Bayesian and Classical Forecasting of Canadian Macroeconomic Time Series: A Comparison Study

by
Erik Drysdale

An essay submitted to the Department of Economics
in partial fulfillment of the requirements for
the degree of Masters of Arts

Queen's University
Kingston, Ontario, Canada
August 2014

Acknowledgements

The author wishes to thank Marco Cozzi for all his help and guidance. He also wishes to thank Johannes Pfeifer for his advice provided on the Dynare help forums.

Contents

1 Introduction

Forecasting key macroeconomic time series such as consumption, interest rates, and inflation remains one of the most important responsibilities of macroeconometrics. Central banks, government institutions, and private sector companies all make important decisions based on forecasts of these variables. Forecasters must make two decisions when performing their vocation: choosing a model and a statistical methodology. The goal of this paper was to consider the forecasting accuracy of two forecasting models and two statistical methodologies. The performance of each model was judged on its ability to both forecast different variables and maintain the consistency of its results over time.

Two types of models were considered: constrained and unconstrained. The constrained type was a Dynamic Stochastic General Equilibrium (DSGE) model and the unconstrained type was a Vector Autoregression (VAR) model. The variables of a DSGE model were constrained in the values they could take by the structure imposed from the underlying economic theory of the model.[1] In contrast, a VAR model placed no constraints between its variables, other than a linear structure. Both models were estimated by two statistical techniques: maximum likelihood (ML) and Bayesian. The four models studied and compared in this paper were: Bayesian DSGE (B-DSGE), maximum likelihood DSGE (ML-DSGE), VAR, and Bayesian VAR (BVAR).

Comparing the forecasting ability of different models is important for three reasons. First, the relative ability of different model types and statistical methodologies to accurately predict and describe real world data series is determined. Second, the specific weaknesses of models are highlighted which allows them to be addressed.[2] Third, future researchers are given a benchmark

[1]For example, the relationship between consumption and investment was that their sum must equal output.

[2]A DSGE model which is particularly bad at forecasting inflation can prompt a re-

to see if augmenting features of a specific model improve its performance.

Once model forecasts are generated, a researcher is tasked with deciding on how to measure 'performance.' In this paper, five Canadian time series were forecasted: (i) consumption, (ii) investment, (iii) money-balances, (iv) gross interest rates, and (v) gross inflation rates. Additionally, four aspects of forecasting performance were considered. First, different forecast horizons were used ranging from one- to four-quarters-ahead.[3] Second, an ordinal measure of forecasting performance, mean squared errors (MSEs), allowed models to be relatively ranked. Third, model forecasts were compared head-to-head, and differences in performance were tested for statistical significance. Fourth, a range of forecasting start dates were considered to determine both the robustness of the results and the dynamics of forecasting performance.

This paper differentiated itself from the existing DSGE forecasting literature through several aspects. Forecasting took place over a contemporary time frame, including the recent financial crisis. A range of model types were estimated with Canadian data using two statistical methodologies. Performance was evaluated on a wide series of measures including dynamic start dates which tested for robustness. The final results showed that the unconstrained VAR model, estimated by classical methods, yielded the most accurate forecasts. The evidence for the result came from measures of relative performance, the number of statistically significant tests, and the degree of robustness across start dates.

Background

DSGE models are the main policy tool in macroeconomics for analyzing key policy questions. Examples of such questions include: (i) Under which sit-

searcher to reconsider the model's price formation mechanism.

[3] A k-step-ahead forecast predicts the value of a variable k quarters in the future.

uations does the money supply and consumption positively covary? (ii) How persistent are shocks to the economy? (iii) Which variables react most strongly to a change in interest rates? The advantage of DSGE models are that they are microfounded, able to incorporate new designs in their structure,[4] provide quantitative measures of policy trade-offs, and address the 'Lucas critique'.[5]

DSGE models can be written in a state-space representation which allows them to make forecasts of any of their model's variables. Traditional forecasting models, such as VARs, are also representable in a state-space form, but they impose no structural relationship between variables. This presents a natural research question: does imposing a specific structure derived from economic theory, and constraining the relationship between variables, improve the forecasting results for important macroeconomic time series? Recent literature suggested that, in certain cases, the constraints imposed by the DSGE model improved forecasting accuracy when compared to a range of other benchmark models.

The growth of computer power allows the estimation of increasingly complicated DSGE models by both classical and Bayesian techniques. Classical methods involve finding a vector of parameters which maximize a likelihood function (known as maximum likelihood estimation). Such an approach requires numerical methods which explore high-dimensional spaces to find the optimal point. In contrast to classical methods, the Bayesian approach assumes the vector of parameters is itself a random vector and maps this vector into a probability distribution (the posterior). In Bayesian DSGE models, the

[4]For example, the financial crisis prompted theoreticians to develop models with new 'financial frictions' which reflected the recent business cycle experience.

[5]DSGE models use agents with rational expectations that ensure all variables are simultaneously related to each other and that the solution to the dynamic system is made up of the 'deep' parameters of the economy. See Lucas (1976) for the original 'Lucas Critique' or Tesfatsion (2013) for a contemporary and readable overview.

parameter vector which maximizes the median[6] of the posterior probability distribution is chosen as the optimal vector.

Bayesian estimation has been described as a "bridge" between the calibration techniques used by the early DSGE models[7] and the ML approach (Griffoli, 2008, pg. 83). The specification of a probability distribution for each parameter (a prior) is comparable to calibration. The use of data to inform the parameter values is akin to maximum likelihood. Estimating a DSGE model by Bayesian methods has several advantages. First, priors can ensure that parameter estimates conform to economic theory and empirical evidence.[8] Second, and related, ML estimation can suffer from identification issues when several combinations of parameters achieve the same likelihood or flat subspaces lead to "the dilemma of absurd parameter estimates" (Griffoli, 2008, p. 78). For a more technical discussion of the differences between ML and Bayesian estimation see Sections 9.2.3 and 9.2.4 in the appendix.

Macroeconomists have long known that a constrained model can outperform a more flexible one as long as its constraints are not too inconsistent with the underlying data generating process.

> *... recent theoretical work gives rigorous foundation .... that in high dimensional models restricted estimators can easily produce smaller forecast or projection errors than unrestricted estimators ...* (Sims, 1980, pg. 25)

The relative forecasting performance between constrained and unconstrained

[6]Other moment possibilities include the mean, mode, or minimum variance. However the Bayesian DSGE literature has adopted the median as its moment of choice.

[7]Early dynamic stochastic models were known as real business cycle (RBC) models (see Kydland and Prescott (1982)) and they differed from today's DSGE models in that they did not have 'frictions' which caused nominal variables to impact the real economy.

[8]For example, a variable which is logically bounded between zero and one can take a beta distribution so that there is no probability weight outside this domain. For a further discussion see Chapter 2 of King (1998).

models is determined by the trade-off between the number of constraints imposed on a model and the degree of finite sample bias (Dib et al., 2008, pg. 139). If the constraints implied by the DSGE model are useful representations of reality then this would weight forecasting performance in its favour due to the larger number of variables an unconstrained model, such as a VAR, would need to estimate.[9] However, the constrained model will suffer from misspecification more, relative to an unconstrained model, when the data is generated from a more complex process.

The difference between the Bayesian and classical VAR was that the former addressed the problem of overparameterization by reducing the number of parameters to a handful of hyperparameters.[10] While the BVAR appeared to have the best of both worlds - an unconstrained model unplagued by finite sample bias - it suffered from an increasing dependence on the choice if its priors. See Sections 6.4 and 6.5 for a discussion of the two models.

The post-1992 period of Canadian macroeconomic history was the most conducive to the New Keynesian (NK)[11] model used by this paper, as this was when the Bank of Canada began its inflation targeting regime and successfully disciplined inflation expectations and actual rates close to 2%. The number of pseudo out-of-sample forecasting periods, which took place from 2003-Q1 to 2013-Q4, provided sufficient power to test whether one model's forecasting performance was superior to another's. Forecasting start dates ranged from 2003-Q1 to 2011-Q2 for the dynamic tests.

---

[9]In this paper the VAR model had to estimate 50 parameters compared to only 25 for the DSGE.

[10]A hyperparameter maps its value to multiple parameters, thereby reducing the dimensions of the parameter space.

[11]The NK acronym became associated with dynamic stochastic models which incorporated frictions. Related to a previous footnote, empirical evidence showed that the frictionless RBC models failed to describe important aspects of business cycle behavior and NK models became the norm. Consequently, the acronyms NK and DSGE are used interchangeably throughout this paper.

Researchers remain divided on the use of out-of-sample forecasts for model evaluation.[12] Out-of-sample forecasting accuracy requires that the data be partitioned into estimation and forecasting portions, which means not all available information is used. In contrast, an in-sample model comparison with a likelihood or Bayes ratio uses all the available data. This method has indisputably more statistical power. However, the out-of-sample approach allows the researcher to ask different questions that are of interest to central banks and related institutions. Such questions can include: (i) Since 2003, which model produced the most accurate forecasts of inflation? (ii) Do the results change for yearly rather than quarterly forecasts? (iii) Which model performed best during the financial crisis? (iv) How does model performance differ across the Bayesian and classical methodologies?[13]

The use of out-of-sample forecasts addressed two concerns brought up in previous papers. First, little evidence existed as to the robustness of DSGE models.

> *... DSGE model's ... forecast samples ... do not cover events, such as a deep recession, that are particularly difficult to foresee.* (Christoffel et al., 2010, pg. 5)

The use of dynamic forecast tests with contemporary data showed how DSGE model performance changed from the time of the 'great moderation' to the 'great recession.' Second, in-sample fit could not identify which variables a model was particularly poor at predicting. This was why some researchers, particularly those affiliated with central banks, have focused on out-of-sample forecasts so that,

[12] For a discussion of the advantages and disadvantages see Diebold (2012) and for an argument against see Clements and Hendry (2005).

[13] While Bayes and likelihood ratios can be compared between DSGE and VAR models in certain cases, a Bayesian model cannot be compared to a classical model through in-sample fit.

> *... the success[es] in forecasting individual series may provide more information to help improve the model.* (Edge et al., 2010, pg. 19)

The structure of the rest of the paper is as follows: Section 2 provides a literature review and list of technical sources. Section 3 presents the structure of the New Keynesian model. Section 4 explains the methodology used to construct the data series. Section 5 details model calibration, including the choice of Bayesian priors. Section 6 analyzes the parameter estimates and their implications. Section 7 examines the forecasting results of all four models. Section 8 provides concluding remarks and directions for further research. Lastly, Section 9 contains an appendix with the relevant model equations, methodologies, and figures. Items which are contained in the appendix will be cited as such throughout the paper. All codes, directions to replicate results, and additional figures can be found in the following Dropbox folder: `https://db.tt/FD9eZJpi`.

2 Literature Review

Forecasting macroeconomic time series has a voluminous literature stretching back decades. The use of DSGE models to make forecasts has a more recent history. Empirically estimated DSGE models go as far back as Christiano and Eichenbaum (1992), Smith (1993) and Watson (1993). Historic attempts to compare the fit and forecasting performance of DSGE models to other benchmarks yielded poor results. Bergin (2003) found that a random walk outperformed a DSGE model for most open-economy variables at one-step-ahead forecasts[14] and Schorfheide (2000) found that the Bayes ratio of a DSGE to a BVAR was extremely low. A Bank of England report best summarized the consensus of the time:

> "*The ... desire that a model should be both theoretically and empirically coherent ... has proven impossible to satisfy ... and therefore a trade-off is perceived to exist.*" (Pagan, 2003, pg. 68)

In their breakthrough paper, Smets and Wouters (2003) were the first to show that the forecasting accuracy of a large-scale DSGE model was comparable to, or better than, a VAR and BVAR for inflation, exchange rates, and interest rates in the euro-area. The results were intriguing and widely recognized. The fit of a DSGE model improved when there were more frictions and shocks. Frictions allowed the DSGE model to better match the data by estimating more parameters. The large-scale model's ten autoregressive shocks helped to pick up autocorrelations in the data.

Since 2003, a significant amount of DSGE forecasting research emerged. Papers differentiated themselves through several features including, but not limited to, the type of model, estimation method, measure of performance, and

[14] A model whose MSE is higher than a random walk should not be used for forecasting.

data set used. Ireland (2004) was able to show that a small-scale DSGE model estimated by classical methods could beat VAR models over a long period of US post-war data in out-of-sample forecasting.[15] Many central banks have developed their own DSGE models to match features of their country's economy. The Swedish central bank's (Sveriges Riksbank) open-economy model proved a competent forecaster for real exchange rates, exports and imports (Adolfson et al. (2007)). The paper also considered forecasting performance beyond point estimates through density forecasts and predictive densities. See Christoffel et al. (2010) for a similar but more contemporary paper.

Edge et al. (2010) showed that the Federal Reserve's large-scale closed-economy 'Edo' model had forecasts comparable, if not better than, those made by Federal Reserve staff. The paper used real-time data to ensure the DSGE model forecasted with the non-revised data that was available to the Federal Reserve staff at the time. Kolasa et al. (2012) also used real-time data to forecast US time series but included external information, through conditional forecasting, and compared the performance of the DSGE model to the Survey of Professional Forecasters (SPF). The inclusion of external information was able to make the forecasts comparable if not superior to the SPF.[16] Other examples of conditional forecasting include Julliard and Maih (2010).

The most common NK model used in the DSGE forecasting literature was the closed-economy model from the Smets and Wouters (2007) paper originally used to forecast US time series. For an overview of the forecasting ability of several DSGE models for different measures of forecasting performance see Del Negro and Schorfheide (2012). Table 1 tabulates key features of existing research and their relation to this paper.

[15] While the DSGE model only had one shock, three additional shocks were added through measurement errors.

[16] In contrast, earlier research found that DSGE models were not able to beat the SPF (Rubaszek and Skrypcyzynki (2008)).

Table 1: Sample of the DSGE Forecasting Literature

	Linearized	Out-of-sample	In-sample	Method	Type	Real-time & Survey*
Adolfson et al. (2007)	✓	✓	✓	Bayesian	Open-economy	✗
Christoffel et al. (2010)	✓	✓	✓	Bayesian	Open-economy	✗
Del Negro and Schorfheide (2012)	✓	✓	✓	Bayesian	Multiple	✓
Dib et al. (2008)	✗	✓	✗	Classical	Closed	✗
Drysdale (2014)	✓	✓	✗	Classical/Bayesian	Closed	✗
Edge et al. (2010)	✓	✓	✗	Bayesian	Closed	✓
Ireland (2004)	✓	✓	✗	Classical	Closed	✗
Kolasa et al. (2012)	✓	✓	✗	Bayesian	Closed	✓
Rubaszek and Skrypcyzynki (2008)	✓	✓	✗	Bayesian	Closed	✓
Smets and Wouters (2007)	✓	✓	✓	Bayesian	Closed	✗

	Location	Hybrid†	Success‡	Benchmark¶	Period	# of shocks
Adolfson et al. (2007)	Europe	✗	✓	(B)VAR, VECM, RW	1980-Q1 to 2002-Q4	12
Christoffel et al. (2010)	Europe	✗	✓	(B)VAR, RW	1985-Q1 to 2006-Q4	12
Del Negro and Schorfheide (2012)	US	✗	✓	AR(2)	1992-Q1 to 2011-Q2	6, 12
Dib et al. (2008)	CAN	✗	✓	(B)VAR	1981-Q3 to 2004-Q1	4
Drysdale (2014)	CAN	✗	✗	(B)VAR	1992-Q1 to 2013-Q4	5
Edge et al. (2010)	US	✗	✓	AR(2), (B)VAR	1984-Q4 to 2004-Q4	24
Ireland (2004)	US	✗	✓	VAR	1948-Q1 to 2002-Q2	4
Kolasa et al. (2012)	US	✓	✓	DSGE-VAR	1964-Q1 to 2008-Q2	6
Rubaszek and Skrypcyzynki (2008)	US	✗	✗	VAR, BVAR, SPF	1972-Q2 to 2006-Q2	3
Smets and Wouters (2007)	US	✗	✓	(B)VAR	1966-Q1 to 2004-Q4	7

*Compared DSGE forecasts, using real-time data, to historic forecasts made by professional forecasters.

†Used a DSGE-VAR model.

‡DSGE model was best forecaster in at least one variable.

¶Model acronyms are as follow: Vector error correction model (VECM), random walk (RW), and univariate autoregressive model of order 2 (AR(2)).

The majority of empirically estimated DSGE models forecasted American and European data series. This paper was interested in forecasting Canadian time series. There was a small literature of Canadian-estimated DSGE models. The Bank of Canada developed a large-scale open-economy model, titled 'Totem II' (Dorich et al. (2013)), and estimated it by classical techniques. While the model is used by the Bank of Canada for quarterly forecasts, its accuracy has not been analyzed in a public document. Lubik and Schorfheide (2007) estimated a small-scale open-economy model with Bayesian methods, for several countries including Canada, with a monetary policy rule that included exchange rates.[17]

Within the literature, Dib et al. (2008) was the most relevant to this paper. The authors estimated a medium-scale DSGE model by classical techniques to forecast Canadian time series and compared its forecasting accuracy with a VAR and BVAR. This paper expanded upon the work done by Dib, Gammoudi, and Moran by using: more recent data (including the financial crisis),[18] Bayesian techniques, and dynamic tests. There were many similarities in model design between this paper, Ireland (2003), and Dib et al. (2008). A discussion of the differences between the models can be found in Section 3.5.

An alternative approach to estimating and forecasting with DSGE models was the hybrid DSGE-VAR approach. The following thought experiment, paraphrased from Del Negro and Schorfheide (2006), best explained the link between DSGE and VAR models. A VAR of order p, when provided sufficient data generated from a DSGE system would have parameters which yielded autocorrelations that matched those of the DSGE model's up until the p^{th} autocorrelation. Therefore, a mapping between the parameters of the DSGE

[17]Their empirical result suggested that the policy rate in Canada was affected by exchange rate movements.

[18]The aforementioned paper used data up until 2004-Q1 whereas this paper has observations until 2013-Q4.

and VAR model existed. The DSGE-VAR literature used a hyperparameter (λ) to capture how far the estimated VAR parameters (resulting from the actual data) were from the parameters implied by the DSGE mapping.

The actual forecasting model used by the DSGE-VAR was a BVAR whose prior mean and covariance was set to the implied DSGE values and the inverse of λ, respectively. The DSGE-VAR approach was both a highly innovative way of specifying priors and measuring DSGE model fit. This hybrid method was similar to comparing likelihood or Bayes ratios rather than the measures of forecasting performance that this paper was interested in. Other hybrid DSGE models, such as Bekiros and Paccagnini (2013)'s factor-augmented DSGE-VAR model, can be found in the literature.

After the financial crisis, many commentators wondered why macroeconomic models failed to forecast the largest economic downturn since the Great Depression. While DSGE models that incorporated financial 'frictions' and 'accelerators' were around since the new millennium (see Bernanke et al. (1999)), a flurry of research has been underway to develop new microfounded mechanisms that capture the dynamics of financial markets. A full discussion of these models is beyond the purview of this literature review and readers are directed toward recent literature surveys by Brunnermeier et al. (2012) and Brazdik et al. (2012) as well as a growing database[19] of DSGE models (Wieland et al. (2012)) for more information.

Despite the general impression that DSGE models 'failed to predict' the crisis, there has been a paucity of published research into their relative forecasting performance during this period.[20] Del Negro and Schorfheide (2012) found that standard DSGE models saw their forecasting performance deteri-

[19]A database of DSGE models organized by their distinctive paradigms can be found at: `www.macromodelbase.com`.

[20]However, several working papers, including Rubaszek and Kolasa (2014), are currently investigating this topic.

orate during the crisis years. A further contribution of this paper was to add needed evidence of DSGE forecasting accuracy during the financial crisis for a range of measures.

Bayesian models have become increasingly common due to the growth of computer power. Markov Chain Monte Carlo (MCMC) methods are algorithms which sample from complicated probability spaces and approximate the posterior. Such spaces, which are multi-dimensional integrals, cannot usually be analyzed by standard numerical methods. An and Schorfheide (2007), Fernandez-Villaverde (2010), and Guerron-Quintana and Nason (2012) provided literature surveys of Bayesian methods and their applications to DSGE models.

Six technical documents were particularly useful in the creation of this paper. The mathematical appendix to Ireland (2003)[21] provided robust model explanations and a suite of Matlab files. The material for Lawrence Christiano's course on 'Estimation, Solution and Policy Analysis using Equilibrium Monetary Models' had an informative and technical overview of Bayesian techniques and DSGE models (see Christiano (2014)). An introductory textbook to Bayesian techniques can be found in Lancaster (2004). Three documents were especially helpful for the implementation of Dynare (the software used by this paper to estimate the DSGE models). Pfeifer (2014) showed how observation equations were best specified in Dynare. The Dynare User Guide from Griffoli (2008) was an indispensable overview about both the methodologies and practical uses of the software. Den Haan (2011) explained the Bayesian approach to DSGE models and its implementation in Dynare. The next section discusses the structure of the New Keynesian model.

[21]The mathematical appendix is cited as Ireland (2002) in the References section.

3 Closed Economy Model

Numerous DSGE models were used in the literature to forecast macroeconomic time series. This paper's specific NK model was chosen as it was very similar to the model used by Dib et al. (2008). Choosing a model close in design to the aforementioned paper had two advantages. First, if similar results were discovered, and the Bayesian DSGE was found to be the best forecaster, this result could be attributed to the statistical methodology rather than a different model design. Second, the parameter estimates between the models could be contrasted.

The remainder of this section details the four agents of the model and their respective optimization problems as well as modifications to the NK model. The exact order is as follows: Subsections 3.1 to 3.4 discuss the representative household, representative finished goods-producing firm, intermediate goods-producing firms, and the central bank, respectively. Lastly, Subsection 3.5 examines model differences between this paper and similar ones found in the literature. Additional equations and a general summary of the model can be found in Section 9.1 in the appendix.

3.1 Representative Household

During period t, the representative household supplied to each of the i intermediate goods-producing firms $H_t(i)$ units of labour at a wage rate of W_t and $K_{t-1}(i)$ units of capital at a rental rate of Q_t. The household received total nominal factor payments of $W_t H_t + Q_t K_{t-1}$. The household's endowment of labour was normalized to 1.

$$H_t = \int_0^1 H_t(i)di$$
$$K_{t-1} = \int_0^1 K_{t-1}(i)di$$

At the end of period t, the representative household made consumption and investment decisions with the factor payments it earned over the period and the cash balances it had from the previous period. The representative household purchased the finished good Y_t at a price of P_t which it could consume as a consumption good C_t or use as an investment good I_t to generate a capital good K_t. However, investments were transformed into productive capital goods at an additional cost when the growth rate of capital from periods t to $t+1$ differed from the deterministic growth rate g. For every unit of investment the household paid an adjustment cost as shown in equation (1). The magnitude of the friction was determined by the level of the capital stock and the parameter $\phi_K \geq 0$.

$$\Phi(K_t, K_{t-1}) = \frac{\phi_K}{2}\left(\frac{K_t}{gK_{t-1}} - 1\right)^2 K_{t-1} \tag{1}$$

The parameter g was the deterministic growth rate of capital along the balanced growth path. Equation (2) presents the law of motion of the capital stock, where δ was the depreciation rate, and χ_t was a shock to the marginal efficiency of capital.

$$K_t = (1-\delta)K_{t-1} + \chi_t I_t \tag{2}$$

$$\ln(\chi_t) = \rho_\chi \ln(\chi_{t-1}) + \varepsilon_t^\chi \tag{3}$$

The household's budget constraint for period t is shown in equation (4), where D_t were firm profits accruing to the household, and R_t was the gross interest rate.

$$\begin{aligned} &M_{t-1} + T_t + B_{t-1} + W_t H_t + Q_t K_{t-1} + D_t \\ \geq &P_t(C_t + I_t + \Phi(K_t, K_{t-1})) + \frac{B_t}{R_t} + M_t \end{aligned} \tag{4}$$

The household had preferences over an infinite time horizon as shown in equation (5). Utility was a non-separable function of a composite good (made up of consumption and real money balances) and leisure: $u(C_t, M_t/P_t, 1 - H_t)$. A non-separable utility function ensured that changes to the nominal interest rate would have impacts on the real variables.

$$E_0 \sum_{t=0}^{\infty} \beta^t \left\{ a_t \frac{\gamma}{\gamma - 1} \ln\left(C_t^{\frac{\gamma-1}{\gamma}} + e_t^{\frac{1}{\gamma}} (M_t/P_t)^{\frac{\gamma-1}{\gamma}}\right) + \eta \ln(1 - H_t) \right\} \tag{5}$$

There were two shocks in the utility function. First, there was a general preference shock (a_t) to the composite good. Second, there was a specific money-demand shock (e_t) within the composite good.

$$\ln(a_t) = \rho_a \ln(a_{t-1}) + \varepsilon_t^a \tag{6}$$

$$\ln(e_t) = (1 - \rho_e)\ln(e) + \rho_e \ln(e_{t-1}) + \varepsilon_t^e \tag{7}$$

The household faced the following maximization problem.

$$\max_{C_t,H_t,M_t,B_t,I_t,K_t} E_0 \sum_{t=0}^{\infty} \beta^t u(C_t, M_t/P_t, 1-H_t)$$

Subject to:

$$\frac{M_{t-1}+T_t+B_{t-1}+W_tH_t+Q_tK_{t-1}+D_t}{P_t} \geq C_t + I_t + \Phi(K_t, K_{t-1}) + \frac{B_tR_t^{-1}+M_t}{P_t}$$

$$K_t = (1-\delta)K_{t-1} + \chi_t I_t$$

The first-order conditions (FOCs) that resulted from the above constrained maximization problem are found in the appendix in Section 9.1.1. Combining the relevant-first order conditions yielded equation (8), the money demand equation. The parameter $\gamma > 0$ was, by a first-order Taylor approximation, the interest elasticity of money demand.[22] Even though the utility function was non-separable, there would always be a positive demand for both consumption and real money balances as long as the net interest rate was not zero.

$$C_te_t = \left(\frac{M_t}{P_t}\right)(1-R_t^{-1})^{\gamma} \tag{8}$$

3.2 Finished Goods-Producing Firm

The representative finished goods-producing firm used $Y_t(i)$ units of each intermediate good to produce a single finished good Y_t. In any one period t the representative finished goods-producing firm sought to maximize its profits by choosing the level of output and the amount of each input i, subject to a technological constraint.

[22]Section 9.1.1 provides details on the derivation of the money demand equation and the interpretation of γ.

$$\max_{Y_t, Y_t(i)} \quad P_t Y_t - \int_0^1 P_t(i) Y_t(i) di$$

$$\text{Subject to:} \quad \left(\int_0^1 Y_t(i)^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}} \geq Y_t$$

Combining the three FOCs of the above constrained optimization yielded the demand curve for each intermediate good, as shown in equation (9), where θ represents the price elasticity of demand.

$$Y_t(i) = Y_t \left(\frac{P_t(i)}{P_t} \right)^{-\theta} \tag{9}$$

3.3 Intermediate Goods-Producing Firms

The i^{th} intermediate goods-producing firm would hire $H_t(i)$ units of labour, rent $K_{t-1}(i)$ units of capital from the representative household, and set prices $P_t(i)$ to maximize the present value of the firm, subject to the constraint that the output must lie on the demand curve (seen in equation (9)). Formally, the firm sought to maximize its market value over an infinite horizon.

$$\max_{H_t(i), K_{t-1}(i), P_t(i)} E_0 \sum_{t=0}^{\infty} \beta^t \Lambda_t \frac{D_t(i)}{P_t}$$

Subject to:

$$\frac{D_t(i)}{P_t} = \left(\frac{P_t(i)}{P_t} \right)^{1-\theta} Y_t - \frac{W_t H_t(i) + Q_t K_{t-1}(i)}{P_t} - \frac{\phi_p}{2} \left(\frac{P_t(i)}{\Pi P_{t-1}(i)} - 1 \right)^2 Y_t$$

$$\left(\frac{P_t(i)}{P_t} \right)^{-\theta} Y_t = Y_t(i) \leq A_t K_{t-1}(i)^{\alpha} [g^t H_t(i)]^{1-\alpha}$$

$$\ln(A_t) = \rho_A \ln(A_{t-1}) + (1-\rho_a)\ln(A) + \varepsilon_t^A$$

Where $D_t(i)$ was the firm's profits in time t, Π was the steady-state rate of inflation growth, and $[\beta^t \Lambda_t / P_t]$ was the marginal utility of an additional dollar of profit in time t.[23] The second constraint shows that output assumed a Cobb-Douglas production mechanism with a total factor productivity shock A_t and a deterministic rate of growth g. The resulting first-order conditions are found in the appendix in Section 9.1.2. Firm i paid a cost to change its prices at a rate different than the gross steady-state rate of inflation, as shown below. The magnitude of the pricing friction was determined by the level of output and the parameter $\phi_p \geq 0$.

$$\frac{\phi_p}{2}\left(\frac{P_t(i)}{\Pi P_{t-1}(i)} - 1\right)^2 Y_t \tag{10}$$

3.4 Central Bank

The central bank set the gross nominal interest rate (R_t) in response to deviations of gross inflation (Π_t) and detrended output ($y_t = Y_t/g^t$) about their steady states. The monetary policy rule is defined in equation (11).[24] The strength of the central bank response to deviations of output and inflation was determined by the relative magnitude of ω_R to ω_y and ω_Π.

$$\omega_R \tilde{R}_t = \omega_\pi \tilde{\Pi}_t + \omega_y \tilde{y}_t + \ln(v_t) \tag{11}$$

$$\ln(v_t) = \rho_v \ln(v_{t-1}) + \varepsilon_t \tag{12}$$

[23] As the representative consumer owned the profits of the firm, her marginal utility determined the firm's net present value.

[24] Tilde's denote percent deviations about the steady state, where $\tilde{\Pi}_t = \frac{\Pi_t - \Pi}{\Pi}$,$\tilde{R}_t = \frac{R_t - R}{R}$, $\tilde{y}_t = \frac{y_t - y}{y}$.

Given the Bank of Canada's explicit mandate to maintain low, stable, and predictable inflation about the 2% inflation target, it may seem unnecessary to include the output gap in the central bank response function. However, this paper was agnostic as to whether the Bank of Canada indirectly targeted inflation through the output gap, and the parameters which best fit the data were chosen. As was standard in the literature, there was a first-order autoregressive monetary policy shock (v_t).

3.5 Comparison to the Literature

The most important difference between this paper, Dib et al. (2008), and Ireland (2003) was that the latter two made use of a money growth variable, $\mu_t = M_t/M_{t-1}$, which was important for two reasons. First, it pinned down the steady-state inflation rate. Second, it was included in the central bank response function. This paper chose not to include money growth as it seemed unlikely that the Bank of Canada would consider the growth rate of the money supply when making policy rate decisions.

Central Bank Response Function

$$\text{Drysdale (2014)}: \quad \omega_R \tilde{R}_t = \omega_\pi \tilde{\Pi}_t + \omega_y \tilde{y}_t + \ln(v_t)$$

$$\text{Dib et al. (2008)}: \quad \tilde{R}_t = \omega_\Pi \tilde{\Pi}_t + \omega_y \tilde{y}_t + \omega_\mu \tilde{\mu}_t + \ln(v_t)$$

$$\text{Ireland (2003)}: \quad \omega_R \tilde{R}_t = \omega_\Pi \tilde{\Pi}_t + \omega_y \tilde{y}_t + \omega_\mu \tilde{\mu}_t + \ln(v_t)$$

While all papers made use of a capital adjustment friction, there were subtle differences between them. Dib et al. (2008) applied the penalty to the square of the difference in the gross capital growth rate and the gross economic growth rate. In contrast, Ireland (2003) applied the penalty to the square of the percentage difference of the gross capital growth rate relative to the economic growth rate. This paper employed the approach of Ireland (2003).

Capital Adjustment Friction

Dib et al. (2008) : $$\frac{\phi_k}{2}\left(\frac{K_{t+1}}{K_t}-(1+g)\right)^2 K_t$$

Ireland (2003)/Drysdale (2014) : $$\frac{\phi_k}{2}\left(\frac{K_{t+1}}{gK_t}-1\right)^2 K_t$$

The production function of this paper differed from the other two in that the total factor productivity shock (TFP) did not solely apply to labour productivity. This type of Cobb-Douglas production function was employed by other papers such as Ireland (2004).

Production Function

Drysdale (2014) : $$A_t K_t^{\alpha}[g^t H_t]^{1-\alpha}$$

Dib et al. (2008)/Ireland (2003) : $$K_t^{\alpha}[A_t g^t H_t]^{1-\alpha}$$

This paper shared the same pricing friction as Ireland (2003) whereas Dib et al. (2008) used the more common Calvo pricing friction which yielded the familiar New Keynesian Phillips Curve. The variable $1-\varsigma$ was the probability that the firm was able to reset its prices. The Calvo mechanism and ς variable were not used in this paper.

Price Adjustment Friction

$$\text{Dib et al. (2008)}: \qquad \tilde{\Pi}_t = \beta\tilde{\Pi}_{t+1} + \frac{(1-\varsigma)(1-\beta\varsigma)}{\varsigma}\tilde{mc}_t$$

$$\text{Ireland (2003)/Drysdale (2014)}: \qquad \frac{\phi_p}{2}\left(\frac{P_t(i)}{\Pi P_{t-1}(i)} - 1\right)^2 Y_t$$

4 Data

Five data series were used to estimate the model: (i) consumption, (ii) investment, (iii) money-balances, (iv) gross interest rates, and (v) gross inflation rates. The data series were of quarterly frequency, seasonally adjusted (where relevant), spanned from 1992-Q1 to 2013-Q4, and totaled 88 observations. Consumption, investment, and money-balances were in real per-capita terms. Quarterly population estimates were used to construct the per capita variables.[25]

The DSGE model in this paper contained neither an international sector nor a government. The most appropriate data series from the national accounts for investment was the sum of capital spending and changes in inventories. Real per capita money balances used M2 as a measure of the money supply.[26] Interest rate data came from the yield on 3-month Canadian government treasury bills (this particular quarterly maturity imitated B_t in the model).

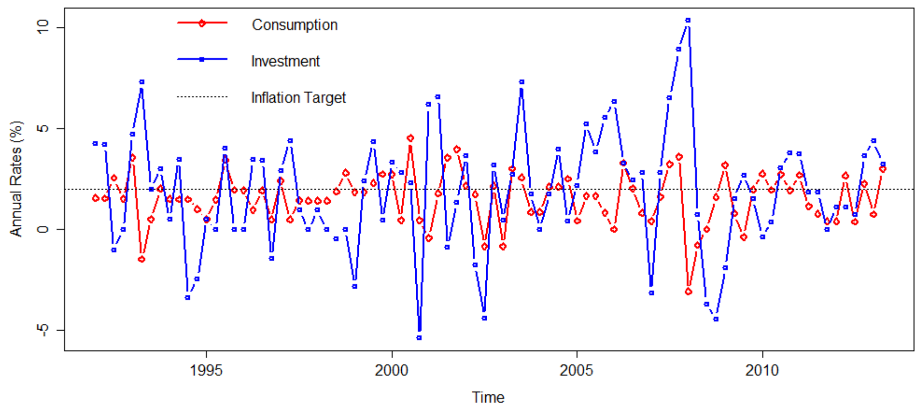
Inflation was calculated as the quarterly gross change in a weighted average of the consumption and investment price deflator series, which are shown in Figure 1. The weights were determined by the relative share of each series to total output. As consumption ranged from 72-82% of total output over the sample, the data series used to estimate the model most resembled the consumption price deflator series.

The variables of the log-linearized DSGE model were measured in percentage deviations from the steady state. A Hodrick-Prescott (HP) filter was able to separate the data series into a trend and cyclical component. The cyclical

[25]Note that the population estimates were for the total Canadian population whereas Dib et al. (2008) used the working age population.

[26]M2 is comprised of: currency outside banks, chartered bank demand and notice deposits, chartered bank personal term deposits, and inter-bank demand and notice deposits.

Figure 1: Inflation rates from price deflators



component was the estimate of the variable's percentage deviation from its steady state. However, two-sided HP filters estimate the trend and cyclical component at time t with both future and past observations. This makes them unsuitable for the recursive state-space nature of estimated DSGE models (Pfeifer, 2014, pg. 30). A solution to this problem was to use a one-sided HP filter, first implemented by Stock and Watson (1999) to forecast inflation. The advantage of this filter was that the trend and cyclical components were estimated using the data available only until that point in time.[27]

Figures 2(a) and 2(b) present the five Canadian time series. In 1992 the Bank of Canada adopted inflation targeting and by 1995 the 2% target was firmly entrenched in the Bank of Canada's joint inflation-control agreement with the Government of Canada.[28] As Figure 2(b) shows, since 1992, inflation has been consistently mean-reverting about the 2% level.

Figures 3(a) and 3(b) present the cyclical component of the data after the one-sided HP filter was applied.[29] The only difference between the data used to

[27]See `http://ideas.repec.org/c/dge/qmrbcd/181.html` for the relevant Matlab files to implement this filter.

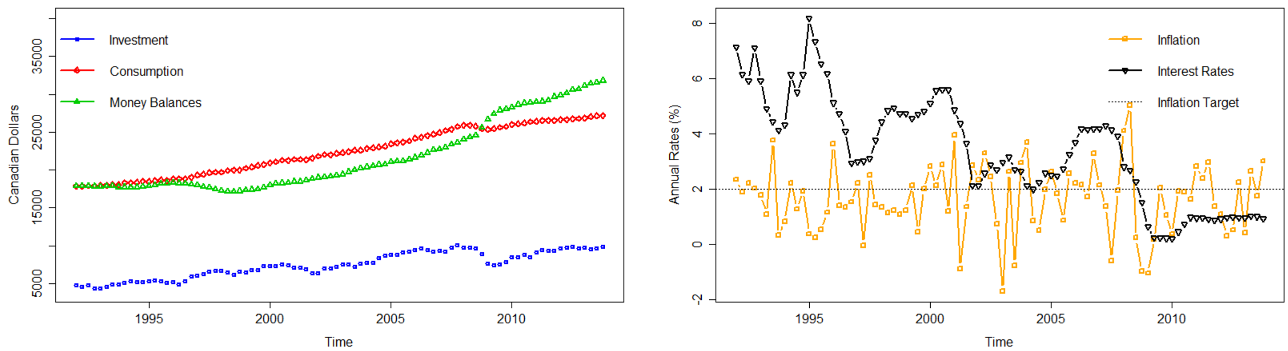
[28]See `http://www.bankofcanada.ca/wp-content/uploads/2010/11/inflation_control_target.pdf`.

[29]The parameter which controls the sensitivity of the trend component growth rate, λ, was set to 1600 (the literature standard for quarterly data).

Figure 2: Time series in levels

(a) Invesment, consumption, and money-balances (per capita in 2007 dollars)

(b) Net inflation and interest rates

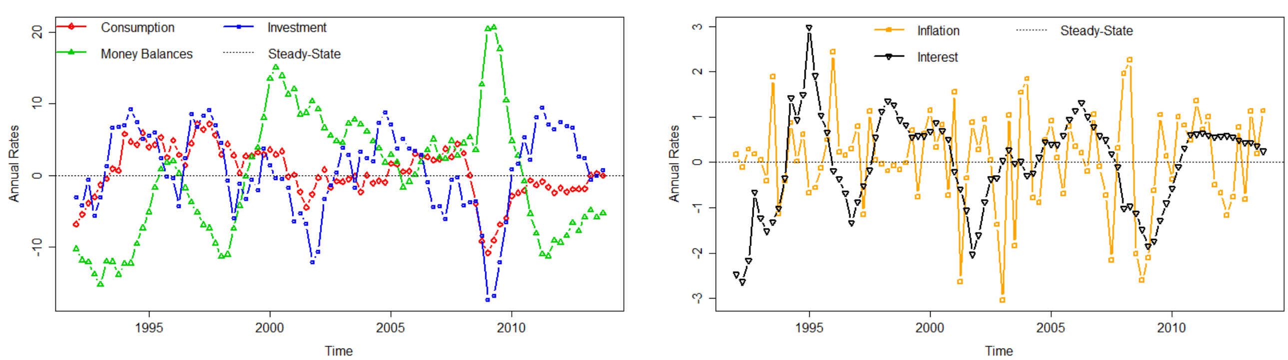


estimate the model and the figures seen in this section is that the visual data is presented in annualized percentage form. Figure 3(a) shows that consumption was fairly stable except during the financial crisis when it fell 10% below its potential. Investment, as could be expected, was more volatile and saw two downward swings since 2000. Money demand appeared countercyclical and saw a positive spike during the financial crisis. This could be explained by the very low interest rates that occurred during this period which reduced the opportunity cost of holding cash.

Figure 3: Detrended variables using one-sided HP filter

(a) Consumption, investment, and money-balances

(b) Interest and inflation rates



5 Calibration

Before the DSGE model could be estimated by Bayesian and ML techniques three things were required. First, a certain number of the parameters were fixed due to a lack of data and identification issues. Second, the algorithm which estimated the maximum likelihood model required the specification of an initial point. Third, Bayesian estimation required the specification of a prior distribution for each variable. The three subsections found below discuss each of these issues in turn.

5.1 Fixed Parameters

A DSGE model estimated with likelihood-based methods cannot have more observables than shocks (an issue known as 'stochastic singularity'). The model used in this paper had five shocks and five observables. The large number of parameters (twenty-five) relative to observables, combined with a lack of data needed to estimate certain variables (like the capital stock), meant that not all parameters could be estimated. Table 2 shows the eight parameters which were fixed.

Table 2: Fixed Parameters

Parameters		Value
Elasticity of demand	θ	6
Capital share	α	0.33
Depreciation rate	δ	0.025
Weight on leisure	η	1.35
Time preference	β	0.99
Growth rate	g	1.004
Steady-state inflation	Π	1.005
Composite good shock STD	σ_a	0.1

The elasticity of demand (θ), the capital share of production (α), the depreciation rate (δ), and the utility weight of leisure (η) were fixed due to a lack of data on profits, the capital stock, and hours worked. The standard deviation (STD) of the composite good utility shock (σ_a) was fixed in order to pin down the relative size of the other shocks and aid in their identification.[30] δ was set to the literature standard of 0.025, and implied an annual depreciation rate of 10.4%. θ was set to 6 which implied a 20% markup of prices over marginal costs in the steady state[31] and was the value set by Ireland (2003). α and η were set to 0.33 and 1.35, respectively, as was done in Dib et al. (2008). The ex-post result suggested that the representative household spent 51.7% of their time endowment working which contrasted to Dib et al. (2008)'s estimate of 33%.[32]

The values of β, Π, and g were fixed for three reasons. First, the steady-state value of the natural interest rate, $R = g\Pi/\beta$, did not conform well to the data (a result also found by Rubaszek and Skrypcyzynki (2008)). The parameter β was therefore fixed to a plausible value. Second, the growth rates of money, consumption, and investment differed. This was confirmed by a simple OLS regression of the logged variables on a constant and a trend term. The estimate suggested that consumption, investment, and money-balances grew at an annualized rate of 1.019%, 1.029%, and 1.026%, respectively. Following Ireland (2003), heterogeneous growth rates were ignored in order to keep the conceptual framework of a single growth rate (g) intact. The trend rate of growth was fixed to 1.021% (annualized) – the estimate of the output growth rate.[33]

---

[30]The estimation suffered from stability issues when the standard deviations of all five shocks were estimated.

[31]The intermediate goods-producing firms' ratio of prices to marginal costs in the steady-state were $\frac{\theta}{\theta-1}$.

[32]The calculation of H can be found in the Matlab files located in the Dropbox link.

[33]In Ireland (2003) the author detrended the series with a simple OLS procedure instead of fixing g.

5.2 Maximum Likelihood Initialization

A model estimated by maximum likelihood techniques requires a specification of initial values that conform to the Blanchard-Kahn (BK) conditions which ensure dynamic determinacy (see Blanchard and Kahn (1980)). Different initial values can lead to different ML results in high-dimensional spaces in the presence of local maxima. The initial values were mainly chosen to match either economic theory or previous papers.

Table 3 presents the initial values and upper and lower bounds (if any) of all estimated parameters. γ was set to 0.014 which was Atta-Mensah (2004)'s empirical estimate of money demand elasticity in Canada. ϕ_k, e, and ω_y closely resembled the estimates of Ireland (2003). The standard deviations of the shock variables were set to the initial value of the fixed shock (σ_a). The autocorrelation parameters, except monetary policy persistence, were set to be highly autoregressive. The other coefficients were set to ensure that the BK stability conditions were met and were adjusted through trial and error. DSGE models with many parameters have a higher chance that an initial parameter combination will yield dynamic indeterminacy. It is plausible that a different set of initial values that conformed to the BK conditions could have yielded different results.[34]

5.3 Bayesian Priors

Each variable estimated by Bayesian methods required the specification of a prior distribution.[35] A distribution which best conformed to the likely probability space of the parameter was chosen. This paper followed precedents

[34]This fact highlights one of the weaknesses of ML estimation for medium- and large-scale DSGE models.

[35]In Bayesian estimation it is possible to have a 'non-informative' prior, but this is not the convention with Bayesian DSGE use.

Table 3: Initial Values – Maximum Likelihood Estimation

Parameters		Initial Value	Lower Bound	Upper Bound
Money demand elasticity	γ	0.014		
Capital adjustment friction	ϕ_k	30		
Price adjustment friction	ϕ_p	30		
Steady-state TFP	A	7		
Steady-state money demand	e	2.7		
Interest responsiveness	ω_R	3		
Inflation responsiveness	ω_Π	3.1		
Output gap responsiveness	ω_y	-0.022		
Composite good persistence	ρ_a	0.99	0.00001	0.9999
Money demand persistence	ρ_e	0.90	0.00001	0.9999
TFP persistence	ρ_A	0.90	0.00001	0.9999
Investment productivity persistence	ρ_χ	0.90	0.00001	0.9999
Monetary policy persistence	ρ_v	0.01	0.00001	0.9999
Money demand STD	σ_e	0.10	0.00001	1
TFP STD	σ_A	0.10	0.00001	1
Investment productivity STD	σ_χ	0.10	0.00001	1
Monetary policy STD	σ_v	0.10	0.00001	1

in the literature and linked specific types of parameters to one of four prior distributions.[36] Columns 3 to 5 of Table 4 present the prior distribution type, mean, and standard deviation of each parameter.

The standard deviation of the shocks (σ) followed an inverse gamma distribution with a mean of either 0.01 or 0.10 and a standard deviation of 2. Inverse gamma distributions have a form similar to income or wealth distribution: left-truncated at zero, a high peak about the median, and a long right-tail. The autoregressive coefficients (ρ) followed a beta distribution with a mean and standard deviation of 0.5 and 0.2, respectively. Beta distributions are bounded between $(0, 1)$ and, with a mean of 0.5, are symmetric in structure. The standard deviations and autoregressive coefficients were similar to the prior values chosen by Smets and Wouters (2007). The elasticity of money

[36]The priors of this paper were similar to those of Smets and Wouters (2003), Del Negro and Schorfheide (2004), Smets and Wouters (2007), and Kolasa et al. (2012).

demand was also given a beta distribution as empirical estimates have found it to be inelastic. The capital and price adjustment frictions, as well as monetary policy coefficients, followed a normal distribution as their effects were influenced by the relative values of other parameters and therefore required the most flexible distribution. Lastly, the intercepts of the autoregressive shock processes (e and A) were given gamma distributions that were left-truncated at zero and had a slight right-skew.

Table 4: Prior Distributions and Posterior Parameter Estimates

Parameter		Prior			Posterior			
		Type	Mean	STD	Mode	Mean	STD	ML Result
Composite good persistence	ρ_a	Beta	0.5	0.2	0.9418	0.9687	0.0143	0.9281
TFP Persistence	ρ_A	Beta	0.5	0.2	0.9414	0.9512	0.0082	0.9429
Monetary policy persistence	ρ_v	Beta	0.5	0.2	0.0783	0.0761	0.0409	0.2703
Money demand persistence	ρ_e	Beta	0.5	0.2	0.9626	0.9582	0.0164	0.9632
Investment productivity persistence	ρ_χ	Beta	0.5	0.2	0.9463	0.9525	0.0170	0.5174
Money demand elasticity	γ	Beta	0.5	0.2	0.4029	0.3813	0.0050	0.4662
Steady-state TFP	A	Gamma	7	3	2.2414	2.8007	0.3407	6.4928
Steady-state money demand	e	Gamma	2.7	0.675	0.9760	1.0233	0.0355	1.1394
Interest responsiveness	ω_R	Normal	3	1.5	4.5489	4.6612	0.0545	3.9987
Inflation responsiveness	ω_Π	Normal	3.1	1.55	3.8185	3.9036	0.0940	3.3682
Output gap responsiveness	ω_y	Normal	-0.022	0.011	-0.0007	0.0019	0.0013	-0.0608
Capital adjustment friction	ϕ_k	Normal	30	15	5.7124	6.8052	1.5866	7.4206
Price adjustment friction	ϕ_p	Normal	30	15	6.1997	5.3008	0.5854	32.4575
TFP STD	σ_A	Inv. gamma	0.01	2	0.0098	0.0099	0.0011	0.0107
Money demand STD	σ_e	Inv. gamma	0.10	2	0.0323	0.0309	0.0033	0.0354
Investment productivity STD	σ_χ	Inv. gamma	0.01	2	0.0024	0.0030	0.0019	0.0017
Monetary policy STD	σ_v	Inv. gamma	0.10	2	0.0166	0.0174	0.0006	0.0120

ML results in column 10 are taken directly from Table 5.

6 Estimation Results

Analyzing parameter estimates is important for several reasons. First, the degree of robustness in results across the literature can be determined. Second, the parameter values have implications for the structure of the Canadian economy. Third, the ML and Bayesian methods can be contrasted. Subsections 6.1 and 6.2 discuss the parameter estimates of the DSGE model estimated by maximum likelihood and Bayesian techniques, respectively.

While differences in parameter estimates can have a quantitative impact through forecasting, there are also qualitative implications via impulse response functions (IRFs). In Subsection 6.3 the differences and similarities between the Bayesian and ML impulse responses are highlighted. Lastly Subsections 6.4 and 6.5 outline the VAR and BVAR models, respectively.

6.1 Maximum Likelihood DSGE

Table 5 presents the parameter estimates alongside previous findings from Ireland (2003) and Dib et al. (2008). Not all the parameters values were directly comparable though. The steady-state value of the TFP shock impacted the steady-state output differently due to the varying construction of the Cobb-Douglas production function (see Section 3.5). Additionally, some parameter values were only interpretable relative to others (such as monetary policy coefficients or frictions) so that different results may have still implied similar impulse responses. The majority of shock processes, autoregressive coefficients, and (relative) monetary policy coefficients were similar to at least one of the two other papers. However, there were strong differences in the estimates of money demand elasticity, the intercepts of the shock processes, and the capital and price adjustment frictions.

Table 5: Maximum Likelihood Parameter Estimates

Parameters		Ireland (2003)		Dib et al. (2008)		Drysdale (2014)	
θ	Elasticity of demand	6	-	N/A	N/A	6	-
α	Capital share	0.220	(0.008)	0.33	-	0.33	-
δ	Depreciation rate	0.025	-	0.025	-	0.025	-
γ	Money demand elasticity	0.036	(0.008)	0.067	(0.022)	0.4662	(0.0008)
η	Weight on leisure	1.5	-	1.35	-	1.35	-
β	Time preference	0.992	(0.001)	0.993	(0.001)	0.99	-
ϕ_k	Capital adjustment friction	32.135	(2.258)	15	-	7.4206	(0.0128)
ϕ_p	Price adjustment friction	161.835	(78.578)	N/A	N/A	32.4575	(0.0422)
ρ_a	Composite good persistence	0.903	(0.030)	0.925	(0.036)	0.9281	(0.0101)
σ_a	Composite good STD	0.019	(0.002)	0.023	(0.004)	0.1	-
ρ_A	TFP persistence	0.947	(0.023)	0.899	(0.055)	0.9429	(0.0025)
σ_A	TFP STD	0.020	(0.005)	0.014	(0.003)	0.0107	(0.0008)
A	Steady-state TFP	7090.8	(650.0)	3.258	(0.116)	6.4928	(0.0084)
ρ_χ	Investment productivity persistence	0.954	(0.030)	N/A	N/A	0.5174	(0.0189)
σ_χ	Investment productivity STD	0.0213	(0.082)	N/A	N/A	0.0017	(0.0001)
ρ_v	Monetary policy persistence	0.198	(0.89)	0.270	(0.067)	0.2703	(0.0147)
σ_v	Monetary policy STD	0.01	-	0.008	(0.001)	0.0120	(0.0009)
ρ_e	Money demand persistence	0.979	(0.018)	0.989	(0.015)	0.9632	(0.0103)
σ_e	Money demand STD	0.008	(0.001)	0.012	(0.002)	0.0354	(0.0024)
e	Steady-state money demand	3.896	(0.144)	0.468	(0.041)	1.1394	(0.0019)
ω_R	Interest responsiveness	2.297	(0.273)	N/A	N/A	3.9987	(0.0053)
ω_y	Output gap responsiveness	-0.0157	(0.028)	0.009	(0.036)	-0.0608	(0.0002)
ω_Π	Inflation responsiveness	2.2102	(0.403)	0.747	(0.230)	3.3682	(0.0058)
g	Growth rate	1	-	1.004	(0.001)	1.004	-
Π	Steady-state inflation	1.011	(0.006)	1.008	(0.003)	1.005	-

Standard errors in parenthesis.

6.2 Bayesian DSGE

Columns 6 to 9 of Table 4 show the posterior estimates of the mode, mean, and standard deviation for each of the parameters. The mode and mean were within two standard deviations of each other for every parameter (except γ and ω_R) suggesting a fairly tight convergence in the posterior distribution. The Bayesian estimates strongly differed from the maximum likelihood results (shown in Column 10) in four instances. The price adjustment friction (ϕ_p), persistence of monetary policy shocks (ρ_v), and the intercept of the TFP process (A) were all fractions of the ML estimates. In contrast, the Bayesian estimate of investment productivity persistence (ρ_χ) was closer to the result found in Ireland (2003) which was much higher than the ML estimate.

Another difference was that the Bayesian estimate of the central bank response coefficient to the output gap was insignificant, whereas the ML estimate had a small and negative, but significant result. Whether these differences in estimates had important effects on the models' policy implications could be seen by comparing their impulse responses.

6.3 Impulse Response Functions

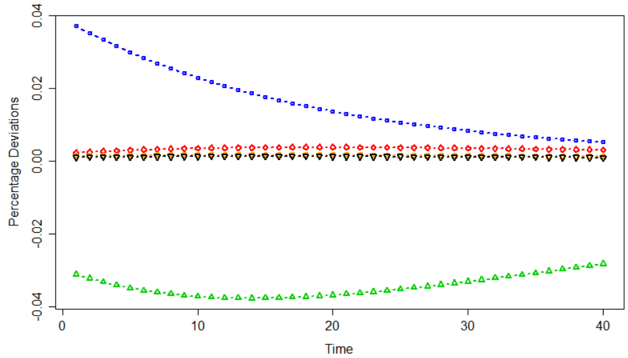
Figure 4 presents the impulse responses to a 1% deviation in three shock terms over 40 quarters at the estimated parameter values. Subfigures 4(a), 4(c), and 4(e) on the left column, and Subfigures 4(b), 4(d), and 4(f) on the right column, display the impulse responses for the ML and Bayesian models, respectively. The Bayesian IRFs are presented at the median of the estimates. The impulse responses of the variables to the shocks $\tilde{A}_t$ and $\tilde{\chi}_t$ were similar for both Bayesian and ML estimates and were thus not included.[37]

Shocks to the composite good and money demand had the same impact,

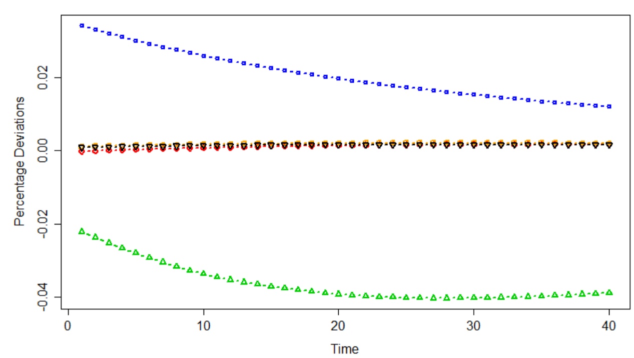
[37] However, these figures can be found in the Dropbox link.

Figure 4: Impulse response functions - 1% Shock

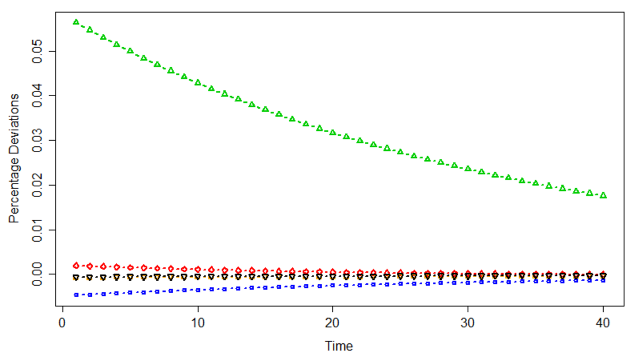
(a) ML-DSGE: Composite good shock ($\tilde{a}_t$)



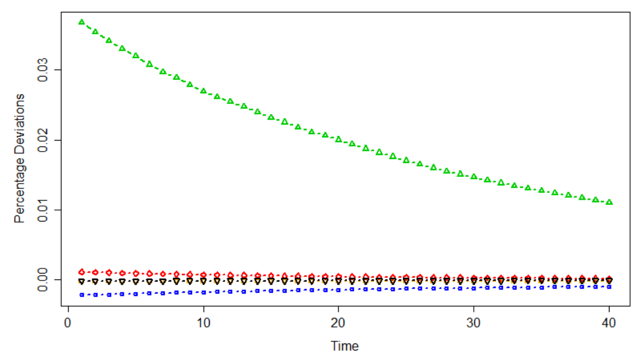
(b) Bayesian DSGE: Composite good shock ($\tilde{a}_t$)



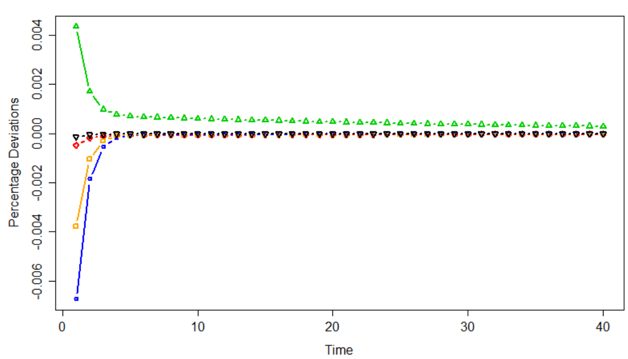
(c) ML-DSGE: Money demand shock ($\tilde{e}_t$)



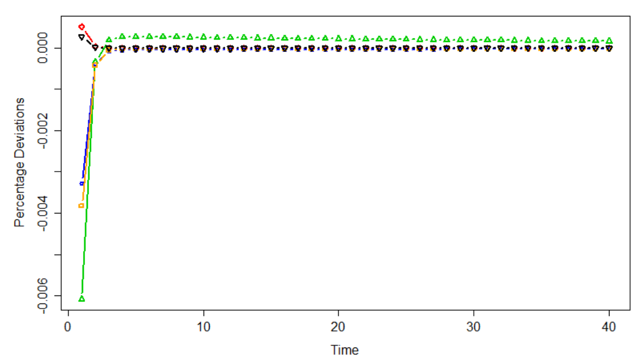
(d) Bayesian DSGE: Money demand shock ($\tilde{e}_t$)



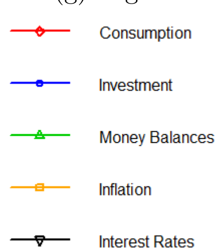
(e) ML-DSGE: Monetary policy shock ($\tilde{v}_t$)



(f) Bayesian DSGE: Monetary policy shock ($\tilde{v}_t$)



(g) Legend



qualitatively, between the Bayesian and ML models. Investment and money-balances had the largest magnitude of responses as can be seen in the subfigures throughout Figure 4. Money-balances were the most persistent, and after 40 periods remained more than 1% below and above their state-state values for the composite good and money demand shock, respectively. Interest rates and inflation had very small deviations when compared to other variables. The exception was when a positive shock to monetary policy (a ‘tightening’) caused a sharp deflation for the first few quarters, as Subfigures 4(e) and 4(f) show.

The relationship between money-balances and consumption differed when there was an inter- and intra-composite good shock. A shock making the composite good more desirable relative to leisure ($\tilde{a}_t$) increased, or had no impact on, consumption while decreasing money-balances, as Subfigures 4(a) and 4(b) show. The money demand equation (8) states that money-balances will increase when consumption increases (and vice versa) unless the interest rate increases by an offsetting amount. The small increase of $\tilde{R}_t$ above its steady-state from a shock to $\tilde{a}_t$ was able to decrease the demand for money, even though the composite good became more desirable. In contrast, a shock making money-balances more desirable within the composite good ($\tilde{e}_t$) increased both money-balances and consumption because the interest rate fell below its steady state, as shown in Subfigures 4(c) and 4(d).

The only qualitative difference in impulse responses between the ML and Bayesian models occurred for a monetary policy shock. This was likely due to the different estimates of the price adjustment friction and persistence of the monetary policy shock, as noted in Subsection 6.2. A tightening of monetary policy caused money-balances to increase above their steady-state values in the ML model, whereas they decreased at first in the Bayesian model.

6.4 Vector Autoregressions

Equation (13) presents a VAR of order p without a trend or a constant, as this fits the nature of the data used. The vector $\boldsymbol{y}_t$ is a 5×1 cross-section of data, $\boldsymbol{y}_t = [\tilde{c}_t \ \tilde{i}_t \ \tilde{m}_t^d \ \tilde{R}_t \ \tilde{\Pi}_t]'$, and $\boldsymbol{A}_k \in \{1, .., p\}$ is a 5×5 coefficient matrix.

$$\boldsymbol{y}_t = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + ... + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{u}_t \tag{13}$$

An order of lag length $p = 2$ was determined by the final prediction error (FPE) criterion. Rubaszek and Skrypcyzynki (2008) also found a lag length of 2 when using the FPE criterion with US data. Equation (13) can be written as a first-order process by stacking the vectors and defining the appropriate matrix.

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{I}_5 & \boldsymbol{0}_5 \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{0}_5 \end{bmatrix} \quad \longleftrightarrow$$

$$\boldsymbol{\xi}_t = \mathcal{A} \boldsymbol{\xi}_{t-1} + \boldsymbol{v_t} \tag{14}$$

If the estimate of $\mathcal{A}$ had moduli of eigenvalues which were less than one then the process was stable.[38] After a stable estimate had been found, the VAR model could forecast future values k-steps-ahead based on recursive methods.

$$\boldsymbol{y}_{t+k|t} = \boldsymbol{A}_1 \boldsymbol{y}_{t+k-1|k} + \boldsymbol{A}_2 \boldsymbol{y}_{t+k-2|t} \tag{15}$$

[38]With a maximum lag length of 8 the FPE criterion chose 8 lags, but this yielded an estimate of $\mathcal{A}$ that was not stable. A second round of tests with a maximum lag length of 4 then yielded an order of 2.

6.5 Bayesian VAR

One criticism of the classical VAR model was that it suffered from overparameterization, as the number of parameters grew exponentially with the order of lags. The Bayesian VAR avoided the potential problem of overfitting by shrinking the parameter space through the use of hyperparameters. The Minnesota prior (see Doan et al. (1984)) was a formula for determining the prior distribution of all parameters through three hyperparameters. First, the elements of the coefficient matrix A_k, as seen in equation (13), were given a Gaussian prior of the following form.

$$a_{iik} \sim \mathcal{N}(1, \sigma_{iik}^2) \tag{16}$$

$$a_{ijk} \sim \mathcal{N}(0, \sigma_{ijk}^2), \quad \text{if } i \neq j \tag{17}$$

The expected value of the coefficients of the i^{th} and j^{th} variable in equation i were assumed to be one and zero, with a standard deviation determined by the order of the lag and whether or not the element was on the diagonal of the coefficient matrix. This implied a prior which was a random walk, $E[y_{i,t}] = y_{i,t-1} + ... + y_{i,t-p}$, for the i^{th} row of equation (13). The prior standard deviation of the coefficient for variable j in equation i of order k was determined by the following formula.

$$\sigma_{ijk} = \theta \cdot w(i,j) \cdot k^{-\phi} \cdot \left(\frac{\hat{\sigma}_{kj}}{\hat{\sigma}_{ki}} \right) \tag{18}$$

In Bayesian estimation a smaller prior standard deviation implies more 'confidence' about the prior mean. This is also known as the 'tightness' of the prior. The hyperparameter θ determined overall tightness, ϕ determined the decay rate, and $w(i,j)$ was element ij of the weighting matrix. A higher value of θ implied a more diffuse prior. The higher the lag order, the smaller the standard deviation and hence the tighter the prior. Lastly, the weight matrix determined the relative tightness of variable j in equation i. The parameters $\hat{\sigma}_{ki}$ and $\hat{\sigma}_{kj}$ were estimates from running the following univariate autoregression of order k with n variables and calculating the standard error for the matching coefficient.

$$y_{it} = a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + ... + a_{1n}y_{n,t-1} + ... + a_{p1}y_{1,t-p} + a_{p2}y_{2,t-p} + ... + a_{pn}y_{n,t-p}$$

The BVAR was estimated using a Gibbs sampler with 10,000 runs and a 10% burn-in rate. The Minnesota prior specifies the following values for the hyperparameters.

$$W = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}, \quad \theta = 0.1, \quad \phi = 1$$

7 Forecasting

The data series for each variable had 88 observations from 1992-Q1 to 2013-Q4. All models had at most 44 forecasts from 2003-Q1 to 2013-Q4 for all variables at each forecast horizon. The Bayesian DSGE model used the median of its forecast values, as is the convention in the literature. This section considers four dimensions of forecast evaluation. The first dimension displays the model forecasts alongside the actual time series. These images provide intuition into the forecasting ability, model dynamics, and relationship between forecasting horizon and accuracy.

The second dimension compares models head-to-head using the Diebold-Mariano (DM) test[39] for forecast accuracy over 44 of the 88 periods (2003-Q1 to 2013-Q4). Models had 44 periods (1992-Q1 to 2002-Q4) to estimate their parameters followed by 41-44 periods of pseudo out-of-sample forecasts at each k-step-ahead.[40] The third dimension considers how the DM test statistic changed, if at all, with the forecasting start date. The fourth dimension illustrates how the mean-squared errors (MSEs) for each variable and model changed with the forecasting start state. Subsections 7.1 to 7.4 address these four dimensions of forecast evaluation in turn.

Visual examples in this section are drawn from various forecasting horizons and variables in order to highlight the most interesting findings. However, a full range of figures can be found in the appendix for the interested reader. Figures 15 to 18 contains all subfigures of actual forecast data. Figures 19 to 23 contain the dynamic DM test statistics for five head-to-head model comparisons. Lastly, Figures 24 to 28 show the MSE plots.

[39]For a full description of the DM test see Section 9.3 in the appendix.

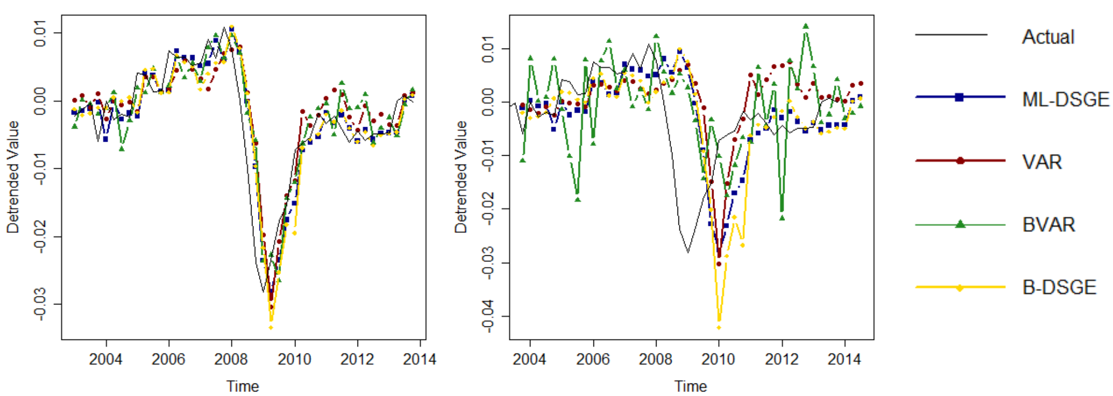
[40]Whenever the forecasting start date occurred, the k-step-ahead forecast would lose $k-1$ observations.

7.1 Time Series Forecasts

All models had similar forecasts for consumption, investment, and money-balances (with the exception of the Bayesian DSGE for money-balances) at the one-step-ahead horizon. Figure 5 shows the one- and four-step-ahead forecasts for consumption. The one-quarter-ahead forecasts, unsurprisingly, lagged exactly one-quarter behind the trough of the business cycle in 2009-Q1. Divergences between model forecasts were seen in the year-ahead-forecasts. However, the forecasts of 2009-Q1, which occurred in 2008-Q1, had all models predicting consumption above its steady-state value.[41]

Figure 5: Forecasts of consumption

(a) One-step-ahead (b) Four-step-ahead (c) Legend



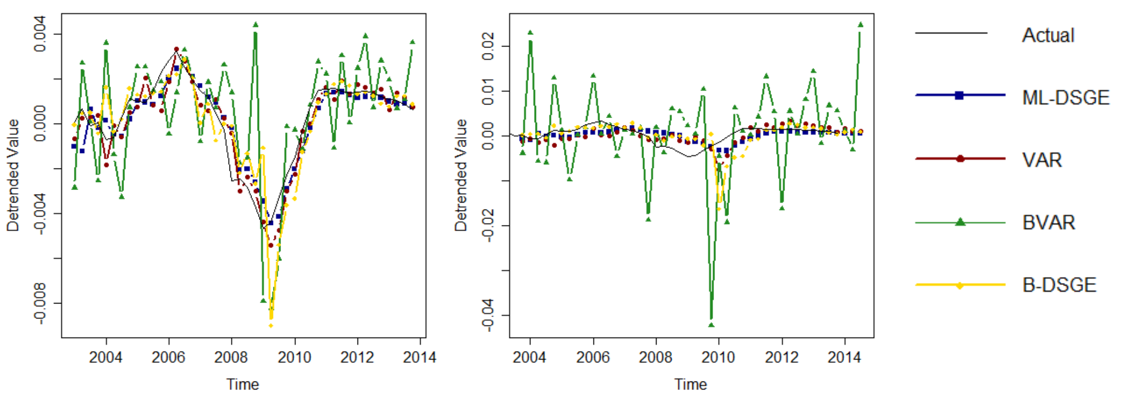
Whereas the forecasts for consumption and investment were similar across models at the shortest horizon, the forecasts of interest rates exposed the weaknesses in some models. Figure 6 presents the one- and four-step-ahead forecasts for the interest rate data. Two observations stand out. First, the volatility of the BVAR was immense, and suggested the model could not credibly forecast interest rate data. Second, the BVAR and Bayesian DSGE models

[41]This result is emblematic of the failure of standard macroeconomic models, and concomitantly macroeconomists, to predict the financial crisis, as discussed in Section 2.

both predicted a much larger fall in interest rates than actually occurred at the trough of the business cycle.

Figure 6: Forecasts of interest rates

(a) One-step-ahead (b) Four-step-ahead (c) Legend



Two other interesting visual phenomena occurred in the forecast results. The first phenomenon was that the Bayesian DSGE model wildly over-predicted the upswing in money-balances during the financial crisis, as Figure 7 shows. Even ignoring this massive error in 2009, the Bayesian DSGE model had the most volatile predictions for money-balances and was a poor forecaster for this variable in general. The result was surprising because of both the scale of the forecast error and its difference to the ML-DSGE (an identical model different only in estimation technique).

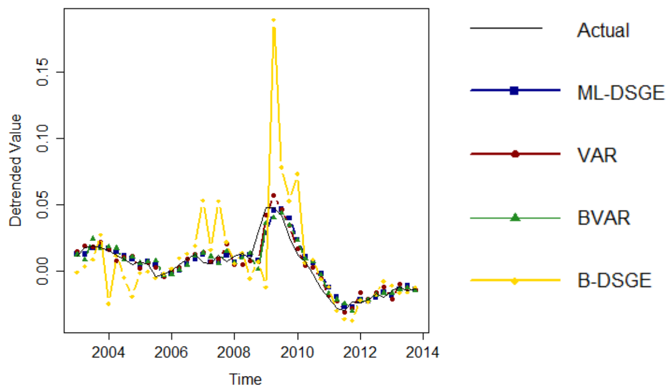
Bayesian forecasts took into account both parameter and shock uncertainty.[42] The combination of these two uncertainties caused a much different estimate of the state of the DSGE system with Bayesian estimation. This explains why the interest rate and money-balance forecasts, as seen in Figures 6

[42]In a linearized state-space system, all deviations in steady-state values are driven by the vector of shocks (the σ's) - the 'state' of the system. A state-space observer system uses a matrix of observations in order to estimate these shocks (see Section 9.2.1 in the appendix).

and 7, were starkly below and above their steady-states values, respectively.[43]

Figure 7: Forecasts of money-balances

(a) One-step-ahead (b) Legend

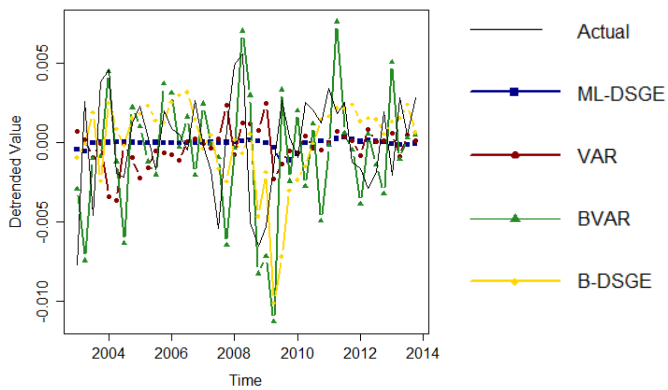


The second phenomenon was the flat inflation forecasts of the ML-DSGE model. Figure 8 presents the one-step-ahead forecasts of inflation. Even at this short time-scale, accuracy was lacking for most models. What this figure revealed, alongside Figures 6 and 7, was that there was a flawed mechanism between interest rates, money-balances, and inflation in the DSGE model. An advantage of DSGE models is that their system output and forecasts are internally consistent and explainable. In order to achieve a variability close to the true level of inflation (as the Bayesian DSGE did) the forecasts of interest rates and money-balances needed to be implausibly large. In contrast, the ML-DSGE had more accurate money-balance forecasts, but at the expense of credible inflation rates. While the out-of-sample forecasts of the DSGE models were lacklustre, their errors provided a clear diagnosis of the faulty mechanisms within the NK model.

[43]Recall the tight link between money demand and interest rates from the money demand equation (8).

Figure 8: Forecasts of inflation rates

(a) One-step-ahead (b) Legend



7.2 DM Test - Maximum Forecast Range

The DM test had a null hypothesis that the expected difference in the squared forecasting errors between two models was zero. Table 6 presents the DM test statistic and p-values for all model comparisons. A negative test statistic implied that the first listed model had smaller forecast errors than the second. If the absolute value of the test statistic was large enough, then the p-value was sufficiently small to reject the null hypothesis. As the DM test was symmetric, a p-value less (greater) than 0.05 (0.95) implied that the first (second) listed model was superior to the second (first) at the 5% level.

Subtables 6(a) and 6(b) present the results of the ML-DSGE model against the VAR and BVAR. At the 5% level, the ML-DSGE outperformed the VAR for consumption at the one-step-ahead horizon. However, the VAR had statistically significant results over the ML-DSGE for money-balances, inflation, and interest rates at one- and two-step-ahead forecasts. The BVAR did not have any significant results against the ML-DSGE. However for investment, money-balances, inflation, and interest rates the ML-DSGE outperformed the BVAR, especially at three- and four-step-ahead forecasts. A similar result emerged for the Bayesian DSGE against the VAR and BVAR, as seen in Subtables 6(c) and 6(d). The VAR outperformed the NK model in investment, money-balances, inflation, and interest rates. The Bayesian DSGE beat the BVAR with no losses against, but did not score as many wins as the ML-DSGE did against the BVAR (10 significant test statistics compared to 4).

As transitivity would suggest, and Table 6(e) shows, the VAR was significantly better than the BVAR at all forecast horizons for money-balances, inflation, and interest rates. However, for consumption forecasts, no clear winner emerged. Lastly, the head-to-head comparison of the DSGE model by ML and Bayesian approaches can be seen in Table 6(f). The Bayesian DSGE was outperformed by the classical model for consumption and investment in six of the eight cases. Median forecasts from the Bayesian model had only one significant result: one-step-ahead inflation forecasts.

7.3 DM Test - Moving Forecast Range

In the previous subsection, half of the observations were used for the out-of-sample forecasting period. The division was arbitrary, and model performance could have been conditional on when forecasting started. Delaying a forecasting start date provides more observations for parameter estimation which should, in theory, improve forecasting performance. However, DSGE models are highly stylized and their forecasting ability may be variant to both underlying policy and structural conditions.

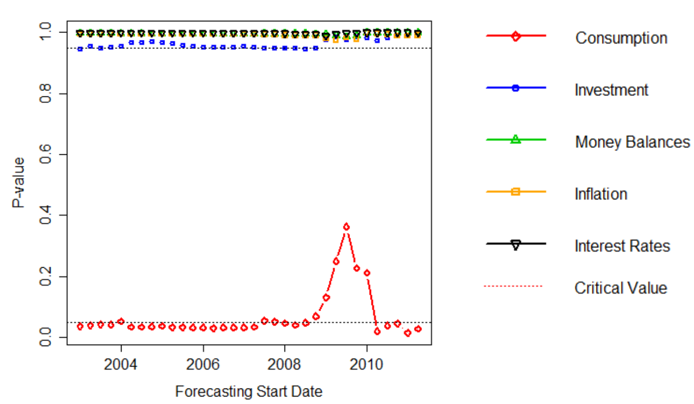
This subsection analyzes how the p-values of the DM test changed alongside the forecasting start date. Three questions are specifically addressed: (i) Were the statistically significant results of 2003-Q1 (which corresponds to the findings in Table 6) robust over time? (ii) Did performance become significant at future dates? (iii) How did the financial crisis affect test results?

While the power of the DM test decreased with later forecasting start dates, there was no apparent decrease in the number of significant results.

In Table 6(a) the ML-DSGE outperformed the VAR in consumption whereas the VAR had better forecasts in the other three variables, at the one-step-ahead horizon. Figure 9 shows that this result was highly stable. The only exception was when the consumption forecast from the ML-DSGE became insignificant for six quarters during the financial crisis. This subfigure is emblematic of the dynamic relation in test statistics for the ML-DSGE and VAR. In this case, the third dimension of forecast evaluation did not add any more information than is contained in Table 6(a). During the financial crisis, the p-values of many variables spiked upwards which indicated a relative decline in the forecasting performance of the ML-DSGE (see Figure 19 in the appendix). However, the effect was not strong enough to cause the p-values to move above the critical value.

Figure 9: ML-DSGE versus VAR

(a) One-step-ahead (b) Legend



P-values below (above) 0.05 (0.95) denote a statistically significant result for the ML-DSGE (VAR) model.

Table 6: Model comparison by DM test (p-values in parenthesis and k denotes the forecast horizon)

(a) ML-DSGE†† versus VAR**

	$k=1$	$k=2$	$k=3$	$k=4$
$\tilde{c}$	-1.844††	-1.032	-0.604	-0.486
	(0.036)	(0.154)	(0.275)	(0.315)
$\tilde{i}$	1.650	-0.633	-1.367	-1.524
	(0.947)	(0.265)	(0.090)	(0.068)
$\tilde{m}^d$	3.071**	1.762**	1.158	-0.233
	(0.998)	(0.957)	(0.873)	(0.408)
$\tilde{\Pi}$	2.693**	1.709**	1.455	1.372
	(0.995)	(0.953)	(0.923)	(0.911)
$\tilde{R}$	3.211**	1.859**	1.533	1.401
	(0.999)	(0.965)	(0.934)	(0.916)

(b) ML-DSGE†† versus BVAR$^{\oplus}$

	$k=1$	$k=2$	$k=3$	$k=4$
$\tilde{c}$	-1.019	-0.801	-0.466	-0.388
	(0.157)	(0.214)	(0.322)	(0.350)
$\tilde{i}$	-2.718††	-2.312††	-2.224††	-2.543††
	(0.005)	(0.013)	(0.016)	(0.008)
$\tilde{m}^d$	0.400	-0.070	-0.704	-2.419††
	(0.654)	(0.472)	(0.243)	(0.010)
$\tilde{\Pi}$	0.614	-0.209	-1.846††	-2.665††
	(0.729)	(0.418)	(0.036)	(0.006)
$\tilde{R}$	-0.413	-1.814††	-4.789††	-3.885††
	(0.341)	(0.038)	(0.00002)	(0.0002)

(c) Bayesian DSGE‡‡ versus VAR**

	$k=1$	$k=2$	$k=3$	$k=4$
$\tilde{c}$	-0.086	0.440	0.654	0.715
	(0.466)	(0.669)	(0.742)	(0.761)
$\tilde{i}$	1.876**	0.083	-0.725	-0.964
	(0.966)	(0.533)	(0.236)	(0.170)
$\tilde{m}^d$	1.682**	1.265	1.157	1.099
	(0.950)	(0.894)	(0.873)	(0.861)
$\tilde{\Pi}$	2.727**	1.756**	1.476	1.451
	(0.995)	(0.957)	(0.926)	(0.923)
$\tilde{R}$	3.048**	1.815**	1.505	1.441
	(0.998)	(0.962)	(0.930)	(0.921)

(d) Bayesian DSGE‡‡ versus BVAR$^{\oplus}$

	$k=1$	$k=2$	$k=3$	$k=4$
$\tilde{c}$	0.380	0.789	0.488	0.324
	(0.647)	(0.783)	(0.686)	(0.626)
$\tilde{i}$	0.593	1.401	1.511	1.492
	(0.722)	(0.916)	(0.931)	(0.928)
$\tilde{m}^d$	1.337	1.216	1.133	1.079
	(0.906)	(0.884)	(0.868)	(0.856)
$\tilde{\Pi}$	-0.746	-1.149	-2.255‡‡	-2.240‡‡
	(0.230)	(0.128)	(0.015)	(0.015)
$\tilde{R}$	-0.378	-	-5.090‡‡	-4.269‡‡
	(0.354)	(-)	(0.00001)	(0.0001)

(e) VAR** versus BVAR$^{\oplus}$

	$k=1$	$k=2$	$k=3$	$k=4$
$\tilde{c}$	0.710	0.409	-0.032	-0.218
	(0.759)	(0.658)	(0.488)	(0.414)
$\tilde{i}$	-1.893**	0.164	1.070	1.306
	(0.033)	(0.565)	(0.855)	(0.901)
$\tilde{m}^d$	-3.146**	-2.084**	-1.902**	-2.062**
	(0.002)	(0.022)	(0.032)	(0.023)
$\tilde{\Pi}$	-3.565**	-2.535**	-1.993**	-1.954**
	(0.0005)	(0.008)	(0.026)	(0.029)
$\tilde{R}$	-3.110**	-2.000**	-2.214**	-2.501**
	(0.002)	(0.026)	(0.016)	(0.008)

(f) Bayesian DSGE‡‡ versus ML-DSGE††

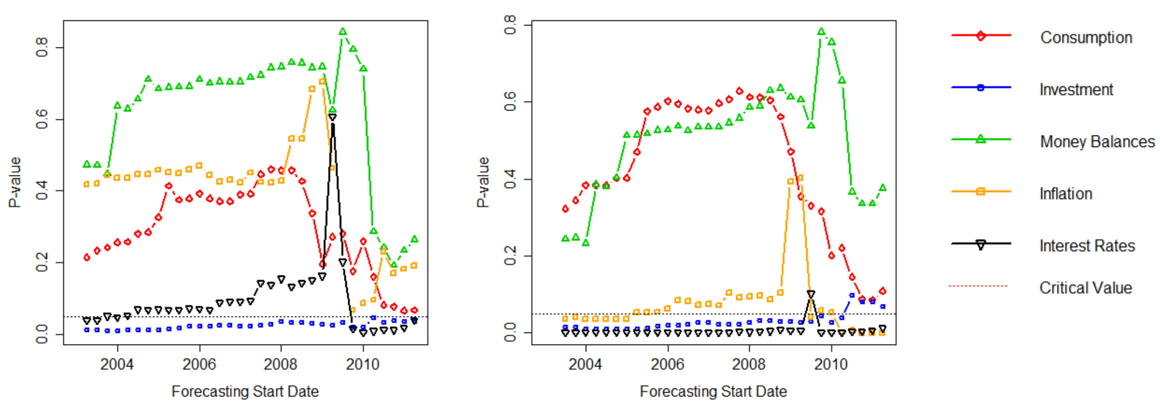
	$k=1$	$k=2$	$k=3$	$k=4$
$\tilde{c}$	2.060††	1.907††	1.497	1.123
	(0.977)	(0.968)	(0.929)	(0.866)
$\tilde{i}$	2.528††	2.076††	1.879††	1.885††
	(0.992)	(0.978)	(0.966)	(0.967)
$\tilde{m}^d$	1.352	1.224	1.152	1.110
	(0.908)	(0.886)	(0.872)	(0.863)
$\tilde{\Pi}$	-1.732‡‡	-1.223	-0.964	-0.557
	(0.045)	(0.114)	(0.170)	(0.290)
$\tilde{R}$	-0.026	0.419	0.366	1.245
	(0.490)	(0.661)	(0.642)	(0.890)

Symbols: $\dagger\dagger$ (ML-DSGE), $\ddagger\ddagger$ (Bayesian DSGE), $**$ (VAR), and $\oplus$ (BVAR) denote that the relative forecasting errors of the respective model were statistically significant at the 5% level.

In Table 6(b) the classical NK model was the clear winner against the BVAR with 10 of 20 tests in its favour and none against. However, an analysis of its dynamic results showed that its performance in many variables deteriorated over time. Figure 10 shows that at the two- and three-step-ahead forecasts the ML-DSGE model lost its statistically significant performance for interest rates and inflation within a year, respectively. This result was informative: the NK model was not as good at forecasting certain variables as the estimates from 2003 would have suggested. The upward spike in p-values between 2008-2010 showed that the relative performance of the BVAR increased during this period.

Figure 10: ML-DSGE versus BVAR

(a) Two-step-ahead (b) Three-step-ahead (c) Legend

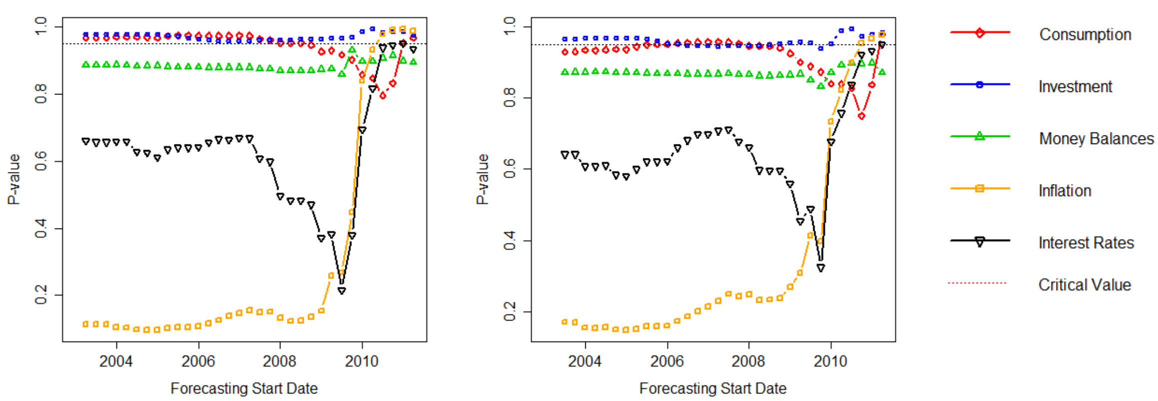


P-values below (above) 0.05 (0.95) denote a statistically significant result for the ML-DSGE (BVAR) model.

In other instances, forecast differences which were insignificant converged to regions of significance. In Subtable 6(f) the classically estimated DSGE model outperformed its Bayesian counterpart with 6 of the 20 tests in its favour and only one against. However, by 2010, the forecasting results were even more in its favour. Figure 11 shows that at 2003-Q1, the classical model had three significant results at the two- and three-step-ahead forecast horizons. By the end of the sample the ML-DSGE had significant results for: consumption, investment, and inflation at the 5% level as well as money-balances and interest rates at the 10% level, for both forecast horizons. The visual evidence revealed a clear decline in the Bayesian model's forecasting performance near the end of the sample, as indicated by the upward surge in p-values.

Figure 11: Bayesian DSGE versus ML-DSGE

(a) Two-step-ahead (b) Three-step-ahead (c) Legend



P-values below (above) 0.05 (0.95) denote a statistically significant result for the Bayesian DSGE (ML-DSGE) model.

7.4 MSEs - Moving Forecast Range

The MSE from any vector of forecasts is simply the average of the sum of squared forecasting errors: $(1/N)\sum_{t=1}^{N} e_t^2$. This measure is one of the many metrics used to evaluate forecasting performance.[44] MSEs weight forecasting errors symmetrically and penalize larger variances. The MSE measure cannot be compared across variables. However, MSEs are effective at ranking models for a specific variable and forecast horizon. A model with a smaller MSE is a 'better' forecaster. Comparing MSEs for a given variable and forecast horizon

[44]Other methods include root mean squared error or absolute error loss.

provides a more intuitive understanding of performance across models. In contrast, the DM test can only compare two models at a time. The logarithm of the MSE was used to facilitate an easier visualization.

Figure 12: Consumption - (log) MSE

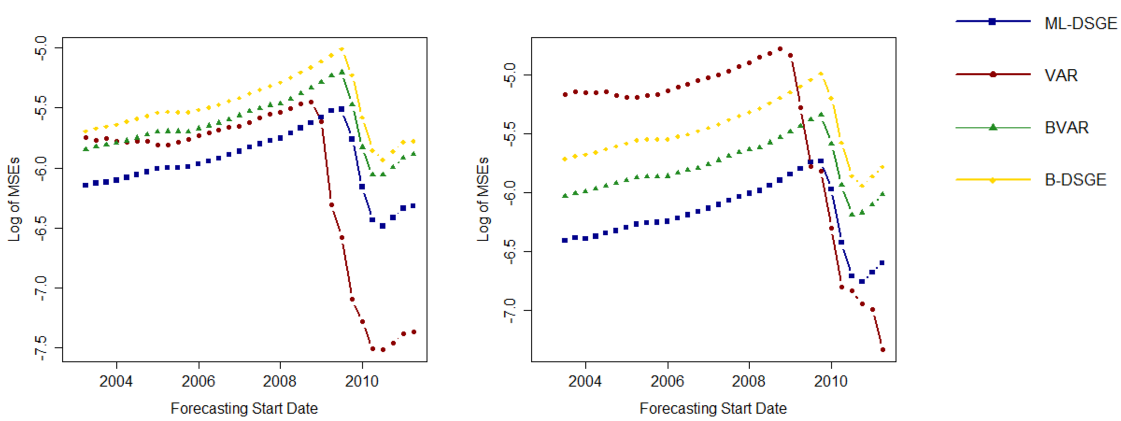
(a) One-step-ahead (b) Two-step-ahead (c) Legend

Figure 12 shows that the ML-DSGE model had the smallest MSEs for all forecasting start dates (except for three quarters in 2011). The NK model had statistically significant results in 3 of the 6 possible cases at the one- and two-step ahead horizons for the DM tests in 2003-Q1, as seen in Tables 6(a), 6(b), and 6(f). However, the evidence contained in Figure 12 presents a stronger endorsement for the ML-DSGE model due to its consistent performance across start dates in forecasting consumption at a short-term horizon.

Comparing MSEs across time showed that two models could switch positions for having the smallest MSEs. Figure 13 shows the two- and three-step ahead forecasts for investment. From 2003-2009, the ML-DSGE had the smallest errors, but afterwards the classical VAR became the most accurate forecaster, and maintained this performance until the end of the sample. This added further evidence that the unconstrained models had better performances during the financial crisis, whereas the stylized NK models did best at fore-

Figure 13: Investment - (log) MSEs

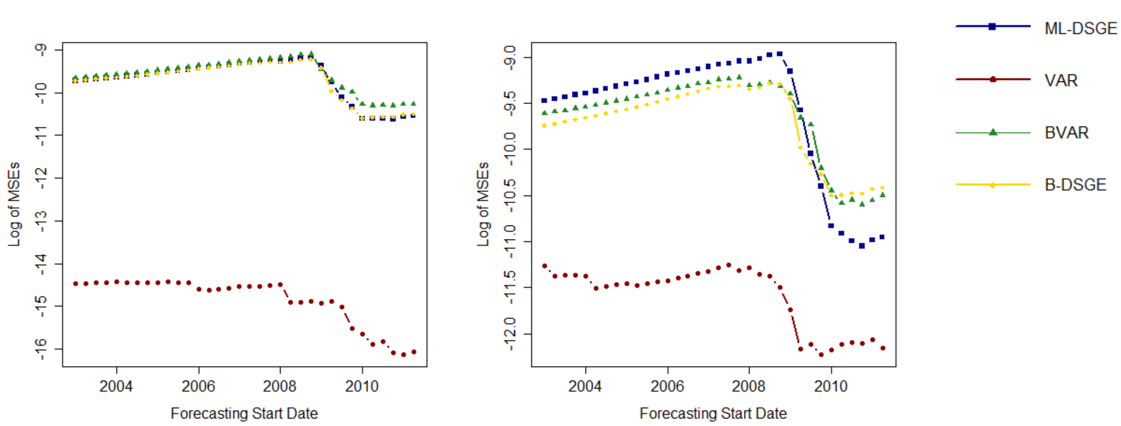
(a) Two-step-ahead (b) Three-step-ahead (c) Legend



casting variables during more ‘normal’ times of economic activity.

Figure 14: Interest rates and inflation - (log) MSEs

(a) Interest rates: one-step-ahead (b) Inflation: one-step-ahead (c) Legend



The poor performances in interest rate and inflation forecasts by the NK and BVAR models were dramatically confirmed by the visual evidence of the dynamic MSEs. Figure 14 displays the MSEs for the one-step-ahead interest rate and inflation forecasts. Given that the MSEs are shown in logarithms, the performance of the classical VAR was exceedingly better than the other models. The classical and Bayesian DSGE models alongside the BVAR model all had similar (and large) MSEs. The result confirmed the need to improve this NK model's ability to accurately match the inflation and interest rates processes of the Canadian economy.

8 Conclusions

An unconstrained VAR model can have superior forecasting results compared to a DSGE model for two reasons. First, changes in the underlying structure of the data can be accounted for by new parameter estimates.[45] Second, the lag structure and unspecified relationship between variables provides powerful short-term forecasts by, in effect, overparameterizing the model. However, the advantages of a VAR are not always sufficient to ensure superior forecasts. Empirical results showed that DSGE models could outperform VAR models (see discussion in Section 2). In small sample sizes, which most macroeconomic time series are, cross-equation restrictions may improve the fit of the model. For example, imposing the restriction that output is the sum of consumption and investment, which is almost certainly true, reduces the number of parameters needed to estimate this relationship.

The results of this paper were threefold. The first result was that the classical VAR model outperformed the New Keynesian models and its Bayesian counterpart in forecasting important Canadian macroeconomic time series.[46] The evidence was comprehensive. The VAR won 25 of the 60 possible head-to-head tests against all other models, while losing only one. The significant test statistics found in 2003 were robust over time, and more significant results were achieved at later dates. The MSEs of the classical VAR for inflation and interest rates were fractions of the other models at all forecast horizons. The trade-offs that could have benefited the restricted model were overpowered by the flexible specification of the VAR. This finding was unfortunate as DSGE models are powerful tools for policy analysis and there is a natural desire amongst their users that they also be competent forecasters.

[45]It is well known that VAR coefficients have not remained stable over time, especially the relationship between unemployment and inflation.

[46]The one exception to this dominance was in consumption.

The second result was that the use of Bayesian techniques to estimate either the DSGE or VAR model did not improve forecasting accuracy. The BVAR was the worst performer, with no tests in its favour. The Bayesian DSGE both underperformed and had results that were not as robust as its classically estimated counterpart. The third result was that the financial crisis proved deleterious for the forecasting performance of both the classical and Bayesian DSGE models. The poor performances of the NK models during 2008-2010 were, unfortunately, in keeping with the generally perceived notion that macroeconometric models 'missed' the crash.

Areas for further research

The uneventful findings of this paper likely speak more to its own weaknesses than those of Bayesian techniques or DSGE model forecasting. An array of options for improving this DSGE model and its estimation exist for future research. As the choice of model matters, it may be that this specific NK model did not conform well to Canadian time series. An open-economy model that interacts with US monetary policy seems a more plausible, if not complicated, representation of the Canadian economy. A panoply of DSGE models that include financial frictions, heterogeneous agents and firms, and learning mechanisms may contain more realistic restrictions and improve forecasting results.

The use of more complicated models advantages the Bayesian over the classical approach as it is much easier to integrate functions of the model's outcome than to maximize them (see Fernandez-Villaverde (2010)).[47] The MCMC algorithm is generally able to explore more complex posterior distributions by simply increasing the number of draws, whereas even the most sophisticated

[47] This claim is not universally accepted though, and researchers have found examples where numerical integration is superior to Monte Carlo integration (see Judd et al. (2011)).

maximization algorithms have trouble exploring complex spaces periled with flat surfaces and local maximas.[48]

Large-scale DSGE models were generally the best forecasters in the literature, suggesting more data series may be required to improve performance. A larger number of time series used to estimate the DSGE model would mean fewer fixed parameters. Removing some of the fixed parameters may also ameliorate the faulty relationship discovered between interest rates, inflation, and money-balances in the DSGE model. Including a vector of measurement errors that followed a first-order autoregressive process, as in Ireland (2004), would allow five more observables while only adding 10 parameters to the DSGE model.[49] In contrast, the number of VAR parameters would increase by 150! Such a change would presumably tilt the forecasting balance in favour of the more parsimonious model.

An existing weakness with Canadian macroeconomic time series is that there is neither a consistent survey (such as the Survey of Professional Forecasters in the US) nor a real-time database accessible to public researchers. Many data series, such as unemployment and output, are subject to ex-post revisions. If model forecasts of Canadian time series are ever to be contrasted to historic forecasts, a real-time database would be necessary so that the models can use the data available to forecasters at the time they were historically made.

A final area of improvement would be to specify the model in a detrended rather than a detrended *and* a log-linearized form.[50] This would have the advantage that the forecasted data series could be mapped directly to their

[48]As an example, the attempt to use Canadian time series for the exact model specification of Ireland (2003) yielded indeterminate results.

[49]Put another way, the ratio of DSGE parameters to observables would change from 10:2 to 7:2.

[50]Dib et al. (2008) was one of the few papers in the forecasting literature to use a non-linear form.

level form. The use of HP filters to extract the cyclical component does not allow a reverse mapping. It seems reasonable to believe that policy makers would prefer forecasts in level form rather than in deviations from the 'steady-state.' Log-linearized systems are often used as they are easier to estimate. However, this only reinforces the advantage of Bayesian techniques which are more conducive to complicated dynamic systems.

The opportunities for future research are abundant and the development of a NK model which can accurately and robustly forecast key Canadian macroeconomic time series should be an important objective for researchers. The emphasis on DSGE modeling stems from its advantages: a solid theoretical foundation, an internally consistent structure, and the ability to provide policy analysis.

References

M. Adolfson, J. Linde, and M. Villani. Forecasting performance of an open economy DSGE model. *Econometric Reviews*, 26(2-4):289–328, 2007.

S. An and F. Schorfheide. Bayesian analysis of DSGE models. *Econometric Reviews*, 26(2-4):113–172, 2007.

J. Atta-Mensah. Money demand and economic uncertainty. *Bank of Canada Working Paper*, 2004.

S. Bekiros and A. Paccagnini. Bayesian forecasting with a factor-augmented Vector Autoregressive DSGE. *Remini Centre for Economic Analysis Working Paper*, 22(13), 2013.

P. R. Bergin. Putting the 'New Open Economy Macroeconomics' to a test. *Journal of International Economics*, 60(1):3–34, 2003.

B. Bernanke, M. Gertler, and S. Gilchrist. The financial accelerator in a quantitative business cycles framework. *Handbook of Macroeconomics Volume 1C*, 1999.

O. J. Blanchard and C. Kahn. The solution of linear difference models under rational expectations. *Econometrica*, 48(5):1305–1311, 1980.

F. Brazdik, M. Hlavacek, and A. Marsal. Survey of Research on Financial Sector Modelling within DSGE Models: What Can Central Banks Learn from It? *Czech Journal of Economics and Finance*, 62(3):252–277, 2012.

M. Brunnermeier, T. Eisenbach, and Y. Sannikov. Macroeconomics with financial frictions: A survey. *Working Paper*, 2012. URL `https://www.princeton.edu/~sannikov/survey_macrofinance.pdf`.

L. Christiano. Econometric methods for the analysis of dynamic general equilibrium models. http://faculty.wcas.northwestern.edu/~lchrist/course/estimationhandout.pdf, 2014.

L. Christiano and M. Eichenbaum. Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations. *American Economic Review*, 82(3): 430–50, 1992.

K. Christoffel, G. Coenen, and A. Warne. Forecasting with DSGE Models. *European Central Bank Working Paper Series*, 1185, May 2010.

M. P. Clements and D. F. Hendry. Evaluating a Model by Forecast Performance. *Oxford Bulletin of Economics and Statistics*, 67(1):931–956, 2005.

M. Del Negro and F. Schorfheide. Priors from general equilibrium models for VARs. *International Economic Review*, 45(2):643–673, 2004.

M. Del Negro and F. Schorfheide. How good is what you've got? DSGE-VAR as a toolkit for evaluating DSGE models. *Federal Reserve Bank of Atlanta Economic Review*, Second Quarter:21–37, 2006.

M. Del Negro and F. Schorfheide. DSGE model-based forecasting. Staff Reports 554, Federal Reserve Bank of New York, 2012. URL http://ideas.repec.org/p/fip/fednsr/554.html.

W. Den Haan. Dynare and Bayesian estimation. www.wouterdenhaan.com/numerical/slidesdynareestimation.pdf, 2011.

A. Dib, M. Gammoudi, and K. Moran. Forecasting Canadian time series with the New Keynesian model. *Canadian Journal of Economics*, pages 138–165, 2008.

F. Diebold. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. *National Bureau of Economic Research, Inc*, 18391, September 2012. URL `http://ideas.repec.org/p/nbr/nberwo/18391.html`.

F. Diebold and R. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.

T. Doan, R. Litterman, and C. Sims. Forecasting and conditional projections using realistic prior distributions. *Econometric Review*, 3(1):1–100, 1984.

J. Dorich, M. Johnston, R. Mendes, S. Murchison, and Y. Zhang. ToTEM II: An Updated Version of the Bank of Canada's Quarterly Projection Model. *Bank of Canada Technical Report*, No. 100, 2013.

E. Drysdale. Bayesian and Classical Forecasting of Canadian Macroeconomic Time Series: A Comparison Study. *Queen's University MA Essay*, 2014.

R. M. Edge, M. T. Kiley, and J. Laforte. A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model. *Federal Reserve Board Finance and Economics Discussion Series*, 2010.

J. Fernandez-Villaverde. The econometrics of DSGE models. *Journal of the Spanish Economic Association*, 1:3–49, 2010.

T. Griffoli. Dynare user guide. `http://www.dynare.org/documentation-and-support/user-guide/Dynare-UserGuide-WebBeta.pdf`, 2008.

P. Guerron-Quintana and J. Nason. Bayesian estimation of DSGE models. *Federal Reserve Bank of Philadelphia Working Paper*, 12(4), 2012.

D. Harvey, S. Leybourne, and P. Newbold. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13:281–291, 1997.

P. Ireland. Endogenous money or sticky prices: Mathematical appendix. http://irelandp.com/progs/endogenous.zip, 2002.

P. Ireland. Endogenous money or sticky prices? *Journal of Monetary Economics*, 50:1623–1648, 2003.

P. Ireland. A method for taking models to the data. *Journal of Economic Dynamics and Control*, 28:105–1226, 2004.

K. Judd, L. Maliar, and S. Maliar. One-Node Quadrature Beats Monte Carlo: A Generalized Stochastic Simulation Algorithm. *NBER Working Paper*, 2011. URL http://www.nber.org/papers/w16708.pdf.

M. Julliard and J. Maih. Estimating DSGE Models with Observed Real-Time Expectation Data. *Federal Reserve Bank of Kansas Working Paper*, 2010. URL http://www.kansascityfed.org/publicat/events/research/2010CenBankForecasting/Maih_paper.pdf.

G. King. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. The University of Michigan Press, 1998.

M. Kolasa, M. Rubaszek, and P. Skrzypczynski. Putting the New Keynesian DSGE model to the real-time forecasting test. *Journal of Money, Credit, and Banking*, 44(7):1301–1324, 2012.

F. Kydland and E. C. Prescott. Time to Build and Aggregate Fluctuations. *Econometrica*, 50(6):1345–1370, 1982.

J. Lagarias, J. Reeds, M. Wright, and P. Wright. Convergence properties of the Nelder-Mead Simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147, 1998.

T. Lancaster. *Introduction to Modern Bayesian Econometrics*. Wiley-Blackwell, 2004.

T. Lubik and F. Schorfheide. Do central banks respond to exchange rate movements? A structural investigation. *Journal of Monetary Economics*, 54:1069–1087, 2007.

R. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1):19–46, January 1976.

A. Pagan. Report on modelling and forecasting at the Bank of England. *Quarterly Bulletin Economic Modelling Articles*, 2003. URL `http://www.bankofengland.co.uk/archive/Documents/historicpubs/qb/2003/qb030106.pdf`.

J. Pfeifer. A guide to specifying observations equations for the estimation of DSGE models. `https://sites.google.com/site/pfeiferecon/Pfeifer_2013_Observation_Equations.pdf?attredirects=0`, 2014.

M. Rubaszek and M. Kolasa. Forecasting with DSGE models with financial frictions. *Dynare Working Paper Series*, 40, 2014. URL `http://www.dynare.org/wp-repo/dynarewp040.pdf`.

M. Rubaszek and P. Skrypcyzynki. On the forecasting performance of a small-scale DSGE model. *International Journal of Forecasting*, 24:498–512, 2008.

F. Schorfheide. Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics*, 15(6):645–670, 2000.

C. Sims. Macroeconomic and reality. *Econometrica*, 48:1–48, 1980.

F. Smets and R. Wouters. An estimated dynamic stochastic general equilibrium model for the euro area. *Journal of the European Economic Association*, 1(5):1123–1175, 2003.

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

A. A. Smith. Estimating non-linear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:63–84, 1993.

J. Stock and M. Watson. Forecasting inflation. *Journal of Monetary Economics*, 44:293–335, 1999.

L. Tesfatsion. Notes on the Lucas Critique, Time Inconsistency, and Related Issues. *Iowa State Macro Theory - Econ 502*, November 2013. URL `http://www2.econ.iastate.edu/tesfatsi/luccrit.pdf`.

M. Watson. Measures of fit for calibrated models. *Journal of Political Economy*, 101(6):1011–1041, 1993.

V. Wieland, T. Cwik, and G. Muller. A New Comparative Approach to Macroeconomic Modeling and Policy Analysis. *Working Paper*, 2012.

9 Appendix

The appendix is organized into five subsections. The first, Subsection 9.1, presents a summary of the model, cited equations, and the system of differential equations. The second, Subsection 9.2, details the state-space representation of the model and includes information on the algorithms and techniques used to estimate the ML and Bayesian DSGE models. The third, Subsection 9.3, explains the statistics behind the Diebold-Mariano test. The fourth, Subsection 9.4, discuses the data sources. The fifth, Subsection 9.5, shows the full range of figures.

9.1 New Keynesian Model

In this section, a general overview of the model is found below followed by six further subsections which contain the first-order conditions of the agents and several systems of equations. The first two subsections, 9.1.1 and 9.1.2, detail the first-order conditions from the household and intermediate goods-producing firm, respectively. Subsections 9.1.3 and 9.1.4 show the system of equations for the model in its nonstationary and intensive form, respectively. In the penultimate subsection, 9.1.5, the time subscripts are dropped and the intensive form model's steady-state is solved. Lastly, Subsection 9.1.6 derives the log-linearized system of equations.

The economy was comprised of four economic agents: (i) a representative household, (ii) a representative finished goods-producing firm, (iii) a continuum of intermediate goods-producing firms, and (iv) a central bank. The continuum of symmetric intermediate goods-producing firms were indexed by $i \in [0, 1]$. These intermediate firms produced an intermediate good i with labour and capital supplied by the household. The households were paid an economy-wide wage rate W_t and a rental rate of capital Q_t for their labour and capital, respectively. Households had a utility function which valued consumption of the finished good (purchased from the representative finished goods-producing firm), leisure, and real cash balances.

The representative finished goods-producing firm purchased an intermediate good from each intermediate goods-producing firm. Due to intermediate goods being imperfect substitutes, there existed a downward sloping demand curve for each intermediate good i. Firm i was therefore able to set its price in a monopolistically competitive environment.[51] There were two frictions within this economy. First, the representative household faced a cost of changing the capital stock at a rate different than the deterministic growth rate g. Second, the intermediate goods-producing firms faced a cost of adjusting their prices at a rate different from the steady-state of inflation.

The order of economic activity was as follows. Households entered period t with M_{t-1} units of money, B_{t-1} bonds, and K_{t-1} units of capital. K_{t-1} denotes the capital stock used in period t and K_t denotes the capital stock chosen through the consumption and investment decisions of that period.[52] The household received a lump-sum transfer of T_t from the monetary authority and the B_{t-1} units of bonds matured into the equivalent amount of currency.

9.1.1 Household's First-Order Conditions

By rewriting capital's law of motion as a function of K_t and K_{t-1}, I_t could be substituted out of the constraint and thus the representative household needed to optimize five variables: C_t, H_t, B_t, M_t, K_t. Equations (19) to (24) present the first-order conditions for these five variables and the Lagrange multiplier (Λ_t).

[51] As there were no fixed costs these monopoly rents did not dissipate over time.

[52] This was why the capital stock was a predetermined endogenous variable.

$$\frac{\partial \mathcal{L}}{\partial C_t}: \quad a_t = [C_t^{(\gamma-1)/\gamma} + e_t^{1/\gamma}(M_t/P_t)^{(\gamma-1)/\gamma}]C_t^{1/\gamma}\Lambda_t \tag{19}$$

$$\frac{\partial \mathcal{L}}{\partial H_t}: \quad \eta = \Lambda_t(W_t/P_t)(1-H_t) \tag{20}$$

$$\frac{\partial \mathcal{L}}{\partial B_t}: \quad \Lambda_t = \beta R_t E_t\left\{\Lambda_{t+1}\left(\frac{P_{t+1}}{P_t}\right)^{-1}\right\} \tag{21}$$

$$\frac{\partial \mathcal{L}}{\partial M_t}: \quad a_t e_t^{1/\gamma} = [C_t^{(\gamma-1)/\gamma} + e_t^{1/\gamma}(M_t/P_t)^{(\gamma-1)/\gamma}](M_t/P_t)^{1/\gamma}\left(\Lambda_t - \beta E_t\left\{\Lambda_{t+1}\left[\frac{P_{t+1}}{P_t}\right]^{-1}\right\}\right) \tag{22}$$

$$\frac{\partial \mathcal{L}}{\partial K_t}: \quad \Lambda_t\left(\frac{1}{\chi_t} + \frac{\partial \Phi(K_t, K_{t-1})}{\partial K_t}\right) = \beta E_t \Lambda_{t+1}\left\{\frac{Q_{t+1}}{P_{t+1}} + \frac{1-\delta}{\chi_{t+1}} - \frac{\partial \Phi(K_{t+1}, K_t)}{\partial K_t}\right\} \tag{23}$$

$$\frac{\partial \Phi(K_t, K_{t-1})}{\partial K_t} = \frac{\phi_K}{g}\left(\frac{K_t}{gK_{t-1}} - 1\right)K_{t-1}$$

$$\frac{\partial \Phi(K_{t+1}, K_t)}{\partial K_t} = \frac{\phi_K}{2}\left(\frac{K_{t+1}}{gK_t}\right)^2 - \frac{\phi_K}{g}\left(\frac{K_{t+1}}{gK_t} - 1\right)\frac{K_{t+1}}{K_t}$$

$$\frac{\partial \mathcal{L}}{\partial \Lambda_t}: \quad (M_{t-1} + T_t + B_{t-1} + W_t H_t + Q_t K_{t-1} + D_t)P_t^{-1} = \tag{24}$$

$$C_t + (K_t - (1-\delta)K_{t-1})(\chi_t)^{-1} + \Phi(K_t, K_{t-1}) + (B_t R_t^{-1} + M_t)P_t^{-1}$$

Substituting equations (19) and (21) into (22) yielded the money demand equation shown in equation (8). A first-order Taylor approximation of $(1 - R_t^{-1})$ could show that γ represented the interest elasticity of money demand. First, defining $r_t = R_t - 1$ as the net interest rate. Second, expanding $(1-R_t^{-1})$ about the point 1, which yielded: $(1 - R_t^{-1}) \simeq R_t - 1$. Third, taking the log of the money demand equation resulted in equation (25).

$$log(M_t/P_t) = log(C_t) + log(e_t) - \gamma log(r_t) \tag{25}$$

9.1.2 Firm's First-Order Conditions

The four first-order conditions yielded equation (26) to (29), where Ξ_t was the Lagrange multiplier.

$$\frac{\partial \mathcal{L}}{\partial H_t(i)}: \quad \Lambda_t H_t(i)\left(\frac{W_t}{P_t}\right) = \Xi_t(1-\alpha)A_t K_{t-1}(i)^{\alpha}[g^t H_t(i)]^{1-\alpha} \tag{26}$$

$$\frac{\partial \mathcal{L}}{\partial K_{t-1}(i)}: \quad \Lambda_t K_t(i)\left(\frac{Q_t}{P_t}\right) = \Xi_t \alpha A_t K_{t-1}(i)^{\alpha}[g^t H_t(i)]^{1-\alpha} \tag{27}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial P_t(i)}: \quad & \phi_p \Lambda_t\left(\frac{P_t(i)}{\Pi P_{t-1}(i)} - 1\right)\left(\frac{P_t(i)}{\Pi P_{t-1}(i)}\right) = \theta \Xi_t \left(\frac{P_t(i)}{P_t}\right)^{-\theta} + \\ & (1-\theta)\Lambda_t\left(\frac{P_t}{P_t(i)}\right)^{1-\theta} + (\beta\phi_p)E_t\left\{\Lambda_{t+1}\left(\frac{P_{t+1}}{\Pi P_t} - 1\right)\left(\frac{P_{t+1}}{\Pi P_t}\right)\left(\frac{Y_{t+1}}{Y_t}\right)\right\} \end{aligned} \tag{28}$$

$$\frac{\partial \mathcal{L}}{\partial \Xi_t(i)}: \quad A_t K_{t-1}(i)^{\alpha}[g^t H_t(i)]^{1-\alpha} = \left(\frac{P_t(i)}{P_t}\right)Y_t \tag{29}$$

9.1.3 Dynamic System

The combined equations of the model in trending form are presented below. Many of the equations from the FOCs simplify. First, symmetry of intermediate goods-producing firms meant that the prices of firm i simplified to the economy-wide price level: $P_t(i) = P_t$. Second, the symmetry of households meant there was a zero net bond condition: $B_t = 0$.[53] The model had 18 variables: $w_t = \frac{W_t}{P_t}, m_t = \frac{M_t}{P_t}, q_t = \frac{Q_t}{P_t}, d_t = \frac{D_t}{P_t}, \Pi_t = \frac{P_t}{P_{t-1}}, Y_t, I_t, K_t, C_t, H_t, R_t, \Xi_t, \Lambda_t, a_t, v_t, \chi_t, e_t, A_t$ and 18 equations. With the following additional notation: $\varphi_{t,t-1} = \frac{\partial \Phi(K_t,K_{t-1})}{\partial K_t}, \varphi_{t+1,t} = \frac{\partial \Phi(K_{t+1},K_t)}{\partial K_t}$.

[53] Whilst consumers did not end up holding bonds, the FOC had to still be taken with respect to bond purchases as this established the gross nominal interest rate (R_t) needed to ensure the zero net bond condition.

Consumer's FOCs

$$a_t = [C_t^{(\gamma-1)/\gamma} + e_t^{1/\gamma}(M_t/P_t)^{(\gamma-1)/\gamma}]C_t^{1/\gamma}\Lambda_t \tag{30}$$

$$\eta = \Lambda_t w_t (1 - H_t) \tag{31}$$

$$C_t e_t = m_t (1 - R_t^{-1})^{\gamma} \tag{32}$$

$$\Lambda_t = \beta R_t E_t (\Lambda_{t+1} \Pi_{t+1}^{-1}) \tag{33}$$

$$\varphi_{t,t-1} = \frac{\phi_K}{g}\left(\frac{K_t}{gK_{t-1}} - 1\right)$$

$$\varphi_{t+1,t} = \frac{\phi_K}{2}\left(\frac{K_{t+1}}{gK_t} - 1\right)^2 - \frac{\phi_K}{g}\left(\frac{K_{t+1}}{gK_t} - 1\right)\frac{K_{t+1}}{K_t}$$

$$\Lambda_t\left(\frac{1}{\chi_t} + \varphi_{t,t-1}\right) = \beta E_t \Lambda_{t+1}\left\{q_{t+1} + \frac{1-\delta}{\chi_{t+1}} - \varphi_{t+1,t}\right\} \tag{34}$$

Firm's FOCs

$$\Lambda_t H_t w_t = (1 - \alpha)\Xi_t Y_t \tag{35}$$

$$\Lambda_t K_{t-1} q_t = \alpha \Xi_t Y_t \tag{36}$$

$$\phi_p \Lambda_t \left(\frac{\Pi_t}{\Pi} - 1\right)\left(\frac{\Pi_t}{\Pi}\right) = (1-\theta)\Lambda_t + \theta \Xi_t + (\beta\phi_p)E_t\left\{\Lambda_{t+1}\left(\frac{\Pi_{t+1}}{\Pi} - 1\right)\left(\frac{\Pi_{t+1}}{\Pi}\right)\left(\frac{Y_{t+1}}{Y_t}\right)\right\} \tag{37}$$

Technological Constraints and Laws of Motion

$$Y_t = A_t K_{t-1}^{\alpha}[g^t H_t]^{1-\alpha} \tag{38}$$

$$K_t = (1 - \delta)K_{t-1} + \chi_t I_t \tag{39}$$

$$d_t = Y_t - w_t H_t - q_t K_{t-1} - \frac{\phi_p}{2}\left(\frac{\Pi_t}{\Pi} - 1\right)^2 Y_t \tag{40}$$

$$Y_t = C_t + I_t + \Phi(K_t, K_{t-1}) + \frac{\phi_p}{2}\left(\frac{\Pi_t}{\Pi} - 1\right)^2 Y_t \tag{41}$$

$$\omega_R \tilde{R}_t = \omega_\pi \tilde{\Pi}_t + \omega_y \tilde{y}_t + \ln(v_t) \tag{42}$$

Stochastic Processes

$$\ln(a_t) = \rho_a \ln(a_{t-1}) + \varepsilon_t^a \tag{43}$$

$$\ln(\chi_t) = \rho_\chi \ln(\chi_{t-1}) + \varepsilon_t^\chi \tag{44}$$

$$\ln(v_t) = \rho_v \ln(v_{t-1}) + \varepsilon_t^v \tag{45}$$

$$\ln(e_t) = \rho_e \ln(e_{t-1}) + (1-\rho_e)\ln(e) + \varepsilon_t^e \tag{46}$$

$$\ln(A_t) = \rho_A \ln(A_{t-1}) + (1-\rho_A)\ln(A) + \varepsilon_t^A \tag{47}$$

9.1.4 Detrended Dynamic System

In order to find a steady-state, the model needed to be written in its intensive (detrended) form. Each variable which grew along the balanced growth path could be decomposed into its intensive form component and the deterministic growth rate. The following variables were rewritten as: $w_t^d = w_t/g^t, m_t^d = m_t/g^t, d_t^d = d_t/g^t, y_t = Y_t/g^t, i_t = I_t/g^t, k_{t-1} = K_{t-1}/g^t, c_t = C_t/g^t, \xi_t = \Xi_t \times g^t, \lambda_t = \Lambda_t \times g^t$.

Consumer's FOCs

$$a_t = [c_t^{(\gamma-1)/\gamma} + e_t^{1/\gamma}(m_t^d)^{(\gamma-1)/\gamma}]c_t^{1/\gamma}\lambda_t \tag{48}$$

$$\eta = \lambda_t w_t^d (1-H_t) \tag{49}$$

$$c_t e_t = m_t^d (1-R_t^{-1})^\gamma \tag{50}$$

$$g\lambda_t = \beta R_t E_t(\lambda_{t+1}\Pi_{t+1}^{-1}) \tag{51}$$

$$\varphi_{t,t-1} = \frac{\phi_K}{g}\left(\frac{k_t}{k_{t-1}} - 1\right)$$

$$\varphi_{t+1,t} = \frac{\phi_K}{2}\left(\frac{k_{t+1}}{k_t} - 1\right)^2 - \phi_K\left(\frac{k_{t+1}}{k_t} - 1\right)\frac{k_{t+1}}{k_t}$$

$$g\lambda_t\left(\frac{1}{\chi_t} + \varphi_{t,t-1}\right) = \beta E_t \lambda_{t+1}\left\{q_{t+1} + \frac{1-\delta}{\chi_{t+1}} - \varphi_{t+1,t}\right\} \tag{52}$$

Firm's FOCs

$$\lambda_t H_t w_t^d = (1-\alpha)\xi_t y_t \tag{53}$$

$$\lambda_t k_{t-1} q_t = \alpha \xi_t y_t \tag{54}$$

$$\phi_p \lambda_t \left(\frac{\Pi_t}{\Pi} - 1\right)\left(\frac{\Pi_t}{\Pi}\right) = (1-\theta)\lambda_t + \theta\xi_t + (\beta\phi_p) E_t \left\{ \lambda_{t+1} \left(\frac{\Pi_{t+1}}{\Pi} - 1\right)\left(\frac{\Pi_{t+1}}{\Pi}\right)\left(\frac{y_{t+1}}{y_t}\right)\right\} \tag{55}$$

Technological Constraints and Laws of Motion

$$y_t = A_t k_{t-1}^{\alpha} H_t^{1-\alpha} \tag{56}$$

$$g k_t = (1-\delta) k_{t-1} + \chi_t i_t \tag{57}$$

$$d_t^d = y_t - w_t^d H_t - q_t k_{t-1} - \frac{\phi_p}{2}\left(\frac{\Pi_t}{\Pi} - 1\right)^2 y_t \tag{58}$$

$$\frac{\Phi(K_t, K_{t-1})}{g^t} = \frac{\phi_K}{2}\left(\frac{k_t}{k_{t-1}} - 1\right)^2 k_{t-1}$$

$$y_t = c_t + i_t + \frac{\Phi(K_t, K_{t-1})}{g^t} + \frac{\phi_p}{2}\left(\frac{\Pi_t}{\Pi} - 1\right)^2 y_t \tag{59}$$

$$\omega_R \tilde{R}_t = \omega_\pi \tilde{\Pi}_t + \omega_y \tilde{y}_t + \ln(v_t) \tag{60}$$

9.1.5 Steady-State

The steady-state of the intensive form variables of the model could be found by dropping the time subscripts and solving the system of equations.

$$a_t = a = 1 \tag{61}$$

$$\chi_t = \chi = 1 \tag{62}$$

$$v_t = v = 1 \tag{63}$$

$$e_t = e \tag{64}$$

$$A_t = A \tag{65}$$

$$\Pi_t = \Pi \tag{66}$$

$$R_t = R = \Pi g/\beta \tag{67}$$

$$q_t = q = \beta/g - 1 + \delta \tag{68}$$

$$\xi_t = \xi = \left(\frac{\theta - 1}{\theta}\right)\lambda \tag{69}$$

$$m_t^d = m^d = e\left(\frac{R}{R-1}\right)^{\gamma} c \tag{70}$$

$$c_t = c = \left[1 + e\left(\frac{R}{R-1}\right)^{\gamma-1}\right]^{-1} \frac{1}{\lambda} \tag{71}$$

$$y = c + i \tag{72}$$

$$k_t = k = \left(\frac{\theta - 1}{\theta}\right)\left(\frac{\alpha y}{q}\right) \tag{73}$$

$$i_t = i = (g - 1 + \delta)k \tag{74}$$

$$y_t = y = c\left[1 - (g - 1 + \delta)\left(\frac{\theta - 1}{\theta}\right)\left(\frac{\alpha}{q}\right)\right]^{-1} \tag{75}$$

$$d_t^d = d = y - w^d H - qk \tag{76}$$

$$H_t = H = \left(\frac{y}{Ak^{\alpha}}\right)^{\frac{1}{1-\alpha}} \tag{77}$$

$$w_t^d = w^d = (1 - \alpha)\left(\frac{\theta - 1}{\theta}\right)\left(\frac{y}{h}\right) \tag{78}$$

$$w_t^d = w^d = (1 - \alpha)\left[A\left(\frac{\alpha}{q} \cdot \frac{\theta - 1}{\theta}\right)^{\alpha}\right]^{\frac{1}{1-\alpha}} \tag{79}$$

$$w^d H = c \cdot (1 - \alpha)\left[\frac{\theta}{\theta - 1} - \frac{\alpha}{q} \cdot (g - 1 + \delta)\right]^{-1} \tag{80}$$

$$\lambda_t = \lambda = \frac{\eta + (1 - \alpha)\left[1 + e\left(\frac{R}{R-1}\right)^{\gamma-1}\right]^{-1}\left[\left(\frac{\theta}{\theta-1}\right) - \frac{\alpha}{q}(g - 1 + \delta)\right]^{-1}}{(1 - \alpha)\left(\frac{\theta-1}{\theta}\right)^{\frac{1}{1-\alpha}}\left(\frac{\alpha A}{q}\right)^{\frac{\alpha}{1-\alpha}}} \tag{81}$$

9.1.6 Log-Linearized System

Lastly, a first-order Taylor expansion of the log of each equation about the steady-state of the variables was performed on the system of equations detailed in Subsection 9.1.4. Steady-state variables are denoted without time subscripts and tildes denote percentage deviations ($\tilde{z}_t = \%\Delta z_t = (z_t - z)/z$). In equation (82) the term $(a - c\lambda)$ was replaced by $\lambda m(R-1)R^{-1}$ which was derived by substituting the money demand equation into the first-order condition for consumption.

$$aR\gamma\tilde{a}_t = aR\gamma\tilde{\lambda}_t + (R-1)(\gamma-1)\lambda m\tilde{m}_t + (R-1)\lambda m\tilde{e}_t + R[c\lambda(\gamma-1)+a]\tilde{c}_t \tag{82}$$

$$0 = \eta\tilde{\lambda}_t + \eta\tilde{w}_t^d - \lambda w H\tilde{H}_t \tag{83}$$

$$0 = (R-1)\tilde{m}_t^d + \gamma\tilde{R}_t - (R-1)\tilde{c}_t - (R-1)\tilde{e}_t \tag{84}$$

$$\tilde{\lambda}_t = \tilde{R}_t + E_t\tilde{\lambda}_{t+1} - E_t\tilde{\Pi}_{t+1} \tag{85}$$

$$g\tilde{\lambda}_t = \phi_k\tilde{k}_{t-1} + g\tilde{\chi}_t + gE_t\tilde{\lambda}_{t+1} + (\beta q)E_t\tilde{q}_{t+1} - \beta(1-\delta)\tilde{\chi}_{t+1} + (\beta\phi_k)E_t\tilde{k}_{t+1} - \phi_k(1+\beta)\tilde{k}_t \tag{86}$$

$$\tilde{y}_t = \tilde{\lambda}_t + \tilde{H}_t + \tilde{w}_t^d - \tilde{\xi}_t \tag{87}$$

$$\tilde{y}_t = \tilde{\lambda}_t + \tilde{k}_{t-1} + \tilde{q}_t - \tilde{\xi}_t \tag{88}$$

$$\phi_p\tilde{\Pi}_t = (1-\theta)\tilde{\lambda}_t + (\theta-1)\tilde{\xi}_t + (\phi_p\beta)E_t\tilde{\Pi}_{t+1} \tag{89}$$

$$\tilde{y}_t = \tilde{A}_t + \alpha\tilde{k}_{t-1} + (1-\alpha)\tilde{H}_t \tag{90}$$

$$gk\tilde{k}_t = (1-\delta)k\tilde{k}_{t-1} + i\tilde{\chi}_t + i\tilde{i}_t \tag{91}$$

$$d\tilde{d}_t^d = y\tilde{y}_t - wH\tilde{H}_t - wH\tilde{w}_t^d - qk\tilde{q}_t - qk\tilde{k}_{t-1} \tag{92}$$

$$y\tilde{y}_t = c\tilde{c}_t + i\tilde{i}_t \tag{93}$$

$$\omega_R\tilde{R}_t = \omega_\pi\tilde{\Pi}_t + \omega_y\tilde{y}_t + \tilde{v}_t \tag{94}$$

$$\tilde{a}_t = \rho_a\tilde{a}_{t-1} + \varepsilon_t^a \tag{95}$$

$$\tilde{\chi}_t = \rho_\chi\tilde{\chi}_{t-1} + \varepsilon_t^\chi \tag{96}$$

$$\tilde{v}_t = \rho_v\tilde{v}_{t-1} + \varepsilon_t^v \tag{97}$$

$$\tilde{e}_t = \rho_e \tilde{e}_{t-1} + \varepsilon_t^e \tag{98}$$

$$\tilde{A}_t = \rho_A \tilde{A}_{t-1} + \varepsilon_t^A \tag{99}$$

9.2 Structural Econometrics

In this section, the details of how DSGE models are written in a state-space representation and estimated are presented in four further subsections. In Subsection 9.2.1, the equations of the state-space model are detailed. In Subsection 9.2.2, the state-space observer form of the model is displayed. Lastly, Subsections 9.2.3 and 9.2.4 discuss the estimation of the ML and Bayesian state-space systems with the relevant algorithms and computational details.

9.2.1 State-Space Representation

The variables of the model were organized into three categories: (i) state variables - $\boldsymbol{s_t}$, (ii) idiosyncratic error terms - $\boldsymbol{\varepsilon_t}$, and (iii) all other variables - $\boldsymbol{f_t}$.

$$\boldsymbol{s_t} = \begin{bmatrix} \tilde{k}_t & \tilde{a}_t & \tilde{\chi}_t & \tilde{v}_t & \tilde{e}_t & \tilde{A}_t \end{bmatrix}$$

$$\boldsymbol{\varepsilon_t} = \begin{bmatrix} \varepsilon_t^a & \varepsilon_t^\chi & \varepsilon_t^v & \varepsilon_t^e & \varepsilon_t^A \end{bmatrix}$$

$$\boldsymbol{f_t} = \begin{bmatrix} \tilde{c}_t & \tilde{m}_t & \tilde{w}_t^d & \tilde{d}_t^d & \tilde{H}_t & \tilde{R}_t & \tilde{\Pi}_t & \tilde{q}_t & \tilde{y}_t & \tilde{\lambda}_t & \tilde{\xi}_t & \tilde{i}_t \end{bmatrix}$$

In order to estimate the parameters of the model, the log-linearized system was written in state-space notation as shown in equations (100) and (101). The vector $\boldsymbol{s_t}$ contained the capital stock and exogenous variables. The vector $\boldsymbol{\varepsilon_t}$ contained the idiosyncratic error terms. Lastly, the vector $\boldsymbol{f_t}$ contained all the other variables of the model.

$$\boldsymbol{s_t} = \boldsymbol{A}\boldsymbol{s_{t-1}} + \boldsymbol{B}\boldsymbol{\varepsilon_t} \tag{100}$$

$$\boldsymbol{f_t} = \boldsymbol{C}\boldsymbol{s_t} \tag{101}$$

The matrices $\boldsymbol{A}(\boldsymbol{\Omega}), \boldsymbol{B}(\boldsymbol{\Omega}), \boldsymbol{C}(\boldsymbol{\Omega})$ were of the appropriate dimensions and had elements that were non-linear combinations of the 25 deep parameters of the model.

$$\boldsymbol{\Omega} = [\Pi, g, \gamma, \beta, \theta, \alpha, \delta, \eta, \phi_k, \phi_\Pi, \rho_a, \sigma_a, \rho_A, \sigma_A, A, \rho_\chi, \sigma_\chi, \rho_v, \sigma_v, \rho_e, \sigma_e, e, \omega_R, \omega_\Pi, \omega_y]'$$

9.2.2 Observation Equation

The 5×1 vector of the observation data, $\boldsymbol{J}_t^{dat}$, contained the log of the five Canadian data series. After applying the one-sided HP filter, the vector could be directly linked to the observation equation from (101).

$$\begin{aligned}
\boldsymbol{J}_t^{dat} &= \begin{bmatrix} \ln C_t^{dat} & \ln I_t^{dat} & \ln \frac{M_t^{dat}}{P_t^{dat}} & \ln R_t^{dat} & \ln \Pi_t^{dat} \end{bmatrix}' \\
\tilde{\boldsymbol{J}}_t^{dat} &= \begin{bmatrix} \tilde{c}_t^{dat} & \tilde{i}_t^{dat} & \tilde{m}_t^{dat} & \tilde{R}_t^{dat} & \tilde{\Pi}_t^{dat} \end{bmatrix}' \\
\tilde{\boldsymbol{J}}_t^{dat} &= \boldsymbol{D}\boldsymbol{f}_t
\end{aligned}$$

Where,

$$\boldsymbol{D} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

This yielded a new state-space observer system shown in equations (102) and (103) which only differed qualitatively from equations (100) and (101) as the observation equation contained a vector of real data as opposed to the flow variables of the model.

$$\boldsymbol{s_t} = \boldsymbol{A}\boldsymbol{s_{t-1}} + \boldsymbol{B}\boldsymbol{\varepsilon_t} \tag{102}$$

$$\tilde{\boldsymbol{J}}_t^{dat} = \boldsymbol{F}\boldsymbol{s}_t \tag{103}$$

$$\boldsymbol{F} = \boldsymbol{D}\boldsymbol{C}$$

9.2.3 Maximum Likelihood Estimation

Define the likelihood function as: $\mathcal{L}(\boldsymbol{\Omega}|\tilde{\boldsymbol{J}}_t^{dat}, \mathcal{M}) = p(\tilde{\boldsymbol{J}}_t^{dat}|\boldsymbol{\Omega}, \mathcal{M})$. Where $\mathcal{M}$ and $p(\cdot)$ denote the structure and probability distribution of the model, which comes from the state-space observer system described in equations (102) and (103). As the shocks which drove this system ($\boldsymbol{\varepsilon}_t$) were assumed to be *i.i.d* then the likelihood function could be written in the following recursive form.

$$p(\tilde{\boldsymbol{J}}_{\boldsymbol{T}}^{dat}|\boldsymbol{\Omega}, \mathcal{M}) = \prod_{t=1}^{N} p(\tilde{\boldsymbol{J}}_t^{dat}|\tilde{\boldsymbol{J}}_{\boldsymbol{t-1}}^{dat}, \boldsymbol{\Omega}, \mathcal{M})$$

Where $\tilde{\boldsymbol{J}}_{\boldsymbol{T}}^{dat}$ is the $T \times 5$ matrix of actual data observations. The ML estimate was a point estimate ($\hat{\boldsymbol{\Omega}}$) which maximized the objective function.

$$\underset{\hat{\boldsymbol{\Omega}}}{\text{argmax}} \quad \mathcal{L}(\hat{\boldsymbol{\Omega}}|\tilde{\boldsymbol{J}}_t^{dat}, \mathcal{M})$$

While this function was not trivial to optimize, Dynare provided a suite of algorithms (implemented through Matlab) to do so. Two algorithms were used and confirmed the same result.[54] The Nelder-Mead Simplex algorithm (see Lagarias et al. (1998)) was initially used to estimate the parameters. A function tolerance of 1.0×10^{-8} was set and the algorithm achieved convergence after 12,625 iterations. The model was run a second time using a modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm developed by Chris Sims and achieved convergence to the same point.[55]

9.2.4 Bayesian Estimation

Denote the prior of the model as: $p(\boldsymbol{\Omega}|\mathcal{M})$. The probability density function $p(\cdot)$ was a combination of the multivariate prior distributions (normal, gamma, inverse gamma, and beta) found in Table 3. By Bayes rule, the posterior distribution could be written as,

$$p(\boldsymbol{\Omega}|\tilde{\boldsymbol{J}}_{\boldsymbol{T}}^{dat}, \mathcal{M}) = \frac{\mathcal{L}(\boldsymbol{\Omega}|\tilde{\boldsymbol{J}}_t^{dat}, \mathcal{M}) \times p(\boldsymbol{\Omega}|\mathcal{M})}{p(\tilde{\boldsymbol{J}}_{\boldsymbol{T}}^{dat}|\mathcal{M})} \tag{104}$$

Where the numerator was the likelihood function times the prior and the denominator was the marginal density of the data conditional on the model

[54]Estimation time was about ten minutes.
[55]For more information see http://sims.princeton.edu/yftp/optimize/.

structure. The marginal density ensured that the posterior density integrated to one. However, it did not need to be considered when finding the vector of parameters which maximized the median (or any moment) of the posterior probability distribution. Only the kernel of the posterior distribution needed to be simulated.[56]

$$
\begin{aligned}
p(\boldsymbol{\Omega}|\tilde{\boldsymbol{J}}_{\boldsymbol{T}}^{dat},\mathcal{M}) &\propto \mathcal{L}(\boldsymbol{\Omega}|\tilde{\boldsymbol{J}}_{t}^{dat},\mathcal{M}) \times p(\boldsymbol{\Omega}|\mathcal{M}) \\
&\propto \mathcal{K}(\boldsymbol{\Omega}|\tilde{\boldsymbol{J}}_{t}^{dat},\mathcal{M})
\end{aligned}
$$

The posterior density function $\mathcal{K}(\cdot)$ was not estimable by either the simplex or BFGS algorithm that yielded convergence in the ML-DSGE model. Instead, a Monte-Carlo based optimization routine was used.[57] The Metropolis-Hastings algorithm was designed to simulate the (initially) unknown and highly complex posterior distribution. The four steps of the algorithm are qualitatively described below.

1. A starting point $\boldsymbol{\Omega_{t-1}}$ (the vector of parameters of the model) was chosen.

2. A nearby point was randomly sampled from a 'jumping distribution': $\boldsymbol{\Omega^*} \sim \mathcal{N}(\boldsymbol{\Omega}_{t-1}, j\boldsymbol{\Sigma})$.

3. The ratio of the likelihoods (the 'acceptance ratio') was calculated: $R = \mathcal{K}(\boldsymbol{\Omega}^*, \boldsymbol{Y}_t)/\mathcal{K}(\boldsymbol{\Omega}_{t-1}, \boldsymbol{Y}_t)$.

4. The sampled $\boldsymbol{\Omega^*}$ either became the new starting point or the process was

[56]In the equation below, $\propto$ means 'proportional to'.

[57]See `http://www.dynare.org/DynareWiki/MonteCarloOptimization` for the details.

repeated based on the following rule:

$$\mathbf{\Omega}_t = \begin{cases} \mathbf{\Omega}^* & \text{with probability min}\{\text{R,1}\} \\ \mathbf{\Omega}_{t-1} & \text{otherwise.} \end{cases}$$

The goal of this algorithm was to sample as much of the probability space domain as possible while developing a simulated density. The parameter j, seen in step 2, scaled the covariance matrix ($\mathbf{\Sigma}$) from the jumping distribution used to draw a sample for the next candidate point. If j was too large then the acceptance rate (R) would be too low and the model would not converge quickly enough to the regions of heavier density. If j was too small, the acceptance rate would be too high, and the algorithm may have converged to a local maxima. The consensus was that the average acceptance rate should be between 25-33%.[58]

The algorithm was specified for 20,000 draws, two parallel chains, a 50% burn-in rate, and a $j = 0.2$. The two chains averaged a 22% acceptance rate. The process was then repeated 44 times for each period in the forecasting exercise.[59]

9.3 Diebold-Mariano Test

Consider two series of k-step-ahead forecasting errors $\{e_{1t}\}_{t=1}^{n}$ and $\{e_{2t}\}_{t=1}^{n}$. Define $d_t = e_{1t}^2 - e_{2t}^2$. Therefore d_t is simply the difference in MSEs between the two series. The original Diebold and Mariano (DM) test (see Diebold and Mariano (1995)) was based on the null hypothesis that the expected difference in forecast errors was zero: $E[d_t] = 0$. Assuming this null hypothesis, it can

[58] See the Dynare Reference Manual: `www.dynare.org/wp-repo/dynarewp001.pdf`.

[59] The computational process was not trivial and took about 275 hours of computing time. However, the algorithm may take substantially less time to implement depending on the computing facilities available.

be shown that,

$$S_{DM} = [\hat{V}(\bar{d})]^{-1/2}\bar{d} \sim \mathcal{N}(0,1) \tag{105}$$

Where $\bar{d}$ is the sample mean of the series d_t, $\hat{V}(\bar{d})$ is the estimate of the variance of the sample mean, and S_{DM} is the DM test statistic. A slightly modified version of the DM test can be found in Harvey et al. (1997). The authors proposed two adjustments. First, they changed the test statistic by using the exact error term, rather than its approximation. Second, they used the Student-t, rather than a normal, distribution to best approximate their test statistic. This paper used the modified test statistic developed by Harvey, Leybourne, and Newbold.

$$S_{HLN} = \left[\frac{n + 1 - 2k + k(k-1)/n}{n}\right]^{1/2} S_{DM} \sim t(n-1) \tag{106}$$

9.4 Data Sources

The following is a list of data sources used and the relevant CANSIM table and series numbers. A description of how the data was imported, handled, and transformed is in the R code found in the Dropbox link.

(1) Output (Y_t): Gross domestic product at 2007 constant prices, expenditure-based, quarterly, seasonally adjusted at annual rates (CANSIM Table 3800084)

   (a) Household final consumption expenditure (V62306859)

   (b) Business gross fixed capital formation (V62306871)

(c) Investment in inventories (V62306882)

(2) Consumption (C_t): Household final consumption expenditure at 2007 constant prices (CANSIM Table 3800084 - Series V62306859)

(3) Money supply (M_t): M2 (gross) (CANSIM Table 1760025 - Series V41552796)

(4) Interest rates (R_t): 3 Month Treasury bills (V122531)

(5) Price index (P_t): Implicit price indexes (CANSIM Table 3800066)

(a) Implicit price indexes (2007=100); Household final consumption expenditure (V62307259)

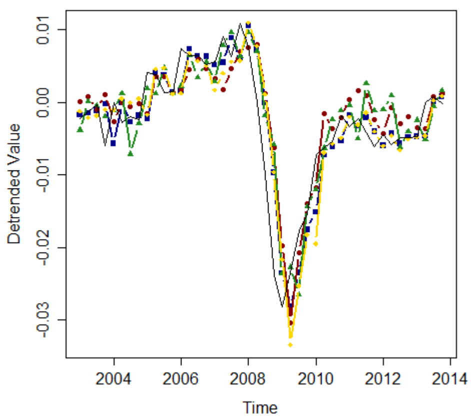
(b) Implicit price indexes (2007=100); Business gross fixed capital formation (V62307259)

(6) Population: Quarterly Canadian population estimates (CANSIM Table 051-0005)

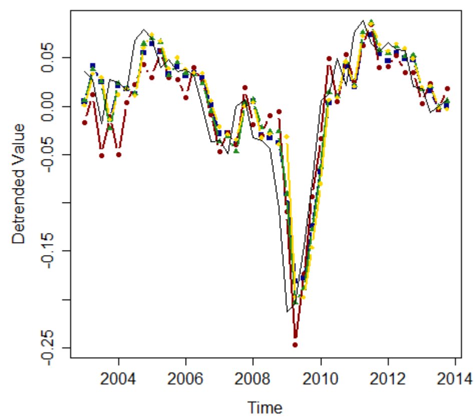
9.5 Figures

Figure 15: One-step-ahead forecasts

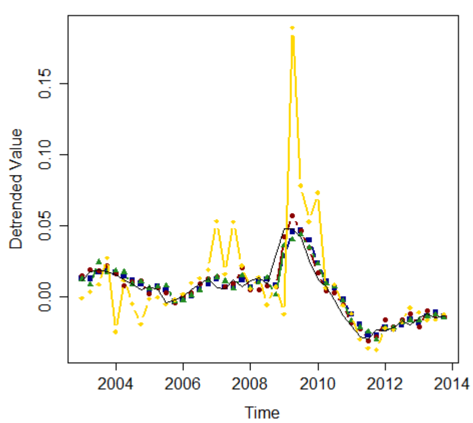
(a) Consumption



(b) Investment



(c) Money-balances



(d) Legend

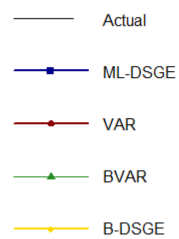
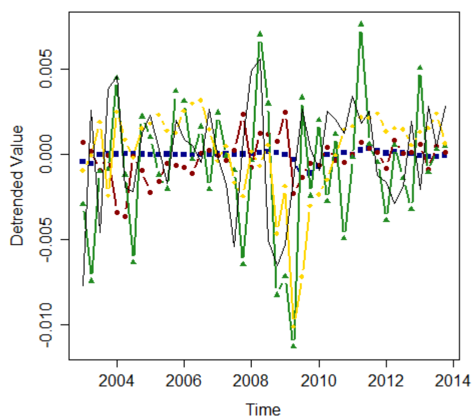
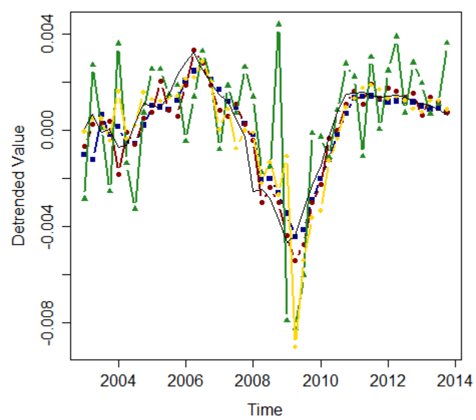


Figure 15: One-step-ahead forecasts cont'd

(e) Inflation



(f) Interest rates



(g) Legend

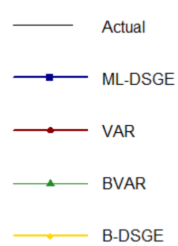
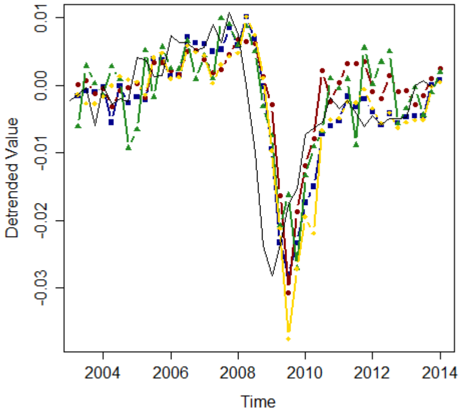
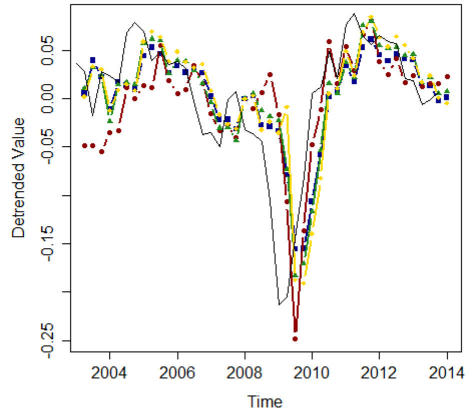


Figure 16: Two-step-ahead forecasts

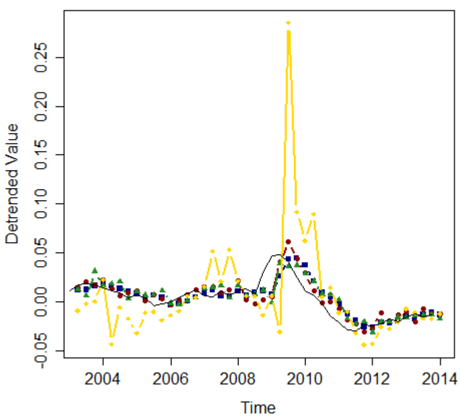
(a) Consumption



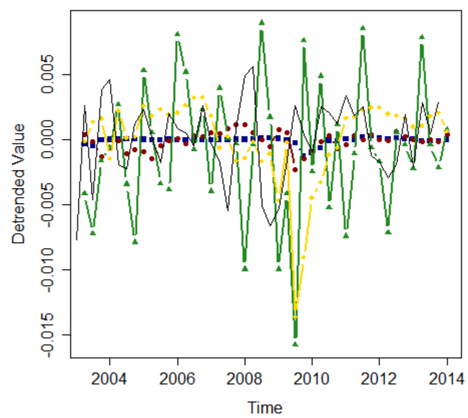
(b) Investment



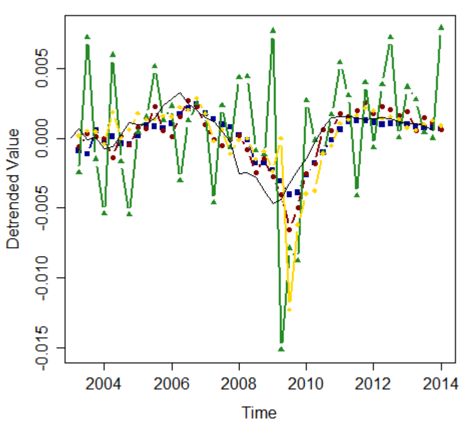
(c) Money-balances



(d) Inflation



(e) Interest rates



(f) Legend

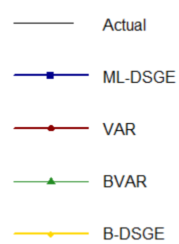
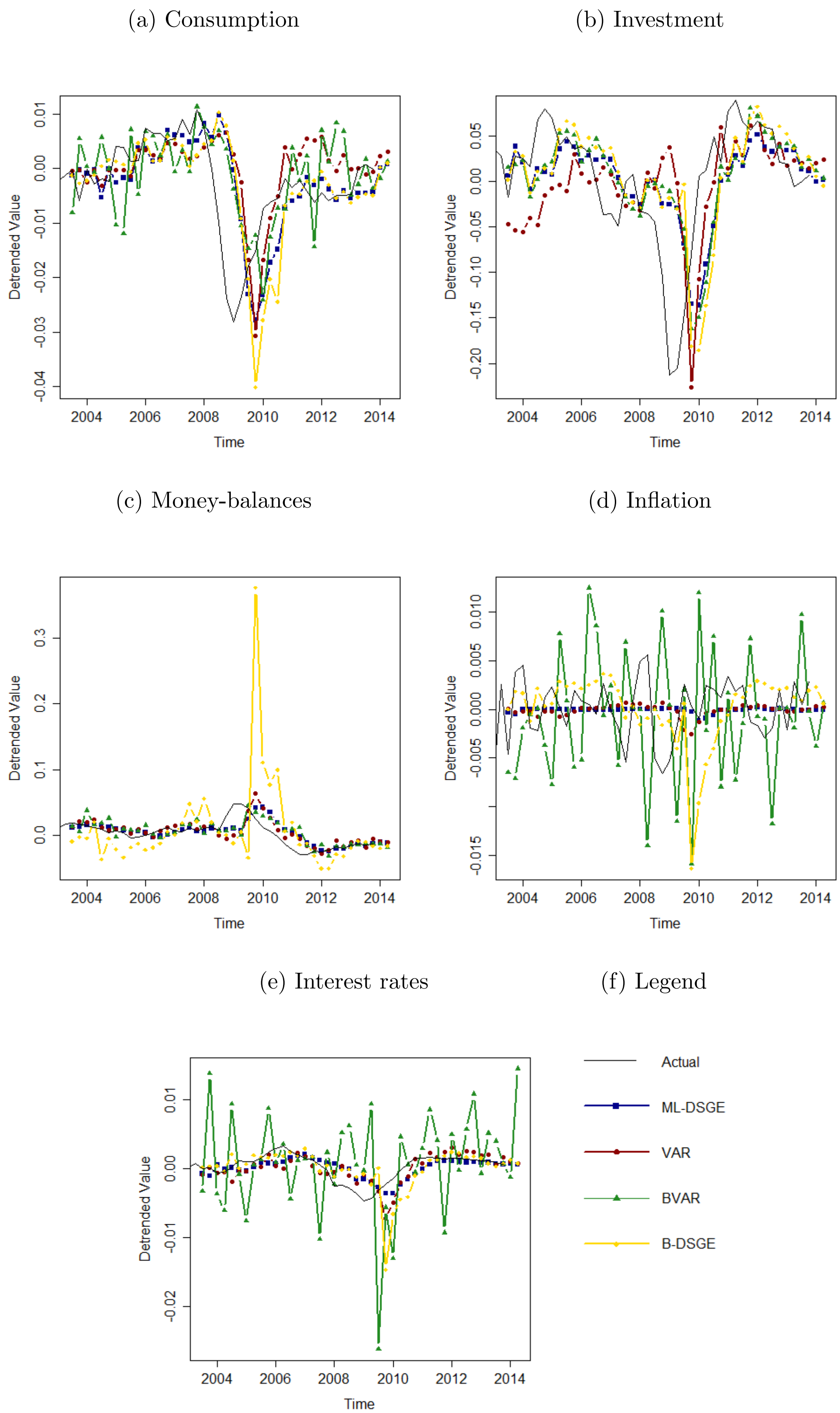


Figure 17: Three-step-ahead forecasts



(a) Consumption

(b) Investment

(c) Money-balances

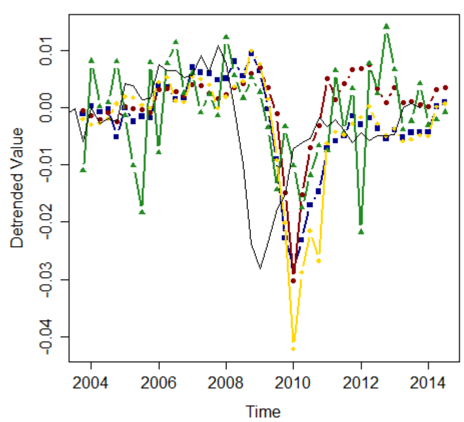
(d) Inflation

(e) Interest rates

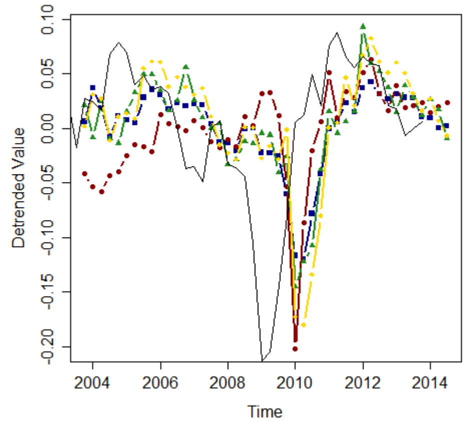
(f) Legend

Figure 18: Four step-ahead forecasts

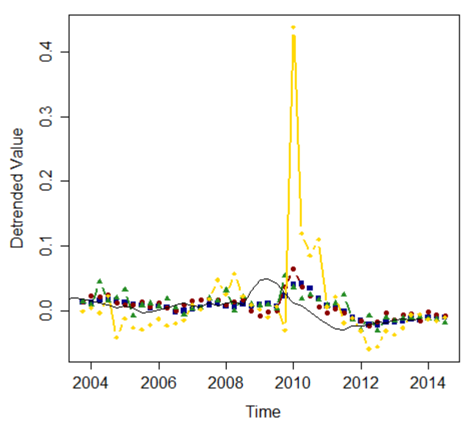
(a) Consumption



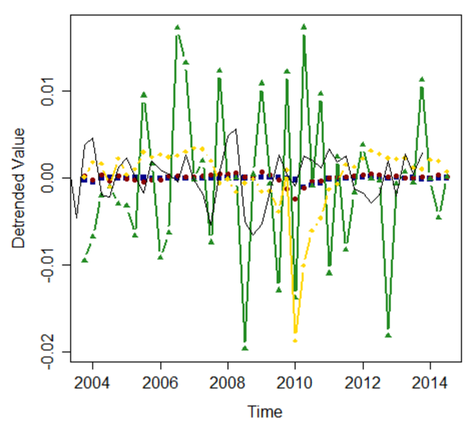
(b) Investment



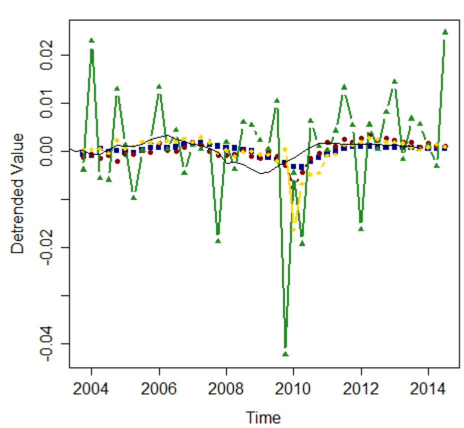
(c) Money-balances



(d) Inflation



(e) Interest rates



(f) Legend

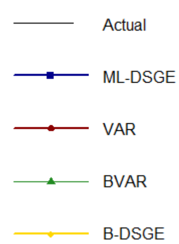
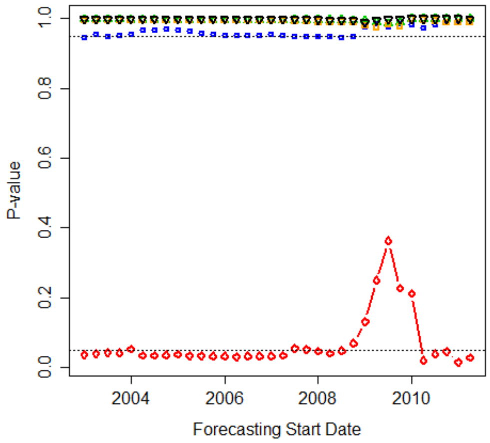
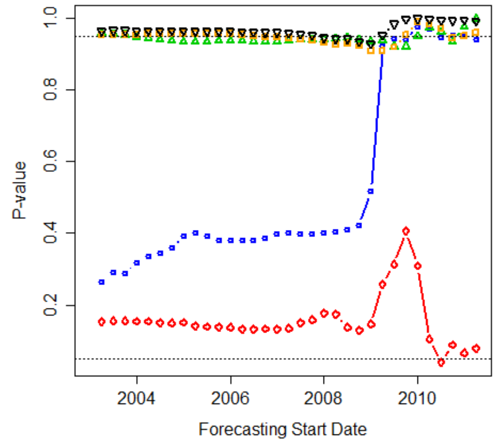


Figure 19: ML-DSGE versus VAR - Dynamic DM test results

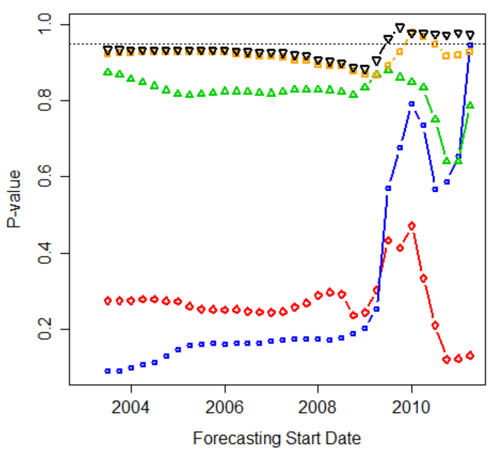
(a) One-step-ahead



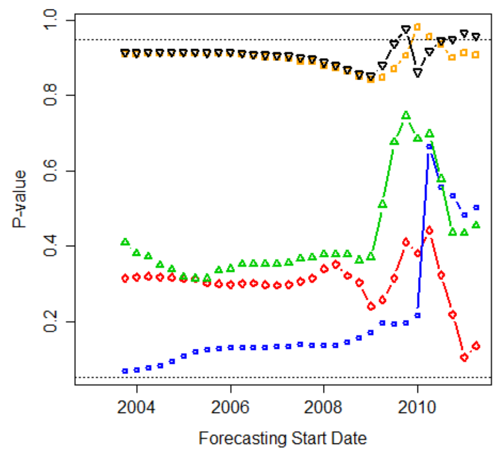
(b) Two-step-ahead



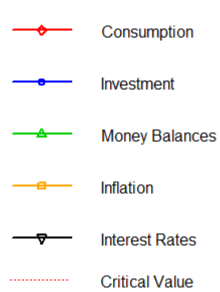
(c) Three-step-ahead



(d) Four-step-ahead



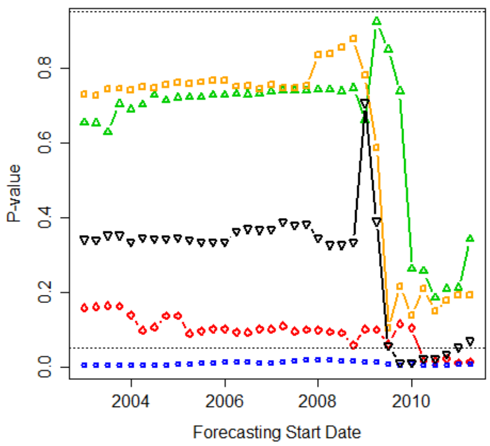
(e) Legend



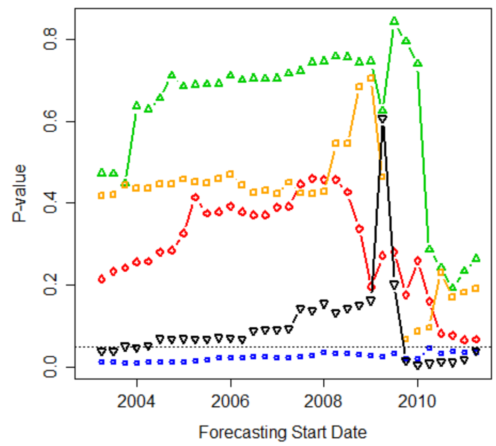
P-values below (above) 0.05 (0.95) denote a statistically significant result for the ML-DSGE (VAR) model.

Figure 20: ML-DSGE versus BVAR - Dynamic DM test results

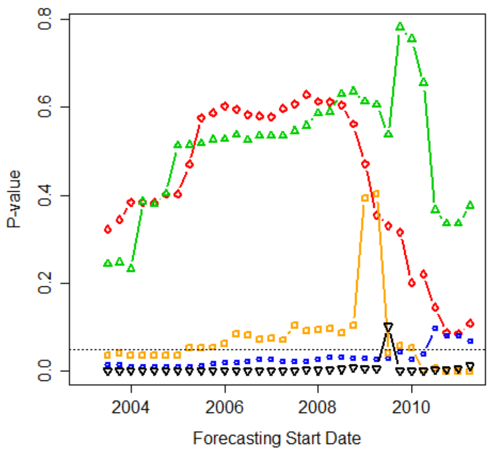
(a) One-step-ahead



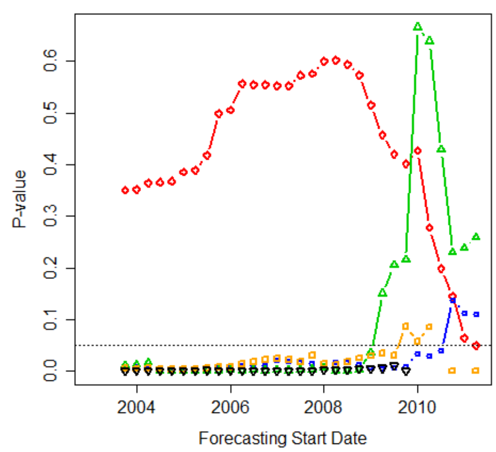
(b) Two-step-ahead



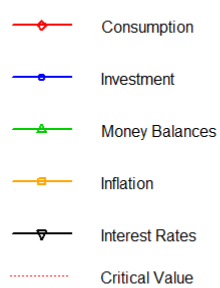
(c) Three-step-ahead



(d) Four-step-ahead



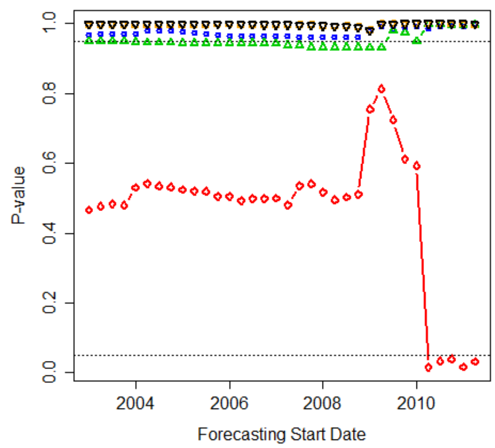
(e) Legend



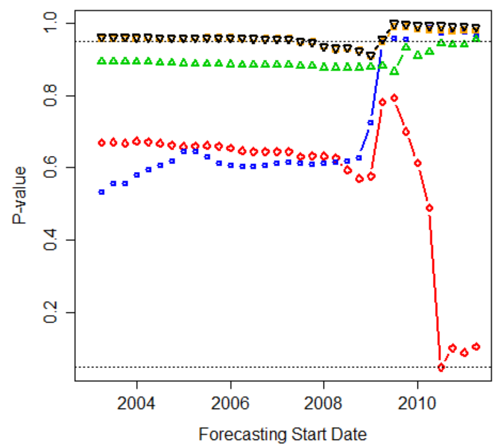
P-values below (above) 0.05 (0.95) denote a statistically significant result for the ML-DSGE (BVAR) model.

Figure 21: Bayesian DSGE versus VAR - Dynamic DM test results

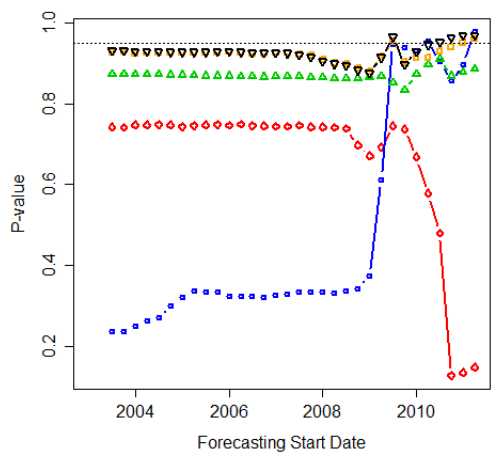
(a) One-step-ahead



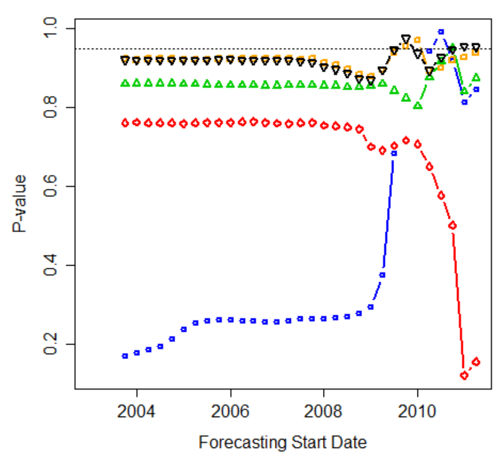
(b) Two-step-ahead



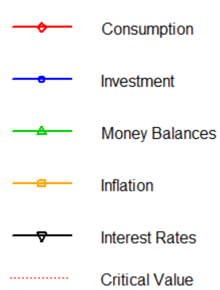
(c) Three-step-ahead



(d) Four-step-ahead



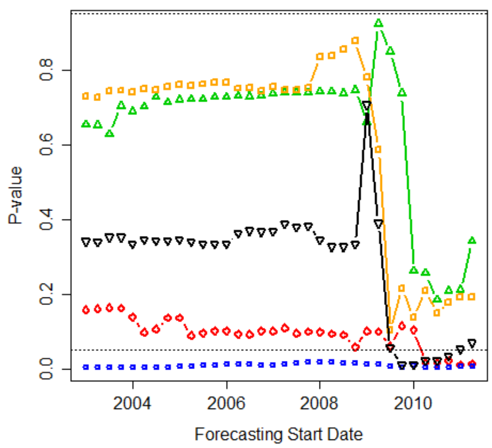
(e) Legend



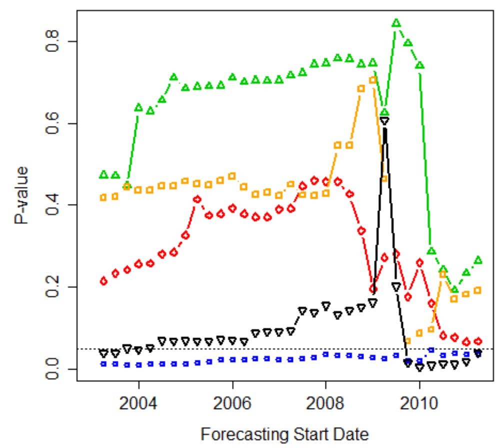
P-values below (above) 0.05 (0.95) denote a statistically significant result for the Bayesian DSGE (VAR) model.

Figure 22: Bayesian DSGE versus BVAR - Dynamic DM test results

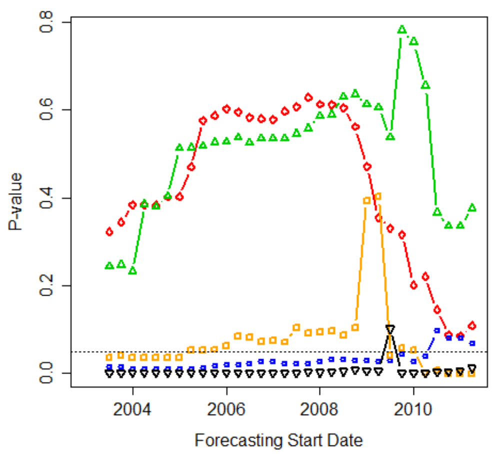
(a) One-step-ahead



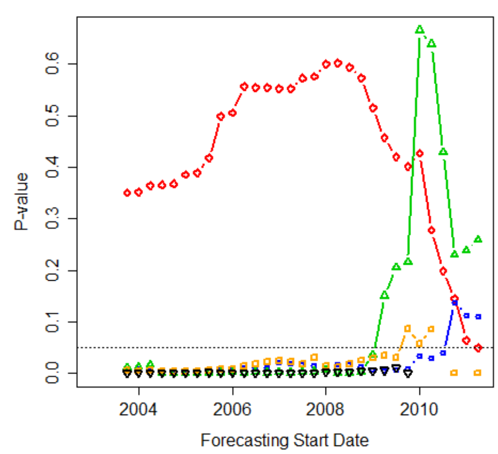
(b) Two-step-ahead



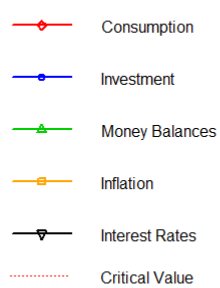
(c) Three-step-ahead



(d) Four-step-ahead



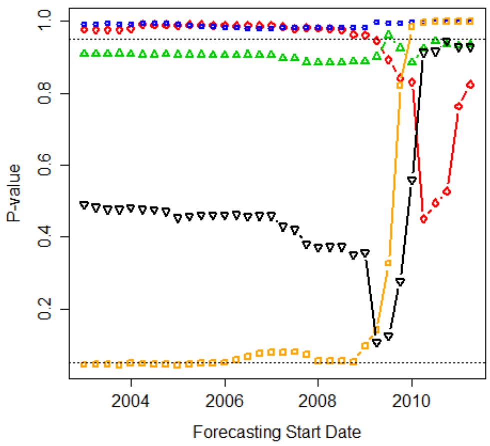
(e) Legend



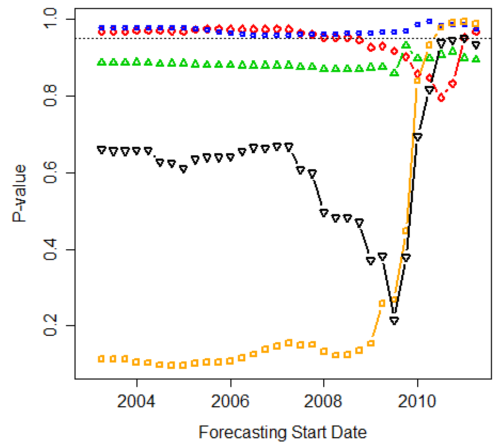
P-values below (above) 0.05 (0.95) denote a statistically significant result for the Bayesian DSGE (BVAR) model.

Figure 23: Bayesian DSGE versus ML-DSGE - Dynamic DM test results

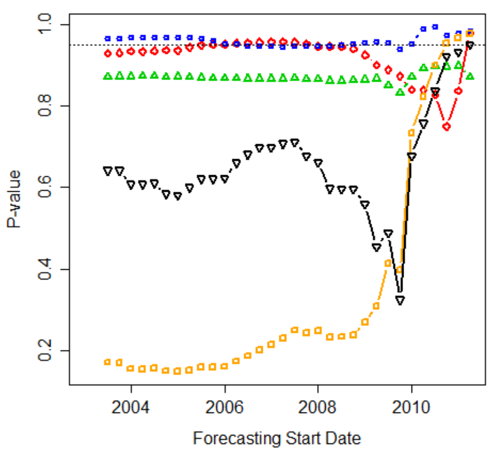
(a) One-step-ahead



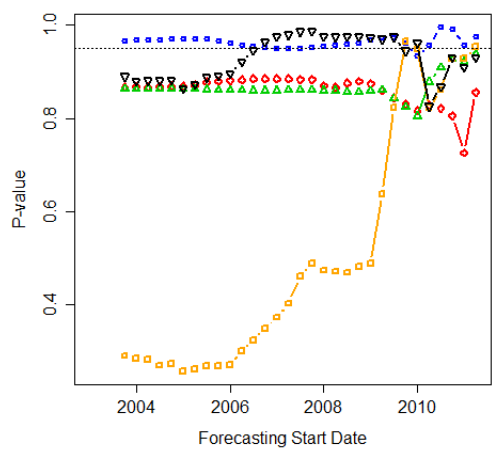
(b) Two-step-ahead



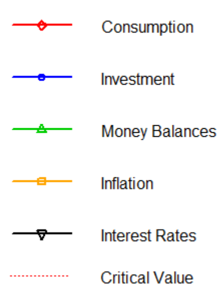
(c) Three-step-ahead



(d) Four-step-ahead



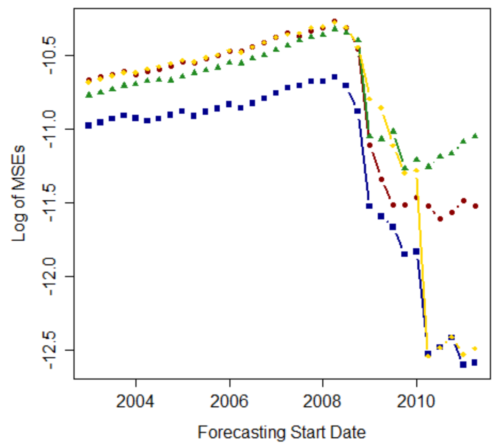
(e) Legend



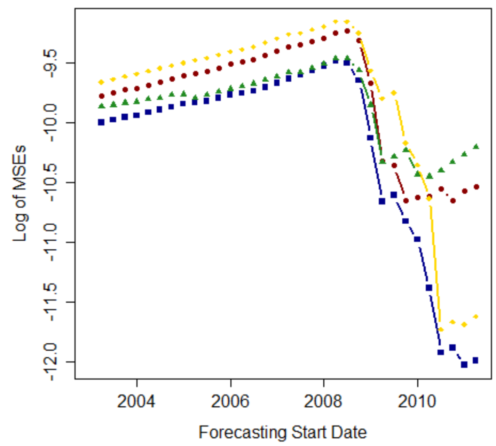
P-values below (above) 0.05 (0.95) denote a statistically significant result for the Bayesian DSGE (ML-DSGE) model.

Figure 24: Consumption - Dynamic (log) MSE results (k denotes the forecast horizon)

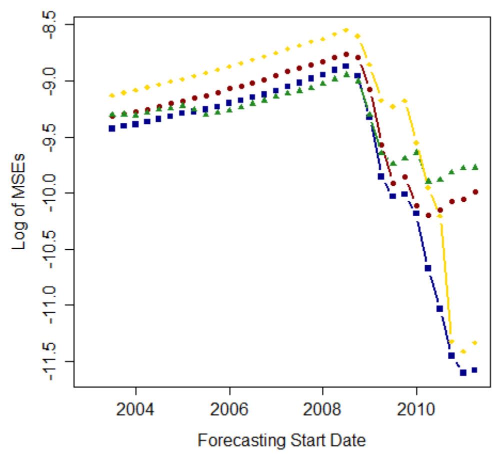
(a) k=1



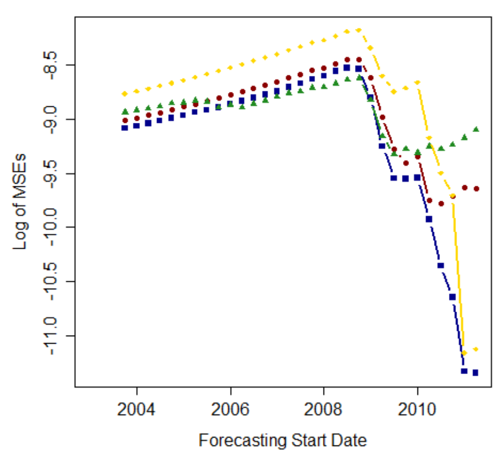
(b) k=2



(c) k=3



(d) k=4



(e) Legend

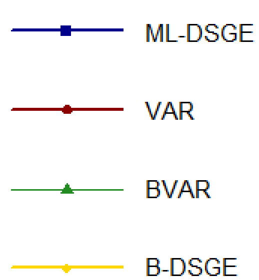
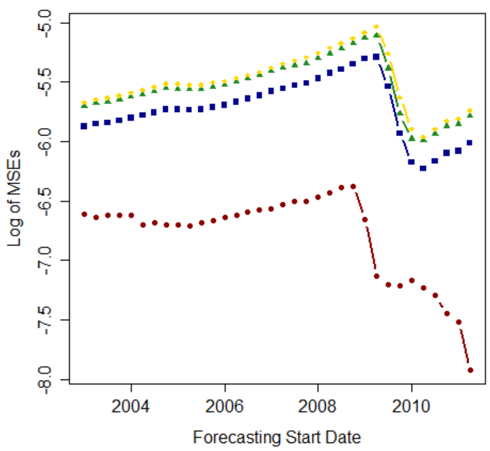
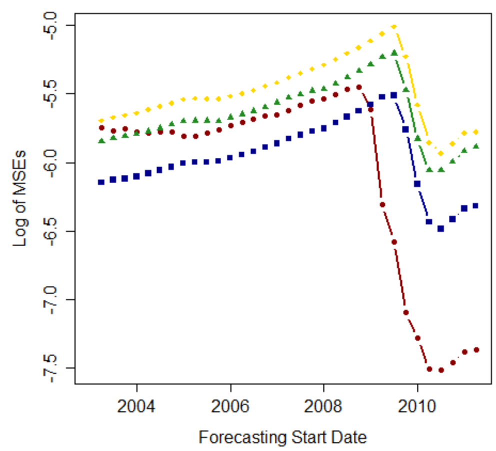


Figure 25: Investment - Dynamic (log) MSE results (k denotes the forecast horizon)

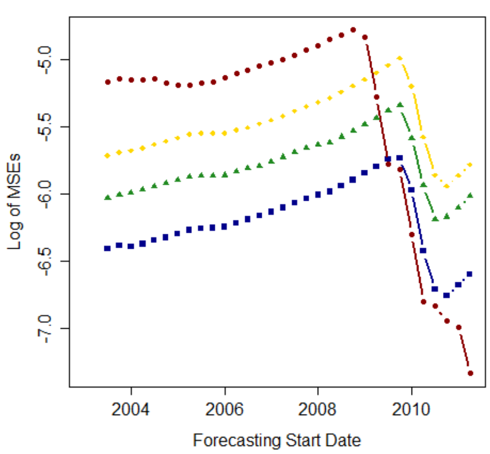
(a) k=1



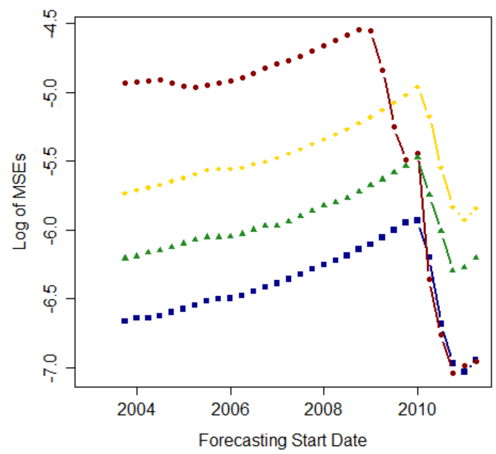
(b) k=2



(c) k=3



(d) k=4



(e) Legend

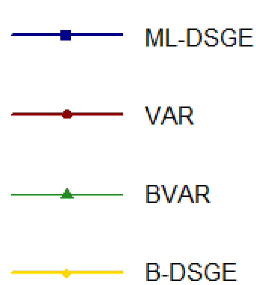
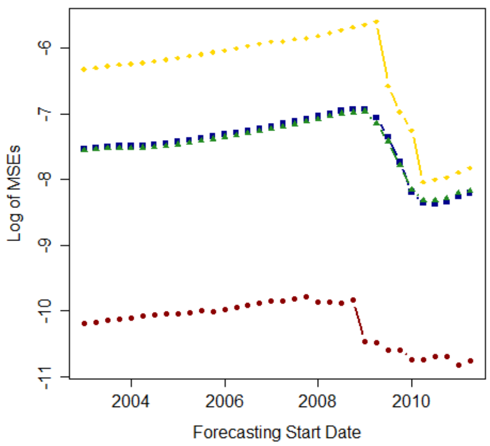
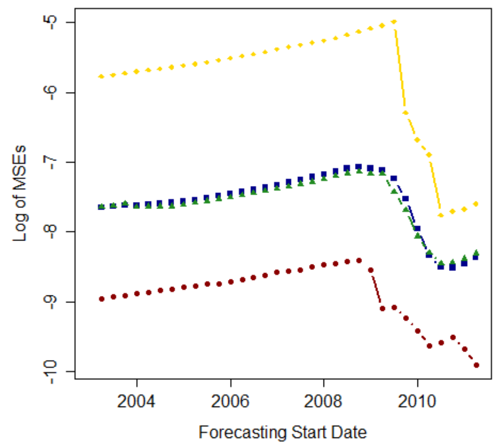


Figure 26: Money-balances - Dynamic (log) MSE results (k denotes the forecast horizon)

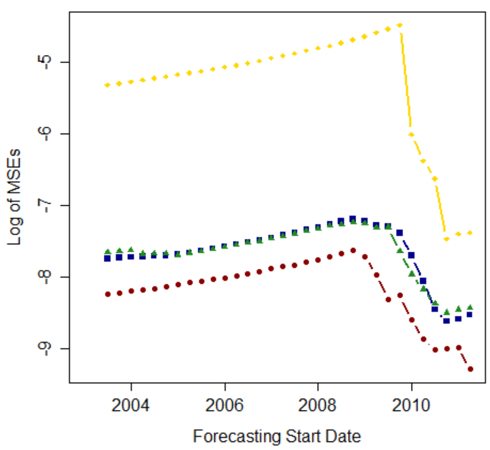
(a) k=1



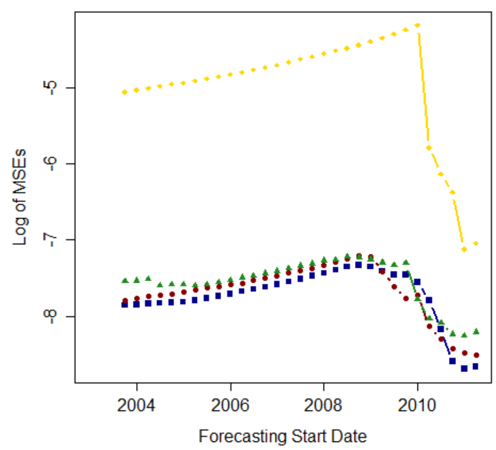
(b) k=2



(c) k=3



(d) k=4



(e) Legend

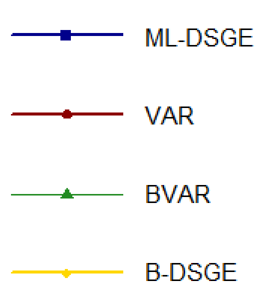
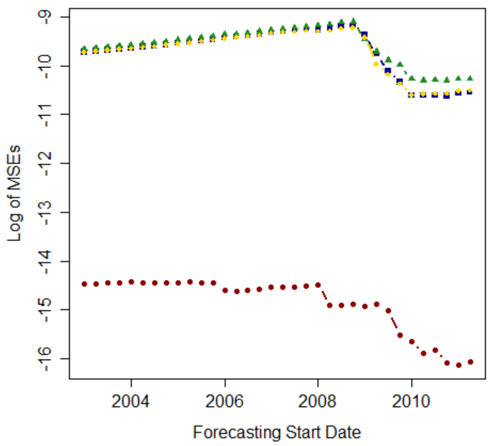
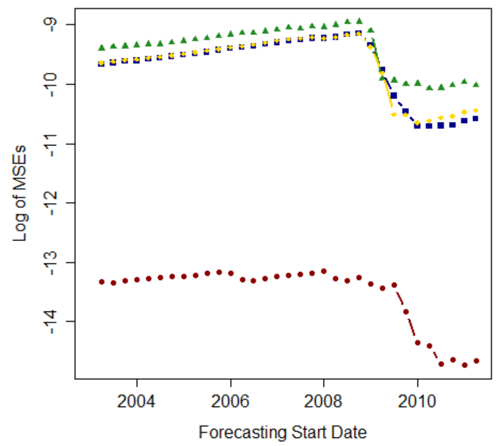


Figure 27: Interest rates - Dynamic (log) MSE results (k denotes the forecast horizon)

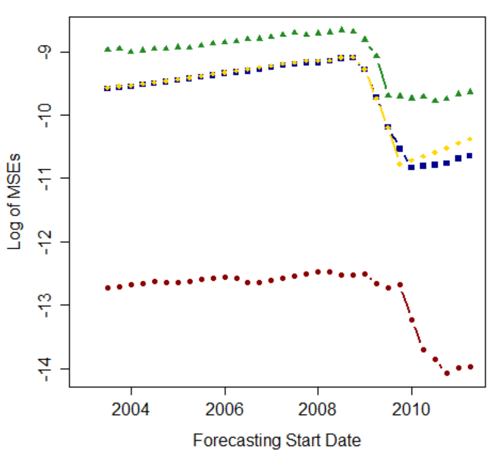
(a) k=1



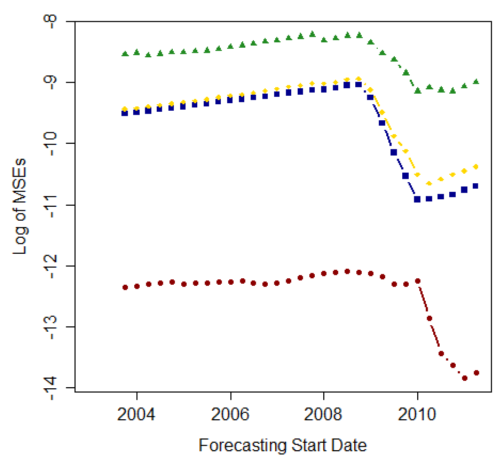
(b) k=2



(c) k=3



(d) k=4



(e) Legend

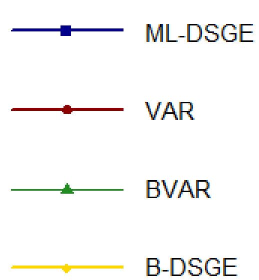
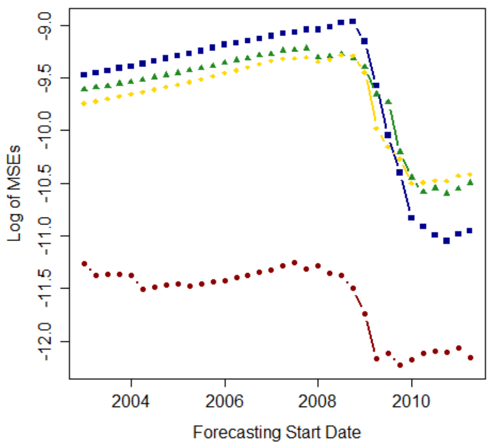
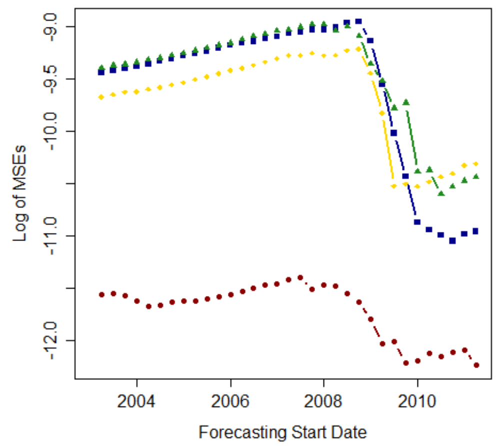


Figure 28: Inflation - Dynamic (log) MSE results (k denotes the forecast horizon)

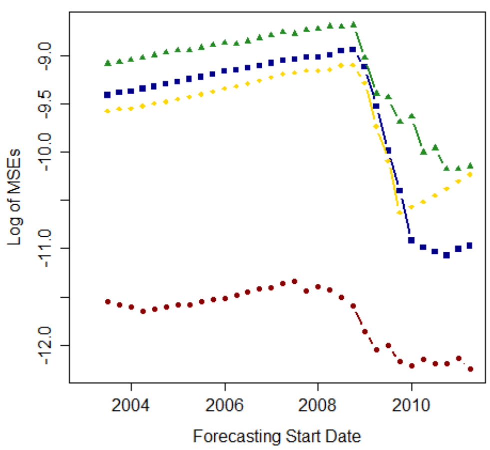
(a) k=1



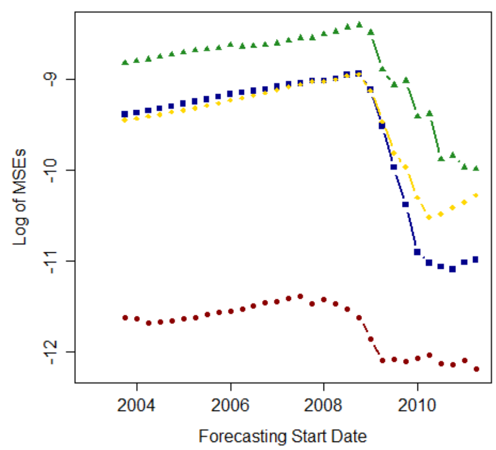
(b) k=2



(c) k=3



(d) k=4



(e) Legend