] Figure 2 illustrates structurally how we intend to proceed with the construction of our joint model. If the business cycle is to be the link between the term structure of interest rates and the government's financial position, we need a good model for the evolution of the business cycle. This leads us to inquire how we might, in a modelling sense, determine the current state of the economy. One obvious, but also quite intuitive, step would be to use the rate of output growth in the economy. In situations of economic expansion, we would expect to observe positive growth of 2 to 4 per cent in proxies such as gross domestic product (GDP) and industrial production. Conversely, we anticipate that recessionary periods will be typified by zero or negative output growth rates.

While output growth might be a convenient way to determine the state of the economy, it is certainly not a model of the business cycle. At first glance, modelling the evolution of the business cycle may seem to be a daunting task. Indeed, it is quite involved. Fortunately, there is a large amount of literature on the subject. Hamilton (1989, 1990, 1996), Kim (1994), Filardo (1993, 1998), Filardo and Gordon (1993, 1994), Durland and McCurdy (1994), and Diebold, Lee, and Weinbach (1993) examine a range of hidden-Markov models that describe the non-linear relationship between output and the underlying state of the economy. Their models generally assume that the economy can be in one of two possible states: recession or expansion. They then specify probabilities that govern the transition of the economy over time from one state to another. For example, if the economy is currently in expansion we must specify the probability of moving into a recession in the current period as well as the probability of remaining in an expansion. A similar set of transition probabilities must be determined for a situation where the economy is currently in a recession. It is upon this conceptual framework that we will build the foundation of our model of the macroeconomy.

In addition, consider the flexibility of the hidden-Markov model. We can adapt this framework to improve the breadth of our analysis in a consistent manner. What does this mean? The stochastic simulation

[7] Another, more practical, reason for modelling the business cycle is the length of the time interval under examination. The typical time horizon for this simulation analysis is 10 years. We would reasonably expect to experience two or more business cycles over a given 10-year period. In the short-term simulation exercises used to compute quantities such as *value-at-risk*, one can usually safely ignore the importance of the business cycle. Given the high probability of experiencing multiple business cycles, however, it would be difficult to defend ignoring this element in our analysis.

Figure 2: **The Basic Framework**: To establish the link between the government's financial position and the term structure of interest rates, we will use the economic business cycle. To proxy the unobservable movement of the business cycle, we use the growth rate of output as measured, for example, by an official statistic such as the gross domestic product.



framework proposed in this paper attempts to model the volatility of key macroeconomic variables under normal market conditions. It is prudent, however, to also examine the impact of events that might occur under abnormal market conditions. This is called *stress testing* and it represents an important element in our debt strategy analysis. Using our hidden-Markov model, we could add a third state with a very small transition probability that leads to extreme outcomes. This could be a large parallel shift in the term structure of interest rates, substantially increased interest rate volatility, or term-structure inversion. It could also involve negative outcomes for the government's financial position. This is similar in spirit to the *peso problem*. That is, the reason for a large, black-market depreciation of the Mexican peso during the late 1970s was attributed to the small probability of a hidden regime with huge consequences.[8] Why is this sort of analysis useful? Standard stress-testing methodologies do not make any statement about the probability of their occurrence. This is a weakness because, in the absence of any sense of their likelihood, stress-test results are difficult to interpret. In this setting, we could assign some arbitrary but small probability of occurrence to this third state. These negative outcomes would then be reflected in the summary measures of the associated debt charge distribution. This methodology would, therefore, provide us with the ability to determine how large the probability of occurrence must be for it to have an important impact on the

[8]In this case, the event was the devaluation of the peso by Mexican authorities. It eventually occurred in 1982. For more details see De Grauwe (1989, page 129).

associated debt charge distribution.

We have established that, within a business cycle, a constant parameterization of the term structure of interest rates is *not* appropriate. That is, the term-structure cycle and the business cycle do not coincide. To capture this relationship between the term structure of interest rates and the business cycle, we propose a simple technique that uses a transformation of one of the key outputs of Hamilton's (1989) model, which plays a crucial role in the discussion to follow. In fact, it is of such critical importance to our analysis that section 3 is dedicated entirely to examining its logic and intuition.

3 Modelling the Economic Cycle

Ultimately, we require a model that describes the time-series evolution of output growth over some interval of time. The model should have a structure that permits output growth to exhibit different behaviour during recessionary and expansionary periods. Hamilton (1989) did the original work in this area and this section is based almost entirely on his work. To motivate his path-breaking approach to this problem, consider the following extremely simple time-series model over the predefined time interval, $[0, T]$,

$$y_t - \phi y_{t-1} = \begin{cases} \mu_0 + \epsilon_t : t \in [0, s) \\ \mu_1 + \epsilon_t : t \in [s, T] \end{cases}, \tag{2}$$

where $\mu_1 > \mu_0$. This model has a structural break over the interval $[0, T]$ occurring at time s. Clearly, one could generalize equation (2) to incorporate multiple structural breaks, but the analysis would be fundamentally unchanged. The first thing to note is that s is a deterministic point in time. This means that, to use this model, one must exogenously specify the point, or points, in time where the mean in this process shifts back and forth between μ_0 and μ_1. This is not necessarily a problem unless, of course, one would like to use this model to forecast future outcomes. If that is the case, then it is not entirely obvious how one should proceed. A very reasonable alternative would be to treat the shift in mean as a random event. One must then specify the probability distribution that governs the shift from μ_0 to μ_1.

To accommodate the movement to a stochastic mean, we need more structure. To accomplish this, we define a new random variable, S_t. This random variable takes on one of two integer values, depending on the state of the process. More precisely, we have

$$S_t = \begin{cases} 0 : \text{process in state 0} \\ 1 : \text{process in state 1} \end{cases}. \tag{3}$$

This would permit us to rewrite the model as

$$y_t - \phi y_{t-1} = \mu_{S_t} + \epsilon_t, \tag{4}$$

for all $t \in [0, T]$ where,

$$\mu_{S_t} = \begin{cases} \mu_0 : S_t = 0 \\ \mu_1 : S_t = 1 \end{cases}, \tag{5}$$

and,

$$\epsilon_t \sim \mathcal{N}(0, \sigma^2). \tag{6}$$

This is quite a good start, but it introduces a new challenge. Namely, we need to find some way to describe our unobserved state variable, S_t. Hamilton (1989) solved this problem by permitting S_t to follow a Markov chain. The theory of Markov chains is quite rich but, fortunately for our current application, we need understand only a handful of their properties. In section 3.1, therefore, we will briefly consider some basic facts about Markov chains to facilitate our discussion.

3.1 Some Markov chain background

A *Markov chain* is a discrete-state stochastic process that can be defined in either continuous or discrete time.[9] Our application will focus on discrete-time Markov chains. Conceptually, this class of processes is similar to the more general (and subtle) theory relating to continuous-time, continuous-state Markov processes.[10] To get started, let us consider an integer-valued, time-indexed sequence of random variables, $\{S_t, t \in \{1, ..., T\}$. Moreover, the probability that S_t takes on a value j given that its current value is i depends on that previous value alone. That is,

$$\begin{aligned} \mathbb{P}[S_t = j \mid S_{t-1} = i, S_{t-2} = k, ...] &= \mathbb{P}[S_t = j \mid S_{t-1} = i], \\ &= p_{ij}. \end{aligned} \tag{7}$$

This is the seminal property of a Markov chain. Loosely speaking, it is a process that has a very limited memory. At any given point in time, the distribution of outcomes in the subsequent state depends only on its current state. As stated in equation (7), we let p_{ij} represent the probability that the process finds itself in state i given that it previously found itself in state j. These are called the transition probabilities of our

[9]This discussion follows, in more or less equal parts, from the excellent analysis in Hamilton (1994), Brémaud (1999), and Meyn and Tweedie (1993). This section is, by construction, fairly terse. Appendix A, however, expands somewhat on these results.

[10]The theory of Markov chains can also be presented in a measure-theoretic framework. Nevertheless, as our use of these processes will be very straightforward, this extra mathematical machinery does not seem necessary or even desirable.

Markov and we define them in the following matrix form,

$$\underbrace{\{p_{ij}\}_{i,j=1,2,\ldots,N}}_{\substack{\text{Set of}\\ \text{transition}\\ \text{probabilities}}} \equiv \underbrace{\begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}}_{\text{Transition matrix: } P} \tag{8}$$

where,

$$p_{i1} + p_{i2} + \cdots + p_{iN} = \sum_{j=1}^{N} p_{ij} = 1 \tag{9}$$

for all $i = 1, ..., N$. Or, more simply, the columns of the transition matrix must sum to unity.

One key feature of Markov chains is that the probability that an observation from state i will be followed, in m periods, by an observation from state j is given by the jth row and ith column from the matrix P^m.[11] Or,

$$\mathbb{P}[S_{t+m} = j \mid S_t = i] = \bar{p}_{ij}, \tag{10}$$

where,

$$P^m = \begin{bmatrix} \bar{p}_{11} & \cdots & \bar{p}_{i1} & \cdots & \bar{p}_{N1} \\ \bar{p}_{12} & \cdots & \bar{p}_{i2} & \cdots & \bar{p}_{N2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{p}_{1N} & \cdots & \bar{p}_{iN} & \cdots & \bar{p}_{NN} \end{bmatrix}. \tag{11}$$

This feature of Markov chains allows us to describe the future dynamics of the process using only its transition matrix. Additional details on why this is actually true are given in Appendix A.

A Markov chain is termed *irreducible* if all of its states can be reached for any given starting point. In the simple first-order, two-state Markov chains that we will be examining, this means that every element in the transition matrix lies in the open interval, $(0, 1)$. To see this more clearly, consider the following transition matrix,

$$P = \begin{bmatrix} 0.75 & 0 \\ 0.25 & 1 \end{bmatrix}. \tag{12}$$

The problem here is that once our stochastic process enters into the second state it will never exit. This is termed an *absorbing* state and it is not a desirable property in a two-state model of the macroeconomy.

[11] This notation—which we have borrowed from Hamilton (1994)—is the opposite from the standard for matrix algebra and could potentially lead to some confusion.

A Markov chain that does not have any absorbing states is, in fact, irreducible. In a sense, these are well-behaved Markov chains. If we desire irreducibility in our first-order, two-state setting, we merely require the following inequalities to hold,

$$\mathbb{P}[S_t = 0 \mid S_{t-1} = 0] \text{ and } \mathbb{P}[S_t = 1 \mid S_{t-1} = 1] < 1. \tag{13}$$

One additional feature of a Markov chain worth reviewing relates to the first property we examined. Specifically, one might reasonably wonder whether, as $m \to \infty$, does the transition matrix, P^m, become stable? The answer to this rather vague question is yes. Slightly more formally, an irreducible Markov chain eventually converges to a limiting distribution.[12] For an N-state Markov chain the $N \times 1$ vector of ergodic (or steady-state) probabilities, denoted π, satisfies the following,

$$P\pi = \pi. \tag{14}$$

It is the vector, π, that describes the steady-state probabilities of a Markov chain. Moreover, we have the following result,

$$\lim_{m\to\infty} P^m = \pi\vec{1}, \tag{15}$$

where $\vec{1}$ is a $1 \times N$ row vector of ones. This result is discussed in Appendix A.

3.2 Hamilton's (1989) filter

We are now armed with the necessary theory to examine the hidden-Markov model employed by Hamilton (1989). Specifically, Hamilton suggested an AR(n) model with the following form,

$$y_t - \mu_{S_t} = \sum_{i=1}^{n} \phi_i \left(y_{t-i} - \mu_{S_{t-i}}\right) + \epsilon_t, \tag{16}$$

where,

$$\epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right). \tag{17}$$

We also assume that these error terms are independent. The mean of this process depends on the outcome of an unobservable state variable, S_t, that is defined as,

$$S_t = \begin{cases} 0 : \text{economy is in a } \textbf{recession}, \\ 1 : \text{economy is in an } \textbf{expansion}. \end{cases} \tag{18}$$

[12] Technically, a Markov chain is termed *ergodic* if one of the eigenvalues of the transition matrix is unity and all other eigenvalues lie inside the unit circle.

We further assume that this unobservable state variable, S_t, is governed by a straightforward, first-order, two-state Markov chain. Thus, the transition matrix is defined as,

$$P = \begin{bmatrix} q & 1-p \\ 1-q & p \end{bmatrix}, \tag{19}$$

where,

$$\mathbb{P}[S_t = 1 \mid S_{t-1} = 1] = p, \tag{20}$$

$$\mathbb{P}[S_t = 1 \mid S_{t-1} = 0] = 1 - q, \tag{21}$$

$$\mathbb{P}[S_t = 0 \mid S_{t-1} = 0] = q, \tag{22}$$

$$\mathbb{P}[S_t = 0 \mid S_{t-1} = 1] = 1 - p. \tag{23}$$

Observe that the columns of our transition matrix sum to unity and that $p, q < 1$. Our Markov chain is consequently both irreducible and ergodic. We define our mean as,

$$\mu_{S_t} = \mu_0(1 - S_t) + \mu_1 S_t. \tag{24}$$

Thus, we have now written out the model in its entirety. Conceptually, it is straightforward. The complexity of the technique arises from the parameterization of the model. Indeed, we desperately require an algorithm to estimate the parameter vector,

$$\theta = \begin{bmatrix} \mu_0 & \mu_1 & \sigma^2 & \phi_1 & \cdots & \phi_n & p & q \end{bmatrix}^T. \tag{25}$$

Most, if not all, of the estimation complexity arises because we do not actually observe S_t.[13] Hamilton's (1989) contribution was to construct a filtering algorithm to iterate through the observations ($\{y_t, t = 1, ..., T\}$) while making and updating inferences about the probability of being in a given state. As we do not observe the underlying state variable, the filtering framework is quite natural. Indeed, the estimation algorithm is similar in spirit to the Kalman filter. Nonetheless, a number of differences stem from the non-linearity of the system.

In this discussion, we provide a highly detailed exposition of the steps involved in building the maximum-likelihood function required to determine the parameter set. The heavy detail aims to add as much clarity as possible into the algorithm. We also use a key component of this method in the actual construction of our model. As such, a good understanding of the estimation algorithm is crucial.

For the purposes of expositional and notational clarity, we restrict our attention to an AR(1) process. We can, of course, generalize this approach to handle an AR(p) process or an rth-order, M-state, Markov

[13] If S_t were observable, then this would all collapse into a simple time-series regression with an indicator variable governing the transition from one regime to the next.

chain. The extra generality, however, is foregone as it adds little to our understanding. Indeed, we will see that a clear presentation is challenging enough even in the simplest setting. The discussion here follows the development in Hamilton (1989, 1990, 1994) and Kim and Nelson (1999).

Our goal is to optimize the following joint density,

$$f(y_1, y_2, ..., y_T; \theta). \tag{26}$$

The difficulty is that we cannot work with the joint density in this form. To solve this problem, we rely on the independence of our error terms in equation (17) to rewrite equation (26) as,

$$\begin{aligned} f(y_1, y_2, ..., y_T; \theta) &= f(y_1 \mid \mathcal{F}_0; \theta) f(y_2 \mid \mathcal{F}_1; \theta) \cdots f(y_T \mid \mathcal{F}_{T-1}; \theta), \\ &= \prod_{t=1}^{T} f(y_t \mid \mathcal{F}_{t-1}; \theta). \end{aligned} \tag{27}$$

This yields a more analytically convenient log-likelihood function and allows us to define our optimization problem as,

$$\max_{\theta} \underbrace{\sum_{t=1}^{T} \ln \left(f(y_t \mid \mathcal{F}_{t-1}; \theta) \right)}_{\ell(\theta)}. \tag{28}$$

The first step in our estimation algorithm is to initialize the log-likelihood function at zero,

$$\ell(\theta) = 0. \tag{29}$$

Recall that we have a time series, y_t, that is observed at $t \in \{1, ..., T\}$. The second step involves specifying the unconditional probabilities of being in each state at time 0.[14] For generality, we specify these *steady-state* probabilities as,

$$\mathbb{P}[S_0 = 1 \mid \mathcal{F}_0] = \pi, \tag{30}$$

$$\mathbb{P}[S_0 = 0 \mid \mathcal{F}_0] = 1 - \pi, \tag{31}$$

where,

$$\mathcal{F}_t \triangleq \sigma\{y_t, y_{t-1}, ..., y_1\}, \tag{32}$$

is the σ-algebra generated by the output growth process. It is important to note that S_t is not in our filtration. We are, in fact, making inferences about the underlying state of the economy conditional on the information generated by the path of economic output (i.e., $\mathcal{F}_t$).

[14]We derive these ergodic probabilities for an irreducible Markov chain in Appendix A.

Our goal is to make an inference about the probability of y_1 taking on a given value. If we knew the state of the regime we could write down the conditional density with confidence, but as it is unknown we must condition on the probability of it being in a given state. This means, in an algorithmic sense, that we compute the contribution of y_1 to the log-likelihood function by reweighting the conditional density of being in either state by the appropriate joint probability. We must, therefore, compute the set of joint probabilities. They are determined as follows,

$$\mathbb{P}[S_1 = 1, S_0 = 1 \mid \mathcal{F}_0] = \underbrace{\mathbb{P}[S_1 = 1 \mid S_0 = 1]}_{\text{equation (20)}} \underbrace{\mathbb{P}[S_0 = 1 \mid \mathcal{F}_0]}_{\text{equation (30)}}, \tag{33}$$

$$\mathbb{P}[S_1 = 1, S_0 = 0 \mid \mathcal{F}_0] = \underbrace{\mathbb{P}[S_1 = 1 \mid S_0 = 0]}_{\text{equation (21)}} \underbrace{\mathbb{P}[S_0 = 0 \mid \mathcal{F}_0]}_{\text{equation (31)}}, \tag{34}$$

$$\mathbb{P}[S_1 = 0, S_0 = 0 \mid \mathcal{F}_0] = \underbrace{\mathbb{P}[S_1 = 0 \mid S_0 = 0]}_{\text{equation (22)}} \underbrace{\mathbb{P}[S_0 = 0 \mid \mathcal{F}_0]}_{\text{equation (31)}}, \tag{35}$$

and,

$$\mathbb{P}[S_1 = 0, S_0 = 1 \mid \mathcal{F}_0] = \underbrace{\mathbb{P}[S_1 = 0 \mid S_0 = 1]}_{\text{equation (23)}} \underbrace{\mathbb{P}[S_0 = 1 \mid \mathcal{F}_0]}_{\text{equation (30)}}. \tag{36}$$

As discussed, these equations serve as the inputs into the construction of the conditional density of y_1. The expression for $f(y_1 \mid \mathcal{F}_0)$ merely follows from the law of total probability.

$$\begin{aligned} f(y_1 \mid \mathcal{F}_0) = & \sum_{i=0}^{1} \sum_{j=0}^{1} f(y_1 \mid S_1 = i, S_0 = j, \mathcal{F}_0) \underbrace{\mathbb{P}[S_1 = i, S_0 = j \mid \mathcal{F}_0]}_{\substack{\text{equations (33)-(36):} \\ \text{Weighting terms}}}, \\ = & \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_1 - \phi_1(y_0 - \mu_1))^2}{2\sigma^2}\right.} \underbrace{\mathbb{P}[S_1 = 1, S_0 = 1 \mid \mathcal{F}_0]}_{\text{equation (33)}} + \\ & \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_1 - \phi_1(y_0 - \mu_0))^2}{2\sigma^2}\right.} \underbrace{\mathbb{P}[S_1 = 1, S_0 = 0 \mid \mathcal{F}_0]}_{\text{equation (34)}} + \\ & \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_0 - \phi_1(y_0 - \mu_0))^2}{2\sigma^2}\right.} \underbrace{\mathbb{P}[S_1 = 0, S_0 = 0 \mid \mathcal{F}_0]}_{\text{equation (35)}} + \\ & \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_0 - \phi_1(y_0 - \mu_1))^2}{2\sigma^2}\right.} \underbrace{\mathbb{P}[S_1 = 0, S_0 = 1 \mid \mathcal{F}_0]}_{\text{equation (36)}}. \end{aligned} \tag{37}$$

This permits us to update the log-likelihood,

$$\ell(\theta) = \ell(\theta) + \ln \left(\underbrace{f(y_1 \mid \mathcal{F}_0)}_{\text{equation (37)}} \right). \tag{38}$$

The additive form of the likelihood function arises from the assumption of independently distributed error terms in equation (17).

We are now at the last step in the iteration. At this point, y_1 is observed and is subsequently added to our filtration. Furthermore, we wish to update equations (30) and (31) to aid us in the computation of the weighting probabilities in the next iteration. This is analagous to the updating step in the Kalman filter. These updated probabilities are termed the *filtered* probabilities. In fact, it is this quantity that we will use in section 4 to link the term structure of interest rates and the business cycle. We will consider each in turn,

$$\begin{aligned}
\mathbb{P}[S_1 = 1 \mid \mathcal{F}_1] &= \mathbb{P}[S_1 = 1, S_0 = 1 \mid \mathcal{F}_1] + \mathbb{P}[S_1 = 1, S_0 = 0 \mid \mathcal{F}_1], \\
&= \frac{f(y_1, S_1 = 1, S_0 = 1 \mid \mathcal{F}_0)}{\underbrace{f(y_1 \mid \mathcal{F}_0)}_{\text{equation (37)}}} + \frac{f(y_1, S_1 = 1, S_0 = 0 \mid \mathcal{F}_0)}{\underbrace{f(y_1 \mid \mathcal{F}_0)}_{\text{equation (37)}}}, \\
&= \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_1 - \phi_1(y_0 - \mu_1))^2}{2\sigma^2}\right.}\right) \mathbb{P}[S_1 = 1, S_0 = 1 \mid \mathcal{F}_0]}{\sum_{i=1}^{1} \sum_{j=1}^{1} f(y_1 \mid S_1 = i, S_0 = j, \mathcal{F}_0) \mathbb{P}[S_1 = i, S_1 = j \mid \mathcal{F}_0]} + \\
&\quad \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_1 - \phi_1(y_0 - \mu_0))^2}{2\sigma^2}\right.}\right) \mathbb{P}[S_1 = 1, S_0 = 0 \mid \mathcal{F}_0]}{\sum_{i=1}^{1} \sum_{j=1}^{1} f(y_1 \mid S_1 = i, S_0 = j, \mathcal{F}_0) \mathbb{P}[S_1 = i, S_1 = j \mid \mathcal{F}_0]},
\end{aligned} \tag{39}$$

and,

$$\begin{aligned}
\mathbb{P}[S_1 = 0 \mid \mathcal{F}_1] &= \mathbb{P}[S_1 = 0, S_0 = 0 \mid \mathcal{F}_1] + \mathbb{P}[S_1 = 0, S_0 = 1 \mid \mathcal{F}_1], \\
&= \frac{f(y_1, S_1 = 0, S_0 = 0 \mid \mathcal{F}_0)}{\underbrace{f(y_1 \mid \mathcal{F}_0)}_{\text{equation (37)}}} + \frac{f(y_1, S_1 = 0, S_0 = 1 \mid \mathcal{F}_0)}{\underbrace{f(y_1 \mid \mathcal{F}_0)}_{\text{equation (37)}}}, \\
&= \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_0 - \phi_1(y_0 - \mu_0))^2}{2\sigma^2}\right.}\right) \mathbb{P}[S_1 = 0, S_0 = 0 \mid \mathcal{F}_0]}{\sum_{i=1}^{1} \sum_{j=1}^{1} f(y_1 \mid S_1 = i, S_0 = j, \mathcal{F}_0) \mathbb{P}[S_1 = i, S_1 = j \mid \mathcal{F}_0]} + \\
&\quad \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1 - \mu_0 - \phi_1(y_0 - \mu_1))^2}{2\sigma^2}\right.}\right) \mathbb{P}[S_1 = 0, S_0 = 1 \mid \mathcal{F}_0]}{\sum_{i=1}^{1} \sum_{j=1}^{1} f(y_1 \mid S_1 = i, S_0 = j, \mathcal{F}_0) \mathbb{P}[S_1 = i, S_1 = j \mid \mathcal{F}_0]}.
\end{aligned} \tag{40}$$

We now move to the next time step and repeat. The complete algorithm for the construction of the log-likelihood function is outlined below. We use a standard non-linear optimization algorithm to maximize this function.

Step 1 The first step is to compute the weighting probabilities as,

$$\mathbb{P}[S_t = i, S_{t-1} = j \mid \mathcal{F}_{t-1}] = \mathbb{P}[S_t = i \mid S_{t-1} = j] \mathbb{P}[S_{t-1} = j \mid \mathcal{F}_{t-1}], \tag{41}$$

for $i, j = 0, 1$.

Step 2 Using the results of the previous step, we can compute the reweighted conditional density and add it to our log-likelihood function in the following fashion,

$$\ell(\theta) = \ell(\theta) + \ln\left(f(y_t \mid \mathcal{F}_{t-1}\right) \tag{42}$$

where,

$$f(y_t \mid \mathcal{F}_{t-1}) = \sum_{i=0}^{1}\sum_{j=0}^{1} f(y_t \mid S_t = i, S_{t-1} = j, \mathcal{F}_{t-1})\mathbb{P}[S_1 = i, S_1 = j \mid \mathcal{F}_t]. \tag{43}$$

Step 3 Finally, we observe y_t. We can now update our filtration, which is defined in equation (32), and the second term in the weighting probabilities computed in the first step. This has the following general form,

$$\begin{aligned}\mathbb{P}[S_t = i \mid \mathcal{F}_t] &= \sum_{j=0}^{1} \mathbb{P}[S_t = i, S_{t-1} = j \mid \mathcal{F}_{t-1}], \\ &= \sum_{j=0}^{1} \frac{f(y_t \mid S_t = i, S_{t-1} = j, \mathcal{F}_{t-1})\mathbb{P}[S_t = i, S_{t-1} = j \mid \mathcal{F}_{t-1}]}{f(y_t \mid \mathcal{F}_{t-1})}.\end{aligned} \tag{44}$$

for $i = 0, 1$.

Step 4 Increment t and repeat steps 1 to 3 until the end of the data sample.

This algorithm is used to determine a parameter set that is consistent with a given set of historical data. Given the parameter set, it would be useful to be able to determine the probability that a given historical period was in recession or expansion. This is possible, but it requires an additional algorithm for its computation. That is, one must compute what are termed *smoothed probabilities*. The original approach for the computation of these values, suggested by Hamilton (1989), is quite involved. Fortunately, Kim (1994) offers a substantially abridged methodology for the computation of these quantities.

3.3 Kim's (1994) smoothing algorithm

The idea behind *smoothing* is that, instead of using only $\mathcal{F}_t$ to make inferences about S_t, we use the entire sample. More specifically, we use the filtration generated by the entire sample to compute,

$$\mathbb{P}[S_t = j \mid \mathcal{F}_T], \tag{45}$$

where $t = 1, ..., T$. To determine these probabilities, we need to work backwards towards the terminal date using the updated and forecast weighted probabilities that we computed in the filter as well as the transition probabilities. Consider the following,

$$
\begin{aligned}
\mathbb{P}[S_t = j \mid \mathcal{F}_T] &= \sum_{k=0}^{1} \mathbb{P}[S_t = j, S_{t+1} = k \mid \mathcal{F}_T], \\
&= \sum_{k=0}^{1} \mathbb{P}[S_{t+1} = k \mid \mathcal{F}_T]\mathbb{P}[S_t = j \mid S_{t+1} = k, \mathcal{F}_T], \\
&= \sum_{k=0}^{1} \mathbb{P}[S_{t+1} = k \mid \mathcal{F}_T] \underbrace{\mathbb{P}[S_t = j \mid S_{t+1} = k, \mathcal{F}_t],}_{\substack{\text{This holds approximately.} \\ \text{See pseudo-proof} \\ \text{below.}}} \\
&= \sum_{k=0}^{1} \frac{\mathbb{P}[S_{t+1} = k \mid \mathcal{F}_T]\mathbb{P}[S_t = j, S_{t+1} = k \mid \mathcal{F}_t]}{\mathbb{P}[S_{t+1} = k \mid \mathcal{F}_t]}, \\
&= \sum_{k=0}^{1} \frac{\mathbb{P}[S_{t+1} = k \mid \mathcal{F}_T]\mathbb{P}[S_t = j \mid \mathcal{F}_t]\mathbb{P}[S_{t+1} = k \mid S_t = j]}{\mathbb{P}[S_{t+1} = k \mid \mathcal{F}_t]},
\end{aligned}
\tag{46}
$$

for $j = 0, 1$. The important question in the derivation of the previous expression is how the following two expressions are equivalent,

$$
\mathbb{P}[S_t = j \mid S_{t+1} = k, \mathcal{F}_T] \stackrel{?}{=} \mathbb{P}[S_t = j \mid S_{t+1} = k, \mathcal{F}_t]. \tag{47}
$$

In actual fact, equation (47) is not true. It is nevertheless a reasonable approximation that follows from the Markov property.[15] Consider the following definition,

$$
\vec{y}_{t+1,T} \triangleq \begin{bmatrix} y_{t+1} & y_{t+2} & \cdots & y_T \end{bmatrix}. \tag{48}
$$

The heart of the issue relates to the approximate equality of the two subsequent density functions,

$$
f\left(\vec{y}_{t+1,T} \mid S_{t+1} = k, S_t = j, \mathcal{F}_t\right) \approx f\left(\vec{y}_{t+1,T} \mid S_{t+1} = k, \mathcal{F}_t\right). \tag{49}
$$

The point is that conditioning on $\{S_{t+1} = k, S_t = j, \mathcal{F}_t\}$ is approximately equivalent to conditioning on $\{S_{t+1} = k, \mathcal{F}_t\}$. Essentially, assuming equation (49) to be true implies that if we somehow received information about S_{t+1} and were told $\mathcal{F}_t$, then we would have all the information we need about S_t. Were this the case, therefore, y_{t+1} would not provide any new information about S_t above and beyond that given by S_{t+1} and $\mathcal{F}_t$. This assumption is the key element in the derivation of equation (47). To see the exact logic,

[15] Equation (47) holds with strict equality in the Hamilton (1989) model when no lags of y_t are introduced.

consider the following manipulation, which assumes that equation (47) holds with equality,

$$
\begin{aligned}
\text{Left-hand side of equation (47)} &= \mathbb{P}[S_t = j \mid S_{t+1} = k, \mathcal{F}_T], \qquad (50)\\
&= \frac{f(S_t = j, \vec{y}_{t+1,T} \mid S_{t+1} = k, \mathcal{F}_t)}{f(\vec{y}_{t+1,T} \mid S_{t+1} = k, \mathcal{F}_t)},\\
&= \frac{f(S_t = j \mid S_{t+1} = k, \mathcal{F}_t) f(\vec{y}_{t+1,T} \mid S_{t+1} = k, S_t = j, \mathcal{F}_t)}{f(\vec{y}_{t+1,T} \mid S_{t+1} = k, \mathcal{F}_t)},\\
&= \frac{f(S_t = j \mid S_{t+1} = k, \mathcal{F}_t) \overbrace{f(\vec{y}_{t+1,T} \mid S_{t+1} = k, \mathcal{F}_t)}^{\text{By equation (49)}}}{f(\vec{y}_{t+1,T} \mid S_{t+1} = k, \mathcal{F}_t)},\\
&= f(S_t = j \mid S_{t+1} = k, \mathcal{F}_t) = \text{Right-hand side of equation (47).}
\end{aligned}
$$

As a final note, we look at the steps involved in computing the set of smoothed probabilities for a two-period example (i.e., $T = 2$).[16]

Step 1 At time 2 (i.e., let $t = T$), we already have the desired probabilities from the Hamilton filter. That is,

$$
\begin{aligned}
\mathbb{P}[S_T = 0 \mid \mathcal{F}_T] &= \mathbb{P}[S_t = 0 \mid \mathcal{F}_t], \qquad (51)\\
&= \mathbb{P}[S_2 = 0 \mid \mathcal{F}_2],
\end{aligned}
$$

and,

$$
\begin{aligned}
\mathbb{P}[S_T = 1 \mid \mathcal{F}_T] &= \mathbb{P}[S_t = 1 \mid \mathcal{F}_t], \qquad (52)\\
&= \mathbb{P}[S_2 = 1 \mid \mathcal{F}_2].
\end{aligned}
$$

Step 2 We now move backward in time to period 1 (i.e., let $t = T - 1$) and use equation (46) to compute $\mathbb{P}[S_1 = 1 \mid \mathcal{F}_2]$,

$$
\begin{aligned}
\mathbb{P}[S_1 = 1 \mid \mathcal{F}_2] =& \frac{\overbrace{\mathbb{P}[S_2 = 1 \mid \mathcal{F}_2]}^{\text{equation (52)}} \cdot \mathbb{P}[S_1 = 1 \mid \mathcal{F}_1] \cdot \mathbb{P}[S_2 = 1 \mid S_1 = 1]}{\mathbb{P}[S_2 = 1 \mid \mathcal{F}_1]} + \qquad (53)\\
& \frac{\overbrace{\mathbb{P}[S_2 = 0 \mid \mathcal{F}_2]}^{\text{equation (51)}} \cdot \mathbb{P}[S_1 = 1 \mid \mathcal{F}_1] \cdot \mathbb{P}[S_2 = 0 \mid S_1 = 1]}{\mathbb{P}[S_2 = 0 \mid \mathcal{F}_1]},
\end{aligned}
$$

[16] Please note that these computations are performed using the parameter set estimated using the previously discussed Hamilton (1989) non-linear filtering algorithm.

and, in an exactly analogous way, we compute $\mathbb{P}[S_1 = 0 \mid \mathcal{F}_2]$,

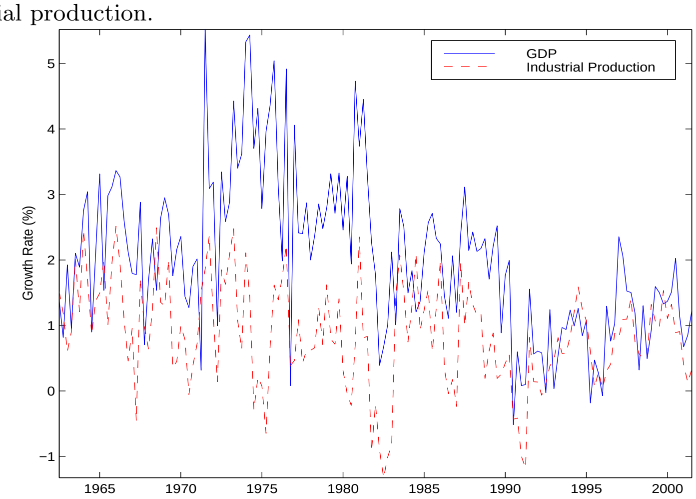
$$\mathbb{P}[S_1 = 0 \mid \mathcal{F}_2] = \frac{\overbrace{\mathbb{P}[S_2 = 0 \mid \mathcal{F}_2]}^{\text{equation (51)}} \cdot \mathbb{P}[S_1 = 0 \mid \mathcal{F}_1] \cdot \mathbb{P}[S_2 = 0 \mid S_1 = 0]}{\mathbb{P}[S_2 = 0 \mid \mathcal{F}_1]} + \frac{\overbrace{\mathbb{P}[S_2 = 1 \mid \mathcal{F}_2]}^{\text{equation (52)}} \cdot \mathbb{P}[S_1 = 0 \mid \mathcal{F}_1] \cdot \mathbb{P}[S_2 = 1 \mid S_1 = 0]}{\mathbb{P}[S_2 = 1 \mid \mathcal{F}_1]}. \tag{54}$$

That completes the iteration. In a case where $T > 2$, we would set $t = T - 2$ and repeat step 2. This would, of course, continue until all of the required smoothed probabilities are calculated.

3.4 Results

In this section, we apply the previously described algorithms to the data. We use two different quarterly time series to approximate Canadian output spanning the period from the first quarter of 1961 to the second quarter of 2001. The two series include expenditure-based GDP and industrial production (or rather, GDP at factor cost). The log differences of these two series are summarized in Figure 3.

Figure 3: **Output Growth Series**: This figure summarizes the log differences of two different quarterly output series: GDP and industrial production.



In our analysis, we consider versions of Hamilton's (1989) model with autoregressive lags of $n = 1, ..., 4$. The actual estimation of Hamilton's (1989) model is more complicated for an AR process where $n > 1$. This is the equivalent of implementing a higher-order Markov chain. One can, however, transform an nth order Markov process with m states into a first-order Markov process with mn states. This allows us to use the previously described estimation process to determine the optimal parameter set.[17]

Applying the model described in equations (16) to (24) and its associated estimation algorithm yields the parameter estimates described in Table 1. Let us first consider the transition matrix. One would expect that an expansion would exhibit a relatively higher degree of persistence than a recession. That is, we believe that there is a substantial probability that the economy will remain in an expansion, given that it is in an expansion during the current period. Encouragingly, both data series indicate that the probability of persistence in an expansionary period is quite high. In particular, $p > 0.95$ for both the GDP and industrial production data. The recessionary states, conversely, appear to have relatively less persistence compared with the expansionary states. We observe that $q = 0.86$ when industrial production data are used and $q = 0.53$ when GDP data are used. Overall, the results for the estimation of the transition matrix seem to be quite reasonable in light of our understanding of the business cycle.

Table 1: **Parameter Estimates**: This table outlines the parameters for the AR(4) models using both quarterly GDP and industrial production data for the Hamilton (1989) constant transition probability model.

Parameter	**Industrial production data**		**GDP data**	
	Estimate	**Std. error**	**Estimate**	**Std. error**
$p = \mathbb{P}[S_t = 1 \mid S_{t-1} = 1]$	0.9800	0.0170	0.9592	0.0198
$q = \mathbb{P}[S_t = 0 \mid S_{t-1} = 0]$	0.8656	0.0972	0.5348	0.1437
ϕ_1	0.3976	0.0905	0.1773	0.0687
ϕ_2	-0.0188	0.0909	0.4735	0.0844
ϕ_3	0.1023	0.0866	0.3068	0.0658
ϕ_4	-0.1458	0.0801	-0.0965	0.0684
σ	0.6225	0.0383	0.7247	0.0481
μ_0	0.1121	0.2569	0.2818	0.4989
μ_1	0.9697	0.0938	2.1261	0.4143

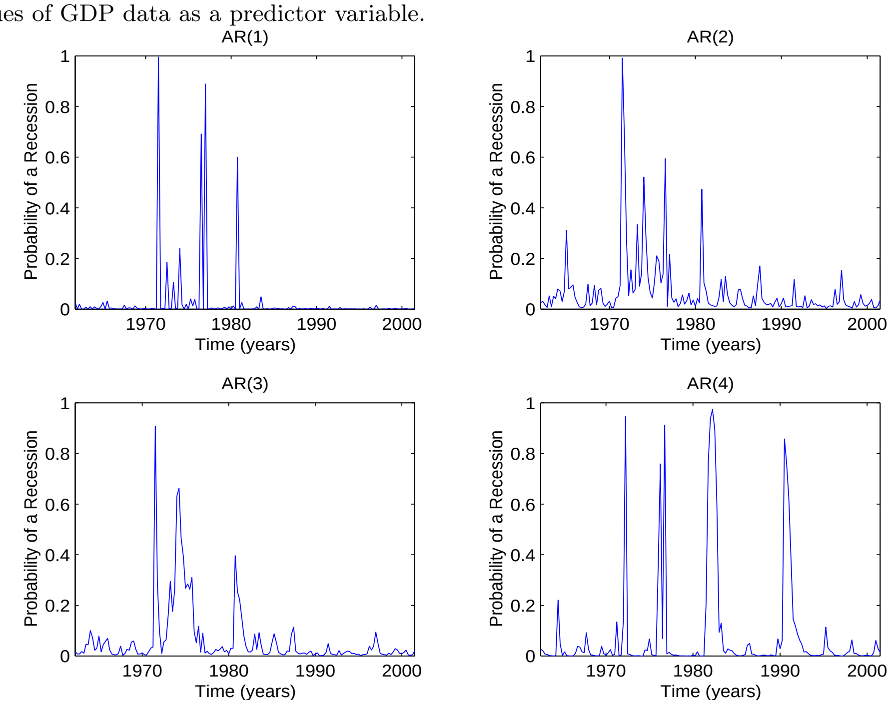
The mean levels for output growth also coincide with our prior beliefs about the macroeconomic cycle. Specifically, we would expect the mean growth rate in an expansionary period to be positive and on the order of 2 per cent. Conversely, a recessionary period should be typified by very low or negative output growth rates. This is evident in the results. In particular, $\mu_0 = 0.28$ per cent and $\mu_1 = 2.12$ per cent when using the GDP data to estimate Hamilton's (1989) model. A similar, if somewhat less pronounced, trend is evident in the industrial production data.

One useful test of the success of a given parameterization and number of lags is the smoothed probabilities. Loosely speaking, the Canadian economy was in recession during the early part of the 1970s, 1980s, and 1990s. We would hope that our model parameterizations would indicate, using the smoothed probabilities,

[17] We wish to thankfully acknowledge the use of a base set of computer routines as described in Kim and Nelson (1999). These routines served as the basis for the estimation work in this section.

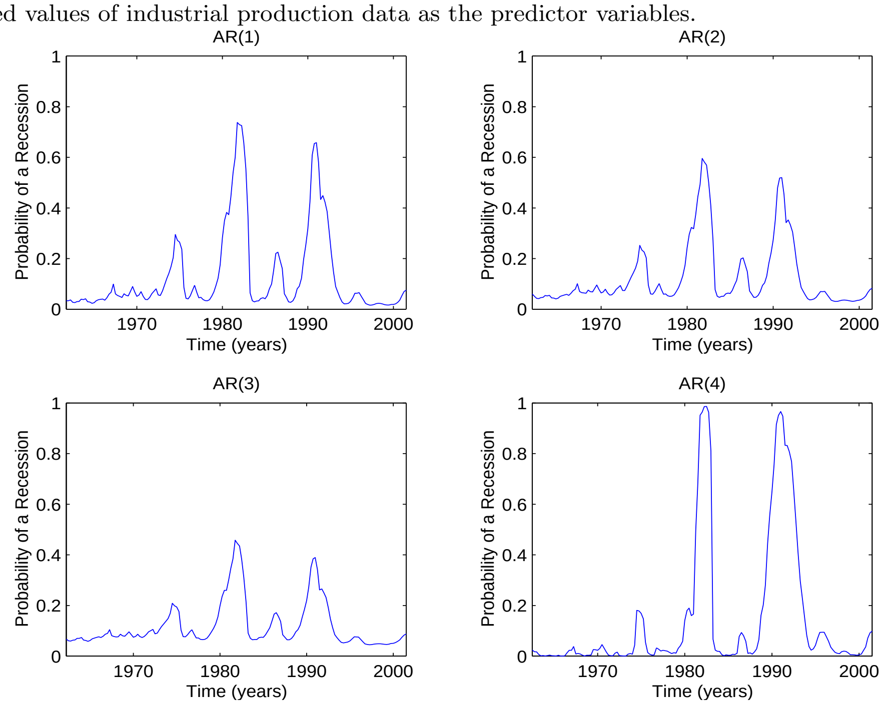
that the Canadian economy was indeed in a recession during these periods. Figure 4 describes Kim's (1994) smoothed probabilities for $n = 1, ..., 4$ using the GDP data, while Figure 5 does the same for the industrial production data. The AR(4) model using the GDP data appears to do the best job of highlighting historical recessions. That is, it does a reasonable job of indicating that the Canadian economy is in a recession during the previously mentioned periods. As a consequence, we have elected to use the parameter estimates stemming from the GDP data. In addition to the more reasonable smoothed probabilities, we feel that the parameter estimates appear to be more reasonable.

Figure 4: **Smoothed Probabilities (GDP Data)**: This graph describes the probability—at each given point in time in the sample—of actually being in a recession following Hamilton (1989). In this model, we consider one to four lagged values of GDP data as a predictor variable.



The smoothed probabilities of recession are quite binary. In other words, the probability of recession is typically either quite close to zero or quite close to one, and it does not tend to take intermediate values in the unit interval. It is much the same with filtered probabilities of recession; as one might imagine, these are quite highly correlated quantities. The reason for this binary behaviour of the smoothed and filtered probabilities is the non-continuous nature of the underlying business cycle model. The economy does not gradually ease into a recession, but rather is assumed to transition from expansion into recession over the course of a single time step. It is important to keep this feature of the smoothed and filtered probabilities in mind as it is responsible for one of the drawbacks of our joint model.

Figure 5: **Smoothed Probabilities (Industrial Data)**: This graph describes the probability—at each given point in time in the sample—of actually being in a recession following Hamilton (1989). In this model, we have used one to four lagged values of industrial production data as the predictor variables.



One potential drawback with our use of Hamilton's (1989) model is the fact that the transition probabilities are constant through time. A number of extensions to Hamilton's (1989) model address this problem by permitting the transition probabilities to be a function of some other economic information, such as consumer confidence or the steepness of the term structure of interest rates.[18] The consequence is a model that permits time-varying transition probabilities. We seriously considered using these more descriptive models and a more detailed examination of their properties is presented in Appendix D. Ultimately, however, we elected to remain with the constant-parameter Hamilton (1989) model, primarily because the time-varying models require the simulation of an additional macroeconomic variable, or variables, to construct a sample path for the business cycle. We felt that it was more important to preserve the parsimony of our model—through the constant specification of transition probabilities—than to permit the transition matrix to vary through time.

[18]This extension originated with Filardo (1993).

3.5 Some practicalities

We have worked our way through a substantial amount of theory regarding Markov chains and considered the model used by Hamilton (1989) to model the evolution of output growth. In our applicaton to debt strategy, however, we need to forecast future states of the economy—over some predetermined time interval—and use those states to determine interest rate behaviour and the government's surplus/deficit position. The question then becomes how to actually simulate a sample path for our Markov chain process. The answer is relatively simple. At each point in time, the probability of transitioning from one state to the next is merely the outcome of a Bernoulli trial with the probability of success determined from the transition matrix. We can replicate a series of Bernoulli trials quite simply by using uniform random variates.[19] To see more clearly how this algorithm works, let us consider a concrete example. This will also permit us to reinforce the concepts discussed in section 3.1. First, we begin with the estimated transition probabilities outlined in Table 1,

$$\begin{aligned} \mathbb{P}[S_t = 0 \mid S_{t-1} = 0] &= q = 0.53, \\ \mathbb{P}[S_t = 1 \mid S_{t-1} = 1] &= p = 0.96. \end{aligned} \tag{55}$$

These transition probabilities, in turn, yield the following transition matrix,

$$P = \begin{bmatrix} q & 1-p \\ 1-q & p \end{bmatrix} = \begin{bmatrix} 0.53 & 0.04 \\ 0.47 & 0.96 \end{bmatrix}. \tag{56}$$

There are, in fact, only two steps in the algorithm and they are very similar. To see exactly what is going on, we will consider each in detail.

Step 1 While each step is similar, the first step is somewhat different, in that we must find a way to launch the process. The most reasonable way to approach this is to use the ergodic, or unconditional probabilities. To obtain these values, we recall equation (15) and apply the result from equation (91)

[19] There are alternative methods—for example, see Fishman (1995, page 235)—but this approach is by far the simplest and is not terribly computationally expensive.

in Appendix A,

$$\begin{aligned}
\lim_{m\to\infty} P^m &= A \lim_{m\to\infty} \left(\Lambda^m\right) A^{-1}, && (57)\\
&= \begin{bmatrix} -0.7071 & -0.0848 \\ 0.7071 & -0.9964 \end{bmatrix} \lim_{m\to\infty} \begin{bmatrix} 0.49 & 0 \\ 0 & 1 \end{bmatrix}^m \begin{bmatrix} -1.3033 & 0.1109 \\ -0.9249 & -0.9249 \end{bmatrix},\\
&= \begin{bmatrix} -0.7071 & -0.0848 \\ 0.7071 & -0.9964 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1.3033 & 0.1109 \\ -0.9249 & -0.9249 \end{bmatrix},\\
&= \begin{bmatrix} 0.0784 & 0.0784 \\ 0.9216 & 0.9216 \end{bmatrix},\\
&= \begin{bmatrix} 0.0784 \\ 0.9216 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix},\\
&= \pi \vec{1}.
\end{aligned}$$

The consequence is that the unconditional probabilities for this example are,

$$\begin{aligned}
\mathbb{P}[S_1 = 0 \mid \mathcal{F}_0] &= \frac{1-0.96}{2-0.53-0.96} = 0.0784, && (58)\\
\mathbb{P}[S_1 = 1 \mid \mathcal{F}_0] &= \frac{1-0.53}{2-0.53-0.96} = 0.9216.
\end{aligned}$$

Thus, we are finally able to write down the approach to determining the initial state of our Markov chain for a given simulation. Select the following standard uniform random variate, $U_1 \sim \mathcal{U}[0,1]$, and use the following logic, outlined here in pseudo-code,

if $U_1 \in [0, 0.0784]$ **then**

  $S_1 = 0$

else if $U_1 \in (0.0784, 1]$ **then**

  $S_1 = 1$

end

This is the equivalent of conducting a Bernoulli trial with probability of success equal to 0.0784.

Step 2 The next task is to determine S_2, which will, of course, depend on the value of our Markov chain at S_1. The algorithm we intend to employ will use the following control structure and the transition matrix. To begin, we select another standard uniform random variate, $U_2 \sim \mathcal{U}[0,1]$, and again use pseudo-code,

if $S_1 = 0$ **and** $U_2 \in [0, 0.53]$ **then**

  $S_2 = 0$

```
else if S_1 = 0 and U_2 ∈ (0.53, 1] then
        S_2 = 1
else if S_1 = 1 and U_2 ∈ [0, 0.96] then
        S_2 = 1
else if S_1 = 1 and U_2 ∈ (0.96, 1] then
        S_2 = 0
end
```

We need only to repeat the second step until the end of the forecast period and it can be extended to the general two-state Markov chain model quite simply.[20]

4 The Model

To this point, we have worked through a fairly well-known model that describes the evolution of the business cycle. We can now construct our link between the term structure of interest rates and the financial position of the government. We will begin with the term structure of interest rates because it is the most complicated aspect of the model.

4.1 The term structure

The term-structure model currently used in the Government of Canada's debt strategy analysis is a two-factor Cox-Ingersoll-Ross (CIR) model.[21] The class of affine term-structure models and its application to debt strategy is described in detail in Bolder (2001). We will, however, briefly introduce the basics of this model. The underlying state variables, y_1 and y_2, that drive the dynamics of the evolution of the term structure of interest rates are assumed to follow two independent continuous-time, continuous-state,

[20] In fact, the general case is as follows,

```
if S_{t-1} = 0 and U_2 ∈ [0, p] then
        S_t = 0
else if S_{t-1} = 0 and U_2 ∈ (p, 1] then
        S_t = 1
else if S_{t-1} = 1 and U_2 ∈ [0, q] then
        S_t = 1
else if S_{t-1} = 1 and U_2 ∈ (q, 1] then
        S_t = 0
end
```

[21] These models originated with the path-breaking work of Cox, Ingersoll, and Ross (1985a,b).

stochastic processes of the following form,

$$dy_1(t) = \kappa_1 \left(\theta_1 - \left(\frac{\kappa_1 + \lambda_1}{\kappa_1} \right) y_1(t) \right) dt + \sigma_1 \sqrt{y_1(t)} dW_1(t), \tag{59}$$

$$dy_2(t) = \kappa_2 \left(\theta_2 - \left(\frac{\kappa_2 + \lambda_2}{\kappa_2} \right) y_2(t) \right) dt + \sigma_2 \sqrt{y_2(t)} dW_2(t), \tag{60}$$

where $\{W_i(t), t \in [0, \infty)\}_{i=1,2}$ are independent standard, scalar Wiener processes defined on the filtered probability space, $\{\Omega, \mathcal{F}, \mathbb{P}\}$. The parameters $\lambda_i, i = 1, 2$ represent the market price of risk for each state variable. Through an arbitrage argument, one can construct a bond price function, $P(t, s)$ for $s \geq t$. The variable t is interpreted as the current point in time, while s represents the term to maturity of an arbitrary zero-coupon bond. For our purposes, we can easily restrict our attention to values of s in the interval $\left[\frac{1}{12}, 30\right]$. The bond price function is a deterministic function of the state variables and is defined as follows,

$$P(t, s, y_1(t), y_2(t)) = e^{\sum_{i=1}^{2} (A_i(t,s) - B_i(t,s) y_i(t))}. \tag{61}$$

The functions $A_i(t, s)$ and $B_i(t, s)$ have the underlying deterministic structure,

$$\begin{aligned} B_i(t, s) &= \frac{2(e^{\gamma_i(s-t)} - 1)}{a_i(e^{\gamma_i(s-t)} - 1) + 2\gamma_i}, \\ A_i(t, s) &= \ln \left[\left(\frac{2\gamma_i e^{\frac{a_i(s-t)}{2}}}{a_i(e^{\gamma_i(s-t)} - 1) + 2\gamma_i} \right)^{\frac{2\kappa_i \theta_i}{\sigma_i^2}} \right], \end{aligned} \tag{62}$$

where,

$$a_i = \gamma_i + \kappa_i + \lambda_i,$$

and,

$$\gamma_i = \sqrt{(\kappa_i + \lambda_i)^2 + 2\sigma_i^2},$$

for $i = 1, 2$. This bond price function is then transformed into the zero-coupon term structure of interest rates by,

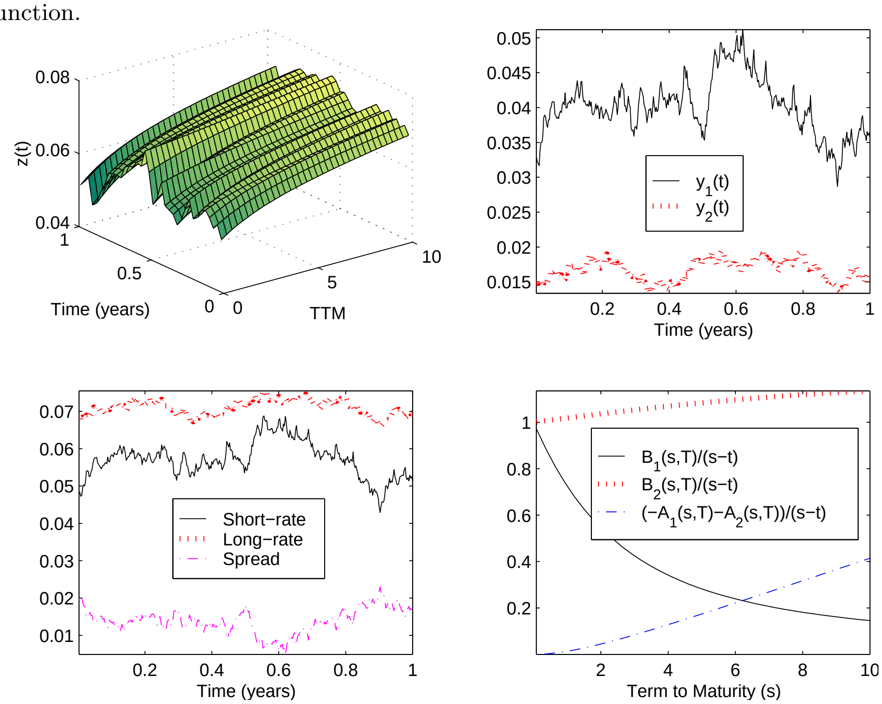
$$z(t, s) = -\frac{\ln P(t, s)}{s - t} = \sum_{i=1}^{2} -\frac{A_i(t, s)}{s - t} + \frac{B_i(t, s)}{s - t} y_i(t). \tag{63}$$

Using the estimation procedure, described in Bolder (2001), we obtain the parameter values of this model. They are summarized in Table 2.[22] We generate a single term-structure sample path and decompose it into a number of components in Figure 6. We will use this decomposition to better understand the fundamental properties of this model. The first quadrant of Figure 6 illustrates the actual evolution of the term structure over a one-year period; the second quadrant outlines the state variable dynamics; the third quadrant traces

[22]Please note that these parameter estimates where slightly altered to ensure that the so-called *Feller* condition would hold. When this condition is satisfied, zero-coupon rates generated by this model cannot be negative.

the movement of the three-month rate, the 10-year rate, and the spread between them; and the fourth quadrant describes the factor loadings outlined in equation (63).[23]

Figure 6: **A Decomposition of the Two-Factor CIR model**: This figure decomposes a single sample path for the two-factor CIR model. It includes the evolution of the underlying state variables, the associated short and long zero-coupon rates, and the factor loadings—$\frac{B_1(t,s)}{s-t}$, $\frac{B_2(t,s)}{s-t}$, and $\frac{-A_1(t,s)-A_2(t,s)}{s-t}$)—used in equation (61) to construct the bond price function.



To this point, we have been rather vague about the identity of the state variables, $y_1(t)$ and $y_2(t)$. We can, however, infer their identity by using the decomposition in Figure 6, the parameters described in Table 2, and the type of analysis presented in Geyer and Pichler (1998) and de Jong (2000). In particular, we observe that the first state variable exhibits strong mean reversion ($\kappa_1 = 0.993$) and sizable variability ($\sigma_1 = 0.101$). It also has a downward-sloping factor loading $\left(\frac{B_1(t,s)}{s-t}\right)$. This implies that our first state variable does *not* impact all zero-coupon yields in a similar fashion. Finally, we observe that $y_1(t)$ is highly correlated with the spread between the short, three-month interest rate and the longer-term, 10-year rate. Based on this evidence, we may conclude that the first state variable can be identified as the *slope*, or steepness, of the term structure of interest rates.

The second state variable, conversely, demonstrates weak mean reversion ($\kappa_2 = 0.065$), is relatively less variable ($\sigma_2 = 0.060$), and is highly correlated with the 10-year zero-coupon rate. Moreover, the factor

[23]We set the time step in the generation of the graphics in Figure 6 at approximately one day.

loading $\left(\frac{B_2(t,s)}{s-t}\right)$ is quite flat across maturities. This means that the second state variable influences all zero-coupon bond maturities in, more or less, the same manner. As a consequence, we interpret the second state variable as the *level* of interest rates.

The interpretation of the state variables of a two-factor affine term-structure model as the level and slope of the term structure is standard. This realization originated in the literature with the work of Litterman and Scheinkman (1991). They demonstrated, using an eigenvalue decomposition of zero-coupon bond returns, that the majority of term-structure variability can be explained by three orthogonal factors: level, slope, and twist. Of these three factors, the level and slope are the most important in summarizing term-structure evolution.

Our interest in this paper is in the steepness of the term structure of interest rates. In particular, the steepness of the term structure relates to the term premium demanded by investors for holding bonds of longer maturities. One can, in fact, derive the form of the term premium assumed by the two-factor CIR model.[24] Let us denote the instantaneous risk premium as $\theta(t,s)$ for a bond with term to maturity of s years. It has the following form,

$$\theta(t,s) = r(t) - \sum_{i=1}^{n} \lambda_i \sigma_i \sqrt{y_i(t)} B_i(t,s). \tag{64}$$

The derivation of this expression follows de Jong (2000) and is outlined in greater detail in Appendix C.[25] In equation (64), we know that the instantaneous standard deviation of our state variables ($\sigma_i \sqrt{r(t)}$) and the function $B_i(t,s)$ are both strictly positive. This implies that, for the term premium to be positive, we require negative market price of risk parameters. Or, at least, we require that one state variable have a market price of risk parameters that is sufficiently large and negative to offset a positive market price of risk parameter in the other state variable. As shown in Table 2, both of our market price of risk parameters are negative.

In this particular exercise, the most important model parameter is λ_1. As previously discussed, the strongly mean-reverting, highly variable first state variable represents movement in the steepness of the term structure. Moreover, from equation (64), the parameter that exerts primary control over the term premium is λ_1. Clearly, both market price of risk parameters (λ_1 and λ_2) influence the steepness of the term structure, but λ_1 is the most important and thus the most efficient candidate for our analysis.[26] As the absolute value of the typically negative λ_1 increases in size, the simulated term structure of interest rates will increase in

[24] The key assumption is that investors have logarithmic utility and thus constant relative risk aversion.

[25] In actuality, we have not been entirely successful in establishing this derivation. Specifically, there may be a fundamental error in the derivation or, perhaps less likely, the actual form of equation (64) may not be precisely correct. It will, however, be some close variation on this form and, as such, this is not a terribly important point.

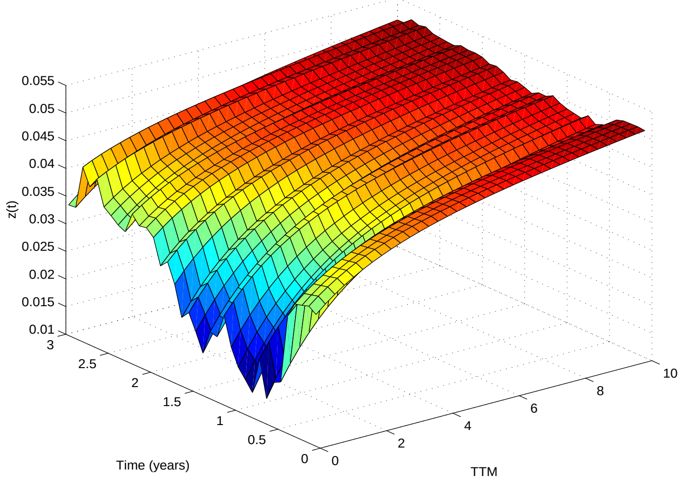
[26] This approach would still work with a single-factor model. In that case, the slope of the term structure is influenced by a single market price of risk parameter.

Table 2: **Parameter Estimates**: This table summarizes the parameter set for our two-factor CIR model estimated using Canadian term-structure data from 1994 to 2001.

Parameter	Estimate	Std. error
κ_1	0.993	0.032
κ_2	0.065	0.041
θ_1	0.033	0.004
θ_2	0.015	0.005
σ_1	0.101	0.010
σ_2	0.060	0.007
λ_1	-0.315	0.038
λ_2	-0.103	0.047

steepness. To see this more clearly, consider Figures 7 and 8. In Figure 7, we have set $\lambda_1 = -0.315$ and the corresponding term-structure outcomes are upward sloping. In Figure 8, however, we have reduced λ_1 to -0.05. The result is a sequence of relatively flat and inverted term structures of interest rates.

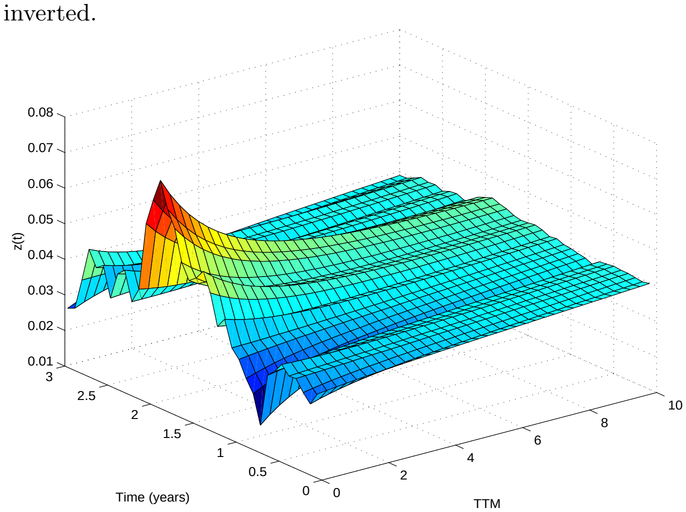
Figure 7: **A Steep Two-Factor CIR Term Structure**: This figure illustrates a single sample path of a two-factor CIR model with the parameter set specified in Table 2. Observe that it is generally upward-sloping.



The first market price of risk parameter, therefore, is the key to controlling the relative steepness of the term structure of interest rates. Using the logic described in Figure 1, we want the term structure of interest rates to be relatively flat roughly four quarters prior to a recession. That is, we require that λ_1 be relatively small in absolute value over this period. Conversely, the term structure of interest rates should steepen at the end of a recession.

To see specifically how we intend to govern the steepness of the term structure of interest rates in concert

Figure 8: **A Flat Two-Factor CIR Term Structure**: This figure illustrates a single sample path of a two-factor CIR model with the parameter set specified in Table 2 and the λ_1 parameter reset to equal -0.05. Observe that it is generally flat or inverted.



with the business cycle, we need to return to Hamilton's (1989) model. The best way to demonstrate our suggested approach is to consider an example. First, let us use our previously estimated transition matrix, P, to generate a quarterly sample path for the business cycle for the next 10 years.[27] The actual sequence of expansionary and recessionary outcomes is illustrated in the first quadrant of Figure 9.

By itself, this sequence of business cycle outcomes is not terribly helpful. Using our estimated AR(4) model described in equation (16), however, we can generate a sample path for output growth in the economy that is consistent with the previously generated business cycle outcome. It is important to note that these output growth outcomes are computed with noise. The relationship between the current state of the economy and output growth is *not* deterministic and is subject to small shocks that are orthogonal to the current macroeconomic state. This output growth realization is summarized in the second quadrant of Figure 9.

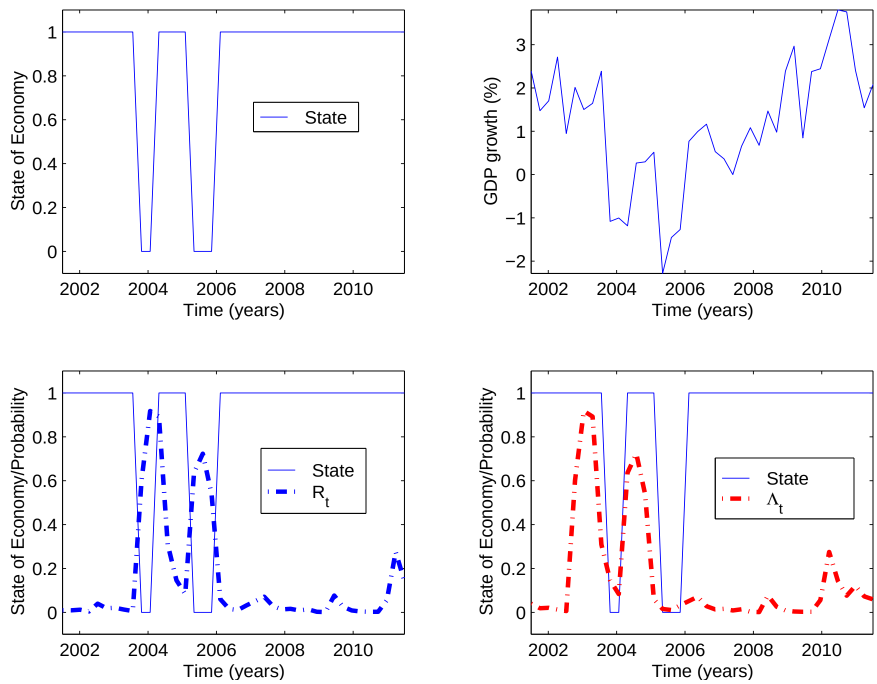
Using this sequence of output growth outcomes, we may proceed to compute the associated conditional probability that the economy is in recession. This is termed the filtered probability and we will denote it as,

$$R_t = f(y_t, y_{t-1}, ..., y_0, \theta) = \mathbb{P}[S_t = 0 \mid \mathcal{F}_t]. \tag{65}$$

For an AR(n) model, this probability is conditioned on the previous n output growth lags. In fact, we have seen this quantity in equation (44) when constructing the non-linear filter used to estimate Hamilton's (1989) model. Inspection of equation (44) reveals that it is a function—denoted f in equation (65)—of the current and past output values and the parameter set. How should we think about this filtered probability? Imagine

[27]The specific mechanics for generating a sample path from a given transition matrix, P, are described in section 3.5.

Figure 9: **Determining the Steepness of the Term Structure**: This figure outlines the steps in the computation of a modified filtered probability series, Λ_t, that is used in determining the steepness of the term structure of interest rates.



that you observe output growth over a number of periods and, using this information, attempt to infer the state of the economy. If recent observations of output growth were weak relative to previous observations, you would likely suggest that the probability of currently being in a recession is high. Conversely, if you observed strong output growth in recent periods, you would infer that the probability of recession was low. R_t performs exactly this inference for each individual time period. We can think of R_t, in a more technical sense, as mapping the current output figure and the previous $n-1$ output growth lags into a real number: the probability that the economy is currently in a recession.

Examining R_t in conjunction with the business cycle sample path, as described in the third quadrant of Figure 9, we observe that it increases contemporaneously with the recessionary periods. Note also that R_t can rise during non-recessionary periods. This behaviour is caused by a smaller-than-anticipated element in the output growth series. The possibility of non-recessionary negative shocks to output growth, such as spikes in oil prices, implies that this is a not-unreasonable feature in the model. It is exactly this process, R_t, that we wish to use to alter the slope of the term structure of interest rates. The problem is that we need to shift this measure *backwards* to capture the forward-looking nature of the term structure of interest rates. To accomplish this, we define the following altered random process,

$$\Lambda_t = R_{t+\tau}, \tag{66}$$

where τ is set to a value from four to six quarters. This is consistent with the literature describing the forward-looking behaviour of term-structure steepness. This process, Λ_t, describes the probability of the economy being in a recession τ quarters from the current point in time. Λ_t is summarized in the final quadrant of Figure 9.

Λ_t is a useful process and forms the basis of our link between the business cycle and the term structure of interest rates. It allows us to look τ periods into the future and determine the associated probability of recession in this future period.[28] Technically, one cannot *peek* into the future. In our model, however, the relationship between the macroeconomy and our other processes is one-directional. That is, the macroeconomy impacts the evolution of the term structure of interest rates and the government's financial position, but the reverse is not true. As such, we believe it is defensible to condition on the entire business cycle sample path to construct our term-structure realizations.

The forward-looking nature of Λ_t allows us to relate the current steepness of the term structure to the probable future state of the economy τ periods into the future. This is accomplished by defining a new λ_1 value in the following manner,

$$\tilde{\lambda}_{1,t} = (1 - \Lambda_t)\lambda_1^e + \Lambda_t \lambda_1^r, \tag{67}$$

where λ_1^e is the market price of risk parameter for expansionary periods and λ_1^r pertains to recessionary regimes. Note that λ_1^e and λ_1^r have not been estimated but selected, given our parameterization, such that λ_1^e implies a relatively steep expected term structure while λ_1^r leads to a generally flat or inverted term structure of interest rates. We can interpret $\tilde{\lambda}_{1,t}$ by focusing on Λ_t. If the probability of recession τ periods hence is close to zero, then the term structure will have a normal slope as determined by λ_1^e. If, however, the probability of the economy being in a recession in τ periods is close to one—as measured by Λ_t—then the market price of risk parameter will be very close to λ_1^r. This will lead to a generally flat or inverted term structure of interest for the current period. Ultimately, using this approach, the steepness of the term structure of interest rates is determined by a convex combination of λ_1^e and λ_1^r with parameter, Λ_t.

The attentive reader might have noticed that the proposed methodology does a reasonable job of creating flat or inverted term structures prior to recession but does not do as well at generating steep term-structure outcomes at the end of a recession. The reasons for this are primarily structural. The proposed technique is

[28] Those familiar with the theory of stochastic processes will have certainly observed that the construction of Λ_t violates a host of measurability conditions. Nevertheless, our intentions in this analysis are pragmatic and the relationship between the business cycle and the term-structure, in our model, is assumed to be one-directional. As a result, we feel that this justifies our actions.

not sufficiently flexible to capture the term-structure dynamics in a symmetric fashion. That is, it generates the requisite flatness τ periods prior to recession, but does not allow for greater-than-average steepness τ periods prior to a recovery. This is a weakness in our approach. The methodology does allow for the term structure to be relatively steep τ periods prior to a recovery as compared with τ periods prior to a recession. While this is a weakness in our technique, we felt that capturing the leading indicator nature of term-structure steepness prior to recession was most important. Moreover, the modifications we considered to rectify this drawback were ultimately rejected owing to their increased complexity.

4.2 The government's financial position

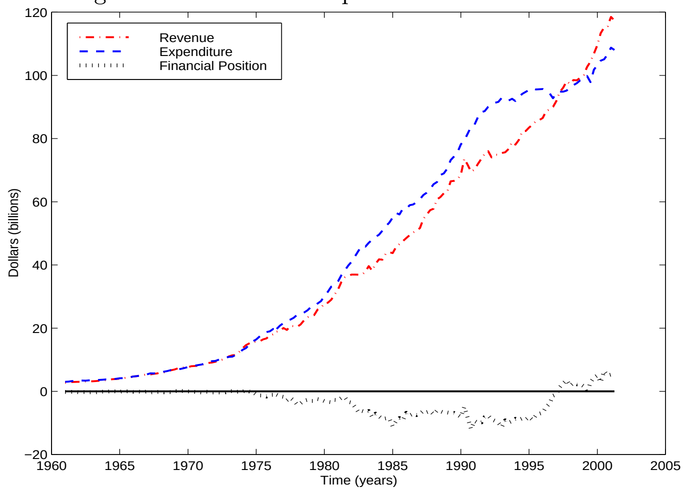
Having worked through the details of our term-structure model and its relationship with our business cycle model, we turn to consider the government's financial position. This is accomplished in a relatively straightforward manner. Recall that we argued in previous sections that the government's financial position is a transformation of the government's revenues and expenditures at any given point in time. These quantities are themselves functions of current economic conditions. We could, therefore, model the government's revenue and expenditure positions separately and then amalgamate them to determine the government's ultimate financial position. We have decided, however, for reasons of conceptual and econometric simplicity, to model the government's financial position explicitly rather than focus on government revenues and expenditures.

Figure 10 illustrates the quarterly revenue, expenditure, and financial position data for the Government of Canada from the first quarter of 1961 until the second quarter of 2001. The source of this data is the Canadian National Accounts quarterly data release. We can make a number of observations from Figure 10. First, the revenue and expenditure series are growing steadily over time. They also exhibit very strong covariance.[29] Initially, we considered a bivariate linear model of the government's revenue and expenditure growth using current and past output growth as well as lagged revenue and expenditure growth observations. While this model appears to be reasonable in a conceptual sense, we experienced problems with explosiveness of these series when simulating future outcomes. The result was a sizable number of outcomes with unreasonably large government surplus or deficit positions over the simulation interval.

Continued inspection of the government financial position series in Figure 10 is less encouraging. Specifically, there appear to be at least three different regimes. The first period, from 1961-75, is characterized by a balanced budget position, while the second period, running roughly from 1975-97, exhibits a sequence of deficit positions. The final period, from 1997 to the second quarter of 2001, demonstrates a trend towards an increasing surplus position. There are, therefore, two fundamental reasons for concern with the use of the presented historical data to estimate the parameter set for any postulated financial position process.

[29] The simple sample correlation of these two series is in excess of 0.98.

Figure 10: **Quarterly Government Revenue, Expenditure and Financial Position**: This figure outlines the evolution—from 1961 until the second quarter of 2001—of the quarterly Government of Canada expenditures, revenues, and their difference: the government's financial position.



The first is that one might reasonably argue that the sequence of observed government financial positions is not representative of future outcomes. As such, any parameterization would not be terribly useful from a modelling perspective. Second, in a more statistical sense, these data are not *stationary*. While there are a number of definitions of non-stationarity, in this context we are referring to a process with a non-constant transition density.[30] One could use any number of statistical techniques to build a model to fit this data. Nevertheless, this would not resolve the first problem of non-representativeness of this data to future outcomes.

We argue, therefore, for the specification of a parsimonious model of the government's financial position. Using this simple model with a relatively small set of parameters, the modelling team may *calibrate* the parameter set to their expectations of future outcomes. That is, they can decide on the qualitative aspects of the given model through the selection of model parameters. This framework also permits examination of the sensitivity of the results to any assumption about the future dynamics of the government's financial position. The point is that, in this analysis, the parameters of our financial position process are *not* estimated but rather calibrated. The parameters of Hamilton's (1989) hidden-Markov model and the two-factor affine term-structure model are actually estimated using historical data.

If we are to build a simple model for the government's financial position, we need to consider the important aspects of this series. In particular, there are rather compelling arguments for mean-reversion in the government's financial position over a sufficiently lengthy time interval. That is, we believe that a random

[30]In Hamilton (1994), this is referred to as *strict* stationarity.

macroeconomic shock could move the government's financial position from its long-term mean value—which one could legitimately argue to be approximately zero—but that these shocks do not persist over time. This would suggest that our initial specification, in equation (1), is actually a good place to start. Recall that the drawback of using this process to describe the government's financial position was the fact that its evolution is orthogonal to the current state of the macroeconomy. To rectify this problem, we will slightly modify the stochastic process suggested in equation (1).

Before we actually describe the model, let us introduce some notation. We denote F_t as the government's financial position in period t. In a mathematical sense, we will consider this to be a continuous-time, continuous-state stochastic process. Operationally speaking, this amounts to the difference between the government's revenues and expenditures over a given three-month period. Some may claim that the use of a continuous-time stochastic process is an inappropriate tool for the modelling of the evolution of quarterly economic series. These arguments are not without merit. Nevertheless, we feel that this is a convenient and intuitive manner in which to represent the dynamics of the government's financial position. Moreover, this approach is consistent with the nature of the processes used to model the evolution of the term structure of interest rates.

The model, therefore, that we will use to describe the evolution of the government's financial position is a slight modification of the Ornstein-Uhlenbeck process suggested in equation (1). It has the following form,

$$dF(t) = \alpha(\beta - F(t))dt + \boxed{\gamma \dot{R}_t} + \xi dW(t), \tag{68}$$

where $\dot{R}_t$ is the *instantaneous* filtered probability of being in a recession at time t. The discrete-time version of this process was first defined in equation (65).[31] Loosely speaking, one can interpret the additional boxed term $\gamma \dot{R}_t$ as a randomly occurring *jump* in the government's financial position process. More specifically, the parameter γ is the size of the jump and $\dot{R}_t$ both governs its arrival and rescales its size. That is, the actual size of the impact on the government's financial position depends on the magnitude of $\dot{R}_t$ which lies in the unit interval. We should note, rather emphatically, that this is *not* a jump diffusion model. Instead, we are postulating a mean-reverting process for the government's financial position that incorporates randomly occurring information about the current economic state.

For clarity, we will discuss each of the model parameters and attempt to reconcile them with our intuition about the influence of the macroeconomy on the government's financial position.

Mean reversion (α) This parameter ultimately governs the speed at which surpluses and deficits return to their long-term mean value (β). In other words, it represents how quickly the government alters discretionary spending to return to its deficit target.

[31] Because equation (68) is a stochastic differential equation, it is necessary that $\dot{R}_t$ also be continuous with respect to time.

Long-term mean (β) Typically, we would expect this value to be zero. That is, in expectation, the government targets a balanced budget. This can, however, be quite flexible. For example, we could model the government's financial position on an *ex-interest charge* basis. Thus, the long-term mean would represent the expected source from government operations before debt servicing costs. The debt servicing costs, therefore, could be computed endogenously in the debt strategy model as a function of the existing portfolio and the financing strategy.

Deficit size (γ) We can think of this parameter as representing the expected impact of a recession on the government's financial position. As discussed earlier, this stems from the influence of increased spending from automatic stabilizers and correspondingly reduced tax revenues.

Volatility (ξ) This final parameter represents the inherent volatility in the financial position process that is independent from the business cycle. This is a reasonable thing to include in the model. A variety of small random elements are involved in determining the government's financial position. A relatively small volatility permits us to incorporate those elements.

To simulate the financial position process, for a given parameter set, we need to discretize the stochastic differential equation in equation (68). It does not seem obvious, if indeed it is possible, to solve this stochastic differential equation explicitly. We assume, therefore, that its transition density is minimally unchanged from the basic Ornstein-Uhlenbeck situation. That is,

$$F_t \mid F_{t-1} \sim \mathcal{N}\left(\beta\left(1-e^{-\alpha\Delta t}\right)+e^{-\alpha\Delta t}F_{t-1}+\gamma R_t, \frac{\beta^2}{2\alpha}\left(1-e^{-\alpha\Delta t}\right)\right), \tag{69}$$

where R_t is the discrete-time filtered probability of being in a recession at time t as defined in equation (65). The derivation of the *exact* transition density for the Ornstein-Uhlenbeck process is outlined in Bolder (2001, Appendix B). We expect that this will not be an unreasonable approximation to the true transition density for this process.

Four alternative calibrations for the government financial position process, outlined in equation (68), are illustrated in Table 3. The actual dynamics of these three potential financial position processes are described in Figure 11. We have generated 50 separate sample paths using each parameter set described in Table 3. The first two cases, outlined in the first two quadrants of Figure 11, have a moderate level of mean reversion ($\alpha = 0.4$), but differ in terms of jump size, volatility, and long-term mean. Observe how the larger jump size leads to a small number of realizations with large negative shocks to the government's financial position. Also note that these large shocks do not persist. Cases 3 and 4 are identical to the first two cases, except that the level of mean reversion has been increased ($\alpha = 0.7$). Note the reduction in the dispersion around the long-term mean value in these two cases compared with the first two cases.

We caution that the suggested modified Ornstein-Uhlenbeck process for the government's financial position is quite simplistic. There may be more sophisticated and realistic ways to capture the random evolution

Table 3: **Various Potential Parameter Calibrations**: This table outlines calibrated parameters for the simple model—based on the Ornstein-Uhlenbeck process—used to model the evolution of the government's financial position. All currency amounts are in billions of Canadian dollars.

Financial position parameters				
Parameter	**Case 1**	**Case 2**	**Case 3**	**Case 4**
F_0 (initial financial position)	\$1	\$1	\$1	\$1
β (long-term mean)	\$0	\$3	\$0	\$3
α (mean reversion)	0.4	0.4	0.7	0.7
γ (jump size)	-\$1	-\$3	-\$1	-\$3
ξ (volatility)	\$1	\$3	\$1	\$3

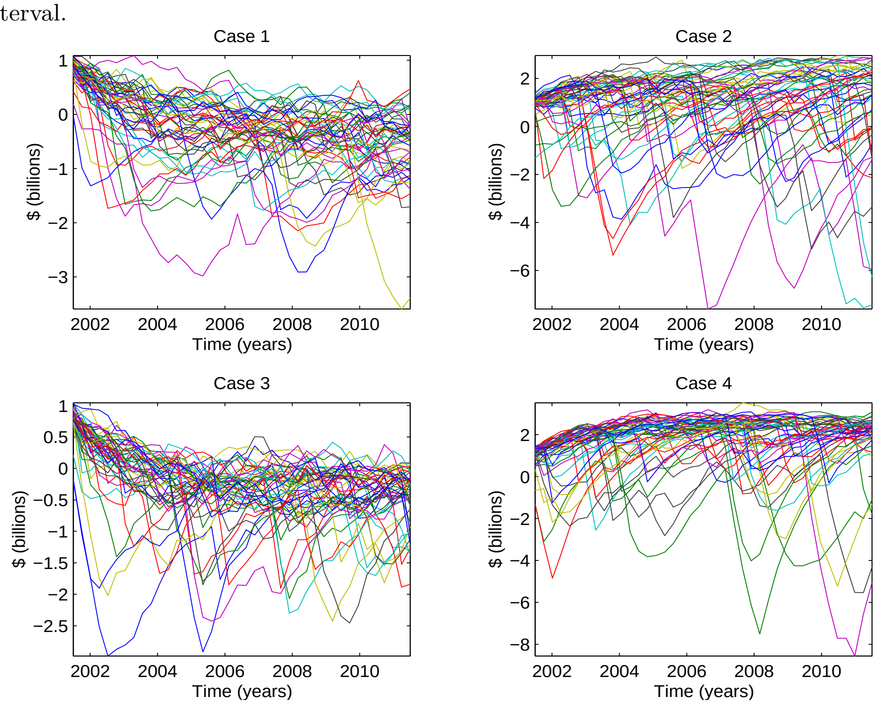
of the government's finances. Nevertheless, we felt that a straightforward first-order approximation was a reasonable starting place for our modelling work. In particular, the mean-reverting nature of this model avoids the rather serious problem of explosive government deficit and surplus positions. In addition to having a well-controlled process, we also have an explicit dependence on the current state of the business cycle. Finally, we have additional financial position uncertainty that is orthogonal to the underlying state of the macroeconomy.

4.3 The unified simulation framework

In this final section dedicated to a discussion of the model, we will briefly review the entire algorithm suggested for generating future joint sample paths of the business cycle, the government's financial position, and the term structure of interest rates. This essentially requires an amalgamation of the discussion in the preceding sections. Before we get into the details, we introduce a final element of notation. Let the superscript $*$ denote a simulated random variable. For example, if $\{X_t, t \in [0, s]\}$ is a generic stochastic process, then $\{X_t^*, t \in [0, s]\}$ represents a simulated sample path from this process. Our model algorithm can be described in the following four steps. That is, the subsequent sequence of steps must be performed to produce a single future realization of our joint processes for our stochastic simulation model.

Step 1: The State of the Economy We discretize our time dimension into T equal time steps, $1, ..., T$. We then use our transition matrix, P, to compute a sample path for the underlying state of the macroeconomy. This information is summarized in our state variable, $\{S_t^*, t = 1, ..., T\}$. The actual approach used to accomplish this is described in section 3.5. We may then use equation (16) to construct a set of output growth realizations, $\{y_t^*, t = 1, ..., T\}$ that are consistent with the generated macroeconomic observations. We have now computed the elements of the macroeconomy. It is important to note that the impact of the macroeconomy in this setting is one-directional. That is, the macroeconomic state influences the evolution of the term structure and the government's financial position, but these

Figure 11: **Financial Position Sample Paths**: This figure outlines 50 Government of Canada financial position sample paths computed using the parameter set described in Table 3. In each case, the outcomes are generated over a 10-year time interval.



processes do not impact the macroeconomic state.

Step 2: Output Growth and Filtered Probabilities The next step is to use our output growth sequence, $\{y_t^*, t = 1, ..., T\}$, to construct the filtered probabilities that are required to control the steepness of the term structure of interest rates. Using equation (44) and the generated output growth, we may compute $\{R_t^*, t = 1, ..., T\} \equiv \{\mathbb{P}[S_t^* \mid \mathcal{F}_t^*], t = 1, ..., T\}$, where $\mathcal{F}_t^* \triangleq \sigma\{y_t^*, t = 1, ..., T\}$ is the σ-algebra generated by our simulated output growth process. Adjusting R_t backwards by τ quarters—following from equation (66)—provides us with the desired process, $\{\Lambda_t^*, t = 1, ..., T\}$.

Step 3: The Term Structure In this step we make use of equation (67) and $\{\Lambda_t^*, t = 1, ..., T\}$, which was created in the previous step. These inputs are used to construct the market price of risk parameter, $\tilde{\lambda}_{1,t}^*$. With $\{\tilde{\lambda}_{1,t}^*, t = 1, ..., T\}$, we can now construct the associated term structure of interest rates in a manner that is consistent with the observed sequence of macroeconomic states.[32] This completes the construction of the term structure of interest rates.

[32] See Appendix B for a description of how to generate the non-central χ^2-variates necessary for the construction of the two-factor CIR model.

Step 4: The Government's Financial Position The final step involves using equation (68) to compute a realization for the government's financial position, $\{F_t^*, t = 1, ..., T\}$.

This algorithm is summarized in Figure 13, which essentially fills in the details around the general presentation in Figure 2.

A single sample path stemming from the previously described four-step algorithm is summarized in Figure 12. Observe, in the first quadrant of Figure 12, that this realization has two recessionary periods: one lasting for a single quarter and another persisting for three quarters. We can see, in the case of the prolonged three-quarter recession, that Λ_t provided $\tau = 4$ quarters of warning of the impending economic slowdown. This contributes to the relative flat set of term-structure outcomes for this period as described in the fourth quadrant of Figure 12. Finally, note that the recessionary periods are associated with weak output growth and troughs in the government financial position realizations.

Figure 12: **Simulation Results**: This graph illustrates the various outputs for the construction of a sample path for output, expenditure, and revenue growth as well as the corresponding term structure of interest rates.

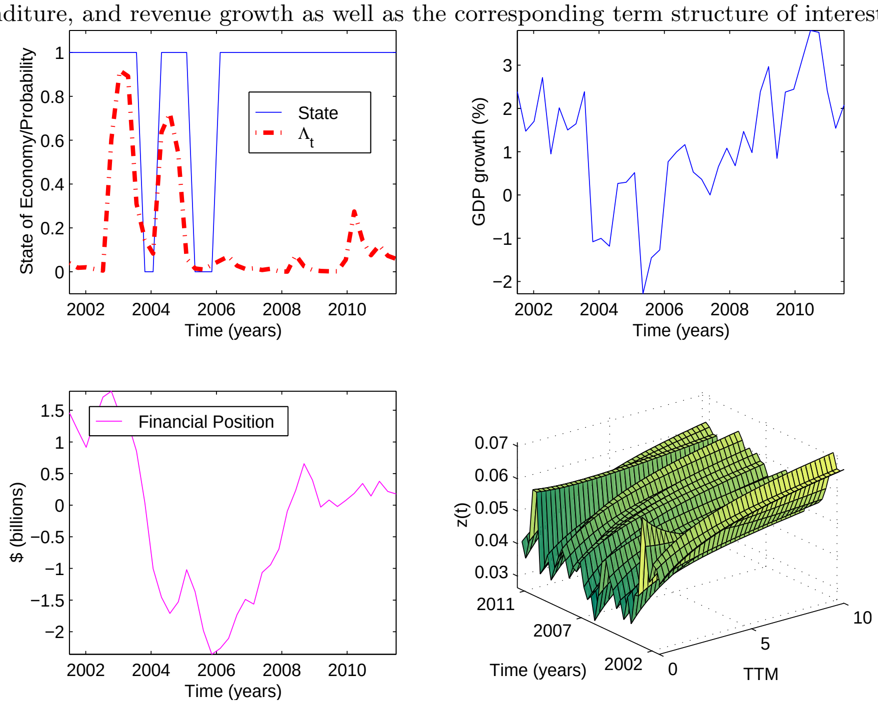
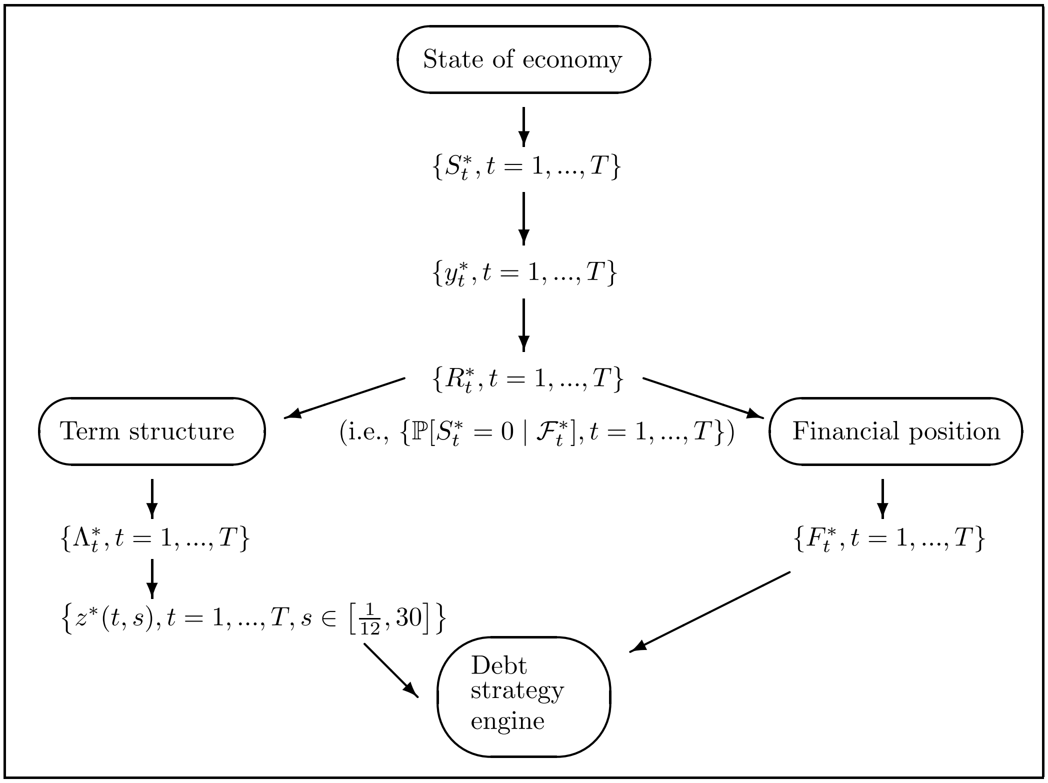


Figure 13: **The Simulation Framework**: This figure summarizes the basic algorithm suggested for the generation of a realization for our joint model of the business cycle, the term structure of interest rates, and the government's financial position. Figure 12 described a single realization in a graphical manner.



5 Conclusion

The objective of this paper was to construct a parsimonious reduced-form model describing the joint evolution of the economic business cycle, the government's financial position, and the term structure of interest rates. To accomplish this goal, we modelled the dynamics of the business cycle with the hidden-Markov model suggested by Hamilton (1989). This allowed us to build our entire stochastic framework on a conceptually straightforward and flexible foundation. We then employed a transformation of the filtered probability of recession to capture the flat or inverted term-structure outcomes observed to occur prior to business cycle downturns. We are able to capture these dynamics—in an admittedly simplistic manner—by constructing a time-varying market price of risk parameter through a convex combination involving the filtered probabilities. Finally, we specify the government's financial position as a modified Ornstein-Uhlenbeck process. The process is modified in the sense that the dynamics of the government's financial position depend importantly on the current state of the business cycle.

How helpful is this model to our debt strategy analysis? To answer this question, we revisit the guidelines established in the first section of this paper. We indicated that our model should permit us to perform *stress* analysis. One of the advantages of the hidden-Markov framework is its potential application to this area. In particular, we could easily add a third state that leads to the occurrence of extreme outcomes with small, but positive, probability. As stated earlier, this is conceptually similar to the peso problem in economics. These negative realizations could involve both the term structure of interest rates and the government's financial position. The specification of the transition probabilities of this additional state would permit us to incorporate some element of the probability of occurrence into our stress-testing framework. This is interesting because standard stress-testing methodologies do *not* make any statement about the probability of their occurrence. This methodology, therefore, represents a welcome addition to this important but difficult area of debt strategy analysis.

Another important guideline is sufficient model flexibility to allow *sensitivity* analysis. The simplicity of the financial position model and the calibration approach permits examination of the sensitivity of the results to any assumption about the future dynamics of the government's financial position. We could, for example, consider the impact on the debt charge distribution of an increase in the impact of a recession on the government's financial position. Conversely, we might want to increase the mean-reversion properties of the financial position process. This is equivalent to saying that the government is quicker to take steps in altering discretionary spending in the face of budget deficits. In general, this model specification is extremely useful in helping us to understand how different financing strategies react to changes in model assumptions.

In our initial guidelines, we asked that the stochastic processes employed capture the general empirical properties of the individual random macroeconomic variables. There are at least two weaknesses in this area. First, the proposed methodology does a reasonable job of creating flat or inverted term structures prior to recession, but it does not perform as well at generating steep term-structure outcomes at the end of a recession. The reasons for this are primarily structural. The proposed technique is not sufficiently flexible to capture the term-structure dynamics in a symmetric fashion. That is, it generates the requisite flatness τ periods prior to recession, but it does not allow for greater-than-average steepness τ periods prior to a recovery. This is a weakness in our approach. The methodology does allow for the term structure to be relatively steep τ periods prior to a recovery, compared with τ periods prior to a recession. Despite the weakness in our technique, we felt that capturing the leading indicator nature of term-structure steepness prior to recession was most important. Moreover, the modifications we considered to rectify this drawback were ultimately rejected owing to their increased complexity.

Second, the suggested modified Ornstein-Uhlenbeck process for the government's financial position is quite simplistic. There may be more sophisticated and realistic ways to capture the random evolution of the government's finances. In addition, our inability to estimate model parameters is a concern. Nevertheless, a number of alternative specifications we examined for this purpose exhibited rather poor behaviour. In particular, every specification without mean-reversion that we considered displayed the rather serious problem of explosive government deficit and surplus positions. Ultimately, we opted for a simple calibrated process for the government's financial position, because it is well-behaved, easy to interpret, and has an explicit dependence on the current state of the business cycle.

This paper, in conjunction with Bolder (2001), has taken a first step towards broadening the stochastic framework used in the analysis of the Government of Canada's debt strategy problem. In the introduction, we outlined a number of guidelines to focus our analysis. Among them, defensibility, parsimony, and flexibility were paramount. The resulting approach, therefore, involves simple models that capture first-order effects often at the expense of statistical and theoretical sophistication. Clearly, more work needs to be done in this area. Nevertheless, we feel that the suggested model serves as a reasonable foundation for our stochastic simulation framework. Moreover, because it represents an improvement upon our current approach, it should help to increase our confidence in the risk and cost measures generated in the course of our debt strategy analysis.

Appendix A: Some Markov Chain Results

In section 3.1, it was claimed that the transition probabilities for a Markov chain m periods in the future are described by the matrix, P^m. To demonstrate this concept, we introduce an $N \times 1$ random vector, ξ_t, which has the following form

$$\xi_t = \begin{cases} \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T, & \text{if } S_t = 1 \\ \begin{bmatrix} 0 & 1 & \cdots & 0 \end{bmatrix}^T, & \text{if } S_t = 2 \\ \vdots & \vdots \\ \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}^T, & \text{if } S_t = N \end{cases}. \tag{70}$$

This is a useful abstraction. It introduces the convenience of the indicator function into our analysis. More specifically, if $S_t = i$, then the jth element of ξ_t is equal to one with probability, p_{ij}, and zero otherwise. Moreover, if we write out the conditional expectation of ξ_{t+1} given that $S_t = i$, we have

$$\mathbb{E}[\xi_{t+1} \mid S_t = i] = \begin{bmatrix} p_{i1} \\ p_{i2} \\ \vdots \\ p_{iN} \end{bmatrix}. \tag{71}$$

This is merely the ith column of the transition matrix. Moreover, the actual realization of ξ_t is merely the ith column of an $N \times N$ identity matrix. To see this more clearly, consider a three-state example where

$$\mathbb{E}[\xi_{t+1} \mid S_t = 1] = \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \end{bmatrix}, \tag{72}$$

where,

$$\xi_t = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Combining the two expressions from equation (72), and using the fact that conditioning on $S_t = 1$ or ξ_t is equivalent, we may conclude with a bit of simple matrix algebra that,

$$
\begin{aligned}
\mathbb{E}[\xi_{t+1} \mid S_t = 1] &= \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \end{bmatrix}, \\
\mathbb{E}[\xi_{t+1} \mid \xi_t] &= \begin{bmatrix} p_{11} & p_{21} & p_{31} \\ p_{12} & p_{22} & p_{32} \\ p_{13} & p_{23} & p_{33} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \\
&= P\xi_t.
\end{aligned} \tag{73}
$$

Now we use the result in equation (73), the Markov property, and a bit of manipulation, to obtain the following useful identity,

$$
\begin{aligned}
\mathbb{E}[\xi_{t+1} \mid \xi_t] &= P\xi_t, \\
\mathbb{E}[\xi_{t+1} \mid \xi_t, \xi_{t-1}, ...] &= P\xi_t, \\
\underbrace{\xi_{t+1} - \mathbb{E}[\xi_{t+1} \mid \xi_t, \xi_{t-1}, ...]}_{\text{Call this } \nu_{t+1}} &= \xi_{t+1} - P\xi_t, \\
\xi_{t+1} &= P\xi_t - \nu_{t+1}.
\end{aligned} \tag{74}
$$

Thus, we have that ξ_t is an AR(1) process and, given that ν_t is a sequence of martingale differences, it has zero expectation (i.e., $\mathbb{E}[\nu_t] = 0$). Our original goal, however, was to see what happens m periods into the future. Consider, therefore, the simple case for $t + 2$,

$$
\begin{aligned}
\xi_{t+2} &= P \underbrace{\xi_{t+1}}_{\substack{\text{equation} \\ (74)}} - \nu_{t+2}, \\
\xi_{t+2} &= P(P\xi_t - \nu_{t+1}) - \nu_{t+2}, \\
\xi_{t+2} &= P^2\xi_t - P\nu_{t+1} - \nu_{t+2}.
\end{aligned} \tag{75}
$$

If we generalize equation (75) for an m period forecast, we have

$$
\begin{aligned}
\xi_{t+m} &= P^0\nu_{t+m-0} + P^1\nu_{t+m-1} + \cdots + P^{m-1}\nu_{t+m-(m-1)} + P^m\xi_t, \\
&= P^m\xi_t + \sum_{i=0}^{m-1} P^i\nu_{t+m-i}.
\end{aligned} \tag{76}
$$

This is a useful recursion relation, but it becomes even more handy when we add the expectation operator.

Recall that the ν_t terms have zero expectation. Therefore,

$$\begin{aligned}
\mathbb{E}[\xi_{t+m} \mid \xi_t, \xi_{t-1}, ...] &= \mathbb{E}[P^m \xi_t + \sum_{i=0}^{m-1} P^i \nu_{t+m-i} \mid \xi_t, \xi_{t-1}, ...], \qquad (77)\\
&= \mathbb{E}[P^m \xi_t \mid \xi_t, \xi_{t-1}, ...] + \mathbb{E}\left[\sum_{i=0}^{m-1} P^i \nu_{t+m-i} \mid \xi_t, \xi_{t-1}, ...\right],\\
&= P^m \underbrace{\mathbb{E}[\xi_t \mid \xi_t, \xi_{t-1}, ...]}_{=\xi_t} + \sum_{i=0}^{m-1} P^i \underbrace{\mathbb{E}\left[\nu_{t+m-i} \mid \xi_t, \xi_{t-1}, ...\right]}_{=0},\\
\mathbb{E}[\xi_{t+m} \mid \xi_t] &= P^m \xi_t.
\end{aligned}$$

This is a useful result. It holds that the transition probabilities for m periods in the future are provided by raising the one-period transition matrix to the mth power.

In section 3.1, we also briefly addressed what happens to the transition probabilities of our Markov chain as m gets very large. Loosely speaking, as we consider the limit of m tending to infinity in equation (15), the importance of the initial value decreases until we essentially lack any conditioning information. Thus, we can interpret these limiting values as the ergodic, or unconditional, probabilities of our Markov chain. Another way of thinking about these probabilities is to imagine a Markov chain achieving a steady set of transition probabilities after a large number of periods. To the extent that these values serve as the starting point for the non-linear filter we use to estimate the parameter values in section 3.2, we will need some method for computing these probabilities. To see how this is done, let us consider a two-state example with the following transition matrix,

$$P = \begin{bmatrix} q & 1-p \\ 1-q & p \end{bmatrix}. \qquad (78)$$

The first step is to find the eigenvalues of this matrix. To do that, we need to find the roots of the characteristic polynomial,

$$\begin{aligned}
\det(P - \lambda I) &= \det\left(\begin{bmatrix} q-\lambda & 1-p \\ 1-q & p-\lambda \end{bmatrix}\right), \qquad (79)\\
&= (q-\lambda)(p-\lambda) - (1-p)(1-q),\\
&= \lambda^2 - \lambda(q+p) + (q+p-1),\\
&= (\lambda - 1)(\lambda + 1 - q - p).
\end{aligned}$$

This implies that the eigenvalues are,

$$\lambda_1 = 1, \qquad (80)$$

$$\lambda_2 = q + p - 1, \qquad (81)$$

where I denotes the identity matrix. Using the definition of ergodicity, introduced in section 3.1, we can place conditions on q and p. In particular, we have one eigenvalue, λ_1, that is unity and thus the second eigenvalue must satisfy

$$\begin{aligned} \mid \lambda_2 \mid &< 1, \\ \mid q+p-1 \mid &< 1, \\ -1 < q+p-1 &< 1, \\ 0 < q+p &< 2. \end{aligned} \tag{82}$$

The next step is to compute the eigenvectors associated with our two eigenvectors. This is done in the usual way.[33] The results are,

$$x_{\lambda_1} = \begin{bmatrix} \frac{1-p}{2-q-p} & \frac{1-q}{2-q-p} \end{bmatrix}^T, \tag{85}$$

$$x_{\lambda_2} = \begin{bmatrix} -1 & 1 \end{bmatrix}^T. \tag{86}$$

It is worth noting that the second eigenvector was normalized such that the following is true,

$$\vec{1} x_{\lambda_2} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1-p}{2-q-p} \\ \frac{1-q}{2-q-p} \end{bmatrix} = 1. \tag{87}$$

At this point, we make the claim that x_{λ_2} is, in fact, the vector of ergodic probabilities. That is, the unconditional probabilities are,

$$\mathbb{P}[S_t = 0] = \frac{1-p}{2-q-p}, \tag{88}$$

$$\mathbb{P}[S_t = 1] = \frac{1-q}{2-q-p}. \tag{89}$$

While we will not prove this result, we will work through a simple exercise to provide some intuition.[34] First, we have shown that our matrix of transition probabilities, P, has two real-valued, distinct eigenvalues and

[33] That is, we must find a basis for the parameterized solution of the following two equations,

$$(P - \lambda_1 I)x_{\lambda_1} = 0, \tag{83}$$

$$(P - \lambda_2 I)x_{\lambda_2} = 0. \tag{84}$$

[34] This discussion is not a proof because we consider only the case where the eigenvalues in our spectral decomposition are real-valued and distinct. While it is possible to generalize the following discussion to these cases, it is a level of generality that is not required in our analysis.

thus we can safely perform the following spectral decomposition,

$$
\begin{aligned}
P &= A\Lambda A^{-1}, \qquad (90)\\
&= \begin{bmatrix} \frac{1-p}{2-q-p} & -1 \\ \frac{1-q}{2-q-p} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & q+p-1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \frac{-(1-q)}{2-q-p} & \frac{1-p}{2-q-p} \end{bmatrix}.
\end{aligned}
$$

We have seen that if we want the m-period-ahead transition probabilities for an ergodic Markov chain, we need only to compute P^m. Clearly, the ergodic probabilities would be the limit of this quantity as m tends to infinity. Consider, therefore, the following,

$$
\begin{aligned}
\lim_{m\to\infty} P^m &= \lim_{m\to\infty} \left(A\Lambda^m A^{-1}\right), \qquad (91)\\
&= A \lim_{m\to\infty} \left(\Lambda^m\right) A^{-1},\\
&= A \lim_{m\to\infty} \left(\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & q+p-1 \end{bmatrix}}_{\substack{\text{Recall } (q+p-1)<1 \\ \text{by definition.}}} \right)^m A^{-1},\\
&= A \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} A^{-1},\\
&= \begin{bmatrix} \frac{1-p}{2-q-p} & \frac{1-p}{2-q-p} \\ \frac{1-q}{2-q-p} & \frac{1-q}{2-q-p} \end{bmatrix},\\
&= \begin{bmatrix} \frac{1-p}{2-q-p} \\ \frac{1-q}{2-q-p} \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix},\\
&= \pi\vec{1},
\end{aligned}
$$

where $\pi = x_{\lambda_2}$, as in equation (15). Thus, we can see that, in a loose sense, the limiting probabilities are given by the second eigenvector.[35]

[35] Furthermore, this gives some intuition as to why, to have an ergodic Markov chain, only one eigenvalue can be unity and the rest must lie inside the unit circle.

Appendix B: The Non-Central χ^2 Distribution

This appendix briefly discusses how one can make random draws from a non-central χ^2 distribution. The appropriate place to start in this discussion is with the χ^2 distribution. The χ^2 distribution is a special case of the gamma distribution. There are, however, two additional facts about the χ^2 distribution that are both interesting and useful in its simulation:

- If $X \sim \mathcal{N}(0,1)$, then $X^2 \sim \chi^2(1)$. That is, the square of a standard normal variate is a χ^2 random variate with one degree of freedom.
- If $X_1, ..., X_n$ are independent random variates and $X_i \sim \chi^2(1)$, then $\sum_{i=1}^{n} X_i \sim \chi^2(n)$. Or, rather, independent χ^2 normal variates sum (as do their degrees of freedom) to a χ^2 random variate.[36]

How, therefore, does a *non-central* χ^2 distribution arise? If $X_1, X_2, ..., X_n$ are standard normal variates and $a_1, ..., a_n \in \mathbb{R}$, then,

$$\sum_{i=1}^{n} (X_i + a_i)^2 \sim \chi^2(n, m), \tag{92}$$

where,

$$m = \sum_{i=1}^{n} a_i^2, \tag{93}$$

and $\chi^2(n,m)$ denotes a non-central χ^2 distribution with n degrees of freedom and a non-centrality parameter, m.[37] Two interesting facts about the non-central χ^2 distribution—related closely to the two previously mentioned facts about the χ^2 distribution—will assist us in our task:

- If $X \sim \mathcal{N}(\sqrt{m}, 1)$, then $X^2 \sim \chi^2(1, m)$. Therefore, a squared standard normal random variate gives rise to a χ^2 random variate, while a squared normal random variate with mean, $\sqrt{m}$, leads to a non-central χ^2 random variate with a non-centrality parameter, m.
- If $X \sim \chi^2(1)$ and $Y \sim \chi^2(1, m)$ are independent random variables, then $X + Y \sim \chi^2(2, m)$.

These two features of the non-central χ^2 provide us with the actual simulation algorithm. In particular, one may simulate a random variate from a $\chi^2(a, b)$ distribution in two steps. First, generate X such that $X \sim \mathcal{N}(\sqrt{b}, 1)$. Then, merely generate an independent Y such that $Y \sim \chi^2(a-1)$. $X^2 + Y$ is, therefore, a draw from a $\chi^2(a, b)$ distribution.

[36] These facts come almost directly from Casella and Berger (1990, page 222).

[37] This definition follows from Johnson, Kotz, and Balakrishnan (1997, page 433).

Armed with this information, we can generate the state variables in equations (59) and (60) from the two-factor CIR term-structure model. The transition density of each state variable, y_i, is as follows,

$$y_i(t) \mid y_i(s) \sim \gamma\chi^2(\alpha, \beta), \tag{94}$$

where,

$$\alpha = \frac{4\kappa_i}{\sigma_i^2}, \tag{95}$$

$$\beta = \frac{4\kappa_i e^{-\kappa_i(t-s)}}{\sigma_i^2\left(1 - e^{-\kappa_i(t-s)}\right)} y_i(s), \tag{96}$$

$$\gamma = \frac{\sigma_i^2\left(1 - e^{-\kappa_i(t-s)}\right)}{4\kappa_i}, \tag{97}$$

for $t \geq s$ and $i = 1, 2$.[38] To reiterate, therefore, we generate X such that $X \sim \mathcal{N}(\sqrt{\beta}, 1)$. Then, we draw independent Y such that $Y \sim \chi^2(\alpha - 1)$. The simulated value of $y_i(t)$, conditioning on $y_i(s)$, is thus $\gamma(X^2 + Y)$.[39]

[38] For more details on the exact nature of this transition density, see Cox, Ingersoll, and Ross (1985b, pp. 391-92).

[39] I am thankful to Mark Reesor of the University of Western Ontario and Antje Berndt of Stanford University for bringing this straightforward algorithm to my attention.

Appendix C: The Term Premium

This appendix relates to a question in Appendix C of de Jong (2000). I have specialized the analysis of the question to the single-factor CIR model because this eases the analysis somewhat.[40] We are given the Itô dynamics of the affine bond price function,

$$\frac{dP(\tau)}{P(\tau)} = \mu_{P(\tau)}dt + \sigma_{P(\tau)}dW(t), \tag{98}$$

where $\{W(t), t \in [0, \infty)\}$ is a standard, scalar Wiener process defined on the filtered probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. The claim is that the coefficients have the following form,

$$\mu_{P(\tau)} = r(t) + \lambda\sigma_{P(\tau)}, \tag{99}$$

$$\sigma_{P(\tau)} = -\sigma^2 r(t)B(\tau). \tag{100}$$

Moreover, one may conclude that the term premium for a longer-term bond is equal to $-\lambda\sigma^2 r(t)B(\tau)$. I have tried to establish this claim directly using Itô's theorem and the actual CIR bond price function, and have not been successful. It is, of course, possible that I have

- formulated the problem incorrectly or have made a serious logical error in the way I have approached the establishment of this claim,
- or, if this is not the case, I have made some fundamental error in my calculations.

Here is my approach. By assumption, we have that,

$$P(\tau) = e^{A(\tau)-B(\tau)r(t)}, \tag{101}$$

where,

$$A(\tau) = \ln\left[\left(\frac{2\gamma e^{\frac{a\tau}{2}}}{a(e^{\gamma\tau}-1)+2\gamma}\right)^{\frac{2\kappa\theta}{\sigma^2}}\right], \tag{102}$$

$$B(\tau) = \frac{2(e^{\gamma\tau}-1)}{a(e^{\gamma\tau}-1)+2\gamma} \tag{103}$$

and,

$$a = \kappa + \lambda + \gamma, \tag{104}$$

$$\gamma = \sqrt{(\kappa+\lambda)^2 + 2\sigma^2}. \tag{105}$$

[40] It is also a member of the affine class and thus this analysis should hold in a two-factor setting.

By Itô's theorem, we have that

$$dP(\tau) = P_t dt + P_r dr(t) + \frac{1}{2}P_{rr}d\langle r\rangle(t), \tag{106}$$

$$= P_t dt + P_r \left(\underbrace{(\kappa\theta - (\kappa+\lambda)r(t))dt + \sigma\sqrt{r(t)}dW(t)}_{\text{dynamics of } r(t) \text{ under } \mathbb{P}} \right) + \frac{1}{2}P_{rr}\sigma^2 r(t)dt, \tag{107}$$

$$= \left(P_t + (\kappa\theta - (\kappa+\lambda)r(t))P_r + \frac{\sigma^2}{2}P_{rr}r(t) \right) dt + \sigma\sqrt{r(t)}P_r dW(t). \tag{108}$$

By equation (101), we have the required partial derivatives,

$$P_t = (-A'(\tau) + B'(\tau)r)P(\tau), \tag{109}$$

$$P_r = -B(\tau)P(\tau), \tag{110}$$

$$P_{rr} = B^2(\tau)P(\tau), \tag{111}$$

where $A'(\tau)$ and $B'(\tau)$ represent the partial derivatives of the functions $A(\tau)$ and $B(\tau)$ with respect to t. If we substitute equations (109) to (111) into equation (108) and simplify, we have

$$\frac{dP(\tau)}{P(\tau)} = \left(-(A'(\tau) + \kappa\theta B(\tau)) + (B'(\tau) + (\kappa+\lambda)B(\tau) + \frac{\sigma^2}{2}B^2(\tau))r(t) \right) dt - \sigma\sqrt{r(t)}B(\tau)dW(t). \tag{112}$$

Thus, we have,

$$\mu_{P(\tau)} = -(A'(\tau) + \kappa\theta B(\tau)) + (B'(\tau) + (\kappa+\lambda)B(\tau) + \frac{\sigma^2}{2}B^2(\tau))r(t), \tag{113}$$

$$\sigma_{P(\tau)} = -\sigma\sqrt{r(t)}B(\tau). \tag{114}$$

To establish the claim, we require that equation (113) be of the same form as equation (99) and also that equation (114) be equivalent to equation (100). The second set of equations are equivalent almost directly. I suspect that equation (C.4) in de Jong (2000) either has a typographical error or I have made a mistake in my algebra. To show the equivalence of the first two equations, we require that the following two expressions hold,

$$B'(\tau) + \kappa B(\tau) + \frac{\sigma^2}{2}B^2(\tau) = 1, \tag{115}$$

$$A'(\tau) + \kappa\theta B(\tau) = -\lambda\sigma\sqrt{r(t)}B(\tau). \tag{116}$$

Let us see whether we can establish these two equalities. We will start with the partial derivatives of $A(\tau)$ and $B(\tau)$ with respect to t. We have,

$$A'(\tau) = \frac{A(\tau)}{\partial t} = \frac{\partial}{\partial t}\left[\underbrace{\ln\left[\left(\frac{2\gamma e^{\frac{a\tau}{2}}}{a(e^{\gamma\tau}-1)+2\gamma}\right)^{\frac{2\kappa\theta}{\sigma^2}}\right]}_{\text{equation (102)}}\right] = \frac{\kappa\theta a}{\sigma^2}\left[\frac{(e^{\gamma\tau}-1)(2\gamma-a)}{a(e^{\gamma\tau}-1)+2\gamma}\right], \tag{117}$$

$$B'(\tau) = \frac{B(\tau)}{\partial t} = \frac{\partial}{\partial t}\left[\underbrace{\frac{2(e^{\gamma\tau}-1)}{a(e^{\gamma\tau}-1)+2\gamma}}_{\text{equation (103)}}\right] = \frac{-4\gamma^2 e^{\gamma\tau}}{(a(e^{\gamma\tau}-1)+2\gamma)^2}. \tag{118}$$

To rearrange equation (115), we also require the following,

$$(\kappa+\lambda)B(\tau) = \frac{2(\kappa+\lambda)(e^{\gamma\tau}-1)}{a(e^{\gamma\tau}-1)+2\gamma}\cdot\frac{a(e^{\gamma\tau}-1)+2\gamma}{a(e^{\gamma\tau}-1)+2\gamma}, \tag{119}$$

$$= \frac{2(\kappa+\lambda)a(e^{\gamma\tau}-1)^2+4(\kappa+\lambda)\gamma(e^{\gamma\tau}-1)}{(a(e^{\gamma\tau}-1)+2\gamma)^2}, \tag{120}$$

and,

$$\frac{\sigma^2}{2}B^2(\tau) = \frac{2\sigma^2(e^{\gamma\tau}-1)^2}{(a(e^{\gamma\tau}-1)+2\gamma)^2}. \tag{121}$$

We can now represent the left-hand side of equation (115) using equations (118) to (121) as follows,

$$1 = B'(\tau)+\kappa B(\tau)+\frac{\sigma^2}{2}B^2(\tau) \tag{122}$$

$$1 = \frac{-4\gamma^2e^{\gamma\tau}+2(\kappa+\lambda)a(e^{\gamma\tau}-1)^2+4(\kappa+\lambda)\gamma(e^{\gamma\tau}-1)+2\sigma^2(e^{\gamma\tau}-1)^2}{(a(e^{\gamma\tau}-1)+2\gamma)^2}, \tag{123}$$

$$1 = \frac{2((\kappa+\lambda)a+\sigma^2)(e^{\gamma\tau}-1)^2+4(\kappa+\lambda)\gamma(e^{\gamma\tau}-1)-4\gamma^2e^{\gamma\tau}}{a^2(e^{\gamma\tau}-1)^2-4\gamma a(e^{\gamma\tau}-1)+4\gamma^2}, \tag{124}$$

$$0 = 2((\kappa+\lambda)a+\sigma^2)(e^{\gamma\tau}-1)^2+4(\kappa+\lambda)\gamma(e^{\gamma\tau}-1)-4\gamma^2e^{\gamma\tau}- \\ \left(a^2(e^{\gamma\tau}-1)^2-4\gamma a(e^{\gamma\tau}-1)+4\gamma^2\right), \tag{125}$$

$$0 = 4(\kappa+\lambda)\gamma(e^{\gamma\tau}-1)-4\gamma^2e^{\gamma\tau}+4\gamma a(e^{\gamma\tau}-1)-4\gamma^2, \tag{126}$$

$$0 = -4\gamma^2e^{\gamma\tau}, \tag{127}$$

which is clearly untrue. Nevertheless, it is all tantalizingly close and I wonder whether I am missing a trick or a manipulation somewhere (or, what is more likely, an error). In particular, the difference relates to the numerator of $B'(\tau)$, which seems suspiciously coincidental (although I cannot locate an error). Finally, we can examine the LHS of equation (116) using equations (117) and equation (103) as follows,

$$\text{LHS of equation (116)} = \frac{\kappa\theta a}{\sigma^2}\left[\frac{(e^{\gamma\tau}-1)(2\gamma-a)}{a(e^{\gamma\tau}-1)+2\gamma}\right] + \frac{2\kappa\theta(\gamma e^{\gamma\tau}-1)}{a(e^{\gamma\tau}-1)+2\gamma}, \tag{128}$$

$$= \frac{\kappa\theta(e^{\gamma\tau-1})(a(2\gamma-a)+2\sigma^2)}{\sigma^2(a(e^{\gamma\tau}-1)+2\gamma)}. \tag{129}$$

A bit of manipulation reveals that,

$$a(2\gamma-a)+2\sigma^2 = ((\kappa+\lambda)+\gamma)(-(\kappa+\lambda)+\gamma)) + 2\sigma^2, \tag{130}$$

$$= (\gamma^2-(\kappa+\lambda)^2)+2\sigma^2, \tag{131}$$

$$= (\kappa+\lambda)^2+2\sigma^2-(\kappa+\lambda)^2+2\sigma^2, \tag{132}$$

$$= 4\sigma^2. \tag{133}$$

Substitution of this result into equation (129) yields,

$$\frac{\kappa\theta(e^{\gamma\tau-1})4\sigma^2}{\sigma^2(a(e^{\gamma\tau}-1)+2\gamma)} = \frac{2\kappa\theta\left(2(e^{\gamma\tau-1})\right)}{(a(e^{\gamma\tau}-1)+2\gamma)} = 2\kappa\theta B(\tau). \tag{134}$$

Again, this is not equal to the RHS of equation (116). Matching coefficients, it is clear that,

$$2\kappa\theta \neq -\lambda\sigma\sqrt{r(t)}. \tag{135}$$

As before, it is unclear where the error in this approach lies.

Appendix D: The Extended Filter

Hamilton's (1989) filter is a very useful way to introduce non-linearity into the study of economic regimes. Indeed, it permits us to simultaneously estimate the AR dynamics of the growth in output and the two distinct regimes for that growth. A drawback of these models, however, is that conditional on being in a given state, the probability of either remaining in that state or transitioning to another state is *fixed*. It depends, in fact, only on the previous state. This is a simple feature of Markov chains. There are, nevertheless, times when it would be useful to make these probabilities a function of some set of leading economic indicators. In this way, the elements of our transition matrix would vary over time. An additional consequence of a time-varying transition matrix is that the expected duration of a given business cycle is not—as in the simple hidden-Markov model—a constant value.

How, therefore, do we implement this type of approach? The basic model is the same as in equation (16), but we must define the transition probabilities in a more complicated manner than the representation in equations (20) to (23).[41] In particular, we give the transition probabilities the following logistic form,

$$\mathbb{P}[S_t = 1 \mid S_{t-1} = 1, z_{t-1}] = p(z_t) = \frac{e^{\alpha_1 + z_{t-1}^T \beta_1}}{1 + e^{\alpha_1 + z_{t-1}^T \beta_1}}, \tag{136}$$

$$\mathbb{P}[S_t = 1 \mid S_{t-1} = 0, z_{t-1}] = 1 - p(z_t) = \frac{1}{1 + e^{\alpha_1 + z_{t-1}^T \beta_1}}, \tag{137}$$

$$\mathbb{P}[S_t = 0 \mid S_{t-1} = 0, z_{t-1}] = q(z_t) = \frac{e^{\alpha_0 + z_{t-1}^T \beta_0}}{1 + e^{\alpha_0 + z_{t-1}^T \beta_0}}, \tag{138}$$

$$\mathbb{P}[S_t = 0 \mid S_{t-1} = 1, z_{t-1}] = 1 - q(z_t) = \frac{1}{1 + e^{\alpha_0 + z_{t-1}^T \beta_0}}, \tag{139}$$

where z_t is a vector of leading economic indicators at time t.[42] As we will see, this has some slight, but important, implications for the form of our non-linear filtering algorithm. First, it implies that our parameter vector is larger. That is,

$$\theta = \begin{bmatrix} \mu_0 & \mu_1 & \sigma^2 & \phi_1 & \alpha_0 & \alpha_1 & \beta_0 & \beta_1 \end{bmatrix}^T. \tag{141}$$

The approach used to determine this parameter vector is—as is the case with Hamilton's (1989) filter—the maximization of the conditional log-likelihood function. Specifically, we numerically solve the following

[41] This approach was introduced in Filardo (1993, 1998), and Filardo and Gordon (1993, 1994). This appendix draws heavily on their works.

[42] This implies that our transition matrix is,

$$P = \begin{bmatrix} q(z_t) & 1 - p(z_t) \\ 1 - q(z_t) & p(z_t) \end{bmatrix}. \tag{140}$$

In a two-state example, such as this, it is possible to give the transition probabilities an inverse standard normal distribution function (i.e., probit) specification.

non-linear maximization problem,

$$\max_{\theta} \sum_{t=1}^{T} \ln\left(f^*(y_t \mid y_{t-1}, z_{t-1}; \theta)\right). \tag{142}$$

The trick, introduced in Filardo (1993), is to decompose f^* in a manner that allows us to apply the non-linear filtering algorithm proposed by Hamilton (1989). How does it work? As usual, we begin with the initialization of the log-likelihood function at zero, $\ell(\theta)$. We then compute the steady-state probabilities introduced in Appendix A using,

$$\mathbb{P}[S_0 = 1 \mid y_0, z_0] = \frac{1 - q(z_0)}{2 - p(z_0) - q(z_0)} = \frac{1 + e^{\alpha_1 + z_0^T \beta_1}}{2 + e^{\alpha_0 + z_0^T \beta_0} + e^{\alpha_1 + z_0^T \beta_1}}, \tag{143}$$

$$\mathbb{P}[S_0 = 0 \mid y_0, z_0] = \frac{1 - q(z_0)}{2 - p(z_0) - q(z_0)} = \frac{1 + e^{\alpha_0 + z_0^T \beta_0}}{2 + e^{\alpha_0 + z_0^T \beta_0} + e^{\alpha_1 + z_0^T \beta_1}}. \tag{144}$$

Once again, the next step is the computation of the joint probabilities. Consider the following:

$$\begin{aligned}\mathbb{P}[S_1 = i, S_0 = j \mid y_0, z_0] &= \underbrace{\mathbb{P}[S_1 = i \mid S_0 = j, y_0, z_0]}_{\substack{\text{Independent of } y_0 \\ \text{by Markov property}}} \cdot \mathbb{P}[S_0 = j \mid y_0, z_0], \\ &= \underbrace{\mathbb{P}[S_1 = i \mid S_0 = j, z_0]}_{\text{equations (136) to (139)}} \cdot \underbrace{\mathbb{P}[S_0 = j \mid y_0, z_0]}_{\text{equation (143) or (144)}},\end{aligned} \tag{145}$$

where $i, j = 0, 1$. Observe that, as in Hamilton's (1989) filter, we will have to update the second expression ($\mathbb{P}[S_t = j \mid y_t, z_t]$) at each iteration of the filter. We now have sufficient ammunition to actually expand f^* into something reasonable. This is accomplished through the following expansion,

$$\begin{aligned}&f^*(y_1 \mid y_0, z_0) \\ &= \sum_{i=1}^{1} \sum_{j=1}^{1} f(y_1, S_1 = i, S_0 = j \mid y_0, z_0), \\ &= \sum_{i=1}^{1} \sum_{j=1}^{1} \underbrace{f(y_1 \mid S_1 = i, S_0 = j, y_0, z_0)}_{\substack{\text{Independent of } z_0 \\ \text{by equation (16)}}} \cdot \mathbb{P}[S_1 = i, S_0 = j \mid y_0, z_0], \\ &= \sum_{i=1}^{1} \sum_{j=1}^{1} f(y_1 \mid S_1 = i, S_0 = j, y_0, z_0) \cdot \underbrace{\mathbb{P}[S_1 = i \mid S_0 = j, z_0] \cdot \mathbb{P}[S_0 = j \mid y_0, z_0].}_{\text{equation (145)}}\end{aligned} \tag{146}$$

In gory detail, we have that,

$$
\begin{aligned}
f^*(y_1 \mid y_0, z_0) = & \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1-\mu_1-\phi_1(y_0-\mu_1))^2}{2\sigma^2}\right.} p(z_0) \left(\frac{1+e^{\alpha_1+z_0^T\beta_1}}{2+e^{\alpha_0+z_0^T\beta_0}+e^{\alpha_1+z_0^T\beta_1}} \right) + \\
& \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1-\mu_1-\phi_1(y_0-\mu_0))^2}{2\sigma^2}\right.} (1-p(z_0)) \cdot \left(\frac{1+e^{\alpha_0+z_0^T\beta_0}}{2+e^{\alpha_0+z_0^T\beta_0}+e^{\alpha_1+z_0^T\beta_1}} \right) + \\
& \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1-\mu_0-\phi_1(y_0-\mu_0))^2}{2\sigma^2}\right.} (1-q(z_0)) \left(\frac{1+e^{\alpha_1+z_0^T\beta_1}}{2+e^{\alpha_0+z_0^T\beta_0}+e^{\alpha_1+z_0^T\beta_1}} \right) + \\
& \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y_1-\mu_0-\phi_1(y_0-\mu_1))^2}{2\sigma^2}\right.} q(z_0) \left(\frac{1+e^{\alpha_1+z_0^T\beta_1}}{2+e^{\alpha_1+z_0^T\beta_1}+e^{\alpha_1+z_0^T\beta_1}} \right).
\end{aligned} \tag{147}
$$

This previous expression holds only for the first iteration of the filter, because we were able to use the steady-state probabilities (as summarized in equations (143) and (144)) in our computations of $\mathbb{P}[S_0 = i \mid y_0, z_0]$ for $i = 0, 1$. For subsequent iterations, we need to update this to determine $\mathbb{P}[S_1 = i \mid y_1, z_1]$ for $i = 0, 1$.[43] In particular, we require an updating step similar to that performed in equation (44). This has the following general form,

$$
\begin{aligned}
\mathbb{P}[S_1 = i \mid y_1, z_1] &= \sum_{j=0}^{1} \mathbb{P}[S_1 = i, S_0 = j \mid y_1, z_1], \\
&= \sum_{j=0}^{1} \frac{f(y_1, S_1 = i, S_0 = j \mid y_0, z_0)}{f^*(y_1 \mid y_0, z_0)}, \\
&= \sum_{j=0}^{1} \frac{\overbrace{f(y_1 \mid S_1 = i, S_0 = j, y_0)\mathbb{P}[S_1 = i, S_0 = j \mid y_0, z_0]}^{\text{Easily computed from equations (145) to (147)}}}{\underbrace{f^*(y_1 \mid y_0, z_0)}_{\text{equation (147)}}}.
\end{aligned} \tag{148}
$$

Finally, let us demonstrate how we compute the smoothed probabilities in the context of the time-varying

[43] Of course, in general we compute $\mathbb{P}[S_t = i \mid y_t, z_t]$.

parameter model. The underlying analysis follows from equation (46),

$$
\begin{aligned}
\mathbb{P}[S_t = j \mid y_T, z_T] &= \sum_{k=0}^{1} \mathbb{P}[S_t = j, S_{t+1} = k \mid y_T, z_T], \\
&= \sum_{k=0}^{1} \mathbb{P}[S_{t+1} = k \mid y_T, z_T]\mathbb{P}[S_t = j \mid S_{t+1} = k, y_T, z_T], \\
&= \sum_{k=0}^{1} \mathbb{P}[S_{t+1} = k \mid y_T, z_T] \underbrace{\mathbb{P}[S_t = j \mid S_{t+1} = k, y_t, z_t]}_{\substack{\text{This holds approximately.} \\ \text{See pseudo-proof} \\ \text{in section 3.3.}}}, \\
&= \sum_{k=0}^{1} \mathbb{P}[S_{t+1} = k \mid y_T, z_T] \left(\frac{\mathbb{P}[S_t = j, S_{t+1} = k \mid y_t, z_t]}{\mathbb{P}[S_{t+1} = k \mid y_t, z_t]} \right), \\
&= \sum_{k=0}^{1} \frac{\mathbb{P}[S_{t+1} = k \mid y_T, z_T]\mathbb{P}[S_t = j \mid y_t, z_t] \overbrace{\mathbb{P}[S_{t+1} = k \mid S_t = j, y_t, z_t]}^{\substack{\text{Independent of } y_t \\ \text{by Markov property}}}}{\mathbb{P}[S_{t+1} = k \mid y_t, z_t]}, \\
&= \sum_{k=0}^{1} \frac{\overbrace{\mathbb{P}[S_{t+1} = k \mid y_T, z_T]}^{\text{From previous iteration}} \overbrace{\mathbb{P}[S_t = j \mid y_t, z_t]}^{\text{equation (148)}} \overbrace{\mathbb{P}[S_{t+1} = k \mid S_t = j, z_t]}^{\text{equations (136) to (139)}}}{\sum_{j=0}^{1} \underbrace{\mathbb{P}[S_t = j \mid y_t, z_t]}_{\text{equation (148)}} \underbrace{\mathbb{P}[S_{t+1} = k \mid S_t = j, z_t]}_{\text{equations (136) to (139)}}}.
\end{aligned}
\tag{149}
$$

Again, as in section 3.3, the movement from the second to the third line of the previous manipulation is only an approximation.

Bibliography

Atta-Mensah, J., and G. Tkacz (1998): "Predicting Canadian Recessions using Financial Variables: A Probit Approach," Bank of Canada Working Paper No. 98-5.

Auditor General of Canada (2000): "Managing Canada's Debt: Facing New Challenges," Discussion paper.

Barro, R. J. (1995): "Optimal Debt Management," National Bureau of Economic Research: Working Paper No. 5327.

Bergström, P., and A. Holmlund (2000): "A Simulation Model Framework for Government Debt Analysis," Riksgälds Kontoret: The Swedish National Debt Office.

Billingsley, P. (1995): *Probability and Measure*. Wiley, Third Avenue, New York, New York, third edn.

Black, R., and C. Telmer (1999): "Liability Management Using Dynamic Portfolio Strategies," *Algo Research Quarterly*, 2, 9–19.

Bolder, D. J. (2001): "Affine Term-Structure Models: Theory and Implementation," Bank of Canada Working Paper No. 2001-15.

Bolder, D. J., and D. Stréliski (1999): "*Yield Curve Modelling at the Bank of Canada*," Technical Report No. 84. Ottawa: Bank of Canada.

Brémaud, P. (1999): *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*. Springer Verlag, 175 Fifth Avenue, New York, NY, Texts in Applied Mathematics.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey.

Casella, G., and R. L. Berger (1990): *Statistical Inference*. Duxbury Press, Belmont, California.

Cassard, M., and D. Folkerts-Landau (1997): "Risk Management of Sovereign Assets and Liabilities," International Monetary Fund: Research Department.

Clinton, K. (1994): "The Term Structure of Interest Rates as a Leading Indicator of Economic Activity: A Technical Note," *Bank of Canada Review* (Winter).

Committee on the Global Financial System (2000): "How Should we Design Deep and Liquid Markets: The Case of Government Securities," Discussion paper, Bank for International Settlements.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985a): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

——— (1985b): "A Theory of Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

Cozier, B., and G. Tkacz (1994): "The Term Structure and Real Activity in Canada," Bank of Canada Working Paper No. 94-3.

Danish National Bank (1998a): "Management of Interest Rate and Refinancing Risk: Cost-at-Risk," Discussion paper, Working Party on Debt Management: Organisation for Economic Co-operation and Development.

Danish National Bank (1998b): "The Interest and Refinancing Risk on Domestic Government Debt—Cost at Risk," Discussion paper, Danish National Bank.

DeGrauwe, P. (1989): *International Money: Post-War Trends and Theories*. Oxford University Press, Walton Street, Oxford.

DeJong, F. (2000): "Time-Series and Cross-Section Information in Affine Term Structure Models," *Journal of Business and Economic Statistics*, 18, 300–314.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Department of Finance Canada (2001): "Debt Management Strategy: 2001-2002," Discussion paper.

Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1993): "Regime Switching with Time-varying Transition Probabilities," *Nonstationary Time-Series Analysis and Cointegration*, ed. C. Hargreaves, 283–302.

Durland, J. M., and T. H. McCurdy (1994): "Duration-Dependent Transitions in a Markov Model of U.S. GNP Growth," *Journal of Business and Economic Statistics*, 12(3), 279–288.

Durrett, R. (1996): *Probability: Theory and Examples*. Duxbury Press, Belmont, California, second edn.

Filardo, A. J. (1993): "Business Cycle Phases and their Transitional Dynamics," Federal Reserve Bank of Kansas City: Working Paper No. 93-14.

——— (1998): "Choosing Information Variables for Transition Probabilities in a Time-Varying Transition Probability Markov Switching Model," Federal Reserve Bank of Kansas City: Working Paper No. 93-14.

Filardo, A. J., and S. F. Gordon (1993): "Business Cycle Durations," Federal Reserve Bank of Kansas City: Working Paper No. 93-11.

——— (1994): "International Co-Movements of Business Cycles," Federal Reserve Bank of Kansas City: Working Paper No. 94-11.

Fishman, G. S. (1995): *Monte Carlo: Concepts, Algorithms, and Applications.* Springer-Verlag, 175 Fifth Avenue, New York, NY, Springer Series in Operations Research.

Garcia, R., and P. Perron (1995): "An Analysis of the Real Interest Rate Under Regime Switches," Centre interuniversitaire de recherche en analyse des organizations: Scientific Series 95-5.

Geyer, A. L. J., and S. Pichler (1998): "A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates," Department of Operations Research, University of Economics Vienna.

Goldfeld, S. M., and R. E. Quandt (1973): "A Markov Switching Model for Switching Regressions," *Journal of Econometrics*, 1, 3–16.

Gray, S. (1996): "The Management of Government Debt," Bank of England.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

——— (1990): "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39–70.

——— (1994): *Time Series Analysis.* Princeton University Press, Princeton, New Jersey, Chapter 22.

——— (1996): "Specification Testing in Markov-Switching Time Series Models," *Journal of Econometrics*, 70, 127–157.

Harvey, A. C. (1990): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge, first edn.

Hermanutz, D., and C. Matier (2001): "Modelling Federal Finances under Uncertainty," Department of Finance Canada.

Holmlund, A. (1999): "The Debt Office's Model for Analyzing Duration Choice for the Swedish Kronor Debt," Riksgälds Kontoret: The Swedish National Debt Office.

Hörngren, L. (1999): "Methods for Analyzing the Structure of the Central Government Debt," Riksgälds Kontoret: The Swedish National Debt Office.

Hostland, D. (2001a): "Monetary Policy and Medium-Term Fiscal Planning: Preliminary Version," Department of Finance Canada.

——— (2001b): "Specification of a Stochastic Simulation Model for the Analysis of Monetary and Fiscal Policy," Department of Finance Canada.

Hostland, D., and C. Matier (2001): "An Examination of Alternative Strategies for Reducing Public Debt in the Presence of Uncertainty," Department of Finance Canada.

Ingves, S., and J. Goldstein (2000): "Draft Guidelines for Public Debt Management," International Monetary Fund and The World Bank.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1997): *Continuous Univariate Distributions: Volume 2*. John Wiley and Sons, New York, NY, Wiley Series in Probability and Statistics.

Judge, G. G., W. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1985): *The Theory and Practice of Econometrics*. John Wiley and Sons, New York, New York, second edn.

Karatzas, I., and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*. Springer-Verlag, Berlin, second edn.

Kim, C.-J. (1994): "Dynamic Linear Models with Markov-Switching," *Journal of Econometrics*, 60, 1–22.

Kim, C.-J., and C. R. Nelson (1999): *State-Space Models with Regime Switching*. The MIT Press, Cambridge, Massachusetts, Chapters 4-5.

Litterman, R., and J. Scheinkman (1991): "Common Factors Affecting Bond Returns," *The Journal of Fixed Income*, 1, 54–61.

Meyn, S. P., and R. L. Tweedie (1993): *Markov Chains and Stochastic Stability*. Springer-Verlag, 175 Fifth Avenue, New York, NY.

Morgan Stanley Dean Witter (1999): "Debt Strategy Analysis," Discussion paper, Working Paper No. 10 - Fixed Income Research.

Oksendal, B. K. (1995): *Stochastic Differential Equations*. Springer Verlag, Berlin, fourth edn.

Pesaran, M. H., and S. M. Potter (1992): "Nonlinear Dynamics and Econometrics: An Introduction," *Journal of Applied Econometrics*, 7, S1–S7.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992): *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Trumpington Street, Cambridge, second edn.

Schaller, H., and S. van Norden (1997): "Regime Switching in Stock Market Returns," *Applied Financial Economics*, 7, 177–191.

Singh, M. K. (1995): "Estimation of Multifactor Cox, Ingersoll, and Ross Term Structure Model: Evidence on Volatility Structure and Parameter Stability," *Journal of Fixed Income*, 4, 8–28.

van Norden, S., and R. Vigfusson (1995): "Analytical Derivatives for Markov Switching Models," Bank of Canada Working Paper No. 95-7.

——— (1996): "Regime Switching Models: A Guide to the Bank of Canada Gauss Procedures," Bank of Canada Working Paper No. 96-3.

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Wheeler, G. (1996): "New Zealand's Experience with Autonomous Sovereign Debt Management," New Zealand Debt Management Office.

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. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue*.**

2002

2002-12	Modelling Financial Instability: A Survey of the Literature	A. Lai
2002-11	Risk, Entropy, and the Transformation of Distributions	R.M. Reesor and D.L. McLeish
2002-10	La fiabilité des estimations de l'écart de production au Canada	J.-P. Cayen et S. van Norden
2002-9	The Microstructure of Multiple-Dealer Equity and Government Securities Markets: How They Differ	T. Gravelle
2002-8	Restructuring in the Canadian Economy: A Survey of Firms	C. Kwan
2002-7	Contribution of ICT Use to Output and Labour-Productivity Growth in Canada	H. Khan and M. Santos
2002-6	Currency Fluctuations, Liability Dollarization, and the Choice of Exchange Rate Regimes in Emerging Markets	P.N. Osakwe
2002-5	The Effects of Bank Consolidation on Risk Capital Allocation and Market Liquidity	C. D'Souza and A. Lai
2002-4	Does Micro Evidence Support the Wage Phillips Curve in Canada?	J. Farès
2002-3	An Introduction to Wavelets for Economists	C. Schleicher
2002-2	Asset Allocation Using Extreme Value Theory	Y. Bensalah
2002-1	Taylor Rules in the Quarterly Projection Model	J. Armour, B. Fung, and D. Maclean

2001

2001-27	The Monetary Transmission Mechanism at the Sectoral Level	J. Farès and G. Srour
2001-26	An Estimated Canadian DSGE Model with Nominal and Real Rigidities	A. Dib
2001-25	New Phillips Curve with Alternative Marginal Cost Measures for Canada, the United States, and the Euro Area	E. Gagnon and H. Khan
2001-24	Price-Level versus Inflation Targeting in a Small Open Economy	G. Srour

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
E-mail: publications@bankofcanada.ca
Web site: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca