ABSTRACT

MAI, ZIYI. Essays on Bayesian Inference with Applications to Open Economy Macroeconomics. (Under the direction of Nora Traum.)

The objective of this dissertation is to use Bayesian methods in time series state-space models and apply them to issues in international macroeconomics and forecasting. Bayesian inference has been increasingly dominant in empirical macroeconomics in the past decade. Economists build more complicated models with high dimension of parameters extensively, while many macroeconomic time series have limited information due to their short sample periods and low frequency. Bayesian inference is thus a complement to statistical models applied to macroeconomic time series because it provides extra information to the estimation process through prior specifications.

In this dissertation, we investigate three different economic problems through the lens of Bayesian statistical models. We first consider a structural vector autoregression (TVP-VAR) with stochastic volatility and mixture innovation to study the time-varying effect of fiscal policy on the U.S. current account. We show that the U.S. current account reacts positively and has a gradual time-varying pattern in the early 1980s and the mid-2000s when the present discounted value of defense spending rises. In the late 1980s and the early 1990s when defense spending decreases, the current account responds negatively, and exhibits less time-varying effect than it does when defense spending increases in terms of impact responses. The maximum impact response of the current account to a government spending shock is 0.2 percentage points in the early 1980s and 0.095 the in mid-2000s. Fiscal policy in these two periods was characterized by dramatic increases in defense spending along with significant tax cuts. We then apply Romer and Romer’s narrative records on unanticipated tax shocks and find that the time-varying effect of a tax shock on the current account is less volatile and consistent with the existing literature. Impact responses of key macroeconomic variables are estimated to be 0.2-0.4 percentage points lower than the impact responses of those in a closed economy. The main results differ from the
existing literature in that the fiscal deficits driven by government spending largely contributed to the current account deficits in the 1980s and the mid-2000s.

We then move to examine stochastic volatilities in macroeconomic models. Many macroeconomic and financial variables usually exhibit a common pattern in their volatilities. We test if there is evidence for a common factor volatility driving the global financial cycle among a large dataset of macroeconomic and financial variables. We propose a Bayesian vectorautoregressive model (BVAR) with common stochastic volatility. The model is computationally feasible as the common factor volatility allows the posterior covariance matrix of the VAR coefficients to have a Kronecker structure. We show that the proposed model is strongly supported by the data and able to capture the common time-varying volatility suggesting the existence of a global financial cycle. We compare the forecasting performance of three BVAR-CSV models and find that the benchmark model produces slightly better point and density forecasts in periods without crisis, while the BVAR-CSV with Student-t innovations outperforms the others when including the 2007-2008 financial crisis periods.

Lastly, we propose an exchange rate model connecting the dynamic movements in the Canadian-U.S. real exchange to a set of macroeconomic fundamentals in the two countries. More specifically, we associate the Taylor rule type monetary policy reaction functions of the two countries and derive a reduced form state-space model with unobserved trend and transitory components. Observed outputs of both countries and the real exchange rate are decomposed into trend and transitory components respectively. We then link the transitory component of the real exchange rate with Taylor rule fundamentals such as the output gap and the inflation gap. The state-space model with parameter restrictions is estimated via maximum likelihood methods. The key findings of this chapter are that the estimated trend of the real exchange rate between the U.S. and Canada has substantial time variation over the sample periods. In addition, the transitory component linked to the Taylor rule fundamentals captures a large portion of the real exchange rate’s fluctuations.
Essays on Bayesian Inference with Applications to Open Economy Macroeconomics

by

Ziyi Mai

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Raleigh, North Carolina 2018

APPROVED BY:

__________________________  ____________________________
Douglas Pearce             Walter Thurman

__________________________  ____________________________
Subhashis Ghoshal           Nora Traum
  Chair of Advisory Committee
DEDICATION

To Jennifer Wei
BIOGRAPHY

Ziyi Mai hails from Guangzhou (Canton), China. He came to the U.S. in 2009 to pursue a master’s degree in economics at North Carolina State University. Upon obtaining his master’s degree, he engaged in research on state related public policy with the John Locke Foundation, a North Carolina think-tank. In 2012, he decided to advance his academic career by pursuing a Ph.D. in economics at North Carolina State University. He is completing his doctoral degree and will be joining the quantitative team at BB&T Corporation in Winston-Salem, NC.
ACKNOWLEDGEMENTS

I would like to express my thankfulness to all my committee members for their support of my research every step along the way. I first owe my largest debt of gratitude to Dr. Nora Traum for her tireless effort of training me to be an economist. Nora has also been extraordinarily supportive in my job search process, providing a lot of constructive guidance. None of my achievements would have been possible without her. I am grateful to Dr. Douglas Pearce for giving his in-depth knowledge and feedback regarding my research, and his opinions on improving my writing. I would also like to appreciate Dr. Walter Thurman for his willingness to listen to my ideas and provide valuable insights that I would not have been noticed. I am also thankful to Dr. Subhashis Ghoshal for his patience guiding me through statistical techniques that at the time I was not familiar with. Finally, the entire journey of doctoral study requires enormous amount of time and energy, so I owe my sincere gratitude to my family: Jennifer, Xitao, Rundi and Zimeng.
# TABLE OF CONTENTS

**LIST OF TABLES** .................................................. vii

**LIST OF FIGURES** ............................................... viii

**Chapter 1** Structural evolution of the fiscal and current account deficit in the U.S. ........................................... 1

1.1 Introduction ...................................................... 1
1.2 Historic aspects of the fiscal policy and the current account deficit in the U.S. ........................................... 6
1.3 The Model ......................................................... 9
  1.3.1 Basic TVP-VAR model ........................................ 9
  1.3.2 Mixture innovation for time varying coefficients and the covariance matrix ........................................... 12
  1.3.3 Priors and computation details ................................ 16
  1.3.4 Sketch of the MCMC algorithm ................................ 17
1.4 Identification schemes and data .................................. 18
  1.4.1 Identification restrictions .................................... 18
  1.4.2 Data .......................................................... 20
1.5 Empirical results ................................................ 20
  1.5.1 Statistical evidence of time-varying parameters .............. 20
  1.5.2 Impulse response analysis .................................... 23
  1.5.3 Time-varying effects of defense spending shocks .............. 25
  1.5.4 Time-varying effects of unanticipated tax shocks ............ 32
1.6 Conclusion ........................................................ 36

**Chapter 2** Common Stochastic Volatility in the Global Financial Cycle . . . . . . 38

2.1 Introduction ...................................................... 38
  2.1.1 Financial volatility in an AR-SV model ......................... 39
2.2 Volatility structure in a Bayesian VAR ................................ 43
  2.2.1 Kronecker structure of the covariance matrix .................. 45
  2.2.2 Common factor volatility in a Bayesian VAR ................... 46
2.3 Bayesian Estimation .............................................. 47
  2.3.1 Posterior Analysis ............................................ 47
  2.3.2 Implementation ............................................... 49
2.4 Data and Empirical Results ..................................... 51
  2.4.1 Data .......................................................... 51
  2.4.2 Common drifting volatility in the global financial cycle .... 51
  2.4.3 Forecasting exercise .......................................... 55
  2.4.4 MCMC Convergence and integrated likelihoods ................ 60
2.5 Concluding remark ............................................... 61

**Chapter 3** Taylor-rule fundamentals and exchange rate dynamics .......................... 65

3.1 Introduction ...................................................... 65
3.2 Taylor rule fundamentals in exchange rate model .................. 69
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Prior choices for key parameters: $\mathcal{N}$ and $\mathcal{IW}$ denote the normal and inverse Wishart</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>distribution. $\hat{A}<em>{OLS}$, $\hat{\alpha}</em>{OLS}$, and $\hat{V}(\hat{\alpha}_{OLS})$ are obtained</td>
<td></td>
</tr>
<tr>
<td></td>
<td>from training sample least square estimators.</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>Priors and restrictions on models</td>
<td>22</td>
</tr>
<tr>
<td>1.3</td>
<td>Priors and restrictions on models</td>
<td>23</td>
</tr>
<tr>
<td>2.1</td>
<td>Root Mean Square Forecasting Error</td>
<td>59</td>
</tr>
<tr>
<td>2.2</td>
<td>Average log predictive density</td>
<td>63</td>
</tr>
<tr>
<td>2.3</td>
<td>Inefficiency Factors for BVAR-CSV</td>
<td>64</td>
</tr>
<tr>
<td>2.4</td>
<td>Estimated log integrated likelihoods</td>
<td>64</td>
</tr>
<tr>
<td>3.1</td>
<td>Selected parameter estimates and confidence intervals</td>
<td>77</td>
</tr>
<tr>
<td>A.1</td>
<td>Description of variables</td>
<td>98</td>
</tr>
<tr>
<td>A.2</td>
<td>Posterior median of selected impact responses (14th/86th percentile in parentheses)</td>
<td>99</td>
</tr>
<tr>
<td>B.1</td>
<td>Description of variables</td>
<td>103</td>
</tr>
</tbody>
</table>


LIST OF FIGURES

Figure 1.1 U.S. current account (red solid, negative for deficit) and Federal primary budget (blue dashed, negative for deficit) as a ratio of GDP. 
Source: Bureau of Economic Analysis and the author’s calculation. ................. 7

Figure 1.2 Components of the current account as percentages of GDP: Goods (solid red), Investment Income (green dotted), Services (blue dotdash), and Unilateral transfer (purple dashed). 
Source: Bureau of Economic Analysis. ......... 9

Figure 1.3 Impulse responses to a defense spending shock with an informative prior. .... 26

Figure 1.4 Impact responses to a defense news shock with an informative prior: these impact response functions measure contemporaneous responses of the four variables following a defense news shock. The horizontal axis measures how these responses vary across time while the vertical axis shows the percentage deviation of the variable. .................. 27

Figure 1.5 Time-varying impulse responses to a defense news shock with a flat prior. 28

Figure 1.6 Impact responses to a government spending shock with a flat prior. ....... 29

Figure 1.7 Time-varying impulse responses to an unanticipated increase in the tax rate with a flat prior. ........................................ 33

Figure 1.8 Impact responses to an unanticipated increase in the tax rate with a flat prior. 35

Figure 2.1 Estimated Volatilities (defined as standard deviations) of selected variables from AR-SV models. Shaded areas are the 5% and 95% confidence bands .... 41

Figure 2.2 Posterior mean of common drifting volatility: BVAR-CSV is the benchmark model. ....................................................... 54

Figure 2.3 VIX and Global Realized Variance: quarterly global realized measured using daily returns of the MSCI index. ........................................ 55

Figure 2.4 Heat map of the VAR coefficients ........................................ 56

Figure 3.1 Trend and transitory components of the real exchange rate: the top panel shows the smoothed estimates (dash) and the filtered estimates (solid). The bottom panel displays the transitory component of the real exchange rate. The filtered estimates are obtained in each time period when a new observation is updated (i.e., using the forward recursion of the Kalman filter). The smoothed estimates are obtained after all observations are observed (i.e., using the backward recursion of the Kalman smoother) ......... 78

Figure 3.2 Actual real exchange rate (dash) and filtered trend real exchange rate (solid) 80

Figure A.1 Standard Deviations of errors in current account equation to a tax shock .... 100
Chapter 1

Structural evolution of the fiscal and current account deficit in the U.S.

1.1 Introduction

In the past three decades, America has been transformed from the world’s largest creditor to the world’s largest debtor nation. This means that Americans borrow abroad more than they lend overseas. The U.S. current account deficit, substantially driven by its trade deficit, has been in a deteriorating state since the 1980s. At the same time, government borrowing has become a norm of fiscal policy except for a short period during the 1990s. The link between fiscal policy and the current account has long triggered interest among economists and policymakers, from a variety of perspectives. For instance, the causal relation between fiscal deficits and current account deficits has motivated abundant research focusing on the "twin deficits" hypothesis. To a large open economy such as the United States whose current account imbalance has been persistent for years, a relevant inquiry is to what extent fiscal policy, either in the form of government spending or tax changes, can account for the current account imbalances.

This paper examines the evolution of fiscal policy and the current account in the U.S. by applying a time-varying vector autoregression with stochastic volatility (TVP-VARSV).
Allowing for the time-varying parameters in the model renders possible answers of how the effect of fiscal shocks on the U.S. current account evolves over the sample period. More specifically, I adopt the dynamic mixture model approach of Gerlach et al. (2000) enabling tests of whether a parameter changes over time. In terms of the VAR identification, I use a variety of shocks to the U.S. fiscal policy introduced by Ramey (2011) and Romer and Romer (2010). Ramey’s defense news series is the present discounted values extracted from war news capturing exogenous disturbances to expected changes in government spending. Romer and Romer’s narrative records documented various legislations and presidential speeches to identify endogenous and exogenous tax changes.

The key finding is that the time-varying impact impulse responses of the current account show positive and gradual changes during the early 1980s and mid-2000s when defense spending rose dramatically. In contrast, the current account responds negatively when defense spending contracted in the late 1980s and early 1990s. In other words, fiscal deficits characterized by dramatic increases in defense spending lead to a current account surplus in the early 1980s and mid-2000s. The effect of a positive unanticipated tax shock on the current account displays fewer time variations where all impact responses across time stay positive. On one hand, the main result indicates that fiscal deficits driven by the defense spending news causes a current account surplus, which contrasts with the existing literature. On the other hand, the effect of fiscal deficits characterized by tax cuts is in line with the existing literature–associating with current account deficits across the sample period.

Identities of the National Income and Product Accounts (NIPA) describe two major channels from fiscal policy to the current account.\(^1\) This first is the intratemporal trade channel

\[^1\text{Following Abbas et.al (2011), relations fiscal policy and current account cast be cast as:}\]

\[
CA = EX(e) - IM(Y, e) + TP \\
CA = (S(Y, r) - I(r)) + (T - G)
\]

where \(CA\) is current account, \(EX\) and \(IM\) represent export and import respectively. They are defined as a function of output \(Y\) and real exchange rate \(e\), with higher \(e\) meaning appreciation. \(TP\) is transfer payments. \(S\) and \(I\) are private savings and investment respectively. They are a function of real interest rate \(r\). \(T\) and \(G\) denote taxes and government spending. The difference between the two is defined as government deficits.
represented by the celebrated Mundell-Fleming model (Mundell, 1963 and Fleming, 1962). The model shows how an expansionary fiscal policy shock raises the demand for home goods and money demand, increasing the nominal and real exchange rates through higher interest rates. Net exports in turn decline. The second channel focuses on an intertemporal approach represented by Frenkel et al. (1996), Backus et al. (1992) and Baxter (1995). These models emphasize how economic agents are aware that current tax cuts or spending increases require an increase in future taxes. If economic agents raise private savings today to offset expected future increases in taxes, private savings can offset public savings and the current account may not worsen. The new open economy models represented by Obstfeld and Rogoff (2000) incorporate both approaches as well as sticky prices and wages, and imperfect competition.²

The empirical analysis of the two deficits is substantially important as a matter of rebalancing the external deficit. However, the empirical literature investigating the relationship between the two deficits has generated contrasting results. Kim and Roubini (2008) was the first to use a structural VAR to study the issue. They find a positive innovation in the budget deficit improves the current account balance. They thus challenge the twin deficit hypothesis and assert it should be "twin divergence" in the U.S. data. In contrast, Monacelli and Perotti (2006) find that the twin deficit hypothesis is supported with data from four OECD countries (U.S., UK, Canada and Australia), using a similar identification approach as Kim and Roubini. They point out that Kim and Roubini orthogonalize the reduced form innovation via Choleski ordering, preventing contemporaneous effects of government deficit shocks on output. Corsetti and Müller (2006) use the same dataset as in Monacelli and Perotti (2006) and add an index to measure openness of trade in their SVAR. Their results confirm what was Kim and Roubini (2003)'s hypothetical claim that countries with relatively less openness and strong home bias have twin divergence. Using the VAR identification of Blanchard and Perotti (2002), Corsetti et al. (2012)²

²The New open economy models attempt to address some puzzles found in the empirical literature. For example, Monacelli and Perotti (2006) find that private consumption rising and real exchange rate depreciating in response to a positive government spending shock. See Obstfeld and Rogoff (2000) for summary of the six major puzzles in international macroeconomics. Lane et al. (2010) argues that news shocks implying governments provide fiscal impetus may lead to a sell-off in the foreign exchange market.
find that net exports drop briefly on impact but quickly turn to surplus after a government spending shock. Their underlying economic interpretation is that agents anticipate a spending reversal, and thus an initial exogenous increase in government spending does not crowd out consumption and depreciates the real exchange rate.

What motivates this paper is that the relationship between fiscal policy and current account deficit may not be constant over time. Shocks and the way they are generated can vary across the sample period. Such consideration drives the use of a time-varying parameter VAR to study the issue. However, the standard TVP-VAR model, as in Primiceri (2005), has the potential problem of overfitting because it allows every parameter to change at every time period. Firstly, this is in fact a strict assumption imposed on parameters as not all of them have gradual changes over time. If some parameters tend to be time-invariant, whereas others would have abrupt changes with large magnitude, the standard TVP-VAR will overestimate some variation in some parameters and underestimate some others. Moreover, the model dramatically raises the bar of computational cost so that models following Primiceri’s usually limit endogenous variables to be fewer than four.

My model is closely related to Koop et al. (2009)’s dynamic model with mixture innovations but extends it in a more flexible way. Like the standard TVP-VAR model in Primiceri (2005) and Cogley and Sargent (2005), I allow the error covariance matrix to change over time. To evaluate the structural change of the two deficits more precisely, I extend Koop et al. (2009)’s TVP-VAR with mixture innovation to allow parameters of the same equations to vary at different points of time, including different blocks. This assumption is built on the idea that parameters of the same equation might change at some point within the span of the sample. A Bernoulli prior is imposed on the latent binary variable to reflect the frequency of structural breaks in a given equation. The introduction of a mixture innovation to the standard TVP-VAR models renders more flexibility in estimating the parameter variations than models in Primiceri (2005), Cogley and Sargent (2001) and Cogley and Sargent (2005). If the true model is one where parameters

---

3To model the error covariance matrix as stochastic volatility is consistent with the time series evidence that volatilities of many macroeconomic variables have been in decline since the 1980s.
have many "small" breaks, the data will deliver the probability of the latent variable for all blocks to be close to one. Otherwise, the probability will tend to be zero if the model appears to have a few "abrupt breaks." The key feature of this dynamic mixture model is expected to bring thorough assessment of the structural evolution of fiscal policy and the current account through prior specifications.

As for the VAR identification, I adopt the narrative approach due to its connection with historic events that could be better interpreted in a time-varying manner. If the current account responded negatively in some periods but positively in other periods, one can trace back to the defense news series and compare the change of present discounted value in those periods. Specifically, I use Ramey (2011)’s defense news series to identify the government spending shock and Romer and Romer (2010) for the tax shock. Shocks to fiscal policy are approximated by innovations to government spending and innovations to tax policy separately. For example, a positive shock to government spending represents an expansion of fiscal policy, whereas a positive shock to a tax change indicates a contraction of fiscal policy, thus generating a fiscal deficit.

This paper delivers two major findings on the effect of fiscal shocks to the current account balance. First, the time-varying effect of government spending shocks on the U.S. current account is mixed. Under the specification with a diffuse prior, impact responses of the current account have positive and gradual changes over time following positive defense news shocks. In contrast, impact parameters have a few abrupt breaks when defense news shocks are negative. More specifically, impact responses reach 0.2 percent deviation from the trend in 1984Q2 and -0.38 percent from the trend in 1987Q3. This result does not support solely any existing empirical literature on the matter. Second, the time-varying effect of tax shocks identified by Romer

---

4Romer and Romer (2010)'s approach has been extended and modified by Mertens and Ravn (2011), Mertens and Ravn (2013), and Mertens and Ravn (2014). The identification in this paper follows part of their contribution.

5The concept of "shock" may be ambiguous in some of the literature. Conventionally, the term shocks mean, for example, the residuals from a reduced form VAR model. Some researchers refer to instruments as "shocks" because of the exogeneity. In this paper I adopt the concept that structural shocks are initial exogenous forces that are uncorrelated with each other and also have economic interpretations. See Bernanke (1986) and Blanchard and Watson (1986).
and Romer(2010) on the current account demonstrates insignificant changes across the sample
periods. The impact responses of the current account to a positive unanticipated tax shock
stays 0.31 - 0.41 percent deviation above trend across all time, confirming that contractionary
fiscal policy in the form of a tax increase improves the current account.

The rest of this paper is organized as follows. Section 2 reviews the background of the
fiscal and current account deficit, along with a brief discussion of the related literature. Section
3 presents the TVP-VAR model, its variation, and the computational algorithm. Section 4
describes the data and identification schemes. Section 5 offers empirical results and structural
analysis. Section 6 concludes and raises extension to future research questions.

1.2 Historic aspects of the fiscal policy and the current account
deficit in the U.S.

The twin deficits first captured macroeconomists’ attention in the 1980s. In 1981 the conse-
quence of Ronald Reagan’s tax cuts was a soaring federal government deficit, from 2.5% of
GDP in 1981 to 4.9% in 1986. The current account swayed into deficit almost simultaneously.
Following this event, the notion of "twin deficits"- in government borrowing and the current
account- became tantalizing to researchers.6 Figure 1.1 presents this co-movement between the
two deficits for the sample period used in this paper.7 Simple observations in the early 1980s
gave birth to the twin deficit hypothesis that the fiscal deficit fuels the current account deficit.
As the economy grew rapidly in the 1990s, the federal government’s budget approached balance,
but the current account deficit continued to grow.

6Chinn (2005) provides a non-technical overviews of the twin deficits starting the 1980s. Other empirical studies
of the fiscal and the current account deficits include Abell (1990) and Evans (1986).
7The Federal primary budget deficit is constructed by the difference between government income receipt and
net government saving less interest payment.
A new hypothesis emerged in the 2000s in an effort to explain the persistent U.S. current account imbalance. Bernanke (2005) proposes the "global saving glut" hypothesis, asserting that the U.S.’ persistent current account deficit may be caused by an overwhelming inflow of financial and capital investment from areas where savings are excessive. Bernanke points out recent evidence showing that the role of fiscal policy (both government spending and changes in taxes) has been declining in influencing the current account. Perhaps the inflow of savings overseas is not a sign that the U.S. is in shortage of savings. Instead, foreign investors see the U.S. as one of the core financial markets in the world, thus purchasing large amounts of
U.S. assets that are considered less risky compared to other countries'. Figure 1.2 displays the components of the current account as a percentage of GDP. Hence, the trade deficit as a major force driving the current account alleviated during the crisis due to a drop of consumption. This hypothesis has sparked a series of empirical analysis focusing on current account imbalances and its adjustment. Fiscal policy, of course, has become a subset of the subjects under this strand of the literature.\(^8\)

The two deficits moving in the same direction reemerged during most of the 2000s as real GDP bounced back leading households to reduce saving during this period. (see, e.g. Valderrama et al. (2007)) After the financial crisis in 2008, there’s little sign that the current account deficit improved, while the fiscal deficit had a significant contraction during the second term of the Obama administration.

It is worth noting that the twin deficit hypothesis is a global issue. Most industrial economies and many countries in the Organization for Economic Co-operation and Development (OECD) improved their fiscal accounts and achieved a surplus in the early 2000s. However, performance of current account deficits are different across countries. For example, Germany experienced a strong negative correlation between the two deficits from 1990 to 2005, whereas Japan, France, UK, Italy, Canada from the G7, all had a positive correlation in the same period. The global features of the twin deficit have led to another branch of the literature comparing the effects across countries using panel data and factor VAR models.\(^9\). However, consensuses are even harder to reach in those models as every country’s economic structure and policy are drastically different. Each country’s fiscal policy and current account might evolve differently over time.

---

\(^8\)Chinn and Ito(2008) study the U.S. fiscal and global current account imbalances, featuring East Asia savings. Gourinchas and Rey (2014) investigate the role of capital flows and exchange rates in current account adjustments.

\(^9\)Cross sectional and panel data approaches are dominating the empirical literature in studying the twin deficit(see Chinn and Ito (2008), Lee and Chinn (2006))
1.3 The Model

1.3.1 Basic TVP-VAR model

Following the notation from Durbin and Koopman (2012), the reduced form time-varying VAR model with order $k$ can be written as
\begin{equation}
y_t = Z_t' \alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, H_t) \tag{1.1}
\end{equation}

\begin{equation}
\alpha_{t+1} = T_t \alpha_t + R_t \eta_{t+1}, \quad \eta_t \sim \mathcal{N}(0, Q_t) \tag{1.2}
\end{equation}

where $y_t$ is a $p \times 1$ vector of observed endogenous variables, $Z_t$ is a $p \times m$ matrix collecting intercepts and lagged variables as $Z_t = I_t \otimes [1, y_{t-1}', \ldots, y_{t-l}']$. Equation (1.1) has the structure of a linear regression model where coefficient vector $\alpha_t$ varies over time. The second equation indicates that the VAR coefficients evolve according to a first order vector autoregressive model, which is a general formulation of state space models. Due to the concern of over-parameterization of TVP-VAR models, however, the VAR coefficient is usually modeled as a random walk process such that the matrix $T_t$ is set to be the identify matrix. The advantage of this modeling is that it greatly reduces the parameters to be estimated. In addition, many economic variables are found to follow random walk processes.\textsuperscript{10}

Applying a triangular reduction to the variance-covariance of (1.1), $H_t$, such that

$$A_t H_t A_t' = \Sigma_t, \quad A_t^{-1} u_t = \epsilon_t, \quad u_t \sim IID. \mathcal{N}(0, I_p)$$

where $A_t$ is a lower triangular matrix with diagonal elements of ones:

\begin{equation}
A_t = \begin{bmatrix}
1 & 0 & \ldots & 0 \\
\alpha_{21,t} & 1 & \ldots & . \\
. & . & \ldots & . \\
. & . & \ldots & 1 \\
\alpha_{p1,t} & \ldots & \alpha_{p(p-1),t} & 1
\end{bmatrix} \tag{1.3}
\end{equation}

Given the structural innovation vector $u_t$, the reduced form TVP-VAR model can be written

\textsuperscript{10}This is consistent with the current literature on TVP-VAR, see Cogley and Sargent (2001), Cogley and Sargent (2005), and Primiceri (2005).
as
\[ y_t = Z_t'\alpha_t + A_t^{-1}u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_t^{1/2}\epsilon_t) \] (1.4)

where \( \Sigma_t \) is a diagonal matrix with diagonal elements \( \sigma^2_{j,t} \) for \( j = 1, \ldots, k \). It is worth noting that using a lower triangular matrix here is not equivalent to an identification assumption. It is a way of re-parameterizing the covariance matrix of the measurement equation. It takes advantage of the symmetric structure of the VAR variance-covariance matrix that any symmetric matrix \( \Sigma_t \) satisfies the decomposition \( \Sigma_t = \Sigma_t^{1/2}(\Sigma_t^{1/2})' \), where \( \Sigma_t^{1/2} \) is a square non-unique matrix. We can then take \( H_t^{1/2} = A_t\Sigma_t^{1/2} \).

The time-varying nature underlying in \( \Sigma_t \) and \( A_t \) must be addressed by an extra state equations. Primiceri (2005) uses:

\[ h_{t+1} = h_t + \zeta_{t+1}, \quad \zeta_t \sim \mathcal{N}(0, W) \] (1.5)

where \( h_t = (h_{1,t}, \ldots, h_{p,t})' \), \( h_{i,t} = ln(\sigma_{i,t}) \). In other words, \( h_t \) has all the diagonal elements of the VAR covariance \( \sigma_t \) in logarithm.

Let \( a_t = vec(A_t) \) be a vector collecting all non-zero elements in matrix \( A_t \),

\[ a_{t+1} = a_t + \nu_{t+1}, \quad \nu_t \sim \mathcal{N}(0, S) \] (1.6)

where \( W, S \) are positive definite and \( S \) is divided into \( p-1 \) blocks corresponding to each equation. Primiceri (2005) sets \( S \) as block diagonal for the main reason that Gibbs sampling can be implemented, as well as assuming the structural parameters are correlated in each equation but not correlated across different equations. This assumption works well if the order of variables is appropriate under a Cholesky decomposition as an identification scheme.\(^{11}\)

\(^{11}\)Primiceri (2005) sampled the blocks of \( S \) without conditioning on other blocks. Del Negro and Primiceri (2015) corrected the algorithm, showing that conditional on different blocks of \( S \) does not affect the use of Gibbs sampling.
1.3.2 Mixture innovation for time varying coefficients and the covariance matrix

The original version of the TVP-VAR with stochastic volatility essentially imposes an assumption that every parameter changes in every period. From the standard point of inference, the consequence is that the model is usually over-parameterized and overfit. Previous methods on shrinking the number of parameters of a constant parameter VAR such as Sims and Zha (1998) cannot reduce the size of parameters effectively when it comes to the time-varying VAR. In addition, TVP-VAR only works well in models with small dimensions and is likely to become computationally unstable once the dimension exceeds four with more than two lags.\footnote{The representative TVP-VAR literature such as Cogley and Sargent (2001, 2005) and Primiceri (2005) all use three variables for their models. Koop and Korobilis (2010) develop a TVP-VAR model that fits large dimensions in order to overcome this difficulty on dimension} With time-varying parameters, the number of state variables could dramatically increase. This is similar to that the covariance matrix could become singular when the number of variables exceeds the number of observations.\footnote{See Cai et al. (2011) for methods on estimating the inverse of covariance matrix in this case.}

Koop et al. (2009) extend the traditional TVP-VAR models and introduce mixture innovations that better represent the structural changes in the VAR coefficient and variance-covariance matrix. Computationally speaking, mixture innovation serves as a method of shrinking some of the parameters in both the coefficients and covariance matrix to zero so that the size of the parameter space can be manageable. The macroeconomic implication of mixture innovation is that the parameters might not have as many small structural breaks as modeled under the original TVP-VAR models. I modify Koop et al. (2009)’s modeling strategy to allow breaks in the VAR coefficients equation by equation, the VAR variance-covariance matrix $\Sigma_t$ and the structural parameters in matrix $A_t$.

Define $k_{jt}$ as a dummy variables such that
\[ k_{jt} = \begin{cases} 
1 & \text{if a change in the parameter occurs} \\
0 & \text{otherwise} 
\end{cases} \]

where \( j \) is an integer and the number of the dummy variables depends on how many restrictions are imposed on specific parameters.

Three parameters blocks depend on the dummy variables, reflecting the time-variation in parameters. Let \( K_{\alpha,t}, K_{a,t}, K_{h,t} \) denote diagonal matrices with elements \( k_{jt} \) controlling the VAR coefficients \( \alpha \), the structural parameters \( a \), and the volatilities \( h \) respectively. For convenient notation, define \( K_t \) as a matrix containing all \( k_{jt} \).

The transition equation (1.2) thus becomes

\[ \alpha_{t+1} = \alpha_t + K_{\alpha,t} \eta_{t+1} \]

There are many ways to model \( K_{\alpha,t} \). In Koop et al. (2009), \( K_{\alpha,t} \) is modeled in a way that all VAR coefficients are allowed to change at the same time. Liu and Morley (2014) specify two cases allowing coefficients in the same equation to vary together and coefficients of the same variable to vary across equations at the same time. Since the model in this paper only uses one lag \( (k = 1) \), it is reasonable to put the restriction that coefficients vary together in the same variable across equation. Furthermore, as discussed below, the identifying restrictions are all in the short run instead of the long run. Therefore cross equation variations on parameters are of greater interest than variations on lags.

An illustration of the model specifications is that the order of endogenous variables is \( y_t = (\text{defense}_t \ \text{rgovt}_t \ \text{rgdp}_t \ \text{curt}_t \ \text{int}_t \ \text{fx}_t)' \) in which they denote Ramey’s change of present value of defense news series, real government spending, real GDP growth, current account as a ratio of GDP, trade weighted foreign exchange rate in logarithm, and annualized interest rates on the 3-month Treasury Bills. Note that the number of elements in the state vector \( \alpha_t \) is \( p(pk + 1) \). In a system with six variables and one lag, the number of parameters in the state vector is
42. In a medium scale VAR like this, the number of lags exceeding 2 may lead to unstable matrices from a computational perspective. The impulse responses could become ill-shaped. Most importantly, long lags do not add predictive power to the model. Therefore the number of lags is chosen to be 1. Further I assume that the coefficients of military news series, the government spending and real GDP growth always move together. The controlling matrix $K_{a,t}$ can be represented by 

$$K_{a,t} = \text{diag}\{K_{1t}, K_{2t}, K_{2t}, K_{2t}, K_{2t}, K_{2t}\}$$

where

$$K_{1t} = \text{diag}\{k_{1t}, k_{1t}, k_{1t}, k_{1t}, k_{1t}\}$$

$$K_{2t} = \text{diag}\{k_{2t}, k_{2t}, k_{2t}, k_{2t}, k_{2t}\}$$

The matrices $K_{1t}$ and $K_{2t}$ control the constant terms and the lag coefficients respectively.

Further, restrictions of breaks are also imposed on the structural parameters and the VAR errors. The evolution of equations (1.4) and (1.5) thus becomes:\textsuperscript{14}

$$a_{t+1} = a_t + K_{a,t} \nu_{t+1}, \ \nu_t \sim \text{IID.} \mathcal{N}(0, S) \quad (1.7)$$

$$h_{t+1} = h_t + K_{h,t} \zeta_{t+1}, \ \zeta_t \sim \text{IID.} \mathcal{N}(0, W) \quad (1.8)$$

where, $K_{a,t} = \begin{bmatrix} k_{3t} & 0 & \ldots & 0 \\ 0 & k_{3t} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & k_{3t} \end{bmatrix}$, $K_{h,t} = \begin{bmatrix} k_{4t} & 0 & \ldots & 0 \\ 0 & k_{4t} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & k_{4t} \end{bmatrix}$.

Similarly, the dummy variables $k_{jt}, j = 3, 4$ are equal to one if a change occurs and zero otherwise. Adding mixture innovations to the standard TVP-VAR model with stochastic volatility makes the model more flexible. The elements in the matrix $K_t$ serve as an indicator for whether a parameter vector does or does not change. In addition, it reduces the size of parameter space and thus lead to a more precise inference.

\textsuperscript{14}The number of contemporaneous parameters is equal to $p(p-1)/2$ due to the Cholesky decomposition of the variance-covariance matrix.
Note that the dummy variable $k_{jt}$ are a type of hierarchical parameters in a statistical sense that each $k_{jt}$ will be estimated. Following much of the literature, I adopt a Bernoulli distribution:

$$k_{jt} \sim Bernoulli(p_j), j = 1, 2, 3, 4$$

To preserve the implementation of Gibbs sampling, a conjugate prior (Beta distribution) is imposed on the hyperparameter $p_j$. The prior assumes chances of parameters experiencing a break independently. For example, the priors for the VAR coefficient $k_{jt}$ are all independent of one another, simultaneously and at all leads and lags.\(^\text{15}\)

The introduction of a mixture innovation to the standard TVP-VAR seems adding more unknown parameters to the model, but in fact it allows achieve a more parsimonious form. If all $k_{jt}$ are zeros, the model becomes a time invariant VAR. When all $k_{jt}$ are ones, it becomes the standard TVP-VAR as in Primiceri (2005). In other words, both the standard TVP-VAR and constant VAR are nested in the one with mixture innovations.

All innovation blocks in the above system are assumed to be jointly normally distributed, with no contemporaneous correlations for all $t$. Denote $V$ as the variance-covariance matrix for all innovation blocks:

$$V = \text{Var}
\begin{bmatrix}
\epsilon_t \\
\eta_t \\
\nu_t \\
\zeta_t
\end{bmatrix}
= \begin{bmatrix}
I_p & 0 & 0 & 0 \\
0 & Q & 0 & 0 \\
0 & 0 & S & 0 \\
0 & 0 & 0 & W
\end{bmatrix}$$

Notice that none of the restrictions are imposed in the covariance matrix $V$. Primiceri (2005) justified this setup because adding off diagonal elements to the matrix would increase the difficulty of recovering the structural innovations.

\(^\text{15}\)This assumption can be relaxed and the computational cost is much more expensive. Gerlach, Carter and Kohn (2002) develop an algorithm to fit the situation, but I do not adopt it in this paper.
Table 1.1: Prior choices for key parameters. $\mathcal{N}$ and $\mathcal{IW}$ denote the normal and inverse Wishart distribution. $\hat{A}_{OLS}$, $\hat{\alpha}_{OLS}$, and $\hat{V}(\hat{\alpha}_{OLS})$ are obtained from training sample least square estimators.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Prior Family</th>
<th>coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>Initial state vector</td>
<td>$\mathcal{N}(\hat{\alpha}<em>{OLS}, k</em>\alpha \times \hat{V}(\hat{\alpha}_{OLS}))$</td>
<td>$k_\alpha = 4$</td>
</tr>
<tr>
<td>$A_0$</td>
<td>Initial structural matrix</td>
<td>$\mathcal{N}(\hat{A}<em>{OLS}, k_A \times \hat{V}(\hat{A}</em>{OLS}))$</td>
<td>$k_A = 4$</td>
</tr>
<tr>
<td>$\log\sigma_0$</td>
<td>Initial log volatility</td>
<td>$\mathcal{N}(\log\hat{\sigma}<em>{OLS}, k</em>\sigma I_p)$</td>
<td>$k_\sigma = 10$</td>
</tr>
<tr>
<td>$Q$</td>
<td>Covariance shocks to $\alpha_{t+1}$</td>
<td>$\mathcal{IW}(k_Q^2 \times pQ \times \hat{V}(\hat{\alpha}_{OLS}), pQ)$</td>
<td>$k_Q = 0.01, pQ = 40$</td>
</tr>
<tr>
<td>$W$</td>
<td>VCV of shocks to $\log\sigma_t$</td>
<td>$\mathcal{IW}(k_W^2 \times pW \times I_p, pW)$</td>
<td>$k_W = 0.01, pW = p + 1$</td>
</tr>
<tr>
<td>$S_j$</td>
<td>VCV of shocks to $A_t$</td>
<td>$\mathcal{IW}(k_S^2 \times pR_j \times \hat{A}_{OLS}, pS_j)$</td>
<td>$k_S = 0.01, pS_j = j + 1$</td>
</tr>
</tbody>
</table>

### 1.3.3 Priors and computation details

The priors chosen in this paper are similar to those in Primiceri (2005), Cogley and Sargent (2005) and Koop et al. (2009) with the exception of prior parameters on the dummy variables. They are proper conjugate priors for computational convenience. Table 1.1 gives a complete description of the chosen priors and their parameters.

The normal prior on the state vector $\alpha_t$ is a training sample prior and is conventional as it gives a posterior from the same family. Sims and Zha (1998) employ a similar prior in a VAR model. The normal prior on $\log\sigma$ is a common practice in the stochastic volatility literature (see Kim et al. (1998); Jacquier et al. (2004)). This is not a conjugate prior but provides the advantage of tractability. Conjugate priors such as inverse-gamma and inverse-Wishart have been applied in the literature of Bayesian state-space models (see, e.g., West (1996) and Gelman et al. (2014)).

The prior for hyperparameter $p_j$ is assumed to be Beta such that, $\text{Beta}(\alpha_j, \beta_j)$, for $j = 1, ..., 4$. From the property of the Beta distribution, the expectation of $p_j$ is $E(p_j) = \frac{\alpha_j}{\alpha_j + \beta_j}$. If $\alpha_j = \beta_j = 1$, the expected probability of being 1 is 50%. This gives intuition to a parameter
that a break might occur in a given period with a 50% chance. If parameters are modeled in a way that only a small number of breaks occur, the prior on those parameters should be informative. (e.g. $\alpha_j = 0.1, \beta_j = 10$). Finally, a least informative prior can be used when modeling it as "many small breaks". Koop et al. (2009) refer to the case when $E(p_j) = 1$ as a benchmark and provide an empirical comparison for different priors. This paper chooses a slightly informative prior on these parameters, indicating that the author’s belief on parameters changes is grounded in the previous literature and economic historic events. Priors on different $K$s vary and are discussed in section 5.

1.3.4 Sketch of the MCMC algorithm

The basic MCMC algorithm in this paper follows Primiceri (2005), with a corrected algorithm by Del Negro and Primiceri (2015). The slight difference is the procedure of drawing the latent variable $k_t$. This subsection briefly describes the algorithm, while the computational details are presented in the Appendix. The advantage of the Gibbs sampler is that posteriors are known distributions that are easy to draw samples from. Drawing state vector $\alpha_t$ and $A_t$ employs the techniques in Carter and Kohn (1994) (henceforth CK) where they developed recursive methods of drawing samples of the state conditional on the data first and then the neighboring states and so on.\textsuperscript{16} Samples of the variance-covariance matrices $V$ are drawn from Inverse-Wishart distributions conditional on the rest of the parameters. The variance-covariance matrix of innovations to the structural matrix $A_t$ consists of several blocks that can also be sampled from an Inverse-Wishart distribution, because the the time-varying patterns are very similar among these blocks. Finally, a stability condition is imposed to make sure that the covariance matrix does not become singular in the process of iteration.

The evolution of the volatility state is a non-linear and non-Gaussian state space model where innovations in the measurement equations are distributed as a $log\chi^2(1)$. With this non-

\textsuperscript{16} Durbin and Koopman (2002) (henceforth DK) present an efficient simulation smoother by first constructing the innovations density and then drawing samples from the state density conditioned on the innovations. Both methods of CK and DK can be implemented through forward filtering and backward sampling of the standard Kalman filter as long as the model is linear and has Gaussian errors.
linearity, the use of the Kalman filter is supposed to be improper. But Kim et al. (1998) develop a method to approximate the \( \log \chi^2 \) distribution with the mixture of seven different normal distributions. Simply put, the approximation is a weighted average of normal distributions. The advantage of this method is that drawing the volatility state can be fully integrated into the Gibbs sampling method, making the Metropolis-Hasting algorithm not necessary.

The choice of the sampling scheme for the mixture innovations in the TVP-VAR model is critically important as it would greatly reduce the computational burden of the whole model. The key issue is how to sample \( k_t \) in an efficient way. Previous sampling methods such as McCulloch and Tsay (1993) and Shephard (1994) generate each \( k_t \) conditional on other \( k_{s \neq t} \), the states and observations. Their sampling scheme is straightforward but inefficient. The \( k_t \) can be highly correlated thus the Markov Chain may move extremely slowly. Gerlach et al. (2000) developed an algorithm in which states can be integrated out analytically so that the posteriors of \( K_t \) are conditioned on the observation and all elements of \( K \) except \( K_t \). As discussed in Gerlach et al. (2000), the matrices \( K_{a,t}, K_{a,t}, K_{h,t} \) can be drawn separately from one another combined with the conventional algorithm used in the state space models for the TVP-VAR model. The details of the MCMC procedure are discussed in Appendix A.

### 1.4 Identification schemes and data

#### 1.4.1 Identification restrictions

To identify government spending shocks, I adopt Ramey (2011)’s narrative approach where military buildups are considered as exogenous shocks to government spending. The narrative approach is generally a better fit for investigating the evolution of causality among macroeconomic variables. The interpretation of the time-varying feature of the model needs to coincide with the identification. Ramey (2011) constructs a new time series called ”defense news series” by sourcing from the major U.S. newspapers starting from the later 19th century to the present. The defense news series is the changes in discounted present value of the U.S. defense spending.
It can be treated as exogenous innovations driving the government spending.\textsuperscript{17}

I follow Ramey (2011) to order the defense news series first, followed by government spending and real GDP growth. This assumption means that shocks to defense news and government spending do not respond to output shocks in the same quarter. The current account scaled by GDP is ordered after real GDP, controlling for the latter. Restrictions in this way make it easier to see the contemporaneous impact of the current account to a government spending shock. I follow the VAR literature to order the real interest rate and the real exchange rate. The real interest rate is ordered following the current account.\textsuperscript{18} The real exchange rate is ordered last, as it is thought to reflect all information available in the economy and respond contemporaneously to macroeconomic conditions.\textsuperscript{19}

The choices of identifying tax shocks are well developed in the existing literature. To be consistent with the narrative approach, I follow Romer and Romer (2010)’s original narrative on tax changes with the extension by Mertens and Ravn (2013) to split the unanticipated part from the anticipated tax component of the tax changes.\textsuperscript{20} Ramey (2016)’s handbook chapter explores a variety of exercises following Mertens and Ravn (2011) and Mertens and Ravn (2014). To facilitate comparison, I conduct Ramey’s exercise on Merten and Ravn’s proxy SVAR approach, adding open economy variables (current account to GDP ratio, real interest rate, and real exchange rate) to the original trivariate VAR system.

The identification scheme is highly relevant to how structural parameters are recovered. From equation (1.4), the relationship between the reduced form and structural form parameters is $A_t^{-1} u_t = \epsilon_t$, and thus $VAR(A_t^{-1} u_t) = A_t^{-1} I_p (A_t^{-1})' = VAR(\epsilon_t) = H_t$, with $A_t$ being lower triangular. The recursive method takes advantage of applying a Cholesky decomposition on the reduced form variance-covariance matrix so that identification is achieved by imposing zeros

\textsuperscript{17}The Ramey (2011)’s narrative approach builds on Ramey and Shapiro (1998) where they create dummy variables to identify the timing of the defense news.
\textsuperscript{18}The real interest rate is constructed from the 3-month Treasury bill. To obtain the real interest rate, I subtract the inflation rate from the nominal rate.
\textsuperscript{19}See, e.g., Sims (1992) for reasoning of ordering financial variables in a VAR.
\textsuperscript{20}By their definition, the criteria to distinguish anticipated and unanticipated tax shocks is whether there is an implementation delay of a tax legislation. If there is a delay, then the tax shock is anticipated.
on some elements of the covariance matrix. Samples from MCMC of $A_t$ and $\sigma_t$ can be used to calculate the impulse responses.

Given the research question of this paper, the most important macroeconomic feature of the model is related to fiscal policy. Once exogenous shocks to fiscal policy are identified, I focus on the impulse responses of the variables that are non-policy related, specifically the current account, interest rate and real exchange rate. Impulse responses are presented in the next section.

### 1.4.2 Data

The data used in this paper span from 1971Q1 to 2014Q4. The length of the sample is chosen to coincide with the floating exchange rate period. Appendix A documents each variable’s information along with its transformation method. All the nominal variables are converted to real terms. The real interest rate is on an annualized basis, and the variables measured by an index (i.e. exchange rates) are also in logarithms. For the defense news series and unanticipated tax series, I keep the same format as Ramey (2011) and Romer and Romer (2010).

### 1.5 Empirical results

I execute 25,000 iterations of the Gibbs sampler, with the first 15,000 draws as "burn-in" to allow convergence of the Markov Chain to the ergodic distribution. The thinning parameter is 10, suggesting the effective number of draws is 1,000. Convergence diagnostics are conducted in the Technical Appendix along with traceplots of selective parameters. The result shows that the convergence check is satisfactory.

#### 1.5.1 Statistical evidence of time-varying parameters

The central idea of the TVP-VAR with mixture innovation is that we can control how flexible the model can be by imposing restrictions on the dummy variable $k_{jt}$ or their related transition probabilities $p_j$. In the model, $K_{\alpha,t}$ controls the breaks in the VAR coefficients along with
the constants. $K_{a,t}$ controls the breaks in the structural parameters in matrix $A_t$, and $K_{h,t}$ the stochastic volatility in the VAR. With different combinations of restrictions, the model could be collapsed to a constant coefficient VAR with homoskedastic errors, or converge to the standard TVP-VAR model as in Primiceri (2005).

As discussed in last section, the choice of priors forms different beliefs about the frequency of structural breaks. In my experiments, I combine different choices of priors and controls on $K_t$ to make comparisons between models. The benchmark model assumes an uninformative prior with no restrictions on the states. That is, all VAR coefficients, structural parameters in matrix $A_t$ and the variances are allowed to vary as the model sees fit. For comparison, I assign an informative prior to all $k_{jt}, j = 1, 2, 3, 4$ in this model, and thus, the hierarchical parameters on the Beta distribution are $\alpha = 0.1$ and $\beta = 10$, for all $j$. This prior suggests that the chance of a changing parameter is near zero ($E(p_j) = 0.01$ with standard deviation 0.01). In Koop et al. (2009), their benchmark priors are $\alpha = \beta = 1$, with $E(p_j) = 0.5$. I also adopt this reference prior on some parameters for which I don’t have firm beliefs of how they will change. If the time series data are persistent, they are less likely to change. My prior for the benchmark model reflects my belief that the structural breaks of the current account dynamic occurred not always gradually but with a few dramatic breaks among the gradual changes. As for the prior on the contemporaneous parameters in $A_t$, I assign the reference prior mentioned above indicating the probability of a change in these parameters in a period is 50 percent. Contemporaneous parameters are important as they are able to tell the effects of a shock on impact. Simply speaking, I put more emphasis on the possible time-varying effect on the contemporaneous parameters than the VAR coefficients. (See Inoue and Kilian (2013)). Though the mixture innovation model cannot choose a specific number of breaks, it can loosen or tighten the prior parameters on the transition probability in order to influence the number of breaks. Table 1.2 lists several combinations of priors and restrictions on $K_t$. I choose relatively informative priors in the VAR coefficients as these are often thought to be subject to change in a TVP-VAR model. I put loose priors on the structural parameters because the contemporaneous
parameters are highly related to economic theory, and they might not change even if there’s a structural break in the data. The MCMC results provide some insights of how frequent parameters change.

Table 1.2: Priors and restrictions on models

<table>
<thead>
<tr>
<th>Model</th>
<th>VAR coefficients: $\alpha_t$</th>
<th>variance: $h_t$</th>
<th>structural para: $A_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{M}_1$: with informative priors</td>
<td>$\alpha_{1,j} = 0.1, \beta_{2,j} = 10$</td>
<td>$\alpha_{1,j} = 0.1, \beta_{2,j} = 10$</td>
<td>$\alpha_{1,j} = 0.1, \beta_{2,j} = 10$</td>
</tr>
<tr>
<td>$\mathcal{M}_2$: with diffuse priors</td>
<td>$\alpha_{1,j} = 1, \beta_{2,j} = 1$</td>
<td>$\alpha_{1,j} = 1, \beta_{2,j} = 1$</td>
<td>$\alpha_{1,j} = 1, \beta_{2,j} = 1$</td>
</tr>
<tr>
<td>$\mathcal{M}_3$: with $A$ constant</td>
<td>$\alpha_{1,j} = 1, \beta_{2,j} = 1$</td>
<td>$\alpha_{1,j} = 1, \beta_{2,j} = 1$</td>
<td>$K_{2t} = 0 \forall t$</td>
</tr>
<tr>
<td>$\mathcal{M}_4$: Constant VAR</td>
<td>$K_{1t}^* = K_{2t}^* = 0 \forall t$</td>
<td>$K_{4t}^* = 0 \forall t$</td>
<td>$K_{3t}^* = 0 \forall t$</td>
</tr>
</tbody>
</table>

The statistical evidence supporting which model fits the data better is determined by calculating the marginal likelihood of the model and the posterior means. Table 1.3 presents these results concerning different models with various priors and restrictions. The marginal likelihood is calculated using the simulation method developed by Chib (1995). The posterior means of the dummy variables in a variety of models with or without restrictions show that there is sufficient evidence to support stochastic volatilities in all models. The posterior mean of $p_4$ exceeds 0.9, indicating high chances of changing. It suggests that the dummy variable controlling the stochastic volatility is not sensitive to the alternative priors.

As for the VAR coefficients and structural parameters, the benchmark model with flat priors ($\alpha_j = \beta_j = 1$) receives better support by the marginal likelihood than the benchmark model with a tighter prior. The odds of both the VAR coefficients and structural parameters in matrix $A_t$ are above 0.5, suggesting evidence that these parameters are likely to change, but not as much as the ones in Primiceri (2005), where all have high probabilities to change.

---

21 In Bayesian statistical literature, justifying a prior scientifically usually is rare. However, one could use as hierarchical priors and derive the marginal likelihood for it to form proper prior choices.

22 Marginal likelihood is a likelihood function where all parameters are integrated out. In the Bayesian context, the higher the marginal likelihood is, the higher the probability that the data fits the model.
### Table 1.3: Priors and restrictions on models

| Model                  | ML  | \(E(p_1|Data)\) | \(E(p_2|Data)\) | \(E(p_3|Data)\) | \(E(p_4|Data)\) |
|------------------------|-----|------------------|------------------|------------------|------------------|
| \(\mathcal{M}_1\): with info priors | 2071.3 | 0.5771(0.3) | 0.5787(0.3) | 0.5785(0.02) | 0.9000(0.02) |
| \(\mathcal{M}_2\): with flat priors | 2124.2 | 0.578(0.3) | 0.5775(0.3) | 0.5776(0.03) | 0.9082(0.02) |
| \(\mathcal{M}_3\): with \(A\) constant | 1940.3 | 0.568 (0.3) | 0.562(0.3) | 0 | 0.976(0.009) |
| \(\mathcal{M}_4\): VAR with SV     | 1880.7 | 0 | 0 | 0 | 0.977(0.03) |

The benchmark model and VAR with stochastic volatilities only have less support relative to the one with flat prior. I should also point out that the priors do not play a significant role influencing the posterior likelihoods. When priors are very informative (second row in Table 1.3), the posterior means of the dummy of the VAR coefficients do not change much compared to the first row. The structural parameters in the matrix \(A_t\) are also not sensitive to prior changes. All these results discussed above deliver a picture that impact responses may have a higher chance of gradual changes than the VAR coefficients on lags. This feature of structural parameters renders macroeconomic implications in terms of policy analysis. To deal with the implied consequences of the change, we need to turn to impulse response analysis in the next subsection.

#### 1.5.2 Impulse response analysis

The nonlinearity of the TVP-VAR with stochastic volatility gives rise to some issues regarding impulse response functions not compatible with a linear VAR. If parameters are not allowed to vary over time, impulse responses are taken directly from a vector moving average (VMA) representation transformed from an invertible VAR, such that

\[
y_t = \sum_{i=0}^{\infty} \phi_i \epsilon_{t-i}
\]
where the MA coefficients $\phi$ are nonlinear function of the VAR coefficients and the structural parameters. An impulse response $h$ step ahead is the proper element of $\phi_h$. A TVP-VAR model has time-varying VMA coefficients:

$$y_t = \sum_{i=0}^{\infty} \phi_{t-i} \epsilon_{t-i}$$

The impulse responses are now a function of time and the coefficients may change at every time period. Since the interest of this paper is to investigate the evolution of the current account’s dynamics across all periods, I plot impulse responses for all time periods. In addition, I choose a few time periods where structural breaks happen for detailed studies and offer economic explanations.

Another issue arising from impulse responses in TVP-VAR model is nonlinearity. Koop et al. (1996) introduce a unified approach to treat impulse responses for both linear and nonlinear models. The impulse responses should be interpreted as the difference of conditional expectations:

$$I_y(h, \epsilon_t, \Omega_{t-1}) = E(y_{t+h}|\Omega_t, \epsilon_t) - E(y_{t+h}|\Omega_t)$$

where $I_y$ is the impulse response function, and $\Omega_t$ is the information set up to time $t$. This function shows the effect on $y_t$ $h$-step ahead conditioned on a unit shock, say, to the forecast error in the measurement equation. In the presence of nonlinearity, the precise calculation of impulse responses should use simulation methods. As Koop et al. (1996) points out, however, simulation methods require heavy computation, and I do not pursue it here. It is easier to compute the conventional impulse responses with the structural parameters in matrix $A_t$ and the states $\alpha_t$. To fully elaborate the impact of fiscal policy shocks, I describe the results of government spending shocks and tax shocks separately.

---

23 That is, to simulate artificial data directly from the joint likelihood function of impulse responses, and subsequently use the data to calculate impulse responses.
1.5.3 Time-varying effects of defense spending shocks

I mainly compare two results between the models under two different prior specifications. The first one is under an informative prior, and the second the reference prior that assume a 0.5 percent probability that a parameter could change in a single period. This comparison reveals how strong the data inform about changes of the VAR coefficients and the structural parameters. The selected impulse responses of four variables in the identified TVP-VAR to a defense news shock are presented in Figure 1.3. These plots are three-dimensional, where the two horizontal axes are time periods and forecast horizon $h$ after a shock hits the system. The vertical axis represents the magnitude of the responses.

To better illustrate the time-varying characteristics of the parameters, Figure 1.4 displays the impact responses across all periods of real GDP growth, the current account, the real exchange rate and the short-term real interest rate. Under the informative prior (a few breaks), the impact responses of the current account to a positive defense news shock are mostly negative, except a small period (3 quarters) in the early 1980s. Regarding the pattern of structural changes, the impact responses help identify that gradual changes are seen from 1983 to 1992, and from 2006 to 2008. The period between 1992 and 2005 exhibits fewer gradual changes but more abrupt changes. The responses of the interest rate are not entirely consistent with what the standard theory predicts, namely that an increase in government spending shock has a crowding-out effect by raising the real interest rate. But it responds negatively on impact during 1983 to 1985, and subsequently from 1998 to 2005. The surprising aspect is the negative and stable impact responses of the real exchange rate. The real exchange rate experiences depreciation against foreign currency in light of a positive defense news shock. This is quite contrary to the results in standard macroeconomic textbooks. A government spending shock generates a negative wealth effect and a crowding-out effect, raising the real interest rate because of the consumption-leisure substitution. The real exchange rate is supposed to rise along with the real interest rate, according to uncovered interest rate parity. Regardless, this result is strikingly similar to Corsetti et al. (2012) and Ravn et al. (2012) where they find real exchange rate
depreciation following a government spending shock in a structural VAR. However, their study does not include the real interest rate, and thus shows no results on responses between the real exchange rate and the real interest rate.

Figure 1.3: Impulse responses to a defense spending shock with an informative prior.

When replacing the informative prior with a flat prior, results with the flat prior illustrate some different patterns of impact responses. Figure 1.5 displays the 3-D version of impulse responses of the four variables and Figure 1.6 presents the impact responses of the same variables.
Figure 1.4: Impact responses to a defense news shock with an informative prior: these impact response functions measure contemporaneous responses of the four variables following a defense news shock. The horizontal axis measures how these responses vary across time while the vertical axis shows the percentage deviation of the variable.

but with the flat prior that assumes an equal probability that each of the parameters will change in a single period. With this uninformative prior, all impact responses display a more frequently changing pattern than under the informative prior. Compared to Figure 1.4, there are two major differences. First, the positive impact responses of the current account are in similar periods as those in Figure 1.4, but the magnitudes are larger in Figure 1.6. From early 1980s to 1988, for instance, impact responses of the current account are gradually changing from positive (0.2 percent deviation from stationary state) to negative 0.38. Note that the abrupt changes in the early 2000s in Figure 1.6 have now become gradual changes. But the 1990s does not experience
much difference than the one under the informative prior.

Figure 1.5: Time-varying impulse responses to a defense news shock with a flat prior.

The findings in both prior specifications do not solely support what the existing literature describes as either "twin deficit" or "twin divergence." Kim and Roubini (2008) find positive effects on the current account from a positive budget deficit shock with U.S. data spanning from 1980 to 2004. Corsetti et al. (2012) use an expanded data set to find the opposite results. Ravn et al. (2012) also discover significant negative responses of the trade deficit to a positive
government spending shock using data from four industrial countries.

Figure 1.6: Impact responses to a government spending shock with a flat prior.

The reversal of the current account responding to a defense news shock in Figure 1.6 is dramatic. In Ramey (2011)'s defense news series, this episode is recognized as the Carter-Reagan Buildup in which defense spending was up considerably in the last moment of the Cold War. According to Ramey’s Defense News Narratives24, there are four defense news shocks from 1983Q3 to 1989Q4. No major defense spending cuts were seen until the fiscal year of 1990. Note that what the defense news series classifies as "shocks" to government spending is anticipated. Ramey’s narrative record acknowledges that delays occur naturally because the Pentagon needs time between decisions and actual implementation. When compared to the timing of the defense

---

24Source: econweb.ucsd.edu/~vramey/research/Defense_News_Narrative.pdf
news shocks, impact responses of the current account appear to be positive at the time when the discounted present value was on the rise, e.g. the early 1980s and mid-2000s. But the opposite occur when the discounted present value was declining, such as in the late 1980s and the 1990s, when Congress tried to balance the budget following the end of the Cold War.

The key question turns to whether shocks identified by defense news series cause current account to have positive impact responses. Ramey (2011) emphasizes the timing of the shock and its impact on other macroeconomic variables. If the defense new shocks get the timing right, it is also crucial to identify in what periods defense spending may cause the current account to become a surplus. Following Ramey’s results using Post-Korean War defense news series, both private consumption, residential investment and nonresidential investment respond negatively to a government spending shock. The average marginal income tax rate rises on impact. The national account identity in Section 1 implies that the current account has to rise to offset the declining investment and increasing government spending, if the increase in private savings is outrun by the declining investment. Note that these results are based on the defense news series getting the timing right. In other words, households and firms anticipated the military buildup several quarters ahead and made intertemporal decisions affecting labor supply, consumption, investment, and savings.\footnote{Neoclassical theory bases on model formulation has mixed conclusion on the effect of fiscal policy. Barro and King (1984), for example, predicts that anticipated fiscal policy changes have no effect on current labor and output because they cannot hedge the current goods against future wealth effects.}

Another episode experiencing positive impact responses is between 2004 Q3 to 2006 Q2. Ramey’s narrative record documenting defense news shocks during this period is rich, as the military spending was anticipated to increase substantially after the September 11 attack and the subsequent wars in Afghanistan and Iraq. In contrast, when defense news shocks are zero, the impact responses of the current account do not exhibit significant time-variations. All are negative and change in small amounts. Relating to the mixture innovation, impact responses of the current account undergo gradual changes during times when defense news shocks are present even with informative priors. When the defense news shocks are zeros, such as the
1990s, the current account demonstrates little variation.

From both Figures 1.4 and 1.5, the real exchange rate depreciates sharply and considerably following a positive defense news shock, which is at odds with the standard neoclassical theory that predicts an appreciation of the real exchange rate through the rising interest rate. Several researchers find similar results of depreciation of real exchange rates following a positive government spending shock. Corsetti et al. (2012) use Ramey (2011) and Blanchard and Perotti (2002)’s approaches in their VAR finding that the real exchange rate depreciates when there is an increase in the exogenous government spending shock. They establish a two-country model with the assumption that debt-financed increases may cause government spending to decline below the steady state level for some time. This leads to decrease in long-term real interest rates and causes exchange rate depreciation in real terms. Kollmann (2010) and Bouakez and Eyquem (2015) also find similar real exchange depreciation but rely on different assumptions to interpret it in their theoretical models. Ravn et al. (2012) develop a model grounded in deep habits to explain exchange rate depreciation: firms lower their markups to capture a higher market share following a government spending shock that raises aggregate demand. This reduction on markups depreciates the home currency. Enders et al. (2011) also find a depreciating real exchange rate using sign restrictions. Note that the nominal exchange rate is defined in prices of home country (U.S. dollar) in terms of a unit of foreign currency.26

Output responses on impact are statistically significantly positive across all time periods. The effect of defense spending on output exhibits enlarging impact multipliers over time, in both specifications with different priors. As shown in Figure 1.5, the recorded impact multiplier is closer to 0.3. This is slightly lower than the one in Blanchard and Perotti (2002) and Gali et al.(2007). The rising trend of impact responses on output shows that government spending as an expansionary fiscal policy has not weakened over time. This contradicts what Pereira and Lopes (2014) find in their time-varying VAR on U.S. fiscal policy. Responses on interest

26 The real exchange rate data are "real effective exchange rate indices" from the Bank of International Settlement. An increase of an index indicates appreciation of the home currency. See https://www.bis.org/statistics/eer.htm for more details.
rates are mixed. They are seen to have negative impact responses in two episodes. One is from 1983Q2 to 1985Q4. The second is from 1998Q2 to 2006Q3. The negative impact responses can be also found in Corsetti et al. (2012), Perotti (2004) and Ramey (2011). They have long been considered as difficult to interpret within the standard theory of a fiscal expansion.27

To summarize, the current account exhibits both positive and negative time-varying impact impulse responses in the model. The periods where positive, gradually changing impact responses are present may be highly associated with the defense news shocks episodes, whereas those periods of negative, abrupt changing impact responses may relate to the times when defense spending remains low. The effect of government spending on real output is consistent with much of the literature, though the impact multiplier is stronger over time. Impact responses of the real exchange rate witness significant depreciation and parameters do not observe time-varying patterns. Finally, the real interest rate responds negatively to a government spending shock in two separate periods, but positively in the remaining periods.

1.5.4 Time-varying effects of unanticipated tax shocks

As discussed in the introduction, there exist several identification approaches on the macroeconomic effect of tax shocks. I follow the narrative approach developed by Romer and Romer (2010) and later extended by Mertens and Ravn (2012), Mertens and Ravn (2013) and Mertens and Ravn (2014) dividing Romer and Romer’s exogenous tax shocks into anticipated and unanticipated shocks. I expand Ramey (2016)’s exercise on Mertens and Ravn (2014) with a specification of a trivariate structural VAR with federal government spending, output and federal tax revenue. I add the current account to GDP ratio, short-term real interest rate and government spending, making it an open economy TVP-VAR.28 Except for the current account to GDP ratio and the short term interest rate, all variables are in real per capita logarithm terms.

27 The effect of fiscal policy on interest rate has been debated in the literature. See Perotti (2004), Favero and Giavazzi (2007), Laubach (2009). While some of these authors find that the impacts may be related to the conduct of monetary policy, I do not discuss monetary policy in this paper.

28 Note that the original method in Mertens and Ravn (2014) is proxy SVAR in which they integrate the external instrument as proxy to the unobservable structural tax shocks.
In addition, the unanticipated tax shock is extracted by Mertens and Ravn from Romer and Romer (2010)’s narrative approach.

Figure 1.7: Time-varying impulse responses to an unanticipated increase in the tax rate with a flat prior.

Figure 1.7 presents the impulse responses to a positive shock of the tax change series mentioned above, with the unanticipated tax change ordered first. Figure 1.8 summarizes the time-varying impact responses for real GDP, the current account, the real interest rate and the real exchange rate to a shock to the tax revenue. Following a positive shock to the unanticipated tax change, few variables indicate a strong pattern of time-varying effects. Real GDP on impact stays approximately 0.01 percent deviation from the trend across all periods. Government spending rises slightly on impact, around 0.015 to 0.017 above trend across periods. While the signs of both government spending and real GDP’s impact responses are consistent with
Mertens and Ravn (2014), the magnitudes are much smaller in an open economy VAR. The second panel on the top in Figure 1.7 shows that real GDP falls to 0.6 percent deviation from the trend immediately in the second quarter after the shock, and it quickly recovers within 2 years. This may be due to the fact that the TVP-VAR uses only one lag while Mertens and Ravn (2014) uses four lags. Impact responses for tax revenue are insignificantly below zero following the shock, which is contrary to Mertens and Ravn’s results. However, tax revenue declines dramatically in their results in the following quarter. A possible reason that tax revenue does not rise on impact as expected is the automatic stabilizer in which tax revenue falls following the real GDP plummets, even if the tax rate increases.

As for the current account, its impact responses display little time variation, compared to the ones following a defense news shock. Note that this result is from the specification with a reference prior, suggesting the data itself convey relatively strong information on the time-varying pattern. Moreover, all impact responses of the current account stay above zero following a positive unanticipated tax shock. By definition, a positive movement in tax changes measures contractionary fiscal policy. With that regard, the positive impact responses of the current account could align with standard theory.
Since no flip of signs of impact responses of the current account to an unanticipated tax shock occurs, the focus here is to explore the gradual changes up until the mid-1990s. The appendix of Mertens and Ravn (2011) documents how they separate unanticipated shocks from anticipated shocks from Romer and Romer (2010). The appendix shows legislations regarding tax changes dominate the 1980s up to the mid-1990s. In addition, several major legislations create significant increases in tax liabilities, though liabilities of different tax categories might vary. For example, the Deficit Reduction Act of 1984 and the Tax Reform Act of 1986 raise tax liabilities by $9.3 billion and $22.7 billion respectively.\(^{29}\) The latter also indicates that a large

\(^{29}\)See the appendix of Mertens and Ravn (2011) for more details
portion of the tax cuts fell on individual income taxes.

The substantial positive impact responses on the current account, as Kim and Roubini (2008) explain, are results from tax shocks that are more likely to be permanent rather than transitory in practice. This is because most of the tax changes are in the form of legislation. These results contradict those theoretical models that assume tax shocks are transitory. For instance, Baxter (1995) argues that negative transitory tax shocks may improve the current account in a calibrated two-country model. A transitory tax cut on labor income increase private savings by more than the fall of public savings, as individuals expect that tax cuts would be financed by lump-sum taxes in the future.

The result on the current account’s responses is largely in line with the empirical literature. Though Romer and Romer (2010) and Mertens and Ravn (2013) do not investigate the effect of tax shocks on the current account, Romer and Romer find that exogenous tax increases are highly contractionary through a negative effect on investment. By using their data, Feyrer and Shambaugh (2012) estimate that one dollar of unanticipated tax cuts in the U.S. worsens the current account deficit by 47 cents. In comparison, the time-varying result shows that the current account improves by a maximum 0.042 percent in 2006 following a one percent positive unanticipated tax shock. The results point to a relatively strong correlation between unanticipated positive tax shocks and the current account improvement, as long as investment spending is an important factor in the current account.

1.6 Conclusion

This paper analyzes the structural evolution of the impact responses of fiscal shocks on the current account in the United States. It does so by applying a time-varying structural vector autoregressive model with mixture innovation. I examine both government spending shocks and tax shocks by adopting narrative measures. My model delivers mixed time-variations on

30 The difference could be due to asymmetric effects of an unanticipated tax change. Put simply, the impulse responses may have different magnitudes between a positive tax change and a negative tax change.
the current account to a positive government spending shock, but little time-variation for the current account as to an unanticipated tax shock. Specifically, the current account on impact has shown positive signs and gradual time-variations when Ramey (2011)’s defense news shocks are positive. There is strong statistical evidence that positive government spending shocks identified by the defense news series lead to a temporary surplus of the current account in the United States, and the surplus represented as impact impulse response functions from the TVP-VAR model shows gradual time-variations instead of abrupt changes. In periods when there are negative shocks to defense news series, the current account displays negative signs on impact and does not show gradual change. In terms of the pattern in time-variation, the current account’s response to the defense news shock is not systematic. The result confirms that positive defense spending shocks cause the current account to have a surplus contemporaneously in the early 1980s and the mid-2000s.

As for unanticipated tax shocks summarized by Romer and Romer (2010), the current account’s impact responses are all consistent with the empirical literature and show little time-variation. The effect of other macroeconomic variables are, overall, smaller in magnitude compared to the existing literature(e.g., fiscal multiplier). The time-variation in the results in turn could be interpreted by those narrative records to identify periods in which responses do not agree with the existing literature.

The time-varying impacts of the current account to fiscal policy shocks leave several directions for future research. My analysis implies that government spending and tax policy have different time-varying patterns on macroeconomic variables. This is important since the transmission mechanism of fiscal policy is an active ingredient of policy evaluation, prediction and other empirical work. The asymmetric time-varying effects of fiscal policy such as Corsetti et al. (2012) is worth further investigation. Given that the state space modeling technique becomes increasingly available, more variables should be included in enriching the model for features that can improve the explanatory and predictive power of empirical models of fiscal policy.
Chapter 2

Common Stochastic Volatility in the Global Financial Cycle

2.1 Introduction

The “Global financial cycle”, first studied by Rey (2015), describes a phenomenon where global financial variables, such as capital flows, asset prices, leverage of global banks and credit growth fluctuate along with VIX, an index measuring the uncertainty of the equity market. The existence of the global financial cycle does not particularly depend on specific countries’ macroeconomic conditions. Rey suggests that one of the determinants of the global financial cycle is the monetary policy of a hegemonic country— the United States.

The availability of data on a global scale facilitates the investigation of the link between global financial variables and the U.S. monetary policy. Miranda-Agrippino and Rey (2018) examine the spillover effects of U.S. monetary policy on cross-border financial flows and the asset market. They estimate the effect of U.S. monetary policy shocks to the global financial market in a Bayesian vector autoregressive model.

This paper uses the dataset of Miranda-Agrippino and Rey (2018) to estimate the common factors of stochastic volatilities in the global financial variables. One of the advantages of this

38
dataset is that it includes credit and bank leverage data that are constructed using cross-border and cross-section statistics involving the world’s major banks. In particular, credit growth, world risky asset prices and bank leverage have shown statistically and economically significant predicting power of financial booms and busts (e.g., Schularick and Taylor (2012), Lund-Jensen (2012)). The availability of a large macroeconomic and financial dataset is certainly beneficial to forecasting and structural analysis. Recently, the literature on vector autoregressions (VARs) has argued that a VAR containing 15-20 variables produces better performance in terms of forecasting and structural analysis than a small system with 3-5 variables.\footnote{Several articles develop different methodologies of coping with the relatively high dimension in a VAR system. See Carriero et al. (2016a), Koop and Korobilis (2013), Chan and Eisenstat (2015).}

There’s convincing evidence that fluctuations in macroeconomic and financial variables are characterized by time-varying volatilities (e.g., Clark and Ravazzolo (2015)). In a Bayesian setting, however, the introduction of time-varying volatility in a large system leads to a loss of symmetry in the posterior likelihood. As Carriero et al. (2016a) points out, each equation in the VAR system possesses a changing volatility, dramatically raising the dimension of the variance-covariance matrix.\footnote{The dimension of the variance-covariance matrix is the number of variables squared times the number of lags (plus one if there is a constant).} With the growing size of the covariance matrix, the computational cost becomes highly expensive.

### 2.1.1 Financial volatility in an AR-SV model

The time-varying volatility is unobservable. To motivate the idea of common factor in the variables’ volatility, we estimate a univariate autoregressive model with stochastic volatility (AR-SV) for 12 selected variables. Consider the following model:

\[
\begin{align*}
\rho(L)y_t &= e^{h_t/2} \epsilon_t, \epsilon_t \sim N(0, 1) \\
    h_t &= \mu + \phi h_{t-1} + \eta_t, \eta_t \sim N(0, \sigma^2_{\eta})
\end{align*}
\]  

\[(2.1)\]  

\[(2.2)\]  

\[\rho(L)y_t = e^{h_t/2} \epsilon_t, \epsilon_t \sim N(0, 1)\]  

\[h_t = \mu + \phi h_{t-1} + \eta_t, \eta_t \sim N(0, \sigma^2_{\eta})\]
with \( E(\epsilon_t \eta_{t+s}) = 0 \) for all \( s \), and \( E(\epsilon_t \epsilon_{t+h}) = E(\eta_t \eta_{t+h}) = 0 \) for all \( h \neq 0 \), where \( h_t \) is the variance in log of observed variables \( y_t \) and it evolves following an autoregressive process with order one. \( L \) denotes the lag length and here we use one lag in the Equation 2.1. The logarithm of the variance is assumed to follow an AR(1) process, which is fundamentally different from the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models where the volatility is a linear function of the past squared innovations. We estimate the model for each of the 12 variables via Markov chain Monte Carlo (MCMC) methods. The estimation is implemented in the R package "stochvol" developed by Kastner (2016).

Figure 2.1 displays the estimated time-varying volatilities from the 12 selected macro and financial variables from the dataset. It can be seen that some variables illustrate a common movement of volatilities over time. For example, the volatilities of U.S. nonresidential investment, Global inflows to non-banks and EU Banking leverage have declining volatilities in the 1990s. Most of the U.S. variables have decreasing volatilities since the 1980s, marking the ”Great Moderation” era. Another common pattern is that most variables witness relatively high movements in volatilities during the 2007-2008 financial crisis. A simple principal component analysis by taking the estimated volatilities of these variables shows that the first principal component accounts for approximately 75\% of the variation among all volatilities.

Taking account of the commonality in volatilities, we propose a Bayesian VAR model with common stochastic volatility (BVAR-CSV) to extract the common factor in the volatility. Following Carriero et al. (2016a), we adopt the idea that each equation’s volatility is decomposed into two components: a common factor and an idiosyncratic variation. When conditioned on the idiosyncratic component, only one factor is driving the time variation in the whole system’s volatility. In addition, we assume the factor loading of conditional volatilities is 1, indicating that the common factor strongly affects all variables. These assumptions allow the model to have a Kronecker structure in the conditional posteriors, therefore facilitating the computational tractability of the model. The assumption of a common factor also corresponds to

---

3The estimated volatilities are estimates from single-equation AR(1) models with stochastic volatility for each variable. The models are estimated based on the MCMC method with 10,000 draws and 2,000 burn-in.
Miranda-Agrippino and Rey (2018)’s effort testing the existence of a global factor summarizing the common variation in the large collection of risky asset prices and banking credits.

Figure 2.1: Estimated Volatilities (defined as standard deviations) of selected variables from AR-SV models. Shaded areas are the 5% and 95% confidence bands.

While the BVAR-CSV has computational advantages over the BVAR with stochastic volatility, it is likely to be misspecified because of its strict restrictions. Carriero et al. (2016a) conduct a comparison between the BVAR-CSV and a general version of it but without the factor loading of one. They find that the loss of the marginal likelihood is small, giving favor to the BVAR-CSV if the dataset is large. Given that the information set in Miranda-Agrippino and Rey (2018) is even larger, we do not repeat the model comparison exercise in Carriero et al’s paper, rather, we estimate three BVAR-CSV models with different innovation specifications. Using
the BVAR-CSV with Gaussian innovation as a benchmark, the alternative models are with Student-t innovations and with a moving average Gaussian innovations. The choice of these models is grounded on both economics and consideration of forecasting performance. On one hand, the model with Student-t innovation has been explored recently in the empirical literature to measure uncertainty of economic and financial crises, as its heavy-tail feature seems to better accommodate extreme movements of financial variables. The non-Gaussian innovations in a VAR are also proven to generate more accurate forecasts. (See, e.g. Chiu et al. (2016)) The VAR model with moving average innovations, on the other hand, has the advantage of identifying fundamental structural shocks in empirical macroeconomics if the moving average process is invertible. (See, e.g. Chen et al. (2017))

The common factor volatility models in this paper are different from dynamic factor models with stochastic volatility. Mumtaz and Surico (2012) propose a dynamic factor model with stochastic volatility to estimate the common factor of inflation rates in a cross-country setting. The latter emphasizes the cross-variable correlations along with time-varying volatilities. The multivariate factor volatility models in this paper investigate how the factors drive common variations in the volatilities only. Though using the same dataset, the primary interest of this paper departs from the objective in Miranda-Agrippino and Rey (2018) where they explore the international transmission mechanism of U.S. monetary policy to the global financial market. We do not address the economic causal relations in our models, noting that there is always a trade-off between model size and its interpretability. Having considerable information allows a model to produce more precise forecasts than a model with less information, but the degree of model misspecification may also be on the rise. We recognize this trade-off and restrain our interest on whether our models can capture the common volatility as a measure of uncertainty and their forecasting performance.

As a dimension reducing approach in multivariate volatility study, the method of common factor stochastic volatility should not be confused with principal volatility component (PVC) analysis. Hu and Tsay (2014) borrow the idea of principal component analysis (PCA) that an
eigenvalue close to zero implies the existence of a stable linear combination between variables. They compute the spectral decomposition of the kurtosis matrix to detect linear relationships of multiple time series having no conditional heteroskedasticity. Note that PCA produces zero contemporaneous correlations of the time series data, ignoring the dynamic dependence between volatilities.

To evaluate model fit and model performance, we compute the integrated likelihoods with three models and compare the loss between the BVAR-CSV and the other two alternative models. We find that the BVAR-CSV with student-t innovations has an edge over the other two models. We also compute the root mean square forecast error (RMSFE) and the average log predictive likelihood to compare these models’ forecasting performance. There is evidence that the data prefer the BVAR-CSV with student-t innovations when we include the recent financial crisis as training periods. The forecasting performance is in favor of the benchmark model when the financial crisis episode is not included. The forecasting result fits the "No free lunch" theorem, argues Wolpert (1996), that there is no single model which always does better than any others without substantive information. Due to this principle, we try a variety of specifications and then determine which model we should focus on.

The remainder of the paper is organized as follows. Section 2 discusses the two competing models. Section 3 presents posterior analysis based on the Kronecker structure of the covariance, and the estimation procedures. Section 4 describes the data and conducts a forecasting exercise between competing models. Section 5 concludes and discusses the future direction of finding the common factor of volatilities.

**2.2 Volatility structure in a Bayesian VAR**

In this section I introduce a general framework for modeling the common factors in the covariance of a large Bayesian VAR. Specifically, I outline the covariance structure with many factors as a general model, and then I consider the Bayesian VAR with a common stochastic volatility that is a special case of the general model.
Let $y_t$ denote the $n \times 1$ vector of observed variables of interest. First consider the generic structural VAR($p$) model:

$$ y_t = B_0 + \sum_{j=1}^{p} B_j y_{t-j} + u_t $$

where $B_0$ is an $n \times 1$ vector of intercepts and $B_j, j = 1, ..., p$ are all $n \times n$ coefficient matrices. Let $u_t$ denote the reduced form shocks, and its relationship with the structural shocks is:

$$ u_t = A^{-1}\Omega_t^{0.5} \epsilon_t, \epsilon_t \sim iid \mathcal{N}(0, I) $$

where $A$ is an $n \times n$ lower triangular matrix with ones on the diagonal, and $\Omega_t$ is a diagonal matrix containing volatilities. In a standard VAR the innovations $u_1, .., u_T$ are assumed to be independent and identically distributed (iid) as $\mathcal{N}(0, \Sigma_t)$, which implies a time-varying variance for the innovations:

$$ var(u_t) = \Sigma_t = A^{-1}\Omega_t A^{-1'} $$

To describe the factor structure in the covariance, an arbitrary element of the diagonal matrix $\Omega_t$ can be written as

$$ \omega_{jt} = f_t^{a_j} h_{jt} $$

where $f_t$ represents a common factor and $h_{jt}$ is the idiosyncratic component related to the $j^{th}$ variable in the VAR. The log volatility is thus given by a linear factor model:

$$ \log \omega_{jt} = \alpha_j \log f_t + \log h_{jt}, \ j = 1, ..., n. $$

The idiosyncratic component $h_{jt}$ is assumed to follow (in logarithm) an autoregressive (AR) process:

$$ \log h_{jt} = \rho_j \log h_{jt-1} + \nu_{jt}, \nu_{jt} \sim iid \mathcal{N}(0, \sigma^2_h) $$

Note that the elements of $\nu_t$ are jointly distributed as i.i.d. Normal $\mathcal{N}(0, \Sigma_h)$ with $\Sigma_h = diag(\sigma^2_{h_1}, ..., \sigma^2_{h_n})$. 

44
2.2.1 Kronecker structure of the covariance matrix

The model presented in Equations (1)-(6) describes a general form in which common factors enter the covariance matrix of a Bayesian VAR. The posterior variance of the VAR coefficients cannot be easily estimated if the dimension of the system is large (usually 10-20 variables). To see this, rewrite Equation (2.1) as the following:

\[ Y_t = B x_t' + u_t \]  \hspace{1cm} (2.9)

where \( B = (B_0, B_1, ... B_p) \) is \( n \times k \) and \( x_t = (1, y_{t-1}, ..., y_{t-p}) \) is a \( k \times 1 \) of lags and an intercept with \( k = np + 1 \). Then, stacking the observation over \( t = 1, ..., T \), we have

\[ Y = XB + U. \]  \hspace{1cm} (2.10)

The matrices \( Y, X \) and \( U \) are respectively of dimensions \( T \times n, T \times k \) and \( T \times n \). In a vectorized form, the VAR innovations can be written as \( vec(U) \sim N(0, \Sigma_t \otimes I_T) \), where \( I_T \) is an identity matrix with dimension \( T \), \( \otimes \) is the Kronecker product and \( vec(.) \) denotes the operator stacking columns of a matrix to a column vector.

Following Chan (2015), replacing \( I_T \) with a more general \( T \times T \) covariance matrix \( \Omega \) such that

\[ vec(U) \sim N(0, \Sigma_t \otimes \Omega) \]  \hspace{1cm} (2.11)

This parametrization has two advantages. First, it allows cross-sectional structure and serial covariance structure to be influenced by \( \Sigma_t \) and \( \Omega \) separately. Chan (2015) specifies non-Gaussian innovations, heteroskedastic innovations and serially dependent innovations through this Kronecker structure of the covariance. Second, this structure alleviates the computational complexity by imposing some simple restrictions on the covariance. The framework of Bayesian VAR with common drifting stochastic volatility that Carriero et al. (2016a) propose takes advantage of the Kronecker structure to improve computational efficiency.
2.2.2 Common factor volatility in a Bayesian VAR

One restriction to facilitate computational gain is to impose the factor loading $\alpha_j$ to be equal to 1 for all variables, and the idiosyncratic component $h_{j,t}$ not to vary over time, capturing scale differences of variance for each variable. Formally,

$$\alpha_j = 1, \ h_{j,t} = 1, \forall t, j$$  \hspace{1cm} (2.12)

Equation (2.4) becomes $\omega_{jt} = f_t$. Note that $\omega_{jt}$ is an arbitrary element of the volatility matrix $\Omega_t$. Equation (2.2) can be written as

$$u_t = A^{-1} f_t^{0.5} \epsilon_t$$  \hspace{1cm} (2.13)

The purpose of the restrictions is to estimate the time-varying common factor $f_t$. It implies that all time variations in the volatilities across variables are governed only by $f_t$. Simply put, the time-varying innovations in Equation (2.3) can be decomposed as the product of the time-varying common factor $f_t$ and a constant variance matrix $\Sigma$:

$$\Sigma_t = f_t \Sigma$$  \hspace{1cm} (2.14)

This factorization needs to redefine the structural matrix $A$ to capture differences in variance scales. Given that $\Sigma$ is a positive matrix, define $\Sigma = D^{-1} D^{-1}$ in which the elements of $D^{-1}$ are structural parameters. Note that the structural parameters in $D^{-1}$ are not allowed to change over time. The fact that $D^{-1}$ is time-varying would add additional $\frac{n(n-1)}{2}$ state equations as in (6). Moreover, Primiceri (2005) detects little time variations in the trivariate system. The main challenge of the estimation lies not in the sampling of structural coefficients and volatilities, but in the inverse covariance matrix of the VAR coefficient $B$.

To complete the state-space system, assume that the common factor $f_t$ follows an AR(1)
where \( \epsilon_t^f \sim \mathcal{N}(0, \sigma_f^2) \).

### 2.3 Bayesian Estimation

In this section we start by discussing the posterior analysis of a standard Bayesian VAR and outline the challenges of estimating a high dimensional system. We then briefly introduce the method of triangularization developed by Carriero et al. (2016b) to estimate different factors in a large Bayesian VAR.

#### 2.3.1 Posterior Analysis

In matrix form, the likelihood implied by Equation (2.8) is given by

\[
l(Y|B, \Sigma_t, \Omega) = (2\pi)^{-\frac{np}{2}} |\Sigma_t|^{\frac{n}{2}} |\Omega|^{-\frac{3n}{2}} \exp\left(-\frac{1}{2} \text{trace}\left(\Sigma_t^{-1}(Y - XB)'\Omega^{-1}(Y - XB)\right)\right)
\]

(2.16)

where \( \exp(.) \) denotes the exponential operator, and \( \text{trace}(\cdot) \) represents the trace of a matrix.

Define the parameter space as

\[
\Psi = \{B, \Sigma_t, \Omega\}
\]

(2.17)

A natural way to assign priors to the parameters is to consider the three parameter blocks are mutually independent such that \( p(B, \Sigma_t, \Omega) = p(B)p(\Sigma_t)p(\Omega) \). But computation can become more efficient if we treat the parameter blocks \( \{B, \Sigma_t\} \) and \( \Omega \) as independent. This is because we can use a standard Normal -Inverse Wishart for \( \{B, \Sigma_t\} \) ( see, e.g. Kadiyala and Karlsson (1997) for examples of constructing such priors )

\[
\Sigma_t \sim \mathcal{IW}(d_s, \Sigma), \quad \text{vec}(B|\Sigma_t) \sim \mathcal{N}(\text{vec}(B_0), \Sigma_t \otimes \Omega_B).
\]

(2.18)

Clark and Ravazzolo (2015) present evidence that AR and Random walk specifications are comparable in terms of out-of-sample forecasting.

---

\(^4\text{Clark and Ravazzolo (2015) present evidence that AR and Random walk specifications are comparable in terms of out-of-sample forecasting.} \)
The prior covariance matrix $\Omega_B$ serves the purpose of shrinkage. (e.g. the Minnesota priors in Litterman (1986)). The specific form is given in the following section. Given the conditional priors, the joint density function is:

$$g(B, \Sigma_t) \propto |\Sigma_t|^{-\frac{d_s+n+k}{2}} \exp\left(-\frac{1}{2} \text{trace}\left[(\Sigma_t^{-1}\Sigma) + \Sigma_t^{-1}(B - B_0)\Omega_B^{-1}(B - B_0)\right]\right)$$

(2.19)

Given the natural conjugate priors, posteriors can be derived from combining the likelihood in Equation (2.14) and the joint density for the prior in Equation (2.17):

$$l(B, \Sigma_t|Y, \Omega) \propto |\Sigma_t|^{-\frac{d_s+n+k+T}{2}} \exp\left(-\frac{1}{2} \text{trace}(\Sigma_t^{-1}\Sigma)\right)$$

$$\times \exp\left(-\frac{1}{2} \text{trace}\left(\Sigma_t^{-1}\left[(B - B_0)\Omega_B^{-1}(B - B_0) + (Y - XB)'\Omega^{-1}(Y - XB)]\right)\right)\right)$$

$$= |\Sigma_t|^{-\frac{d_s+n+k+T}{2}} \exp\left(-\frac{1}{2} \text{trace}(\Sigma_t^{-1}\left[\Sigma + (B_0\Omega_B^{-1}B_0 + Y\Omega^{-1}Y - H'W_B^{-1}H)]\right)\right)$$

$$\times \exp\left(-\frac{1}{2} \text{trace}(\Sigma_t^{-1}(B - \tilde{B})'W_B(B - \tilde{B}))\right)$$

where $H = (\Omega_B^{-1}B_0 + X'\Omega^{-1}Y)$, $W_B = \Omega_B^{-1} + X'\Omega^{-1}X$, and $\tilde{B} = W_B^{-1}H$. Note that the fourth line of Equation (2.18) has a quadratic form. This implies that the posterior $l(B, \Sigma_t|Y, \Omega)$ has a Normal-Inverse Wishart structure:

$$\Sigma_t|Y, \Omega \sim IW(d_s + T, \tilde{\Sigma})$$

(2.21)

$$vec(B)|Y, \Sigma_t, \Omega \sim N(vec(\tilde{B}), \Sigma_t \otimes W_B^{-1})$$

(2.22)

where $\tilde{\Sigma} = \Sigma + (B_0\Omega_B^{-1}B_0 + Y\Omega^{-1}Y - H'W_B^{-1}H)$. The size of the covariance matrix $\Sigma_t \otimes W_B^{-1}$ is $n \times k$ (where $k = np+1$). Sampling the VAR coefficients $B$ requires the inverse of this covariance matrix. The computational difficulty lies in the inverse of it involving $O(n^6p^3)$ operations. The Kronecker structure of the covariance in fact allows us to reduce the computational complexity to $O(n^3p^3)$ by avoiding inversing the covariance matrix directly. Specifically, I adopt Chan (2015)’s computational shortcut to speed up the computation of the covariance matrix.
To avoid direct inverse of the covariance matrix, we can first use the Cholesky factor $L_W$ of the matrix $W_B$ such that $L_W L'_W = W_B$. Then we construct $\tilde{B}$ by employing the forward (backward ) substitution:  

$$ \tilde{B} = W_B^{-1}H = (L_W L'_W)^{-1}H = (L_W^{-1})' L_W^{-1} (\Omega_B^{-1} B_0 + X' \Omega^{-1} Y) \tag{2.23} $$

Let $L_\Sigma$ be the Cholesky factor of $\Sigma_t$, and construct a new statistics $G$ such that

$$ G = \tilde{B} + (L'_W)^{-1} Z ) L'_\Sigma \tag{2.24} $$

where $Z$ is a random matrix of i.i.d. $N(0, I_{k \times n})$. Hence, the purpose of constructing $G$ is to simplify the sampling process of the posterior of $\tilde{B}$. In the appendix, we show that $G$ follows a matrix normal distribution:

$$ G \sim MN(\tilde{B}, \Sigma_t \otimes W_B^{-1}) \tag{2.25} $$

Note that in computing $W_B = \Omega_B^{-1} + X' \Omega^{-1} X$, there is no need to inverse $\Omega$ of dimension $T \times T$. Instead, one can use the Cholesky factor of $\Omega$ to construct $W_B$, given that most of the macroeconomic time series do not have a large $T$.

### 2.3.2 Implementation

Given Equation (2.19) and (2.20), latent variables can be modeled in the covariance matrix $\Omega$. In the case of modeling common factor stochastic volatility, $\Omega = diag(f_1, ..., f_T)$, which is also a diagonal matrix. In this subsection I briefly discuss the implementation of the MCMC algorithm, while the detailed algorithm is documented in the appendix.

For parameters $\phi$ and $\sigma^2_f$, we adopt a normal and inverse-gamma respectively,

$$ \phi \sim N(\phi_0, \Sigma_\phi), \quad \sigma^2_f \sim IG(d_h, V_h) $$

\footnote{The forward and backward substitution is available in computing language such as Matlab and R.}
The posterior draws can be obtained from the following steps:

Step 1: Sample the VAR coefficients and the covariance matrix of from \( l(B, \Sigma | Y, \Omega, \phi, \sigma_f^2) \);

Step 2: Sample the covariance matrix of latent variables from \( l(\Omega | Y, B, \Sigma, \phi, \sigma_f^2) \). Note that from step 1 that the common factor \( f_t \) has been factored out from the relation \( \Sigma_t = f_t \Sigma \), and thus \( f_t \) can enter \( \Omega \). In fact, the posterior in this step can be factored as:

\[
  l(\Omega | Y, B, \Sigma, \phi, \sigma_f^2) = l(\mathbf{f} | Y, B, \Sigma, \phi, \sigma_f^2) \propto l(\mathbf{f} | \phi, \sigma_f^2) l(Y | B, \Sigma, f) 
\]

(2.26)

where \( \mathbf{f} = (f_1, ..., f_T)' \) is a vector column stacking all time-varying factors. And \( l(\mathbf{f} | \phi, \sigma_f^2) \) is a normal density implied by the state Equation (2.13). Given \( B, \Sigma, f \), the density of \( Y \) is evaluated at

\[
  l(Y | B, \Sigma, f) = \prod_{t=1}^{T} l(y_t | B, \Sigma, f_t) = \text{constant} - \prod_{t=1}^{T} \left( -\frac{n}{2} f_t - \frac{1}{2} \exp(-f_t) \mathbf{u}'_t \Sigma^{-1} \mathbf{u}_t \right) 
\]

(2.27)

Sampling \( f \) from Equation (2.24) can be implemented directly using a Metropolis-Hastings step. (See, e.g. Jacquier et al. (2002) ) Differing from Carriero et al. (2016a) that samples \( f \) from a mixture normal developed by Kim et al. (1998), the Metropolis-Hastings (M-H) step is a single move while the former is multi-move. The advantage of the M-H method is that it is not necessary to draw extra state variables as in Kim et al. (1998). Chan and Eisenstat (2015) shows that the two methods are very similar in a large Bayesian VAR in terms of computational time.

Step 3: Sample the AR coefficient of the common factor from \( l(\phi | Y, B, f, \Sigma, \sigma_f^2) \);

Step 4: Sample the variance of the common factor from \( l(\sigma_f^2 | Y, B, f, \Sigma) \).

The last two posteriors fall into the standard normal and inverse-gamma categories and can be easily implemented. The entire algorithm can be seen as "Metropolis within Gibbs" sampling because drawing the states involving nonlinearity.
2.4 Data and Empirical Results

2.4.1 Data

Data used in this paper are collected from Miranda-Agrippino and Rey (2018) that is available on Helene Rey’s website. The sample period is from 1980Q1 to 2012Q4. The original dataset has several components: 1. data sourced from IMF’s International Financial Statistics (IFS) and the Bank for International Settlements (BIS) database, such as domestic credit and cross-border credit. 2. data relating to banking and bank leverage are constructed by the authors. 3. Macroeconomic data that have open access such as the U.S. and U.K. macro variables are available in either IFS or the St. Louis Fed’s website. 4. Data measuring global factor and volatility are estimated by the authors. We select 18 variables according to the need of this paper. Appendix B lists all the variables used in this paper and gives details of their transformation methods. For more details on the constructed data in the original form, see the online appendix of Miranda-Agrippino and Rey (2018).\(^6\)

2.4.2 Common drifting volatility in the global financial cycle

To model a common stochastic volatility is not only a computational advantage, but also shows empirical evidence of whether there is a common factor driving the time-varying volatility, given the large dataset including the global financial variables. Figure 2.2 presents the posterior means of the common stochastic volatility models for comparison: BVAR-CSV as a benchmark, BVAR-CSV with Student-t innovations, and BVAR-CSV with moving average innovations. The time-varying factor volatilities of BVAR-CSV and BVAR-CSV with moving average innovations display a similar pattern. The moving average specification produces higher volatility in major financial crisis periods, e.g., the 2007-2008 financial meltdown. The BVAR-CSV with Student-t innovations, however, shows a smooth pattern of time-variation in the volatility. This is due to the heavy tail nature of the Student-t distribution to a large extent, indicating that in any

given period, the probability of occurrence of extreme values is relatively higher than for the Gaussian innovations. In other words, BVAR models with Student-t innovations are more likely to accommodate periods that variances are highly volatile.

A measure to gauge if the common factor is capable of capturing the commonality in volatility is the first principal component obtained from the estimated volatilities of a VAR model with stochastic volatility. We use all 18 variables to estimate a VAR model with stochastic volatility for each variable, and then extract the volatilities from the covariance matrix to perform a factor analysis. We normalize each variable’s volatility before computing the principal component. We find that the commonality in the stochastic volatility appears to be strong as the first principal component accounts for 83% of the overall variation, with a standard deviation of 3.07. Carriero et al. (2016a) estimate both the BVAR-CSV and BVAR without imposing one common factor using 14 U.S. macro variables. They find that the correlation between the two common volatilities is 0.99. Given the large portion of variation that the first principal component explains, the common factor estimated from the BVAR-CSV thus appears to be an effective estimate capturing the time variation in conditional volatilities.

When comparing with the major volatility index, we calculate the correlation coefficients using the mean of estimated volatilities in the three models and two volatility indices, including the VIX and MSCI Log Realized Variance. The BVAR-CSV and BVAR-CSV with MA innovations seem to better align with the volatility indices, with correlation 0.45 and 0.48 with the VIX, respectively. The correlation of the BVAR-CSV with Student-t innovation is only 0.23. Note that by construction the CBOE VIX is calculated using only options prices, reflecting agents’ expected volatility. Since the dataset contains a significant amount of data in the global banking sector, it is possible to map estimated volatilities to dates when major banking crises happened in the past four decades, including the savings and loans crisis in the U.S. and the U.K., the stock market crash in 1989 and the Asian financial crisis during 1997-1998.7

Despite the structural analysis not being a focus in this article, we still want to have a general

---

7Reinhart and Rogoff (2009) document many global episodes of financial and banking crisis in the past several centuries.
assessment of our model assumptions. Our model only allows the elements in the covariance matrix to be time-varying but not the VAR coefficients. Figure 2.4 displays a heat map of all the VAR coefficient estimates from the BVAR-CSV model. The purpose of the heat map is to visualize numerical values by assigning different colors on each cell that represent a VAR coefficient. If the heat map shows the more "hot" colors, it indicates that the correlations between the time series are evolutionary. In this case, a time-varying parameter model may be a better fit. Fortunately, most of the VAR coefficients in Figure 2.4 are smaller than one, implying that the coefficients do not exhibit a large degree of time variation.

The VAR model is estimated based on four lags. The lag length selection is conducted through AIC and BIC.
Figure 2.2: Posterior mean of common drifting volatility: BVAR-CSV is the benchmark model.
2.4.3 Forecasting exercise

In this subsection we focus on evaluating the forecasting performance of the proposed large Bayesian VARs in terms of both point and density forecasts. We use the full sample (1980Q1 - 2012Q4) including the 18 variables to forecast five variables of interest: Global inflows to banks, Global domestic credit, US domestic credit, EU bank leverage and term spread. These variables are key measures of the global financial cycle and more likely to have a common factor stochastic volatility. Particularly, we compare the forecasting performance of three models: BVAR-CSV,
BVAR-CSV with a student-t innovation, and BVAR-CSV with a first order moving average innovation. Given the general Kronecker covariance structure introduced in Section 3, the latter two models are easily implemented in the \( \Omega \). In the case of common stochastic volatility with
a student-t distribution, Equation (2.12) can be characterized as

$$\Sigma_t = \gamma_t \Sigma$$

where $\gamma_t \sim IG(\lambda/2, \lambda/2)$. This setup allows the reduced form innovation $u_t$ to marginally follow a Student-t distribution with a degree of freedom $\lambda$. We follow Chan (2015) to use a uniform prior on $\lambda \sim U(1, 100)$. (See Chan and Hsiao (2014) for estimation of VAR with heavy tail innovations)

To evaluate conditional density forecasts, we use the sample up to time $t < T$ for each model, denoted as $y_{1:t}$, and construct the joint predictive density $f(y_{t+h}|y_{1:t})$ given the model parameters, where $h$ represents the desired forecast horizon. More specifically, at each $t$-th replication, we use the data $y_{1:t}$ to estimate the model parameters and construct the predictive density. We then extend the data up to $t+1$ and continue the procedure until time $T-h$. After the replications, density forecasts can be obtained for each model from the desired specific period to the end of the sample.

Similar to the marginal likelihood, the joint predictive density $f(y_{t+h}|y_{1:t})$ cannot be computed analytically. For Bayesian VARs, the conditional density of $y_{t+h}$ given the data and the model parameters follows a Gaussian distribution and can be estimated by averaging $y_{t+h}$ over the MCMC draws of the model parameters. We then compute the the predictive mean $E(y_{i,t+h}|y_{1:t})$ as the point forecast for variable $i$. We repeat the whole procedure by moving one period forward with data $y_{1:t}$. These point forecasts are computed and evaluated for the period $t = t_0, ..., T-h$. After obtaining the point forecast, we employ the root mean squared forecast error (RMSFE), a commonly used metric, to evaluate the forecast performance. The metric is defined as

$$RMSFE = \sqrt{\frac{\sum_{t=t_0}^{T-h}(y_{i,t+h}^R - E(y_{i,t+h}|y_{1:t}))^2}{T - h - t_0 + 1}}$$

(2.28)

where $y_{i,t+h}^R$ denotes the realized value of the variable $y_{i,t+h}$ as we pretend that the sample

---

9This point forecast method is similar to n-fold cross-validation, but we only need to divide the full sample into two periods instead of the common five folds.
beyond $t_0$ is unknown to a forecasting practitioner. Table 2.1 compares the point forecasting performance in different models. The entries in the BVAR-CS are actual values of the RMSFE while the entries in the competing models are ratio between the alternative model and the benchmark model. If a ratio entry is larger than one, it indicates that the model has a higher RMSEF than the benchmark model at a given forecast horizon. The BVAR-CS with Student-t innovations gains some advantage compared to the other two models, with RMSFEs of global domestic credit, US domestic credit and EU bank leverage that are less than one. But the differences in performance are not significant. The BVAR-CS with Gaussian moving average innovations has the largest RMSFE at almost all horizons across variables among the three models.

While the point forecasts are not sufficient to have a thorough assessment on models’ performance, we then move to evaluate the density forecasts denoted as $f(y_{t+h}|y_{1:t})$. The metric we use is the average of log predictive likelihoods:

$$
\frac{1}{T - t_0 - h + 1} \sum_{t=t_0}^{T-h} \log f(y_{t+h} = y^R_{i,t+h}|y_{1:t})
$$

(2.29)

Again, the measuring criteria is to gauge how likely the realized outcome $y^R_{i,t+h}$ falls into the predictive likelihood. If the realized value is below the density forecast, the numerical value of the predictive likelihood will be large and vice versa. Table 2.2 summaries the average log predictive likelihoods for the forecasting horizon $h = 1, 2, 3, 4$ for the five key variables. For this metric, larger values suggest better performance of forecasting as the distance of predicted value and realized value is small.

Cumulative sums of the log predictive likelihood in Equation (2.27) can be used as a model comparison through cumulative log predictive Bayes factors. Let $f_t(A)$ denote the predictive density of model $A$ at time $t$, and $f_t(B)$ the corresponding value of model $B$, the cumulative log predictive Bayes factor at time $T - h$ in favor of model $A$ over model $B$ is given as
Table 2.1: Root Mean Square Forecasting Error

<table>
<thead>
<tr>
<th>Variable</th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 3Q$</th>
<th>$h = 4Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BVAR-CSV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global inflows</td>
<td>6.376</td>
<td>7.662</td>
<td>9.903</td>
<td>11.226</td>
</tr>
<tr>
<td>Global domestic credit</td>
<td>6.936</td>
<td>8.646</td>
<td>10.497</td>
<td>11.688</td>
</tr>
<tr>
<td>US Domestic credit</td>
<td>3.562</td>
<td>4.427</td>
<td>6.094</td>
<td>7.736</td>
</tr>
<tr>
<td>EU Bank leverage</td>
<td>0.0854</td>
<td>2.245</td>
<td>4.148</td>
<td>5.908</td>
</tr>
<tr>
<td>Term Spread</td>
<td>4.918</td>
<td>2.574</td>
<td>0.415</td>
<td>1.412</td>
</tr>
<tr>
<td><strong>BVAR-CSV-t</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global inflows</td>
<td>1.001**</td>
<td>0.994**</td>
<td>1.005*</td>
<td>1.002*</td>
</tr>
<tr>
<td>Global domestic credit</td>
<td>0.992**</td>
<td>0.961**</td>
<td>0.971</td>
<td>0.973</td>
</tr>
<tr>
<td>US Domestic credit</td>
<td>0.956**</td>
<td>0.985**</td>
<td>0.992**</td>
<td>0.988**</td>
</tr>
<tr>
<td>EU Bank leverage</td>
<td>2.459</td>
<td>0.936*</td>
<td>0.964*</td>
<td>0.959</td>
</tr>
<tr>
<td>Term Spread</td>
<td>1.023**</td>
<td>1.066**</td>
<td>1.455</td>
<td>0.837</td>
</tr>
<tr>
<td><strong>BVAR-CSV-MA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global inflows</td>
<td>1.028**</td>
<td>1.019**</td>
<td>1.019**</td>
<td>1.018*</td>
</tr>
<tr>
<td>Global domestic credit</td>
<td>1.010**</td>
<td>1.023**</td>
<td>1.024</td>
<td>1.026</td>
</tr>
<tr>
<td>US Domestic credit</td>
<td>0.989**</td>
<td>1.059**</td>
<td>1.047**</td>
<td>1.0356**</td>
</tr>
<tr>
<td>EU Bank leverage</td>
<td>0.287</td>
<td>1.048*</td>
<td>1.027*</td>
<td>1.025</td>
</tr>
<tr>
<td>Term Spread</td>
<td>0.982**</td>
<td>0.959**</td>
<td>0.597</td>
<td>1.163*</td>
</tr>
</tbody>
</table>

Note: Entries of BVAR-CSV are benchmark. Other entries are ratio of the RMSFE of the benchmark model. An entry less than one indicates that the model has a lower RMSFE than the benchmark at a given forecast period. To provide the significant level of the accuracy, we use the Diebold-Mariano $t$-statistic for equal MSE. Differences from zero that are statistically significant are denoted by one or two asterisks corresponding to significance levels of 10% and 5% respectively.
\[
\log \left[ \frac{f_A(y_{t_0:T-h}|y_{1:t})}{f_B(y_{t_0:T-h}|y_{1:t})} \right] = \sum_{t=t_0}^{T-h} \log \left[ \frac{f_t(A)}{f_t(B)} \right] = \sum_{t=t_0}^{T-h} \log [\log f_t(A) - \log f_t(B)] \tag{2.30}
\]

When the cumulative log predictive Bayes factor is positive at a given time period, there is evidence in favor of model A, and vice versa.

Using the BVAR-CSV as a benchmark, the average log predictive density of the BVAR-CSV with Student-t innovations has better forecasting performance across the selected variables than the benchmark model. The BVAR-CSV with a first order moving average Gaussian innovation does not outperform the benchmark model except for predicting Global inflows. It’s worth noting that the training sample used to construct density forecasts includes the 2007-2008 financial crisis. However, if the crisis periods are not included, the BVAR-CSV with Student-t innovations cannot outperform the benchmark model across all variables. One possible explanation is that the model with Student-t innovations are heavy-tails, enabling it to accommodate extreme volatilities. In terms of comparing models, it’s important to recognize that their forecasting performance also depends on the training samples used to evaluate the model parameters.

2.4.4 MCMC Convergence and integrated likelihoods

The MCMC results are produced based on sampling 25,000 draws, discarding the first 5,000. The effective draws are thus 20,000 with no tuning. We evaluate the convergence properties of the MCMC algorithm for BVAR-CSV model by computing the inefficiency factor defined by

\[
1 + 2 \sum_{k}^{K} \rho_k
\]

where \( \rho_k \) is the MCMC sample autocorrelation function at lag length \( k \). \( L \) is usually chosen large enough ensuring the decay of autocorrelations. If the posterior draws after burn-in is independent, the corresponding inefficiency factor is close to 1. That an efficiency factor is 20, for example, indicates that approximately 2,000 draws are needed to be equivalent to have 100
independent draws. The common practice is to sample from the posteriors with a relatively smaller number of draws, say, 1000. Then using these draws to calculate the inefficiency factor that provides a rough guess of what draws in total are desired.

Table 2.3 reports the convergence statistics for the BVAR-CSV model. The convergence results are comparable to Carriero et al. (2016a) where theirs have fewer variables and parameters. The mean inefficient factor for the state variable $f$ is relatively high, indicating that the time-varying common volatility may be highly persistent. As mentioned in Section 2, this common factor could be modeled as a random walk without increasing the computational burden.

To gauge the overall performance of the three models, we compute the log integrated likelihoods of them and present their numerical values along with standard errors in Table 2.4. Compared to the benchmark model, the BVAR-CSV with student-t fits the data better, with the log Bayes factor about 25. The benchmark model and BVAR-CSV with MA innovations are not significantly different with the log Bayes factor 7.

## 2.5 Concluding remark

In this paper, we estimate a Bayesian VAR with common factor stochastic volatility. The proposed model provides evidence that there is a single time-varying factor in the volatility driving the global financial cycle in the past three decades. We then conduct forecasting exercises and compare the performance among three BVAR models with different specifications in the innovation: 1. common factor stochastic volatility with Gaussian innovations; 2. common factor volatility with Student-t innovation; 3. common factor volatility with Gaussian moving average innovations. In terms of economic interpretation, we show that the first and the third specifications are better at revealing the time-variation of volatilities of the global financial variables. The estimated volatilities reflect some major financial turbulence in a global scale. The forecast performance of the three models vary depending on the forecasting periods. When including the recent episode of financial crisis, the BVAR-CSV with Student-t innovation produces point and density forecasts at a relatively more accurate level than the other two models.
Compared to the a Bayesian VAR-SV without common factor in the volatility, the BVAR-CSV specification is able to effectively extract the primary information of a large dataset while accounting for a varying volatility. These models have a substantial range of applications on forecasting macroeconomic and financial variables, given that large datasets are increasingly available. Caveats should be taken, however, understanding that there is a trade-off between prediction and interpretation of this class of models. The unfortunate reality is that as the models become larger, relationships between variables tend to be more complex and their interpretability declines. One future research topic could explore the optimal level of model size and forecast accuracy. Another direction is to relax the common loading of one and apply the method of Hu and Tsay (2014) to study the serial dependence of the volatilities.
Table 2.2: Average log predictive density

<table>
<thead>
<tr>
<th>Variable</th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 3Q$</th>
<th>$h = 4Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BVAR-CSV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global inflows</td>
<td>-20.106</td>
<td>-23.376</td>
<td>-29.078</td>
<td>-34.357</td>
</tr>
<tr>
<td>US Domestic credit</td>
<td>-5.803</td>
<td>-10.343</td>
<td>-15.758</td>
<td>-21.027</td>
</tr>
<tr>
<td>EU Bank leverage</td>
<td>11.585</td>
<td>5.510</td>
<td>0.287</td>
<td>-4.765</td>
</tr>
<tr>
<td>Term Spread</td>
<td>12.203</td>
<td>6.696</td>
<td>1.292</td>
<td>-3.475</td>
</tr>
<tr>
<td>Joint Normal</td>
<td>-85.900</td>
<td>-132.375</td>
<td>-175.108</td>
<td>-234.279</td>
</tr>
<tr>
<td><strong>BVAR-CSV-t</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global inflows</td>
<td>0.156</td>
<td>0.272</td>
<td>0.627</td>
<td>0.628</td>
</tr>
<tr>
<td>Global domestic credit</td>
<td>0.014</td>
<td>0.566</td>
<td>0.962</td>
<td>0.382</td>
</tr>
<tr>
<td>US Domestic credit</td>
<td>0.145</td>
<td>0.298</td>
<td>0.216</td>
<td>0.238</td>
</tr>
<tr>
<td>EU Bank leverage</td>
<td>0.014</td>
<td>0.355</td>
<td>0.574</td>
<td>0.552</td>
</tr>
<tr>
<td>Term Spread</td>
<td>0.004</td>
<td>0.393</td>
<td>0.662</td>
<td>0.655</td>
</tr>
<tr>
<td>Joint Normal</td>
<td>1.630</td>
<td>8.784</td>
<td>-0.369</td>
<td>5.019</td>
</tr>
<tr>
<td><strong>BVAR-CSV-MA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global inflows</td>
<td>3.645</td>
<td>2.586</td>
<td>2.827</td>
<td>2.832</td>
</tr>
<tr>
<td>Global domestic credit</td>
<td>-0.105</td>
<td>-0.561</td>
<td>-0.457</td>
<td>-1.255</td>
</tr>
<tr>
<td>US Domestic credit</td>
<td>-0.102</td>
<td>-0.544</td>
<td>-0.618</td>
<td>-0.825</td>
</tr>
<tr>
<td>EU Bank leverage</td>
<td>-0.233</td>
<td>-0.136</td>
<td>-0.160</td>
<td>-0.463</td>
</tr>
<tr>
<td>Term Spread</td>
<td>-0.149</td>
<td>-0.235</td>
<td>-0.302</td>
<td>-0.447</td>
</tr>
<tr>
<td>Joint Normal</td>
<td>-2.008</td>
<td>5.929</td>
<td>-11.983</td>
<td>3.321</td>
</tr>
</tbody>
</table>

Note: BVAR-CSV as benchmark. Numbers in the other two models are differences between the corresponding model and the benchmark model. A positive number given a forecast period indicates that there’s evidence in favor of the alternative model, and vice versa.
### Table 2.3: Inefficiency Factors for BVAR-CSV

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No. of para</th>
<th>Median</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1314</td>
<td>1.329</td>
<td>1.414</td>
<td>0.563</td>
<td>5.498</td>
</tr>
<tr>
<td>Σ</td>
<td>324</td>
<td>1.746</td>
<td>2.949</td>
<td>0.688</td>
<td>17.716</td>
</tr>
<tr>
<td>f</td>
<td>127</td>
<td>7.848</td>
<td>8.375</td>
<td>0.329</td>
<td>21.339</td>
</tr>
<tr>
<td>φ</td>
<td>1</td>
<td>6.66</td>
<td>6.66</td>
<td>6.66</td>
<td>6.66</td>
</tr>
<tr>
<td>(σ_f^2)</td>
<td>1</td>
<td>2.236</td>
<td>2.236</td>
<td>2.236</td>
<td>2.236</td>
</tr>
</tbody>
</table>

### Table 2.4: Estimated log integrated likelihoods

<table>
<thead>
<tr>
<th>Model</th>
<th>Log integrated likelihood</th>
<th>Numerical SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BVAR-CSV</td>
<td>-1789.1</td>
<td>0.47</td>
</tr>
<tr>
<td>BVAR-CSV-t</td>
<td>-1757.4</td>
<td>0.56</td>
</tr>
<tr>
<td>BVAR-CSV-MA</td>
<td>-1782.8</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Note: the difference between the log integrated likelihood is similar to the log Bayes Factor. For example, the difference of the log integrated likelihood between the Student-t innovation with the benchmark model is 25, indicating that the log posterior odd is 25. This gives strong evidence of the Student-t innovation CSV model.
Chapter 3

Taylor-rule fundamentals and exchange rate dynamics

3.1 Introduction

The seminal work of Meese and Rogoff (1983) has put forth the idea that exchange rate fluctuations are difficult to explain and forecast by standard economic models. Since then, macroeconomic fundamentals have been downplayed in forecasting and explaining exchange rate movements. A strand of the literature has produced mixed results by comparing random walk forecasts of exchange rate movements with the out-of-sample predictive ability of exchange rate models derived from economic theories (See Rossi (2013) for a survey). Observationally, exchange rates are not only non-stationary but extremely hard to distinguish from a simple random walk model. Engel and West (2005) emphasize that the current exchange rate is a function of future expected fundamentals. These expected fundamentals may not be identified if some of them are non-stationary with a discount factor close to unity. When unpredictable shocks

\footnote{If the real data generating process is a random walk plus a persistent component, such as an AR(1) process, observationally the time series still has a resemblance to a random walk. The early empirical literature conducted statistical tests concerning the hypothesis of foreign exchange market efficiency. That is, exchange rates would follow a random walk. See Frenkel (1976), Frenkel and Mussa (1980), and Fama (1984). But their results are unfavorable to the hypothesis.}
to the non-stationary fundamentals are present, the exchange rate appear to be a persistent process and thus is almost observationally equivalent to a random walk.

In this paper, we estimate an exchange rate model for the U.S. and Canada where the bilateral real exchange rate is jointly determined by its own expected values and a set of macroeconomic fundamentals. Note that the exchange rate model in this paper focuses on real as opposed to nominal determinants of the price-level-adjusted exchange rate.\(^2\) To do so we first specify monetary policy reactions function of Canada and the U.S., postulating both countries follow a Taylor-rule type monetary policy. Besides the inflation and output gaps, we assume that the Bank of Canada also includes the real exchange rate in its Taylor rule.\(^3\) Based on the analysis of Engel and West (2006), we decompose the real exchange rate into a trend component and a transitory component. I then linked the transitory component to the Taylor rule fundamentals via Uncovered Interest Rate Parity. This forms a system of equations that can be translated into a reduced form state space model that has a linear Gaussian form. The unobserved components of the model can be estimated via a Kalman filter while the parameters can be estimated by the maximum likelihood method.

Connecting macroeconomic fundamentals to the real exchange rate has been a common practice in evaluating out-of-sample predictability. Starting with Mark (1995), several studies have discovered that exchange rate models linked to macroeconomic fundamentals deliver more solid performance in forecasting out-of-sample exchange rates. But these models have not shown consistency in their performance when applied to various samples across time and countries. Cheung et al. (2005) survey the forecasting performance of economic exchange rate models of different types, such as interest rate parity, monetary, productivity-based, and behavioral equilibrium exchange rate model (BEER), finding none of these models consistently outperform the random walk model at any horizon.\(^4\)

\(^2\)See Chinn (2012) for a discussion of exchange models in nominal and real terms. The models for real exchange rates consist of three approaches: the purchasing power parity, the Balassa-Samuelson productivity based approach, and the Lucas two-good model.

\(^3\)For a small open economy, this assumption is common. Engel and West (2006) apply this model to describe the Mark-Dollar exchange rate behavior.

\(^4\)See Clark and MacDonald (1999) for the description of the BEER model.
The essential mechanism linking the Taylor rule based monetary policy to exchange rates is the Uncovered Interest Rate Parity (UIRP) condition. Under the UIRP with rational expectations, a home country’s central bank’s action on raising (lowering) its policy rate relative to the foreign counterpart, renders an expected appreciation (depreciation) of the home country’s currency, relative to the foreign currency (the Dornbusch (1976) overshooting model). Hence, one can substitute the expected change in exchange rates with the interest rate differential induced by UIRP. But UIRP is not well supported empirically, due to the forward premium and the delayed overshooting puzzles (see, e.g., Gourinchas and Tornell (2004)). I follow Molodtsova and Papell (2009)’s approach, assuming that economic agents have distorted beliefs about future interest rates, and forecast the economic fundamentals based on these beliefs.5

The Taylor rule based exchange rate model in this paper is closely related to Engel and West (2005) that presents a present-value asset pricing framework of the exchange rate model, and Molodtsova and Papell (2009) in which they utilize Engel and West’s framework and specify possible Taylor rules specifications regarding evidence on how central banks react to exchange rates. Following Clarida et al. (1998), there is evidence that small open economies such as Canada include the difference between the real exchange rate and its target rate in their policy reaction functions. Byrne et al. (2018) explore various specifications of the Taylor rule exchange rate models targeting sources of uncertainty.

The key contribution of this paper is to measure how much the trend component of the real exchange rate can explain the fluctuations of the actual real exchange rate, after letting the fundamentals affect the transitory components of the real exchange rate only. Even though those fundamentals are modeled as stationary, unexpected shocks to them may generate strong persistence to the transitory component if some of the coefficients are large. Since the transitory component is a linear function of the fundamentals, persistence in it may dominate the process of the real exchange rate, making it indistinguishable from a random walk. Identifying the

5The assumption is based on Gourinchas and Tornell (2004) using survey data to document the distortion in beliefs among investors. They define distorted beliefs as investors overestimating the relative importance of transitory interest rate shocks compared to persistent interest rate shocks, as measured by the variance of their innovations.
source of real exchange rate fluctuations in this way may help explain why Taylor rule models have stronger performance in out-of-sample predictability than random walk models.

The reduced-form model is similar to Berger and Kempa (2012) where they estimate the equilibrium exchange rate of the US-Canadian dollar by linking fundamentals to the transitory components, but their model does not distinguish the shocks to the transitory components of the real exchange rate from shocks to the observed exchange rates. Our model differs in that we do not allow the Balassa-Samuelson effect to enter the model because recent empirical evidence in favor of the standard Balassa-Samuelson hypothesis is weak when focusing on the developed economies. Our model does not allow the macro fundamentals to influence the trend component of the real exchange rates because the fundamentals are assumed to be stationary, and we attempt to isolate the trend component affected only by random determinants.

The primary finding of this paper has several implications for understanding the roles of fundamentals in real exchange rate dynamics. First, the estimated trend of the real exchange rate between the U.S. and Canada has substantial time variation over the sample periods, as opposed to a smooth depreciation trend in Berger and Kempa (2012). Second, the Canadian dollar experiences overvaluation in most of the periods, except for the recent financial crisis. Third, the transitory component linked to the Taylor rule fundamentals captures a large portion of real exchange rate fluctuations. This component does not show strong persistence, thus there is evidence that the trend component is able to describe the persistence of the real exchange rate.

The remainder of the paper is organized as follows: Section 2 introduces the theoretical framework of the Taylor rule exchange rate model. Section 3 details the estimation methodology. Section 4 presents the main results, and Section 5 draws conclusion.

---

6 In the literature relating to exchange rate economic models, "equilibrium exchange rate" refers to different concepts. Usually, researchers use the term to describe long-run level determined by purchasing power parity (PPP). See, e.g. Engel and Kim (1999). Recent literature on forecasting and unobserved component models also use the term referring to the exchange rate’s trend component. We do not use "equilibrium exchange rate" in this paper as to avoid confusion with the actual equilibrium exchange rate solved from a dynamic equilibrium model.

7 See Lee and Tang (2007) and Ricci et al. (2013) for discussion of productivity growth and real exchange rates.
3.2 Taylor rule fundamentals in exchange rate model

I follow Engel and West (2005) and Engel and West (2006) in connecting the two countries’ monetary policies characterized by Taylor rules, respectively:

\begin{align*}
\text{Home country} : \quad i_t &= \tau_\pi E_t(\pi_{t+1}) + \tau_y \tilde{y}_t + \tau_q \tilde{q}_t + \epsilon_t \\
\text{Foreign country} : \quad i_t^* &= \tau_{\pi^*}^* E_t(\pi^*_{t+1}) + \tau_{y^*}^* \tilde{y}^*_t + \epsilon^*_t
\end{align*}

(3.1)

where asterisks represent foreign variables. \(i\) and \(i^*\) denote the policy rates of home and foreign countries respectively. \(E_t(\pi_{t+1})\) and \(E_t(\pi^*_{t+1})\) denote expected inflation relative to targets at home and abroad. \(\tilde{y}_t\) and \(\tilde{y}^*_t\) are the domestic and foreign output gaps. The innovation terms \(\epsilon_t\) and \(\epsilon^*_t\) are the shocks to monetary policies of the home and foreign country. The term \(\tilde{q}_t\) included in the home country’s Taylor rule is the short-run deviation of the real exchange rate from its long-run level. Define the real exchange rate in logarithm as \(q_t = e_t - p_t + p_t^*\), where \(e_t\) denotes the nominal exchange rate in logarithm, expressed as the home currency price of the foreign exchange. \(p_t\) and \(p_t^*\) represent the domestic and foreign price levels in logarithm.

The UIRP condition links the interest rate differentials of the two countries’ monetary policy:

\[ i_t - i_t^* = E_t(e_{t+1}) - e_t. \]  

(3.3)

Take expectations at time \(t\) of the definition of the real exchange rate and move time one period forward. It can be arranged as \(E_t(q_{t+1}) = E_t(e_{t+1}) - E_t(p_{t+1}) + E_t(p_t^*).\) Substitute \(E_t(e_{t+1})\) to equation (3.3), to obtain:

\[ i_t - i_t^* = E_t(q_{t+1} + p_{t+1} - p_t^*) - (q_t + p_t - p_t^*) \]

\[ = E_t(q_{t+1} - q_t) + E_t(p_{t+1} - p_t) - E_t(p_t^* - p_t^*) \]

(3.4)

\[ = E_t(q_{t+1}) - q_t + E_t(\pi_{t+1}) - E_t(\pi_t^*). \]
Note that in equation (3.4), the observable variables include the policy rates of the two countries, and the real exchange rate $q_t$. The expected real exchange rate and expected inflations are not observable. We follow the common practice in the forecasting literature and approximate them using distributed lags.\(^8\) Modeling the inflation gap may deliver different results depending on the model assumptions. Cogley et al. (2010) study the persistence of the inflation gap in the U.S. within the VAR framework and present evidence of strong persistence of the inflation gap. Pivetta and Reis (2007) run rolling window unit-root tests on inflation rates and find little change of inflation persistence. The real exchange rate $q_t$ can be decomposed as the sum of its trend component, denoted as $\bar{q}_t$, a transitory component $\tilde{q}_t$ and an error term $\eta_1$:

\[ q_t = \bar{q}_t + \tilde{q}_t + \eta_1 t \]  

(3.5)

The trend component $\bar{q}_t$ is specified as a random walk,

\[ \bar{q}_t = \bar{q}_{t-1} + u_{1t} \]  

(3.6)

where $u_{1t}$ is a white noise with mean zero. Subtracting equation (3.2) from (3.1) and substituting into (3.4) delivers the transitory component of the real exchange rate as a function of Taylor rule fundamentals:

\[ \tilde{q}_t = \frac{1}{1 + \tau_q} E_t(\tilde{q}_{t+1}) + \frac{1-\tau_{\pi}}{1 + \tau_q} E_t(\pi_{t+1}) - \frac{1-\tau_{\pi}}{1 + \tau_q} E_t(\pi_{t}^*) - \frac{\tau_y}{1 + \tau_q} \tilde{y}_t + \frac{\tau_y^*}{1 + \tau_q} \tilde{y}_t^* + u_{2t} \]  

(3.7)

where $u_{2t} = \frac{1}{1+\tau_q}(\epsilon_t^* - \epsilon_t)$. It may be interpreted as shocks to the risk premium or difference between monetary shocks of the two countries. Since the transitory component captures a set of relevant macroeconomic fundamentals, the innovation of the trend component, $u_{1t}$, can thus be seen as other relevant exchange rate fundamentals not included in the transitory component.
\( \hat{q}_t = \theta(B)\hat{q}_{t-1} + \theta(B)(1-\tau\pi^*)\pi_{t-1} - \theta(B)(1-\tau\pi)\pi_{t-1} - \theta(B)\tau_y\bar{y}_t + \theta(B)\tau_y^*\bar{y}_t^* + \theta u_{2t} \) (3.8)

with the lag polynomial \( \theta(B) = \theta_1 + \theta_2B + \ldots + \theta_pB^p \) and \( \theta = \frac{1}{1+\tau q} \).

Equation (3.8) assembles the features of the traditional monetary exchange rate models widely documented in Frankel (1979), Obstfeld and Rogoff (1995) and Taylor (1995). According to purchasing power parity (PPP), a higher home inflation relative to foreign inflation leads to a depreciation of the home currency. If the two countries follow the Taylor principle strictly, implying that \( \tau\pi > 1 \) and \( \tau\pi^* > 1 \), a rising real interest rate differential would follow an increase in the home relative to foreign inflation rate. Home currency thus experiences an appreciation.

Finally, the output gap needs to be specified in order to close the model. The observed output in the home country can be decomposed into the transitory component \( \tilde{y}_t \) associated the Taylor rule fundamental, and a permanent component \( \bar{y}_t \), which should not be confused with the potential output in the New Keynesian general equilibrium literature. Instead, it can be understood in a statistical sense as the trend output estimated from this specified model. Therefore, the observed output can be decomposed as

\[ y_t = \bar{y}_t + \tilde{y}_t + \eta_{2t} \] (3.9)

As noted previously, the output gap is postulated to follow a stationary autoregressive process,

\( \tilde{y}_t = \psi(B)\tilde{y}_{t-1} + u_{3t} \) (3.10)

---

9 Engel and West (2006) uses a bivariate VAR to approximate the expected output gap and inflation gap. Basistha (2007) assumes adaptive expectation of the inflation and output gaps in order to model them in distributed lags.

10 In the case of using distributed lags, lag length selection largely depends on the frequency of the data.
where $\psi(B)$ denotes the coefficients of the autoregressive process, and $u_{3t}$ is a Gaussian white noise. The potential output $\bar{y}_t$ is assumed to evolve according to a random walk with drift

$$\bar{y}_t = \bar{y}_{t-1} + a_h + u_{4t} \quad (3.11)$$

The output gap in the foreign country is modeled in a similar way:

$$y_t^* = \bar{y}_t^* + y_t^* + \eta_{3t} \quad (3.12)$$

$$\bar{y}_t^* = \rho(B)\bar{y}_{t-1}^* + u_{5t} \quad (3.13)$$

$$\bar{y}_t^* = \bar{y}_{t-1}^* + a_f + u_{6t} \quad (3.14)$$

Note that the transitory components of real exchange rate, the output gaps of both countries are modeled as AR(2) process.\footnote{The lag length selection is indicated by the AIC using monthly data.} In the next section, equations (3.6) - (3.14) are arranged to a state-space representation and estimated through maximum likelihood and the Kalman filter.

### 3.3 Empirical Model and estimation

#### 3.3.1 State-space representation

Let $Z_t = (q_t, y_t, y_t^*)'$, $t \in [1,\ldots,T]$ denote a vector to store the observed time series data of the U.S.-Canadian bilateral real exchange rate and industrial production from Canada and the U.S., respectively. Equations (6) - (14) can be cast into a Gaussian state space model as the following

$$Z_t = F\alpha_t + R\alpha_{t-1} + \eta_t, \; \eta_t \sim \mathcal{N}(0,H) \quad (3.15)$$

$$\alpha_t = T\alpha_{t-1} + \mu + u_t, \; u_t \sim \mathcal{N}(0,Q) \quad (3.16)$$
where \( Z_t \) is a \( p \times 1 \) vector of observed variables in the measurement equation (3.15). \( x_t \) is a \( k \times 1 \) vector of predetermined variables and \( \alpha_t \) is a \( m \times 1 \) vector of unobserved states, modeled in the state equation (3.15). The innovation terms \( \eta_t \) and \( u_t \) are assumed to be mutually independent Gaussian white noise. Matrices \( F, R, T, H, Q \) are unknown parameters.

Modeling \( \tilde{q}_t \) in the state equation (3.16) requires its dependence on the lagged values of the relative output gap \( \tilde{y}_t \) and \( \tilde{y}^*_t \). Since by construction the transitory component \( \tilde{q}_t \) is stationary, equation (3.7) can be approximated by the following:

\[
\tilde{q}_t = \theta_1 \tilde{q}_{t-1} + \theta_2 \tilde{q}_{t-2} - \phi_y \tilde{y}_t + \phi_{y^*} \tilde{y}^*_t + u_{2t}
\] (3.17)

where \( \phi_y = \theta \tau_y \) and \( \phi_{y^*} = \theta \tau_{y^*} \). The expected value of the transitory component of the real exchange rate is measured by an AR(1) process with parameters \( \theta_1 \) and \( \theta_2 \). The unobserved state vector is defined as \( \alpha_t = (\tilde{q}_t \ \tilde{q}_{t-1} \ \tilde{y}_t \ \tilde{y}^*_{t-1} \ \tilde{y}^*_t \ \tilde{y}^*_t \ \tilde{y}^*_t) \)' and the predetermined vector is defined as \( x_{t-1} = (\pi_{t-1} \ \pi^*_{t-1}) \)'.

\[
F = \begin{bmatrix}
1 & 1 & 0 & 0 & -\theta \tau_y & 0 & 0 & \theta \tau_{y^*} & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0
\end{bmatrix}
\]

\[
R = \begin{bmatrix}
\theta(1 - \tau_\pi) & -\theta(1 - \tau_{\pi^*}) \\
0 & 0 \\
0 & 0
\end{bmatrix}
\]

where \( \theta = \frac{1}{1 + \gamma_q} \).
\[
T = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \psi_1 & \psi_2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho_1 & \rho_2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} ; \quad \mu = \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
a_h \\
a_f \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

The covariances of the measurement equation and the state equations are diagonal matrices such that \( H = \text{diag}\{\sigma^2_{\eta_1}, \sigma^2_{\eta_2}, \sigma^2_{\eta_3}\} \), and \( Q = \text{diag}\{\sigma^2_{u_{1t}}, \sigma^2_{u_{2t}}, 0, \sigma^2_{u_{3t}}, \sigma^2_{u_{4t}}, 0, \sigma^2_{u_{5t}}, \sigma^2_{u_{6t}}, 0\} \).

### 3.3.2 Data

We use monthly data for the U.S. and Canada from Jan. 1974 to Dec. 2017. The data are sourced from *International Financial Statistics* published by the IMF. The real exchange rates are computed using nominal exchange rates and inflation indices of both countries. We use monthly data taking advantage of the relatively high frequency of exchange rates, allowing the sample size to be large and facilitating the maximum likelihood method. Because of the monthly frequency, we use industrial production for the U.S. and Canada instead of real GDP. For inflation, we use the log of the seasonally adjusted CPI for Canada and Personal Consumption Expenditures (PCE-Chain Type Price Index) for the U.S. We use the daily average spot rates for the U.S.-Canadian Dollar nominal exchange rate. Real exchange rates in natural logarithm can thus be calculated following the definition stated in the introduction. Except for inflation rates, all other variables enter the model without differencing transformation as stochastic trends are estimated.

Note that inflation rates in the model are treated as exogenous variables. Alternatively, we could use Survey of Professional Forecasters (SPF) for the U.S. expected inflation. This
alternative setting does not change the model structure because the expected values of inflation
gaps in equation (3.7) are still predetermined.

3.3.3 Estimation Strategy

Given parameters $F, R, T, \mu, H, Q$, the model in equations (3.15) and (16) is a Gaussian linear
state space model with parameter restrictions in those matrices. The likelihood function can be
derived analytically and the model can be estimated by the Kalman filter and smoother with
the maximum likelihood method (see e.g. Durbin and Koopman (2012)). The algorithm used
to compute the maximum likelihood can either be the EM algorithm or Quasi-Newton method,
e.g., BFGS. We are aware that the maximum-likelihood estimates of variances of the model
are fundamentally biased. The bias occurs because variances are constrained to be positive.
Therefore if the elements of the covariance matrices $H$ and $Q$ are essentially zero, the mean
estimate will not be zero and thus the bias will be large. To cure the bias issue, we generate
bias corrected variance estimates by using a bootstrap estimate of the bias.

The multivariate model is a linear state-space model with Gaussian innovation. It can be
easily handled by R packages developed by Holmes et al. (2014) and Petris and An (2010).
The likelihood converges in the 760th iteration when using Quasi-Newton method, achieving a
numerical value in log as -6790.003. The AIC is 13626.077 while the corrected AIC is 13626.77.

3.4 Estimation results

3.4.1 Estimates of reduced form equations

Table 3.1 presents the estimation results of the reduced-form model in (3.15) and (3.16). The
sum of AR coefficients for the real exchange rate is 0.278, implying a low degree of persistence.
But the output gaps of the U.S. and Canada are relatively persistent, which is consistent with
the literature studying industrialized countries (See, e.g. Engel and Kim (1999)). The Taylor
rule parameters play important roles in affecting the transitory component of the real exchange
rate. If we follow the literature (e.g. Engel and West (2005)) and set \( \tau_y \) and \( \tau_y^* \) to 0.5, then the \( \tau_q \) is equal to \(-0.66\).\(^{12}\) This suggests that the Canadian central bank puts more weight on the real exchange rate when setting the interest rate. Similarly, by setting \( \tau_y \) and \( \tau_y^* \) to 1.5, we assume that the transitory real exchange rate is positively correlated with the home inflation but negatively related to foreign inflation. When the Canadian central bank observes the Canadian dollar depreciating, the relative price level of Canada to the U.S. is rising. It can react to the depreciation by setting the monetary reaction function accordingly.

This model specification informs how interest rates (and hence real exchange rates) fluctuate in response to changes of inflation and output gaps. As discussed above, the coefficient of real exchange rate \( \tau_q \) moves inversely with the coefficient of home country’s output gap. Consider what the central bank in Canada would do if the inflation gap rises in the Canada relative to the U.S. The Bank of Canada will raise the policy rate, hence appreciating the currency. Simply put, if inflation rises above the target level, the home currency will appreciate. This aligns with the question asked in Clarida and Waldman (2008) that “Is bad news about inflation good news for the dollar?” The answer is yes. If the real exchange rate drops (home currency depreciates), the central bank will raise the interest rate in an effort to strengthen the home currency. Chinn (2008) specifies a slightly different framework with interest rate smoothing in Equation (3.17) to study the dollar/euro. He finds that the output gaps and inflation gaps enter in the equation with expectations having statistical significance. Overall, the model has the implication that real exchange rate movements are useful in predicting future monetary policy because the exchange rate reflects, if not fully, the market’s expectation of a central bank’s policy.

### 3.4.2 The state estimates: Canada-US real exchange rate

Figure 3.1 gives the estimates of the trend component and the transitory component of the real exchange rate. Note that the downward movement indicates depreciation of the Canadian

\(^{12}\)This can be seen in the relationship between the reduced-form parameter \( \phi_y \) and the structural parameter \( \tau_y \) as in \( \phi_y = \theta \tau_y = \frac{\tau_y}{1+\tau_q} \). Though we are not able to recover all the structural parameters, we can make inference with its significant level on the structural parameters based on their linear relationship.
Table 3.1: Selected parameter estimates and confidence intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>-0.061</td>
<td>0.33</td>
<td>$[-0.724, -0.601]$</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.339</td>
<td>0.080</td>
<td>$[0.180, 0.497]$</td>
</tr>
<tr>
<td>$\phi_y$</td>
<td>1.48</td>
<td>1.67</td>
<td>$[-1.80, 4.76]$</td>
</tr>
<tr>
<td>$\phi_{y^*}$</td>
<td>0.374</td>
<td>0.545</td>
<td>$[-0.694, 1.440]$</td>
</tr>
<tr>
<td>$\phi_{\pi}$</td>
<td>-0.158</td>
<td>3.95</td>
<td>$[-7.75, 7.72]$</td>
</tr>
<tr>
<td>$\phi_{\pi^*}$</td>
<td>1.63</td>
<td>2.34</td>
<td>$[-4.58, 4.61]$</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.001</td>
<td>0.03</td>
<td>$[-0.06, 0.06]$</td>
</tr>
<tr>
<td>$\sigma_u^2$</td>
<td>2.53</td>
<td>6.43</td>
<td>$[1.27, 3.79]$</td>
</tr>
<tr>
<td>Output U.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-1.07</td>
<td>0.806</td>
<td>$[-1.510, 1.251]$</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.325</td>
<td>0.295</td>
<td>$[-0.154, 1.000]$</td>
</tr>
<tr>
<td>$a_h$</td>
<td>-0.048</td>
<td>0.0061</td>
<td>$[-0.0016, 0.0072]$</td>
</tr>
<tr>
<td>$\sigma_u^3$</td>
<td>0.076</td>
<td>0.074</td>
<td>$[-0.0146, 0.0146]$</td>
</tr>
<tr>
<td>$\sigma_u^4$</td>
<td>0.208</td>
<td>0.285</td>
<td>$[-0.35, 0.76]$</td>
</tr>
<tr>
<td>Output Canada</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>1.29</td>
<td>1.21</td>
<td>$[-1.31, 1.45]$</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.438</td>
<td>0.118</td>
<td>$[-0.669, -0.206]$</td>
</tr>
<tr>
<td>$a_f$</td>
<td>-0.0053</td>
<td>0.048</td>
<td>$[-0.01, 0.008]$</td>
</tr>
<tr>
<td>$\sigma_u^5$</td>
<td>0.034</td>
<td>0.004</td>
<td>$[-0.092, 0.092]$</td>
</tr>
<tr>
<td>$\sigma_u^6$</td>
<td>2.32</td>
<td>0.92</td>
<td>$[0.517, 4.13]$</td>
</tr>
</tbody>
</table>

Note: the confidence intervals are constructed by bootstrapping 1000 artificial samples with bias corrected.

dollar. First, the trend component exhibits substantial time variation. In most of the sample periods, the Canadian dollar depreciates except for the early 1990s and the periods after the 2008 financial crisis. The downward movement in the real exchange rate can be explained by a declining trend in the relative price of non-traded to traded goods in Canada (Clark and MacDonald (2004)). Second, the transitory component of the real exchange rate is not persistent and is able to capture the primary fluctuation of the actual real exchange rate. The correlation between the actual and fitted real exchange rate change is as high as 0.52. This is likely due to
the dominant position of the U.S. dollar in international trade. Figure 3.2 compares the actual real exchange rate and the corresponding filtered estimates.

Figure 3.1: Trend and transitory components of the real exchange rate: the top panel shows the smoothed estimates (dash) and the filtered estimates (solid). The bottom panel displays the transitory component of the real exchange rate. The filtered estimates are obtained in each time period when a new observation is updated (i.e., using the forward recursion of the Kalman filter). The smoothed estimates are obtained after all observations are observed (i.e., using the backward recursion of the Kalman smoother)

The deviations of the actual real exchange rate from its trend component are considerable and highly persistent. Most of the sample periods show an overvaluation of the Canadian dollar, but several periods display undervaluations: the late 1970s to the mid-1980s, and the
mid-2000s to recent days. The first period of depreciation of the Canadian currency can be related to a major shift of the U.S. monetary policy under Paul Volcker. The recent episodes of undervaluation can be associated with the strong demand for the U.S. denominated assets around the world since the mid-2000s. Bernanke (2005) summarizes the appreciation of the U.S. dollar along with a worsening of the current account balance as the "saving glut hypothesis". Starting in the late 1980s, the U.S. dollar experienced a substantial depreciation against major currencies. Lane and Milesi-Ferretti (2007) document different episodes starting from 1970 where the U.S. real exchange rate fluctuations were highly correlated with capital gains and losses on external assets and liabilities. From the early 1990s to the onset of the Asian currency crisis in 1997, the real exchange rate was broadly flat, due to the strong performance of the stock markets in other countries. The overvaluation of the Canadian dollar during those periods is likely a direct consequence of the changes in the U.S. external positions.

3.5 Conclusion

This paper constructs and estimates an exchange rate model using Taylor rule fundamentals. Our model builds on the recent growing literature concerning the relation between macroeconomic fundamentals and the predictability of exchange rates. We follow Engel and West (2006) and Berger and Kempa (2012)'s approach by linking the observable macroeconomic aggregates to the transitory component of the real exchange rate only, while modeling the trend component as a random walk with a drift.

Similar to Engel and West (2006), we assume that uncovered interest rate parity holds between Canada and the U.S., and that the Bank of Canada has engaged in exchange rate intervention implicitly. The interest rate differential between the two countries links the Taylor rule fundamentals with the real exchange rates. We then derive a reduced form state-state model and estimate the parameters and unobserved components jointly via maximum likelihood methods.

Using monthly data over 1974-2017, our analysis finds that the trend component of the
Figure 3.2: Actual real exchange rate (dash) and filtered trend real exchange rate (solid)

U.S.-Canadian real exchange rate has substantial time variation. The Canadian dollar follows a trend depreciation until the recent financial crisis. Moreover, the Canadian dollar is overvalued in most of the sample period, suggesting a persistent misalignment of the bilateral exchange rate. The transitory component explained by the Taylor rule fundamentals is able to capture most of the fluctuations of the actual exchange rate.
REFERENCES


J. C. Chan. Large Bayesian VARs: A flexible Kronecker error covariance structure. 2015.


M. West. *Bayesian Forecasting*. Wiley Online Library, 1996.

APPENDICES
Appendix A

Chapter 1

A.1 MCMC Algorithm

The algorithm in this paper implements Koop et al. (2009) based on Del Negro and Primiceri (2015) with slight modification. Being consistent with much of the literature, denote by $\alpha_T$ the entire path of state vector $\{\alpha_t\}_{t=1}^T$ and the similar for covariance matrix $H_T$ and contemporaneous coefficient matrix $A_T$. The following steps summarize the MCMC sampler in this study.

The complete state space model is

$$y_t = Z_t^\prime \alpha_t + \epsilon_t, \ \epsilon_t \sim N(0,H_t) \quad (A.1)$$

$$\alpha_{t+1} = \alpha_t + \begin{bmatrix} K_{1t} \\ K_{2t} \end{bmatrix} \eta_{t+1}, \ \eta_t \sim N(0,Q_t) \quad (A.2)$$

$$a_{t+1} = a_t + K_{a,t} \nu_{t+1}, \ \nu_t \sim N(0,S) \quad (A.3)$$

$$h_{t+1} = h_t + K_{h,t} \zeta_{t+1}, \ \zeta_t \sim N(0,W) \quad (A.4)$$

Note that the disturbances terms of above equations $\epsilon_t, \eta_t, \nu_t$ and $\zeta_t$ are all uncorrelated with each other, neither contemporaneously nor in all leads and lags. The indicator matrices can be denoted as $K_t = diag(K_{1t},K_{2t},K_{a,t},K_{h,t})$ for notation convenience.

93
The sampling treatment needs to be modified after introducing the indicator matrix $K_t$. By lemma 3 and 4 in Gerlach et al. (2000), sampling $K_t$ can become efficient by conditioning on reduced states. They show that $K_t$ can be generated by

$$p(K_t|y_T, K_{(-t)}) = \frac{p(y_t|K_t)p(K_t|K_{(-1)})}{\sum_{K_t=j} p(y_t|K_t=j)p(K_t=j)}$$

where $p(y_t|K_t)$ can be carried out by recursion of lemma 4, such that

$$p(y_t|K_t) = \sum_{K_t} p(y_t|K_t)p(K_t|K_{(-1)})$$

The following are detailed steps of the MCMC:

**Step 1**: Given prior distributions, initialize $A_T$, $\Sigma_T$, $S_T$, and $V$ as stated in table 1, where $V$ is the variance-covariance (henceforth VCV) matrix of the whole system stated in equation (9)

**Step 2**: Sample $\alpha_T$ from the posterior distribution $p(\alpha_T|y_T, A_T, \Sigma_T, V)$.

1. Given prior for $p_j$, calculate hyperparameters of $\alpha_j$ and $\beta_j$, $j = 1,2$ sample $K_{1t}$ and $K_{2t}$ index recursively according to Gerlach et al. (2000)

2. The full history of coefficient state $\alpha_T$ can be drawn conditional on draws of $K_{1t}$ and $K_{2t}$, the data $y_T$, a full history of covariance and volatility states in $H_T$, and the hyperparameters in $V$. Following Carter and Kohn (1994),

$$\alpha_T|y_T, \Sigma_T, A_T, V \sim N(\alpha_{T|T}, P_{T|T})$$

where $\alpha_t|y_T, \alpha_{t+1}, \Sigma_T, V \sim N(\alpha_{t|t+1}, P_{t|t+1})$, for $t = 1, \ldots, T - 1$

The conditional mean and variance of $\alpha_T$ can be obtained using forward recursion of Kalman
filter,

\[
\alpha_{t|t} = \alpha_{t|t-1} + P_{t|t-1} Z_t' (Z_t P_{t|t-1} Z_t' + \Sigma_t)^{-1} (y_t - Z_t' \alpha_{t|t-1}) \\
P_{t|t} = P_{t|t-1} - P_{t|t-1} Z_t' (Z_t P_{t|t-1} Z_t' + \Sigma_t)^{-1} Z_t P_{t|t-1} \\
\alpha_{t|t-1} = \alpha_{t-1|t-1} \\
P_{t|t-1} = \alpha_{t-1|t-1} + Q
\] (A.5) (A.6) (A.7) (A.8)

starting from the initial values shown in table one that \(\alpha_0 \sim \mathcal{N}(\hat{\alpha}_{OLS}, k\hat{\Sigma}(\hat{\alpha}_{OLS}))\) with training sample 1973Q1-1982Q4. The elements in \(\alpha_{T-1}\) are sampled using backward recursion of Kalman filter. \(\alpha_{T-1}\) is sampled conditional on the realization of \(\alpha_T\), and \(\alpha_{T-2}\) conditional on the realization of \(\alpha_{T-1}\) and so on. The conditional mean and variance are given as

\[
\alpha_{t|t+1} = \alpha_{t|t} + P_{t|t}(P_{t|t} + Q)^{-1}(\alpha_{t+1} - \alpha_t) \\
P_{t|t+1} = P_{t|t} - P_{t|t}(P_{t|t} + Q)^{-1}P_{t|t}
\] (A.9) (A.10)

**Step 3:** Sample hyperparameter \(Q\) conditional on the entire history of data \(y_T\) and \(\alpha_T\), that is, \(p(Q|y_T, \alpha_T, V^{-Q})\), which is an Inverse- Wishart (IW) distribution.

**Step 4:**

1. Sample \(K_{a,t}\) index and related probability in a similar manner as in step 2.
2. Sample \(A_T\) from posterior \(p(A_T|y_T, K_{a,T}, \alpha_T, V, \Sigma_T)\), using Carter and Kohn (1994). Algorithms are similar in step 2. Notice that \(A_T\) contains the time-varying contemporaneous coefficients \(a_t\). Thus \(A_T\) can be seen as several block of \(a_t\). As Primiceri (2005) points out, due to the triangle structure of \(A_T\), one can apply the Kalman filter and the backward recursion equation by equation to sample each block and recover the conditional mean and variance of \(a_{i,t}, a_{i,t}\) can be sampled from \(p(a_{i,t}|a_{i,t+1}, y_t, \alpha_T, K_{h,T}, \Sigma_T, V)\), which is \(N(a_{i,t|t+1}, \Delta_{i,t|t+1})\). The computation of \(\alpha_{i,t|t+1}\) and \(\Delta_{i,t|t+1}\) follows those procedures in step 2.
Step 5: Sample $S$ that contains several blocks from posterior $p(S|y_T, \alpha_T, A_T, Q, W)$, which are Inverse-Wishart distribution.

Step 6:

(1) Sample $K_{h,t}$ index and related probabilities $p_3$ as in step 2.

(2) Sample the auxiliary discrete variables $s_T$ from posterior $p(s_T|\Sigma_T, \alpha_T, A_T, V)$, using algorithm developed by Kim et al. (1998). This is a step necessary for sampling the the volatility states. The original system of the structural VAR can be written as

$$A_t y_t = A^+_t X + \Sigma_t^{1/2} \varepsilon_t$$ \hspace{1cm} (A.11)

Further,

$$A_t (y_t - Z_t' \alpha_t) = \Sigma_t^{1/2} \varepsilon_t$$ \hspace{1cm} (A.12)

Here the measurement equation is nonlinear but can be converted to a linear representation such that

$$\log \left[ (A_t(y_t - Z_t' \alpha_t))^2 + c \right] = 2 \log \sigma_{i,t} + \log (\varepsilon_{i,t}^2)$$ \hspace{1cm} (A.13)

where $c$ is a very small constant that can be ignored. The state equation is thus assumed to be a logarithm random walk represented by equation (23). The state space representation here is linear with non-Gaussian innovation. The error term in the measurement equation follows as a $\log \chi^2(1)$ distribution. This distribution can be approximated by a mixture of normal distribution. The choices of 7 mixture normal density in Kim et al. (1998) have become the standard in time-varying VAR literature.

(3) Sample discrete auxiliary variables $s^T$ that is a matrix contains all $s_j, j = 1..T$. A number of $T$ indicator variables are to be selected as the auxiliary variables that facilitate the approximation of the state space system. Denote these indicator as a matrix $s^T = [s_1, ..., s_T]'$.
where each point in time when the mixture of normal approximation is used for each element in $\log(\epsilon_{i,t}^2)$. The underlying conditional distribution of $\log\sigma_{i,t}$ is approximately normal distribution $N(\log\sigma_{i,t|t+1}, V(\log\sigma_{i,t|t+1}))$. The conditional mean and variance of this normal distribution can be computed using Kalman smoother as in the previous steps. After the conditional states are generated, each indicator $s_j$ can be drawn conditional on the states and other parameters using sampling technique in Kim et al. (1998)

**Step 7:** Sample $\Sigma_T$ from distribution $p(\Sigma_T|y_T, B_T, \alpha_T, V, s^T)$, which is an inverse Wishart. Note that in the original algorithm of Primiceri (2005), drawing $\Sigma_T$ did not condition on the indicator matrix $s_T$ but it should have.

1. First draw volatilities conditional on draws of auxiliary variables in (3) of step 6. Given the auxiliary draws, volatilities can be approximated by normal distribution. And they can be generated following Carter and Kohn’s algorithm as the draws of state vectors.

2. Sample hyperparameter $W_T$ to volatility state from $p(W_T|y_T, \Sigma_T)$, which is an Inverse Wishart.

**Step 8:** Go to step 2 until convergence.
### A.2 Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Transformation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real gross domestic product</td>
<td>$400\Delta \log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>PCE Chain-type</td>
<td>$400\Delta \log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>U.S. current account</td>
<td>ratio of real GDP</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>Defense news series</td>
<td>NA</td>
<td>Valerie Ramey’s website</td>
</tr>
<tr>
<td>Government spending</td>
<td>$400\Delta \log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>SPF government spending</td>
<td>$400\Delta \log$</td>
<td>Philadelphia Fed</td>
</tr>
<tr>
<td>Tax shocks</td>
<td>NA</td>
<td>Karel Mertens’ website</td>
</tr>
<tr>
<td>Tax revenue</td>
<td>$400\Delta \log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>3-month Treasury Bill</td>
<td>annualized rate</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>Effective real exchange rate</td>
<td>$400\Delta \log$</td>
<td>Bank of International Settlement</td>
</tr>
</tbody>
</table>
A.3 Empirical appendix

This appendix provides additional empirical results that are not included in the body of the paper. Table 4 contains posterior medians and credible sets (i.e. the 16th and 84th percentile of each posterior median) of the selected impact responses. Note that the TVP-VAR is with one lag. In order to keep the table brief as possible, I do not present all the impulse responses and their measures of uncertainty. Figure A1-A4 present the posterior means of the standard deviation of the errors in the current account equations under the reference prior and informative prior respectively.

Table A.2: Posterior median of selected impact responses (14th/86th percentile in parentheses)

<table>
<thead>
<tr>
<th>Var./shock</th>
<th>1983Q3</th>
<th>1987Q1</th>
<th>1996Q2</th>
<th>2006Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ca_t/\gamma_t$</td>
<td>0.186</td>
<td>-0.328</td>
<td>-0.119</td>
<td>0.0167</td>
</tr>
<tr>
<td>( -0.403, 0.782)</td>
<td>( -0.888, 0.231)</td>
<td>(-0.723, 0.471)</td>
<td>(-0.553, 0.591)</td>
<td></td>
</tr>
<tr>
<td>$yt/\gamma_t$</td>
<td>0.086</td>
<td>0.155</td>
<td>0.206</td>
<td>0.281</td>
</tr>
<tr>
<td>( -0.189, 0.355)</td>
<td>( -0.082, 0.385)</td>
<td>(0.014, 0.447)</td>
<td>(0.018, 0.550)</td>
<td></td>
</tr>
<tr>
<td>$rt/\gamma_t$</td>
<td>-0.122</td>
<td>0.088</td>
<td>0.014</td>
<td>0.04</td>
</tr>
<tr>
<td>( -0.574, 0.337)</td>
<td>( -0.313, 0.490)</td>
<td>(-0.402, 0.432)</td>
<td>(-0.349, 0.424)</td>
<td></td>
</tr>
<tr>
<td>$fx_t/\gamma_t$</td>
<td>-0.279</td>
<td>-0.273</td>
<td>-0.279</td>
<td>-0.286</td>
</tr>
<tr>
<td>( -0.482, -0.064)</td>
<td>( -0.469, -0.079)</td>
<td>(-0.478,-0.086)</td>
<td>(-0.490,-0.087)</td>
<td></td>
</tr>
<tr>
<td>$g_t/\tau_t$</td>
<td>0.013</td>
<td>0.012</td>
<td>0.012</td>
<td>0.01</td>
</tr>
<tr>
<td>( -0.082, 0.108 )</td>
<td>( -0.081, 0.105)</td>
<td>(-0.085,0.103)</td>
<td>(-0.086,0.108)</td>
<td></td>
</tr>
<tr>
<td>$ca_t/\tau_t$</td>
<td>0.032</td>
<td>0.034</td>
<td>0.038</td>
<td>0.400</td>
</tr>
<tr>
<td>( -0.145, 0.206 )</td>
<td>( -0.141,0.203)</td>
<td>(-0.125, 0.199)</td>
<td>(-0.114,0.194)</td>
<td></td>
</tr>
<tr>
<td>$yt/\tau_t$</td>
<td>-0.0105</td>
<td>-0.0108</td>
<td>-0.0102</td>
<td>-0.01</td>
</tr>
<tr>
<td>( -0.112, 0.093 )</td>
<td>( -0.114,0.093)</td>
<td>(-0.114, 0.093)</td>
<td>(-0.116,0.0923)</td>
<td></td>
</tr>
<tr>
<td>$rt/\tau_t$</td>
<td>-0.058</td>
<td>-0.056</td>
<td>-0.056</td>
<td>-0.052</td>
</tr>
<tr>
<td>( -0.217,0.096 )</td>
<td>( -0.207,0.091)</td>
<td>(-0.207, 0.094)</td>
<td>(-0.200,0.089)</td>
<td></td>
</tr>
</tbody>
</table>
Figure A.1: Standard Deviations of errors in current account equation to a tax shock
Appendix B

Chapter 2

B.1 Proof of linear transformation of multivariate normal

To derive Equation (23), we use a corollary in multivariate distribution as the following: Let $X \sim \mathcal{MN}_{p \times q}(M, U, V)$, and $Y = CXD + B$, where $C \in \mathbb{R}^{m \times p}$, rank$(C) = m \leq p$, $D \in \mathbb{R}^{q \times n}$, rank$(D) = n \leq q$, and $B \in \mathbb{R}^{m \times n}$. Then,

$$Y \sim \mathcal{MN}_{m \times n}(CMD + B, D'UD \otimes CV'C')$$

**Proof**: The density of $X$ is given by

$$p^{pq}_{MN}(X|M, U, V) = \frac{1}{(2\pi)^{\frac{p+q}{2}}|P|^\frac{p}{2}|Q|^\frac{q}{2}} \exp\left\{ -\frac{1}{2} \text{trace} \left[ Q^{-1}(X - M)'P^{-1}(X - M) \right] \right\} \quad (B.1)$$

, then use the property of the trace operator,

$$\text{trace} \left[ Q^{-1}(X - M)'P^{-1}(X - M) \right] = \sum_{i=1}^{p} \sum_{j=1}^{q} q^{ij}(x_i - m_i)'P^{-1}(x_j - m_j) \quad (B.2)$$

$$= \left[ \text{vec}(X - M)' \right]^\top (Q \otimes P)^{-1} \left[ \text{vec}(X - M) \right] \quad (B.3)$$

$$= \text{vec}(X - M)' (Q \otimes P)^{-1} \text{vec}(X - M) \quad (B.4)$$

101
Thus, $X$ follows $\mathcal{MN}_{p \times q}(vec(M), Q \otimes P)$. Make use of properties of the Kronecker product,

\[
vec(Y) = vec(CXD + B) = (D' \otimes C)vec(X) + B
\]  \hspace{1cm} (B.5)

The mean of $Y$ can be obtained as $\text{mean}(vec(Y)) = vec(D' \otimes C)M + B$. The variance is

\[
\text{var}[vec(Y)] = \text{var}[vec(CXB)]
\]
\[
= \text{var}[(D' \otimes C)(Q \otimes P)(D' \otimes C)']
\]
\[
= \text{var}[(D'Q \otimes CP)(D \otimes C')]
\]
\[
= \text{var}[(D'QD) \otimes (CPC')]
\]

. Therefore we have

\[
vec(Y) \sim \mathcal{MN}_{qp}(vec(CMD + C), (D'QD) \otimes (CPC')).
\]  \hspace{1cm} (B.6)

It is easy to see that $Y \sim \mathcal{MN}_{p \times q}(CMD + C, D'QD \otimes CPC')$. Note that in Equation (22), $G$ is constructed as the linear combination of $\tilde{B}$ and the i.i.d standard normal matrix $Z$ that could be written as $Z \sim \mathcal{MN}_{k \times n}(0, I_n \otimes I_k)$. Using the result in Equation (33), let $X = Z$, $(L_W^{-1})' = C$, and $L_{\Sigma}' = D$. Then the variance of $G$

\[
\text{var}(G) = (L_{\Sigma}^{-1})I_n(L_{\Sigma})' \otimes (L_W^{-1})'I_k(L_W^{-1}) = \Sigma_t \otimes W_B^{-1}.
\]

Therefore we have $vec(G) \sim \mathcal{N}_{nk}(vec(\tilde{B}), \Sigma_t \otimes W_B^{-1})$

102


## B.2 Data

### Table B.1: Description of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Transformation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. real gross domestic product</td>
<td>$400\Delta log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>U.S. PCE</td>
<td>$400\Delta log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>U.S. residential investment</td>
<td>$400\Delta log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>U.S. non-residential investment</td>
<td>$400\Delta log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>EU real gross domestic product</td>
<td>$400\Delta log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>EU GDP implicit deflator</td>
<td>$400\Delta log$</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>Global Inflows to Banks</td>
<td>$400\Delta log$</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>Global Inflows to non-Banks</td>
<td>$400\Delta log$</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>Global domestic credit</td>
<td>$400\Delta log$</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>U.S. domestic credit</td>
<td>$400\Delta log$</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>EU Leverage Quarterly Mkt Cap Weighted</td>
<td>no transformation</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>US brokers and dealers Financial leverage</td>
<td>no transformation</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>U.S. banking sector leverage</td>
<td>no transformation</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>EU and UK banking sector leverage</td>
<td>no transformation</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>Bank of England official bank rate</td>
<td>annualized rate</td>
<td>Miranda A &amp; Rey</td>
</tr>
<tr>
<td>Term spread (10-1 year CMR)</td>
<td>annualized rate</td>
<td>St. Louis FRED</td>
</tr>
<tr>
<td>EU central bank base rate</td>
<td>annualized rate</td>
<td>ECB Data Warehouse</td>
</tr>
<tr>
<td>UK banks leverage quarterly mkt cap weighted</td>
<td>no transformation</td>
<td>Miranda A &amp; Rey</td>
</tr>
</tbody>
</table>