9-2019

Essays on New Keynesian Term Premium Model with Financial Risks

Weiguo Fu

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Finance and Financial Management Commons, and the Macroeconomics Commons

Recommended Citation

https://academicworks.cuny.edu/gc_etds/3313

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
ESSAYS ON NEW KEYNESIAN TERM PREMIUM MODEL WITH FINANCIAL RISKS

by

Weiguo Fu

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2019
Essays on New Keynesian Term Premium Models with Financial Risks

by

Weiguo Fu

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

__________________________
Date

Thom Thurston
Chair of Examining Committee

__________________________
Date

Wim Vijverberg
Executive Officer

Supervisory Committee:

Thom Thurston
Tao Wang
John Devereux

The City University of New York
Abstract

Essays on New Keynesian Term Premium Models with Financial Risks

by

Weiguo Fu

Advisor: Thom Thurston

This dissertation studies the modeling of U.S. Treasury (UST) yield curve term premia under the New Keynesian (NK) framework. Loosely speaking, term premium is the difference between a government bond’s yield for a specific tenor and the average of the expected short rates up to that tenor. The dissertation is divided into three chapters. The first chapter proposes a New Keynesianism-based macro-finance model estimated by a one-step full information maximum likelihood (FIML) method. The second chapter shows that the one-step FIML method may produce estimation biases, which result in biased expected short rates and term premia. The chapter then presents an alternative estimation strategy. The third chapter addresses the policy rate’s zero lower bond (ZLB) constraint in the NK model by including a shadow rate concept.

The first chapter fills a gap in the macro-finance term premium modeling literature by building a two-way feedback loop between the economy and the yield curve in a micro-founded way. In doing so, the chapter incorporates a latent Financial Risk Index (FRI) in the IS curve and Taylor rule of an NK model with consumption habit formation. Using the Affine Term Structure (ATS) finance theory, it fits the model to macroeconomic and yield data to obtain time-varying term premia. The chapter also replaces the FRI with the UST three-month vs. 10-year yield slope
in the NK system to form Model 2, which offers the central bank and financial market participants an observable market variable to monitor and to communicate with and which thus builds the two-way feedback loop. The two models are both estimated by a one-step FIML method, in which the reduced-form vector autoregression of order 1, or VAR(1), coefficients and the structural NK parameters are estimated simultaneously.

The second chapter reveals that the one-step FIML method employed in the first chapter (and in the class of NK term premium models) may produce negative bias to the reduced-form VAR(1) coefficients, which in turn result in a too stable estimated 10-year average expected short rate series and 10-year term premium whose variations track those of the 10-year yield too closely. The chapter presents a two-step estimation strategy. The first step estimates the reduced-form VAR(1) model using ordinary least squares (OLS) and adjusts the negative small-sample estimation bias to the coefficients. The second step recovers structural NK parameters. The chapter proposes a structural restriction to the IS curve and thus improves the model fit to the data. The new estimation method produces a more cyclical and structural 10-year average expected short rate and a more counter-cyclical 10-year term premium than the first chapter. The method also restores consistency between the NK system and the reduced-form VAR system.

The third chapter addresses the ZLB issue by bringing in the Wu-Zhang Shadow Rate New Keynesian (SR-NK) model (Wu and Zhang 2016) into the first chapter’s macro-finance term premium modeling approach. The chapter points out that a connection between the Wu-Zhang SR-NK model and the yield curve cannot possibly be established and that there is a tenor mismatch between the short rate and the Wu-Xia shadow rate (Wu and Xia 2016). The chapter proposes a new SR-NK model that inherits the NK model of the first chapter with the short rate replaced by the latent shadow rate. The new model assigns the shadow rate and the FRI different
roles of yield curve level and slope drivers. It also dedicates the shadow rate to capture the effect of the Federal Reserve (Fed)’s forward guidance and the FRI to capture the effect of quantitative easing (QE). Thus, my SR-NK model addresses the two issues of the Wu-Zhang SR-NK model.

The chapter then proposes a simple way to replace the latent shadow rate by adjusting the short rate during the ZLB period using the variations of the one- and two-year yields to construct Model 2. The adjusted short rate is shown to reach negative levels similar to the negative rates adopted by other central banks. This adjustment constructs a ZLB constraint-free NK model without latent variables. It avoids the imputed latent variables’ sensitivity to parameter values and preserves the two-way feedback loop between the central bank and market participants.
Acknowledgements

First, I would like to express my enormous gratitude to the members of my dissertation committee – Dr. Thom Thurston, Dr. Tao Wang, and Dr. John Devereux – for their invaluable guidance and advice during my process of writing this dissertation. I am particularly in debt to Dr. Thurston, the chair of my dissertation committee. The two courses of Macroeconomics I and Monetary Theory that he taught opened me the door of New Keynesian monetary economics, which I found highly helpful for me to understand how the Fed’s monetary policies affect the global economy and financial markets from a theoretical perspective. As a positive side effect, it also has greatly facilitated my profession as an emerging market fixed income strategist over the past few years. Gradually and naturally, I came up with the idea of this dissertation – to study the mutual responses of the Fed’s monetary policies and the U.S. Treasury yield curve in a New Keynesian modeling framework. Therefore, it was his courses that have given birth to this dissertation. Furthermore, Dr. Thurston has patiently guided me through how to write clear and concise research papers. Dr. Wang and Dr. Devereux, too, have contributed to this dissertation with numerous extremely sharp comments. Many thanks to the three professors for spending their precious time meeting with me regarding this dissertation over the past two years. I have truly enjoyed those thought-provoking conversations with them.

I am deeply grateful to my wife Dr. Yingjie Li, my little kids Warren and Crystal for their understanding and support when I squeezed my time at home, which was supposed to be devoted to them, to work on this dissertation over the past two years. No words can describe my love and appreciation for them.
I would like to thank Dr. Wim Vijverberg, Dr. Merih Uctum, and Dr. Chun Wang for teaching the courses of Econometrics, International Macro, and Applied Macroeconometrics. These rigorous and fun courses have laid a solid foundation for me to conduct research in the interdisciplinary macro-finance field using rich macroeconomic and financial market data.

Special thanks are also due to the Department of Economics at the City University of New York – Graduate Center, which accepted me as a part-time PhD student in fall 2013 and which granted me a five-year tuition scholarship. I highly appreciate the help and support from all the faculty and administrative staff. It has been a truly pleasant environment over the past six years to take courses, to discuss inspiring research topics with the faculty and fellow PhD students, and to make friends.
# Table of Contents

Abstract............................................................................................................................................. iv

Acknowledgements ............................................................................................................................ vii

List of Tables ....................................................................................................................................... xii

List of Figures ...................................................................................................................................... xiii

Chapter 1  New Keynesian Model and the Term Premium: A New Approach ......................... 1
1.1  Introduction .................................................................................................................................. 1
1.1.1  Motivation and main contribution of this paper ................................................................. 1
1.1.2  Organization of this paper .................................................................................................... 8
1.2  Empirical relationship between the economy and the FRI ..................................................... 8
1.3  Model setup .................................................................................................................................. 12
1.3.1  The extended NK model ........................................................................................................ 12
1.3.2  Micro foundations for $q_t$ ...................................................................................................... 13
1.3.3  Adding the ATS part .............................................................................................................. 15
1.3.4  Extracting the term premium ................................................................................................. 17
1.4  Data .............................................................................................................................................. 17
1.5  Estimation .................................................................................................................................... 19
1.5.1  Model 1: the NK-ATS model keeping the UST 3m10y slope out of the state vector ........ 19
1.5.2  Model 2: the NK model treating the UST 3m10y slope as $q_t$ ............................................. 26
1.6  Impulse responses of the state variables to structural shocks ............................................... 27
1.7  Calculating term premia .............................................................................................................. 31
1.7.1  Term premia in Model 1 ......................................................................................................... 31
1.7.2  Treating $i_t$ and UST 3m10y slope as the level and slope factors of the entire yield curve ... 35
1.8  Conclusions and future research directions ............................................................................. 38

Chapter 2  Correcting Estimation Biases in the New Keynesian Term Premium Model with Financial Risks ................................................................................................................................. 40
2.1  Introduction .................................................................................................................................. 40
2.1.1  Motivations and contributions ................................................................................................. 40
2.1.2  Organization of this paper ....................................................................................................... 46
2.2  Model 2 of Fu (2017) and its estimation biases .......................................................................... 46
2.2.1  The NK-ATS model of Fu (2017) ......................................................................................... 46
Chapter 3 A Shadow Rate New Keynesian Term Premium Model with Financial Risks

3.1 Introduction ................................................................. 80
  3.1.1 Motivations and contributions .................................... 81
  3.1.2 Organization of this paper .......................................... 87
3.2 The Wu-Xia SR-ATS model and Wu-Zhang SR-NK model ............. 88
  3.2.1 The Wu-Xia SR-ATS model ....................................... 88
  3.2.2 The Wu-Zhang SR-NK model ..................................... 91
3.3 My proposed SR-NK-ATS term premium model ......................... 94
  3.3.1 The NK structural system ........................................... 94
  3.3.2 Adding the ATS part ................................................ 95
  3.3.3 The roles of \( s_t \) and \( q_t \) in driving the yield curve ....... 95
  3.3.4 The roles of \( s_t \) and \( q_t \) in explaining the Fed’s unconventional policies ....... 97
  3.3.5 Extracting the term premium .................................... 100
3.4 Data .................................................................................. 100
3.5 Estimation of my proposed SR-NK-ATS model ..................... 101
  3.5.1 Estimation by extended Kalman filter .......................... 101
  3.5.2 Estimation results .................................................... 102
3.6 Model 2: finding a proxy for the shadow rate at the ZLB .............. 110
  3.6.1 Using PCA-weighted variations of the one- and two-year yields as the proxy .... 110
3.6.2 Estimation of Model 2 .................................................................................................................. 116
3.7 Computing the 10-year term premium and assessing its impact on growth .......................... 119
  3.7.1 Computing the 10-year term premium using Model 2 .......................................................... 119
  3.7.2 Assessing whether the 10-year term premium can predict economic growth .................. 120
3.8 Conclusions and future research directions ............................................................................. 124

Appendices ......................................................................................................................................... 126
Appendix 1.A The derivation of the extended NK model ............................................................. 126
  1.A.1 Consumer’s utility maximization ............................................................................................. 126
  1.A.2 The IS curve and the stock pricing equation .......................................................................... 127
  1.A.3 Adding lagged output gap to the IS curve ............................................................................ 130
  1.A.4 The Phillips curve .................................................................................................................. 132
  1.A.5 The Taylor rule ...................................................................................................................... 133
Appendix 1.B The derivation of the IS equation using non-separable utility preference and
consumption habit formation ............................................................................................................. 134
Appendix 3.A The extended Kalman filter algorithm ..................................................................... 135

Bibliography ....................................................................................................................................... 137
## List of Tables

1.1 Cross-correlations of output gap with UST 3m10y slope and AAA 10y spread .................. 9
1.2. Cross-correlations of UST 3m yield with UST 3m10y slope and AAA 10y spread ........... 10
1.3. Cross-correlations of the synthetic FRI with output gap, short rate and inflation .......... 11
1.4. ADF test results and autocorrelations .................................................................... 19
1.5. Structural parameter and market prices of risk estimates of Model 1 ....................... 23
1.6. Structural parameter estimates of Model 1 and of Model 2 ..................................... 27
1.7. OLS regression results of yields on 3-month yield and 10-year slope ....................... 37
2.1. Structural NK parameters estimated by Model 2 of Fu (2017) and by SVAR .............. 62
2.2. Structural NK parameters estimated by Model 2 of Fu (2017), SVAR and grid search ... 76
3.2. NK structural parameter estimates vs. those by the grid search of Fu (2018) .......... 102
3.3. OLS regression results of yields on Adj 3m yield and 3m10y slope ......................... 115
3.4. Structural parameter estimates by SVAR and by Fu (2018) ..................................... 118
3.5. Output growth’s cross-correlation with itself, differenced $E_t [i_t^{10y}]$ and $tp_t^{10y}$ .......... 122
3.6. Regression results of output growth on differences in short rate and term premium .... 123
List of Figures

1.1. CBO output gap, UST 3m10y slope and AAA 10y spread ................................................................. 9
1.2. US inflation, output gap, short rate, UST 3m10y slope and FRI.................................................... 18
1.3. Residuals of fitted vs. actual 1-, 3- and 5-year yields and UST 3m10y slope ..................................... 25
1.4. Synthetic and imputed FRIs vs. UST 3m10y slope ....................................................................... 26
1.5. Impulse responses in Model 1..................................................................................................... 29
1.6. Impulse responses in Model 2..................................................................................................... 30
1.7. 10-year term premium, yield and average short rate ................................................................... 31
1.8. 10-year term premium based on different methods in Rudebusch et al. (2007) ......................... 35
1.9. Level factor vs. 3-month yield and slope factor vs. UST 3m10y slope ........................................ 37
1.10. 10-year term premia in Model 1 and 2 .................................................................................... 38
2.1. 10-year average short rate and term premium estimated by FIML and OLS .......................... 55
2.2. 10-year average short rate and term premium by OLS and bias-corrected OLS ..................... 58
2.3. The norm of solution residuals of Equation (2.48) as a function of $\delta_y$ and $\delta_q$ ........... 66
2.4. The norm of solution residuals of Equation (2.46) as a function of $\beta$ and $\kappa$ .................... 67
2.5. The relationship between the norm of residuals and $\phi_q$ ..................................................... 69
2.6. The norm of residuals as increasing functions of $\phi_r$ and $\phi_y$ when $\phi_q = 0.525$ .......... 70
2.7. The relationship between the norm of residuals and $\alpha$ ....................................................... 72
2.8. The relationship between the norm of residuals and $\psi$ ....................................................... 74
2.9. The norm of residuals as increasing functions of $\alpha$ and $\theta$ when $\psi = 0.995$ ............... 75
3.1. Wu-Xia shadow rate vs. EFFR, UST 1-year and 2-year yields ................................................. 91
3.2. Wu-Xia shadow rate vs. two latent factors during the ZLB period ........................................ 96
3.3. $B_n$ coefficients of forward yields w.r.t. the four state variables of $X_t$ .......................... 103
3.4. Imputed shadow rate vs. three-month, one-year and two-year yields (1998-2016) ............. 104
3.5. Adjusted shadow rate vs. three-month, one-year and two-year yields (1998-2016) .......... 105
3.6. Imputed FRI vs. UST 3m10y slope (1998-2016) ................................................................ 106
3.7. Estimated Wu-Xia shadow rates with different lower bounds .......................................... 109
3.8. Adjusted 3m yield vs. the Wu-Xia shadow rate and actual 3m, 1-year, 2-year yields ......... 112
3.9. UST 3m yield’s fitted and forecast values in 2 Taylor rules (1Q 1985 to 3Q 2008)........... 113
3.10. UST 3m yield’s fitted and forecast values in 2 Taylor rules (1Q 1997 to 3Q 2008)......... 114
3.11. Computed 10-year term premium and average expected short rate ............................. 120
3.12. GDP growth, differenced 10-year term premium and average expected short rate........ 124
Chapter 1

New Keynesian Model and the Term Premium: A New Approach

1.1 Introduction

Term premium, loosely speaking, is the difference between a government bond yield for a specific tenor and the average of the expected short-term rates up to that tenor. Term premium started catching the attention of researchers in macroeconomics and finance from 2004 to 2006, a period during which the Federal Reserve (Fed) had kept raising its policy rate while long-term Treasury bond yields had remained steady, a phenomenon that then-Fed Chair Alan Greenspan called "conundrum". In his July 2005 monetary policy testimony (Greenspan 2005), Greenspan attributed the conundrum to a fall in term premia. Since then, term premium has become a well-studied topic in the macro-finance literature.

1.1.1 Motivation and main contribution of this paper

The purpose of this study is to fill a gap in the literature of macro-finance term premium modeling. Rudebusch et al. (2007) categorize this line of research into the following three classes of models:

(1) Consumption-based asset pricing DSGE models. This class of models starts from the representative consumer’s utility maximization to derive the stochastic discount factor for Treasury bonds, from which the yield curve can be calibrated to macroeconomic variables such as output and inflation. The advantage of this model is that it constructs a two-way feedback relationship between the yield curve and macroeconomic variables.
However, there are two drawbacks: 1) calculating the time-varying term premium from the calibrated yield curve requires a third-order approximation to the solution of the DSGE model because the term premium is zero in a first-order approximation and constant in a second-order approximation. Therefore, this model is computationally intractable; and 2) the term premium’s impulse response to a one percentage point shock in output and policy rate is smaller than one basis point, as produced by a sample DSGE model of Rudebusch et al. (2007). This is unrealistically too small.

(2) VAR-based macro-finance model. A seminal paper is Ang and Piazzesi (2003). It uses the Affine Term Structure (ATS) finance theory to exogenously specify the stochastic discount factor as an affine (constant plus linear term) function of two macroeconomic variables and three latent variables. These five state variables are modeled as a VAR process. Therefore, the yield curve is also an affine function of the VAR process. The advantages of this class of model are: 1) it is straightforward to derive the term premium from the calibrated yield curve; and 2) the model is computationally tractable. The two main drawbacks are: 1) there is no structural relationship among those macro and latent variables; and 2) there is no feedback from the yield curve back to the economy.

(3) New Keynesian (NK) macro-finance model. Representative papers are Hördahl et al. (2006), Rudebusch and Wu (2008) and Bekaert et al. (2010). Like VAR-ATS models, this class of models also uses the ATS theory to define the yield curve as an affine function of a state vector process. The difference is that the state variables are governed by an NK structure rather than by a reduced-form VAR. The two main advantages of this class of model are: 1) it provides a parsimonious structural relationship among state variables; and 2) it is computationally tractable. The drawbacks are: 1) there is no
feedback from the yield curve back to the economy in Hördahl et al. (2006); 2) in Rudebusch and Wu (2008), while the feedback from the yield curve to the economy is implicitly provided through plugging in two latent factors – level and slope, this plug-in is not micro-founded. Furthermore, the paper does not extract term premia from the yield curve; and 3) in Bekaert et al. (2010), there is no direct feedback from the yield curve to the economy even though the model also plugs in two latent variables – natural output gap and stochastic inflation target. Furthermore, term premia are constant rather than time-varying in this model.

As summarized above, the gap in the macro-finance term premium modeling literature is a lack of a tractable micro-founded two-way feedback loop between the economy and the yield curve. The class of consumption-based asset pricing DSGE models provides a two-way feedback loop, but it is computationally intractable and the computed term premia are unrealistically too small. VAR-ATS and NK-ATS models are computationally tractable, their computed term premia are time-varying and look realistic, but there is no micro-founded two-way feedback loop in these two types of models. To fill this gap, I choose to work on an NK-ATS model due to its appealing structural relationship among state variables and due to its tractability.

To provide a two-way feedback loop between the economy and the yield curve, I incorporate a latent financial variable into the NK system. I call this variable the Financial Risk Index and hypothesize that it will affect the real economy and the Fed’s monetary policy. This line of thought is inspired by the fact that the Fed has included words like “The Committee will closely monitor incoming information on economic and financial developments” in its meeting statements over the past few years (FOMC 2012 – 2017). Moreover, the Fed has revised “financial developments” to “financial and international developments” since its meeting in
September 2015, after the Chinese stock market experienced an eye-popping crash in June-August 2015 and the Chinese yuan depreciated suddenly against the dollar on August 11, 2015, which triggered a temporary global stock market crash. The explicit inclusion of “financial developments” and “financial and international developments” into the Fed’s meeting statements implies that the conditions of the domestic and major international financial markets should enter the Fed’s reaction function, or in the context of the NK model, the Taylor rule. It also implies that in the Fed’s eye, adverse financial conditions, if persistent, will negatively impact the real economy. In the NK context, this means the FRI should enter the IS curve.

Following this line of reasoning, I insert the latent FRI into the IS equation and Taylor rule and also specify a separate equation for the FRI to form a four-equation NK model. I explore a list of financial variables that may be correlated with the output gap and inflation. I find that, among those financial variables examined, the U.S. Treasury three-month vs. 10-year yield spread (henceforth the UST 3m10y slope) has the largest absolute correlation with the output gap. Therefore, the UST 3m10y slope may be a good proxy for the latent FRI. If so, the UST 3m10y slope, given that it is now part of the NK system and it is widely deemed to drive the slope of the yield curve, will provide a two-way feedback loop between the economy and the yield curve.

Furthermore, unlike Rudebusch and Wu (2008), which plug yield curve level and slope factors into the NK system in an ad hoc way, I include the FRI into the IS equation and Taylor rule with micro foundations' support. My setup is in essence similar to Nisticò (2012), who incorporates firms’ stock prices into consumers’ budget constraint equation to derive an overlapping generation NK model that adds an aggregate stock market price into the IS equation and Taylor rule and that arrives at a forward-looking pricing equation for the stock market. I
extends Nisticò (2012)’s setup by allowing consumers’ habit formation, thus adding lagged output gap into the IS equation to capture empirical data’s serial correlation. I replace the stock market price in Nisticò (2012) with the FRI in the NK system and incorporate the lagged FRI into the pricing equation. I also include lagged inflation and short rate into the NK system, as done by Hördahl et al. (2006) and Bekaert et al. (2010).

After a micro-founded NK model is established, I use the ATS theory to specify the stochastic discount factor and the yield curve as affine functions of the economy represented by inflation, output gap, monetary policy and financial market. This step allows extracting time-varying term premia tractably, as done by VAR-ATS models such as Ang and Piazzesi (2003) and other NK-ATS models such as Hördahl et al. (2006).

Like Hördahl et al. (2006), I use full information maximum likelihood (FIML) method to estimate my NK-ATS model based on a 32-year dataset that contains U.S. personal consumption expenditures (PCE) inflation, U.S. Congressional Budget Office (CBO)-compiled output gap, U.S. Treasury three-month yield (as a proxy for monetary policy rate) and the UST 3m10y slope. Though the FRI rather than the UST 3m10y slope is a state variable of the NK system, the UST 3m10y slope as an affine function of the four state variables is used in the model estimation because the FRI is assumed latent. To improve the model fit, I also include U.S. Treasury one-, two-, three-, five- and seven-year yields, all of which as an affine function of the state vector, in the estimation. After the NK structural parameters are estimated, I impute the FRI and find it indeed highly correlated with the UST 3m10y slope, confirming my hypothesis that the UST 3m10y slope is a good proxy for the FRI. I then re-estimate the NK model using the UST 3m10y slope as a state variable of the NK system without other Treasury yields and thus without using the ATS theory. I call this Model 2 as opposed to Model 1 that uses the ATS theory. The two
models’ structural parameter estimates and state variables’ impulse responses are highly similar – after all, the imputed FRI in Model 1 and the UST 3m10y slope are highly correlated. As such, the second model that uses the UST 3m10y slope as a state variable of the NK system creates a two-way feedback loop between the yield curve and the real economy.

I decompose the 10-year yield into the 10-year average expected short rate and term premium in Model 1 and 2. To obtain the term premium in Model 2, I treat the three-month yield and UST 3m10y slope as the level and slope factors of the yield curve and regress other yields on these two factors. I then use a fitted yield curve to calculate the term premium. The 10-year term premia in Model 1 and 2 share almost the same historical pattern.

I believe my two models provide more value to monetary policy makers and financial market practitioners than those existing seminal NK macro-finance models. First, my Model 1 brings a financial variable – the FRI – into the NK system, while Hördahl et al. (2006) introduce a latent inflation target into the NK system and Bekaert et al. (2010) add a latent natural output gap together with a latent inflation target, both of which are still macro variables. As such, my Model 1 can capture the Fed’s newly-established focus on financial markets (as evidenced by the inclusion of “financial developments” in the FOMC’s meeting statements since 2012) while Hördahl et al. (2006) and Bekaert et al. (2010) cannot. Second, the FRI is intuitive to market practitioners, but it would be unintuitive for market practitioners to link the latent inflation target and latent natural output gap to financial markets. Last but not least, my Model 2 replaces the FRI with the UST 3m10y slope and thus gives the Fed and market practitioners an observable market variable to monitor. The Fed and market practitioners can then communicate through the UST 3m10y slope, which establishes the two-way feedback loop between the two parties.
Moreover, my model provides a better fit to the yield curve than Bekaert et al. (2010). My Model 1 is a direct comparison to Bekaert et al. (2010) because both models use the ATS theory (my Model 2 does not). In the estimation of my Model 1, the Treasury three-month and 10-year yields are assumed to be measured without errors, while in the estimation of Bekaert et al. (2010), the three-month and five-year yields are assumed to be measured without errors. Other yields all have measurement errors. The standard errors for the residuals of one-, three- and five-year yields in my Model 1 are 19, 22 and 16 basis points, respectively, which are smaller than half of the standard errors of 45 and 54 basis points for the residuals of one- and 10-year yields in Bekaert et al. (2010).

I examine the estimated 10-year average short rate and term premium in my Model 1 and 2. I find that the 10-year average short rate has been very flat over the past three decades, especially after the 2008 financial crisis, when the Fed cut its policy rate to nearly zero and had held it unchanged until December 2015. Accordingly, the 10-year term premium is shown to have moved with the 10-year yield in lockstep over 30 years. In particular, the term premium has turned more and more negative between 2014 and 2016. I suspect that the 10-year average expected short rate might be incorrectly estimated given that structural factors (e.g., demographics, globalization, etc.) should have kept pulling the long-term natural rate lower. Indeed, the Fed has kept lowering its estimated long-term natural rate over the past few years. If this hypothesis is correct, the 10-year term premium might have not declined that much between 2014 and 2016. Since this paper’s focus is on proposing an NK macro-finance model that provides a two-way feedback loop between the economy and the yield curve rather than on term premium estimation and since whether the term premium is plausibly estimated will not
invalidate the two-way feedback loop, I leave this hypothesis to be explored in my future research.

1.1.2 Organization of this paper

The rest of this paper is organized as follows. Section 1.2 empirically investigates the relationship between the economy and financial variables. Section 1.3 presents the mathematical setup of my proposed NK-ATS model. Section 1.4 describes the data used. In Section 1.5, the maximum likelihood estimation procedure is discussed in great detail. This section also explains the QZ method used to solve the structural system. Section 1.6 constructs impulse response functions of macroeconomic variables and the yield curve. Section 1.7 discusses how to obtain term premia in Model 1 and 2. Section 1.8 concludes and points to future research directions.

1.2 Empirical relationship between the economy and the FRI

I examine a list of financial variables that are supposed to be correlated with the real economy and find that the UST 3m10y slope and the U.S. Corporate AAA 10-year yield spread (henceforth AAA 10y spread) have been strongly negatively correlated with output gap in a data sample going back to 1985, as depicted in Figure 1.1. Furthermore, these two financial variables appear to be stationary, which means their levels can enter the model without transformation.
To quantify the relationship between the two financial variables and output gap, I calculate output gap’s cross-correlations with these two variables, as shown in Table 1.1.

Table 1.1 Cross-correlations of output gap with UST 3m10y slope and AAA 10y spread

<table>
<thead>
<tr>
<th>Variable</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>UST 3m10y slope</td>
<td>-0.52</td>
<td>-0.69</td>
<td>-0.73</td>
<td>-0.74</td>
<td>-0.71</td>
<td>-0.66</td>
<td>-0.58</td>
<td>-0.50</td>
<td>-0.39</td>
</tr>
<tr>
<td>AAA 10y spread</td>
<td>-0.22</td>
<td>-0.26</td>
<td>-0.32</td>
<td>-0.41</td>
<td>-0.49</td>
<td>-0.54</td>
<td>-0.55</td>
<td>-0.52</td>
<td>-0.47</td>
</tr>
</tbody>
</table>

Table 1.1 confirms the strongly negative correlations depicted in Figure 1.1. Also, we can see that the UST 3m10y slope actually leads output gap by one quarter as the $r-1$ correlation of $-0.74$ is most negative. This empirical phenomenon was once voiced by many prominent economists such as former Fed Chair Ben Bernanke, who pointed out “Historically, the slope of the yield curve has tended to decline significantly in advance of recessions” in his speech at the Economic
Club of New York on March 20, 2006 (Bernanke 2006). Indeed, the UST 3m10y slope declined rapidly in 2006 and turned negative in the fourth quarter of 2006, but switched back to positive in the second quarter of 2007 and then continued to move up afterwards. In the third quarter of 2007, U.S. GDP growth entered a downward trend and slipped to negative in the first quarter of 2008. This lead-lag causal relationship shows that the UST 3m10y slope is an excellent candidate component of the FRI to be added onto the right-hand side of the IS equation. For the AAA 10y spread, though it is led by the output gap by 1-2 quarters, the contemporaneous correlation of -0.49 appears to be strong enough to be also included in the FRI.

The cross-correlations of the short-term interest rate with the UST 3m10y slope and AAA 10y spread, as depicted in Table 1.2, show that the latter two variables have historically affected the Fed's monetary policy. Interestingly, this time the short rate is more negatively correlated with the AAA 10y spread than with the UST 3m10y slope.

Table 1.2. Cross-correlations of UST 3m yield with UST 3m10y slope and AAA 10y spread

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time lag (quarter)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-4</td>
</tr>
<tr>
<td>UST 3m10y slope</td>
<td>-0.34</td>
</tr>
<tr>
<td>AAA 10y spread</td>
<td>-0.51</td>
</tr>
</tbody>
</table>

I also examine output gap's correlations with other financial variables such as the S&P 500 index and the dollar index (DXY) but find rather weak correlations, whether the levels or returns of the S&P 500 and DXY are used (since the S&P 500 and DXY are nonstationary variables, their returns rather than levels should be used to avoid spurious correlations).
Therefore, though "financial (and international) developments" in the Fed's eye may involve many financial variables, the UST 3m10y slope and AAA 10y spread could be two representative ones. After all, these variables determine the funding costs of the U.S. corporate sector and the UST 3m10y also helps determine the funding costs of the U.S. mortgage market and of many countries in the world. A synthetic FRI may be formed by extracting the first principal component from the UST 3m10y slope and AAA 10y spread. Table 1.3 displays the cross-correlations of this synthetic FRI with output gap, short rate and inflation. This demonstrates that the FRI indeed should enter the IS curve and Taylor rule.

Table 1.3. Cross-correlations of the synthetic FRI with output gap, short rate and inflation

<table>
<thead>
<tr>
<th>Variable</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output gap</td>
<td>-0.61</td>
<td>-0.68</td>
<td>-0.73</td>
<td>-0.75</td>
<td>-0.74</td>
<td>-0.71</td>
<td>-0.64</td>
<td>-0.56</td>
<td>-0.45</td>
</tr>
<tr>
<td>Short rate</td>
<td>-0.42</td>
<td>-0.51</td>
<td>-0.57</td>
<td>-0.61</td>
<td>-0.63</td>
<td>-0.62</td>
<td>-0.59</td>
<td>-0.53</td>
<td>-0.47</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.23</td>
<td>-0.28</td>
<td>-0.31</td>
<td>-0.31</td>
<td>-0.30</td>
<td>-0.29</td>
<td>-0.26</td>
<td>-0.24</td>
<td>-0.21</td>
</tr>
</tbody>
</table>

However, the synthetic FRI’s cross-correlations with inflation appear rather weak. I envision that it could be because financial markets’ boom and bust tend to have a much larger effect on the prices of luxury goods than on those of basic consumption goods while inflation is heavily weighted by basic consumption goods. This should at least partly explain why the S&P 500 index has more than tripled since March 2009 but inflation has been muted. The same puzzle has been observed between skyrocketing residential property prices and weak inflation in China since 2012. As such, the FRI probably should not enter the Phillips curve. After all, aggregate supply should be affected by more structural factors than financial market variables.
1.3 Model setup

1.3.1 The extended NK model

My proposed NK-ATS model begins with a modified three-equation NK system with habit formation (see, e.g., Fuhrer 2000, Dennis 2009):

\[
\pi_t = \beta E_t \pi_{t+1} + (1 - \beta) \pi_{t-1} + \kappa y_t + \epsilon_{\pi,t}, \tag{1.1}
\]

\[
y_t = \alpha E_t y_{t+1} + \psi y_{t-1} - \alpha (i_t - E_t \pi_{t+1}) - \theta q_t + \epsilon_{y,t}, \tag{1.2}
\]

\[
i_t = \mu_i + \rho i_{t-1} + (1 - \rho) (\delta \pi_t + \delta y_t - \delta q_t) + \epsilon_{i,t}. \tag{1.3}
\]

Equation (1.1) is the Phillips curve, (1.2) the IS curve and (1.3) the Taylor rule. As usual, \(\pi_t\), \(y_t\) and \(i_t\) denote inflation, output gap and short rate, respectively. Note that this version of “modified” NK model is different than the “standard” NK model in that it adds lagged variables \(\pi_{t-1}, y_{t-1}\) and \(i_{t-1}\) to capture the empirical evidence of adaptive inflation expectations, consumption habit formation and policy rate smoothing. Another difference is that in the IS equation, the expected output gap at time \(t+1\) and the current real interest rate share the same coefficient.

My NK-ATS model extends the modified NK model by bringing in one latent variable – the FRI, which is denoted by \(q_t\) as shown in Equations (1.2) and (1.3). \(q_t\) is assumed to follow the process below:

\[
q_t = \phi_q E_t q_{t+1} - \phi_y E_t y_{t+1} + \phi_r (i_t - E_t \pi_{t+1}) + (1 - \phi_q) q_{t-1} + \epsilon_{q,t}. \tag{1.4}
\]

Note that \(q_t\) depends on the expected values of \(q_{t+1}\) and \(y_{t+1}\). This fits the reality that the FRI should be driven by the expected future states of itself and the real economy.

Also note that \(q_t\) is not on the right-hand side of the Phillips curve.

In Appendix 1.A, I show how Equations (1.1) to (1.4) in the NK system are derived.
Equations (1.1) – (1.4) can be written in a compact form:

\[ BX_t = \mu + AE_tX_{t+1} + MX_{t-1} + V \varepsilon_t, \]  

(1.5)

where \( V \) is an identity matrix, which means the four shocks in \( \varepsilon_t \) are uncorrelated with one another, and \( \varepsilon_t \sim N(0, D) \) where \( D \) is the diagonal variance matrix. This is a rational-expectation structural system. Using the undetermined coefficients method, a VAR(1) solution can be guessed as:

\[ X_t = c + \Omega X_{t-1} + \Gamma \varepsilon_t, \]  

(1.6)

where \( \Gamma \) is a variance-covariance matrix of \( \varepsilon_t \) with non-zero covariance (and correlation) coefficients.

Equation (1.6) will be solved from (1.5) by QZ method, which will be discussed in Section 1.5.

1.3.2 Micro foundations for \( q_t \)

In an NK model based on an infinitely-lived representative agent setting, including an additional variable such as \( q_t \) onto the right-hand side of the IS equation can be done by changing the separable consumption-leisure utility preference to a non-separable one. Andrés et al. (2006) assume a non-separable preference in a money-in-the-utility (MIU) function model and come up with the IS curve, Phillips curve and Taylor rule all including money on the right-hand side. In the context of my proposed model, the FRI will replace money in the representative agent’s utility function of the MIU model and show up in the IS curve, Phillips curve and Taylor rule. As an illustration, I show how to derive the IS equation in Appendix 1.B. This approach, however, has been challenged by, e.g., McCallum (2001) and Woodford (2003), which argue that the non-separability effects can be negligible.
Nisticò (2012) studies stock prices' wealth effect on the real economy and monetary policy using an overlapping generation NK model. The separability of the consumption-leisure utility preference is preserved, yet the infinitely-lived representative consumer in the standard NK model is replaced by an infinite number of cohorts of consumers with a certain probability of dying in each period. Nisticò (2012)'s other main difference from the standard NK model is that the consumer's budget constraint contains not only government bonds, but also firms’ stocks. The author shows that a cohort’s current consumption is a linear function of the current financial wealth and human wealth. By aggregating across cohorts, the author obtains the same conclusion for the aggregate economy. Since the aggregate financial wealth is mainly driven by the aggregate stock market price and the aggregate human wealth is related to future labor income and future consumption, the economy’s current consumption is affected by the current stock market price and future consumption. As such, Nisticò (2012) obtains an IS equation with the stock market price on the right-hand side and a forward-looking pricing equation for the stock market (see a simplified derivation in Appendix 1.A). The Phillips curve remains the same as it is in the standard NK model. Nisticò (2012) examines two versions of Taylor rule including different forms of stock price – stock price level from the equilibrium and stock price growth.

Nisticò (2012) shows that if the probability that consumers die in each period is set to zero, i.e., if consumers live indefinitely, current consumption will not depend on current stock market price and the IS equation collapses to the standard NK version.

My proposed NK model is in essence very similar to Nisticò (2012)'s in that firm’s stocks are risky and the FRI can be a composite indicator that measures the overall riskiness of risky assets including firms’ stocks. From this angle, firms’ stocks and the aggregate stock market in Nisticò (2012)’s consumer's budget constraint can be replaced by a representative risky asset with the
FRI as the indicator of this asset’s riskiness. In the end, this representative risky asset appears on the right-hand side of the IS equation. Assuming the FRI is a linear function of the representative risky asset's price with the sign of the coefficient being negative, we can replace the risky asset with the FRI in the derived NK system (see Appendix 1.A.2).

At this stage, the NK system does not contain lagged state variables. To add lagged output gap into the IS equation, I modify Nisticò (2012)'s consumer utility function to include consumption levels at time $t$ and $t-1$ with a habit formation parameter attached to the latter (see, e.g., Fuhrer 2000). With this type of utility function, a cohort's consumption at time $t$ becomes a weighted average of financial and human wealth at time $t$ as well as consumption at time $t-1$ (see Appendix 1.A.3). As such, lagged output gap also appears on the right-hand side of the IS equation.

### 1.3.3 Adding the ATS part

ATS models (e.g., Duffie and Kan 1996, Dai and Singleton 2002) start with a short rate that is an affine function of a state vector process $X_t$:

$$i_t = a_0 + a_1' X_t. \quad (1.7)$$

In this paper, $X_t$ is as defined in Equation (1.6).

Since $i_t$ is the third state variable in $X_t$, $a_0 = 0$ and $a_1 = [0 \ 0 \ 1 \ 0]'$.

Next step is to make a zero-coupon Treasury bond (and yield) an affine function of $X_t$. To price a Treasury bond, we need to transform the physical probability measure $P$ on which the historical data of $X_t$ are based to the risk-neutral probability measure $Q$. The Radon-Nikodym derivative $L_t$ serves this purpose: $E^Q [W_{t+1}] = E^P [L_{t+1} W_{t+1}]/L_t$, where $W_t$ is a random variable and $L_t$ follows:
\[ L_{t+1} = L_t \exp \left( -\frac{1}{2} \lambda_t' \lambda_t - \lambda_t' \varepsilon_{t+1} \right). \]  

(1.8)

\( \lambda_t \) measures market prices of risk and also is an affine function of \( X_t \):

\[ \lambda_t = \lambda_0 + \lambda_1 X_t. \]  

(1.9)

The stochastic discount factor \( m_t \) is therefore:

\[ m_{t+1} = \exp \left( -i_t \frac{L_{t+1}}{L_t} \right) = \exp \left( -i_t - \frac{1}{2} \lambda_t' \lambda_t - \lambda_t' \varepsilon_{t+1} \right). \]  

(1.10)

The stochastic discount factor is used to discount the bond price at time \( t+1 \) to time \( t \) as follows:

\[ p_t^{n+1} = E_t [m_{t+1} p_{t+1}^n]. \]  

(1.11)

where \( p_t^n \) is the price of an \( n \)-period zero-coupon bond at time \( t \).

Assume the price of \( p_t^n \) is an exponential affine function of \( X_t \):

\[ p_t^n = \exp(A_n + B_n' X_t). \]  

(1.12)

Plugging Equations (1.10) and (1.12) into (1.11), we can solve for \( A_n \) and \( B_n \) iteratively:

\[ A_{n+1} = A_n + B_n' (c - \Gamma \lambda_0) + \frac{1}{2} B_n' \Gamma' B_n - a_0, \]  

(1.13)

\[ B_{n+1}' = B_n' (\Omega - \Gamma \lambda_1) - a_1'. \]  

(1.14)

The initial values of \( A_n \) and \( B_n \) are \( A_1 = -a_0 \) and \( B_1 = -a_1 \).

An \( n \)-period zero-coupon bond yield \( y_t^n \) is obtained as:

\[ y_t^n = -\frac{A_n}{n} - \frac{B_n'}{n} X_t. \]  

(1.15)

As such, the entire yield curve is an affine function of \( X_t \) with the parameters \( c, \Omega \) and \( \Gamma \) in Equation (1.6) and \( \lambda_0 \) and \( \lambda_1 \) in Equation (1.9). These parameters need to be estimated from data.

An \( n \)-period yield spread \( s_t^n \) is then the difference between the \( n \)-period yield and the three-month yield (also short rate in this paper):

\[ s_t^n = y_t^n - i_t. \]  

(1.16)
1.3.4 Extracting the term premium

To extract the term premium from an \( n \)-period yield, we need to first obtain the forward risk premium embedded in a one-period forward yield. The one-period forward premium, denoted as \( r_t^n \), is the difference between the one-period forward yield \( n \) periods ahead, \( f_t^n \), and the expected one-period short rate \( n \) periods ahead, \( E_t[i_{t+n}] \).

\[
r_t^n = f_t^n - E_t[i_{t+n}]. \tag{1.17}
\]

\( f_t^n \) is calculated as the log difference between an \( n \)-period bond and \( n+1 \)-period bond:

\[
f_t^n = \ln(p_t^n) - \ln(p_t^{n+1}) = (A_n - A_{n+1}) + (B'_n - B'_{n+1})X_t, \tag{1.18}
\]

And

\[
E_t[i_{t+n}] = E_t[a_0 + a'_1X_{t+n}] = a'_1\Omega^nX_t. \tag{1.19}
\]

\( E_t[i_{t+n}] \) is just the \( n \)-period ahead forecast of \( i_t \) using the third equation of the VAR(1) system (Equation 1.6). Hence,

\[
r_t^n = (A_n - A_{n+1}) + (B'_n - B'_{n+1} - a'_1\Omega^n)X_t. \tag{1.20}
\]

The term premium is the average of the forward premia up to time \( t+n-1 \):

\[
t^n = \frac{1}{n} \sum_{i=0}^{n-1} r_t^i. \tag{1.21}
\]

Therefore, the term premium also is an affine function of \( X_t \).

1.4 Data

The dataset contains quarterly PCE inflation, CBO output gap, AAA 10y spread, and Treasury three-month, six-month, one-year through 10-year zero-coupon bond yields. Quarterly average yields are computed and used. According to CBO’s white paper (CBO 2004), the output gap series is constructed using a growth model. The UST 3m10y slope is computed as the difference between the 10-year yield and three-month yield. A synthetic FRI is constructed as the
first principal component of the vector of UST 3m10y slope and AAA 10y spread. The sample period is from the first quarter of 1985 to the fourth quarter of 2016. Data are obtained from Bloomberg. As discussed above, the three-month yield is used as \( i_t \). Figure 1.2 depicts a snapshot of inflation, output gap, short rate, the UST 3m10y slope and the synthetic FRI.

Figure 1.2. US inflation, output gap, short rate, UST 3m10y slope and FRI

From the charts above, it appears that all the variables are stationary. Stationarity is required for the state variables of the NK system because the solution to the NK system is a VAR(1) process; if the state variables are nonstationary, they have to be differenced before entering the VAR(1) solution. I apply augmented Dickey-Fuller (ADF) tests and confirm all the variables are stationary (see Table 1.4):
Table 1.4. ADF test results and autocorrelations

<table>
<thead>
<tr>
<th></th>
<th>( \pi )</th>
<th>( y )</th>
<th>( i )</th>
<th>UST3m10y</th>
<th>AAA 10y</th>
<th>Synthetic FRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF</td>
<td>-4.19**</td>
<td>-3.18**</td>
<td>-3.67*</td>
<td>-3.60**</td>
<td>-2.90**</td>
<td>-3.83**</td>
</tr>
<tr>
<td>Autocorrelation(-1)</td>
<td>0.85</td>
<td>0.95</td>
<td>0.97</td>
<td>0.92</td>
<td>0.91</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Note: * means significant at 5% and ** means significant at 1%

Table 1.4 also shows that all the variables are highly correlated with themselves in a lag order of 1.

1.5 Estimation

I use full information maximum likelihood method to estimate the structural parameters in the four-equation NK system. In the estimation, I entertain two types of models: 1) the NK-ATS model with the UST 3m10y slope and some Treasury yields as affine functions of \( X_t \). This is a state space model containing a measurement equation of yield-related variables and a transition equation of the four state variables; and 2) the NK model with the UST 3m10y slope as a state variable. The measurement equation of yields is not needed in this Model. The purpose is to examine whether the UST 3m10y slope is a good proxy of the latent FRI.

To reduce the parameter space, I use demeaned variables in the estimation. Therefore, \( c = 0 \) in Equation (1.6).

1.5.1 Model 1: the NK-ATS model keeping the UST 3m10y slope out of the state vector

In the NK system, since the FRI \( q_t \) is assumed latent (what financial variables to which the Fed responds is unknown to the public), we have to find an observable substitute for \( q_t \) and impute \( q_t \) in the estimation. Since I hypothesis that the UST 3m10y slope may be a good proxy
of the latent FRI and I also find that the UST 3m10y slope is highly correlated with output gap and short rate (see Section 1.2), I use the UST 3m10y slope as the substitute. However, we do not yet know whether the UST 3m10y slope is truly a good proxy of \( q_t \). Hence, at the moment, we cannot directly replace \( q_t \) with the UST 3m10y slope in Equation (1.4), the pricing equation for \( q_t \). Rather, we have to construct the UST 3m10y slope as an affine function of \( X_t \) in Equation (1.6) using Equations (1.15) and (1.16). To improve the model fit, I also include Treasury one-, two-, three-, five- and seven-year yields, all of which as affine functions of the state vector, into the measurement equation. This is the estimation strategy adopted by Ang and Piazzesi (2003).

### 1.5.1.1 Measurement equation and likelihood function of Model 1

Let \( N = 6 \) be the number of yield variables including the UST 3m10y slope. Following Chen and Scott (1993), I assume the UST 3m10y slope is measured without error and those Treasury yields are measured with error. Let \( K_2 = 1 \) be the number of yields measured without error and \( K = 4 \) be the number of variables in \( X_t \). We have the following measurement equation:

\[
Y_t = A_y + [B_{y}^{o} B_{y}^{u}] \begin{bmatrix} X_t^{o} \\ X_t^{u} \end{bmatrix} + B_{y}^{m} u^m, \tag{1.22}
\]

where \( Y_t \) contains all the yields and the UST 3m10y slope, \( X_t^{o} \) consists of the observable output gap, inflation and short rate, \( X_t^{u} \) is just the unobservable FRI, \( [B_{y}^{o} B_{y}^{u}] \) includes the ATS loading coefficients for \( Y_t \) (see Equations 1.15 and 1.16), and \( u^m \) comprises the measurement errors for \( Y_t \). I assume the measurement errors are IID and are not cross-correlated. In this setup, suppose the last variable of \( Y_t \) is the UST 3m10y slope, the corresponding component series of \( u_t^m \) will be zero and then \( X_t^{u} \) can be recovered by inverting the last equation of (1.22). This is the first step.
Once the latent FRI is imputed, the second step is to plug the FRI into \( X_t \) and use the transition equation (Equation 1.6) to solve for the structural parameters in Equations (1.1) to (1.4).

The maximum likelihood estimation is an iterative optimization procedure that repeats the above two steps until certain pre-specified convergence criteria (e.g., the value of the likelihood function does not increase) are satisfied. The likelihood function is:

\[
L = -(T - 1) \ln|\text{det}(J)| - \frac{T-1}{2} \ln[\text{det}(\Gamma D \Gamma')] - \frac{1}{2} \sum_{t=2}^{T} (X_t - \Omega X_{t-1})' (\Gamma D \Gamma')^{-1} (X_t - \Omega X_{t-1}) \\
- \frac{1}{2} \frac{T-1}{2} \ln \sum_{i=1}^{N-K_2} \sigma_i^2 - \frac{1}{2} \sum_{t=2}^{T} \sum_{i=1}^{N-K_2} \frac{(u_{it}^m)^2}{\sigma_i^2},
\]

where \( T \) is the number of the observations, \( \sigma_i \) is the standard error of the \( i \)th measurement error series, and \( J \) is a Jacobian matrix such that:

\[
J = \begin{bmatrix}
I & 0 & 0 \\
B_y^o & B_y^u & B_y^m
\end{bmatrix}.
\]

The parameters to be estimated are the structural parameter vector

\([\beta \kappa \alpha \psi \theta \rho \delta_\pi \delta_y \delta_q \phi_r \phi_y \phi_q]'\), the standard errors of the four state variables \([\sigma_\pi \sigma_\gamma \sigma_l \sigma_q]'\) and the market prices of risk matrix \( \lambda_1 \). The initial values of the parameters are arbitrarily chosen, some of which are taken from the parameter estimates in Bekaert et al. (2010).

In the implementation, the optimization algorithm is actually to minimize \(-L\).

### 1.5.1.2 Solving the NK model using the QZ method

In each iteration of the optimization algorithm, Equation (1.5) (without the intercept \( \mu \)) needs to be solved to arrive at Equation (1.6) (without the intercept \( c \)).
Cho and Moreno (2003) use Uhlig’s QZ method (Uhlig 1997), or the generalized Schur decomposition, to solve a three-variable NK model similar to Equations (1.1), (1.2) and (1.3). The solution can be extended to the four-variable NK system in this study.

Applying the undetermined coefficients technique to Equations (1.5) and (1.6) (without \( \mu \) and \( c \)), we have:

\[
\Omega = (B - A\Omega)^{-1}M, \tag{1.24}
\]

\[
\Gamma = (B - A\Omega)^{-1}. \tag{1.25}
\]

Equation (1.24) can be rewritten as:

\[
A\Omega^2 - B\Omega + M = 0. \tag{1.26}
\]

The problem is to solve for \( \Omega \) from Equation (1.26). A recipe of the QZ method works as follows:

1. Define two \( 8 \times 8 \) matrices \( G = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \) and \( H = \begin{bmatrix} B & -M \\ I & 0 \end{bmatrix} \).

2. Find the generalized eigenvalue matrix \( \Lambda \) and eigenvector matrix \( S \) of the matrix pair \( (H, G) \) such that \( HS = GSA \). Then:

\[
\Omega = S_{21}A_{11}S_{21}^{-1}, \tag{1.27}
\]

where \( S_{21} \) and \( A_{11} \) are \( 4 \times 4 \) submatrices of \( S \) and \( \Lambda \).

The determinacy condition for Equation (1.26) is that all diagonal elements of \( A_{11} \) must be less than unity in modulus, whether those elements are real or complex numbers. This is because the number of stable generalized eigenvalues in the \( \Lambda \) matrix must be the same as the number of predetermined variables (four lagged endogenous variables in Equations 1.1 – 1.4). If the number of stable generalized eigenvalues is smaller than four, there is no solution. If on the other hand the number of stable generalized eigenvalues is greater than four, there are multiple solutions.
1.5.1.3 Estimation results of Model 1

The structural parameter estimates and their z-values are shown in Table 1.5 (for the $\lambda_1$ matrix, z-values are in the parentheses). We can see that most of the estimates are statistically significant. The fact that $\delta_\pi > 1$ implies the Taylor principle.

Table 1.5. Structural parameter and market prices of risk estimates of Model 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Z-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.675</td>
<td>3.04</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.125</td>
<td>2.28</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.403</td>
<td>1.98</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.200</td>
<td>0.87</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.214</td>
<td>0.70</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.834</td>
<td>5.88</td>
</tr>
<tr>
<td>$\delta_\pi$</td>
<td>1.031</td>
<td>2.83</td>
</tr>
<tr>
<td>$\delta_y$</td>
<td>0.793</td>
<td>0.42</td>
</tr>
<tr>
<td>$\delta_q$</td>
<td>0.628</td>
<td>0.90</td>
</tr>
<tr>
<td>$\phi_r$</td>
<td>0.181</td>
<td>5.24</td>
</tr>
<tr>
<td>$\phi_y$</td>
<td>0.318</td>
<td>0.91</td>
</tr>
<tr>
<td>$\phi_q$</td>
<td>0.478</td>
<td>2.81</td>
</tr>
<tr>
<td>$\sigma_\pi$</td>
<td>1.394</td>
<td>8.81</td>
</tr>
<tr>
<td>$\sigma_y$</td>
<td>1.399</td>
<td>5.68</td>
</tr>
<tr>
<td>$\sigma_l$</td>
<td>1.293</td>
<td>7.81</td>
</tr>
<tr>
<td>$\sigma_q$</td>
<td>1.288</td>
<td>8.08</td>
</tr>
</tbody>
</table>

One thing that is worth mentioning is that Bekaert et al. (2010) pointed out that previous studies failed to obtain reasonably large and statistically significant estimates of $\kappa$, the sensitivity of inflation with respect to output gap. For example, Galí and Gertler (1999) obtains -0.016 (wrong sign) using detrended log GDP and 0.023 using real marginal cost. Bekaert et al. (2010)
obtain a larger $\kappa$ of 0.064 with a standard error of 0.007. My estimate is 0.125 and also is significant.

The statistically significant $\delta_\pi$ but insignificant $\delta_y$ and $\delta_q$ show that the Fed has responded to inflation more aggressively in magnitude than to output gap and the FRI. The insignificant estimate of $\theta$ appears to suggest that the FRI has a relatively weak effect on output gap.

The statistically significant $\beta$, $\alpha$ and $\phi_q$ confirm that inflation, output gap and the FRI are forward-looking.

1.5.1.4 **Goodness of fit of the yield curve in Model 1**

Since I assume the UST 3m10y slope is measured without error, the fitted and observed UST 3m10y slopes should be the same. However, the fitted yields for tenors other than three-month (since the three-month yield is $i_t$) and 10-year should not be the same as their observed counterparts. Figure 1.3 shows that this is indeed the case. The residuals of the UST 3m10y slope are in the magnitude of +/-2e-15. The standard errors for the residuals of one-, three- and five-year yields are 19, 22 and 16 basis points, respectively. As a comparison, the standard errors for the residuals of one- and 10-year yields in Bakeart et al. (2010) are 45 and 54 basis points.
1.5.1.5 Examining the imputed FRI

Figure 1.4 plots the synthetic FRI (the first principal component of the UST 3m10y slope and AAA 10y spread), imputed FRI $q_t$ and UST 3m10y slope. We can see that the three variables are highly correlated. It suggests that the UST 3m10y slope is indeed a good proxy of the FRI, validating my hypothesis. Thus, I go ahead to estimate Model 2 with the UST 3m10y slope replacing the FRI as a state variable.
1.5.2 Model 2: the NK model treating the UST 3m10y slope as \( q_t \)

1.5.2.1 Likelihood function of Model 2

In Model 2, the measurement equation (Equation 1.22) is not needed. The UST 3m10y slope enters Equation (1.2) - (4) as \( q_t \). Only Equations (1.5) and (1.6) need to be solved to recover the structural parameters. Maximum likelihood method is still used, but the parameter space does not contain the market prices of risk matrix \( \lambda_1 \) and the likelihood function is simplified to:

\[
L = -2T(-1)\ln 2\pi - \frac{T-1}{2} \ln | \Gamma D \Gamma' | - \sum_{t=2}^{T} \left[ -\frac{1}{2} (X_t - \Omega X_{t-1})' (\Gamma D \Gamma')^{-1} (X_t - \Omega X_{t-1}) \right] \tag{1.28}
\]

1.5.2.2 Estimation results of Model 2

Table 1.6 displays a comparison of the structural parameter estimates of Model 1 (with ATS) versus those of Model 2 (without ATS). The parameter estimates of the two models are rather
similar. There are three differences: 1) the estimate of $\kappa$ in Model 2 is smaller than in Model 1 and is statistically insignificant. Thus, it is less desirable than in Model 1; 2) in Model 2, the estimate of $\delta_q$, which measures the sensitivity of short rate with respect to the FRI, is significant and more desirable than in Model 1; and 3) the standard errors of the structural shocks in Model 2 are smaller than in Model 1, suggesting Model 2 could be a better fit to the data.

Table 1.6. Structural parameter estimates of Model 1 and of Model 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>With ATS</th>
<th>Without ATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.675 3.04</td>
<td>0.596 8.09</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.125 2.28</td>
<td>0.083 0.97</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.403 1.98</td>
<td>0.515 5.73</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.200 0.87</td>
<td>0.115 1.26</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.214 0.70</td>
<td>0.132 1.52</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.834 5.88</td>
<td>0.874 10.91</td>
</tr>
<tr>
<td>$\delta_\pi$</td>
<td>1.031 2.83</td>
<td>1.378 15.19</td>
</tr>
<tr>
<td>$\delta_y$</td>
<td>0.793 0.42</td>
<td>0.803 9.05</td>
</tr>
<tr>
<td>$\delta_q$</td>
<td>0.623 0.90</td>
<td>0.915 10.45</td>
</tr>
<tr>
<td>$\phi_r$</td>
<td>0.181 5.24</td>
<td>0.135 1.58</td>
</tr>
<tr>
<td>$\phi_y$</td>
<td>0.313 0.91</td>
<td>0.235 2.41</td>
</tr>
<tr>
<td>$\phi_q$</td>
<td>0.478 2.81</td>
<td>0.514 6.46</td>
</tr>
<tr>
<td>$\sigma_\pi$</td>
<td>1.394 8.81</td>
<td>0.497 5.60</td>
</tr>
<tr>
<td>$\sigma_y$</td>
<td>1.399 5.68</td>
<td>1.117 12.40</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>1.293 7.81</td>
<td>0.488 5.35</td>
</tr>
<tr>
<td>$\sigma_q$</td>
<td>1.288 8.08</td>
<td>1.027 8.44</td>
</tr>
</tbody>
</table>

1.6 Impulse responses of the state variables to structural shocks

Figure 1.5 and 1.6 plot the impulse responses of four state variables with respect to four structural shocks in Model 1 and 2. We can see that the impulse responses in the two models are very similar. In Model 1, the FRI’s negative responses to a demand shock makes sense as a rise
in output typically leads to a drop in financial risks. The FRI’s negative response to an unexpected upward move in policy rate could be interpreted as financial markets consider the interest rate shock an indicator of the Fed’s confidence in the economy. However, the FRI’s negative response to a supply shock, e.g., an increase in oil price, is unintuitive since an inflation shock is supposed to hurt productivity and thus raise financial risks. Likewise, inflation’s upward response to a spike in the FRI does not seem intuitive barring a special situation where the jitter in financial markets is caused by geopolitics, which, say, raises oil prices. Output’s negative response to an unexpected increase in financial risks makes sense, but short rate’s positive response to a rise in financial risks is puzzling.

Model 2 provides more room to interpret the above puzzles in Model 1. A drop in the UST 3m10y slope in case of a supply shock typically is because the inflation shock causes a larger increase in the three-month yield than in the 10-year yield. One example is during March and June 2008, when crude oil rose sharply from around $100 to around $145, the three-month yield surged 160 basis points versus a 90 basis-point increase for the 10-year yield. On the other hand, a positive shock to the UST 3m10y slope could raise economic agents’ inflation expectations and thus actual inflation, especially in the economy’s early expansion. Last but not least, if an upward shock to the UST 3m10y slope is because economic agents expect the economy to strengthen in the medium to long term and thus cause the yield curve to steepen, the correct reaction for the Fed may be to raise short rate. In March 2006, then-Fed Chair Ben Bernanke said the relationship between the yield curve and monetary policy needed to be looked at in different angles: if the yield curve flattens because the term premium declines, “a higher short-term rate is required;” if on the other hand the yield curve flattens because investors expect future economic weakness and thus “mark down their projected path of future spot interest rates,” the Fed’s
reaction should be the opposite (see Bernanke 2006). In this regard, the UST 3m10y slope, though historically highly correlated with the imputed FRI, appears to offer monetary policy makers more perspectives to look at the economy.

Model 2 provides a two-way feedback loop between the economy and the yield curve because the three-moth yield and UST 3m10y slope are part of the NK system and also part of the yield curve. In Model 1, there is one-way feedback from the economy to the yield curve since the latter is an affine function of the former, but there is no feedback from the yield curve back to the economy since \( q_t \) is the unobservable FRI and thus the yield curve does not directly drive the economy.

Figure 1.5. Impulse responses in Model 1
A two-way feedback loop between the economy and the yield curve can help economic agents make more informed decisions. In Model 2, the private sector can not only forecast long-term borrowing rates as in Model 1, but also can send their feedback about the economic performance and monetary policy stance back to monetary policy makers by moving longer-term Treasury yield spreads higher or lower, which is not achievable in Model 1. Furthermore, monetary policy makers can take long-term yield spreads up and down in Model 2 to meet their growth and inflation mandates. In fact, the main goal of the Fed’s three rounds of quantitative easing (QE) programs was to bring down long-term yields because policy rate prior to the QE programs had already hit the zero lower bound (ZLB) constraint after the collapse of Lehman Brothers and thus the conventional Taylor rule was rendered ineffective. With Model 2,
monetary policy makers will be able to directly assess the potential effect of the QE programs on economic activities.

1.7 Calculating term premia

1.7.1 Term premia in Model 1

Using the fitted NK-ATS model (Model 1), I compute the 10-year term premium series as depicted in Figure 1.7. We can see that between 2004 and 2006, the term premium declined meaningfully, which confirms Greenspan’s explanation for his own conundrum. The term premium turned negative in 2010 and has since stayed negative.

Figure 1.7. 10-year term premium, yield and average short rate (%)
Interestingly, the 10-year term premium started turning negative in the third quarter of 2010, when the Bernanke-led Fed announced the second round of QE program (QE2). The term premium touched a trough of -1.8% in the third quarter of 2012, coinciding with the arrival of QE3, which was announced on September 13, 2012. The term premium quickly moved back up by more than 100 basis points to around -0.5% in the third quarter of 2013, after the Fed announced on June 19, 2013 a “tapering” (gradual reduction to an end) of QE3, which was scheduled to start in the federal open market committee (FOMC) meeting in September 2013. Witnessing the sharply tightening financial conditions (widely known as “Taper Tantrum” by financial market participants), the Fed postponed the start of the QE3 tapering to January 2014.

The causal interactions between the 10-year term premium and the Fed’s monetary policy operations between 2009 and 2013 show that, after the 2008 U.S. sub-prime crisis, the yield curve and term premia have become an important monetary transmission channel of the Fed. Meanwhile, this message has become more and more widely and deeply received by financial market participants: the 10-year term premium started sliding down in the second quarter of 2014 and broke the previous low of -1.8% to hit -1.9% in the third quarter of 2016. Note that during these two years, unemployment in the U.S. had kept declining and the Fed raised its target rate range in December 2015. The ever declining 10-year term premium coupled with rising employment and policy rate in the U.S. has become another conundrum. I envision the following two interpretations for this conundrum:

(1) The 10-year average short rate may be incorrectly estimated in this model. Figure 1.7 shows that the 10-year average short rate has been rather flat over the past three decades, especially after the 2008 financial crisis. Since policy rate’s future values are forecasted by a VAR model (see Equation 1.19), the longer the forecast horizon, the closer the
forecast value converges to the three-decade mean ($\Omega^n$ degenerates as $n$ becomes larger). It is likely that the private sector’s expectation of the 10-year average short rate had kept falling between 2014 and 2016 in tandem with the descending 10-year yield. If so, the 10-year term premium might not have declined during these two years. This hypothesis is hard to validate since the private sector’s expectation of the 10-year average short rate is unobservable. However, we can use the FOMC’s estimates to do an exercise. The FOMC started releasing its long-run federal funds rate projection of 3.5% in the meeting in September 2015. The projection decreased to 2.9% in September 2016, a 60-basis point drop in a year. While the Model 1-produced 10-year term premium sank from -1.14% in the third quarter of 2015 to -1.91% in the third quarter of 2016, a 77-basis point drop in a year. Now, the question is whether the one-year change in the private sector’s expectation of the 10-year average short rate was larger than the change in the FOMC’s estimates.

(2) Financial market participants, after receiving the message that the yield curve has become an important tool by the Fed to achieve its growth mandate, has turned more and more greedy in demanding profits in financial markets. This argument seems to be supported by the fact that the 10-year yield dropped around 100 basis points and the S&P 500 index rose around 10% between July 1, 2014 and September 30, 2016, which was in stark contrast to the usual phenomenon that the prices of stocks and of bonds move in opposite directions in an economic expansion. If this argument is true, then it raises a question of whether the 60-basis point drop in the FOMC’s projected long-run policy rate in a year was partly induced by the sharp fall in long-term Treasury yields. In other words, the Fed and financial market participants might have entered a Game of Chicken: the Fed needs sufficiently loose financial conditions to keep U.S. economic growth from dipping given
that structural factors such as aging population and outdated infrastructure (e.g., compared to China) are expected to push U.S. growth gradually lower; financial markets receive this message and ask for even looser financial conditions and higher profits in bond and stock markets; the Fed receive the feedback from financial markets and has to decide by how much to tighten financial conditions when the Fed’s growth and inflation mandates are (or are about to be) satisfied. In the first half of 2017, the Fed raised its target rate range twice and set a plan to shrink its balance sheet later in 2017 despite relatively soft activity data and muted inflation readings during most of the first two quarters of 2017. This could be the Fed’s reaction to financial markets.

I believe the above two interpretations both have grounds and thus the 10-year term premium should have dropped between 2014 and 2016 but the drop may not have been as much as 77 basis points. The VAR model may be inadequate in forecasting the 10-year average short rate. Structural factors as exogenous explanatory variables may need to be added into the VAR model. This calls for a more fully-fledged forecasting model of the 10-year average short rate. However, this task is outside the scope of this parsimonious NK model. The second interpretation is largely related to the field of political economics in that if the Fed is the chicken in the game, the ongoing income and wealth inequality in the U.S. can only become worse and worse.

The term premium series up to 2006 in Figure 1.7 looks very similar to those by the Bernanke-Reinhart-Sack and VAR methods as discussed in Rudebusch et al. (2007) (see Figure 1.8, which I copied from the Rudebusch et al. paper).
In Model 1, since the entire yield curve is modeled as an affine function of the state vector $X_t$, term premia are seamlessly linked to the economy. In Model 2, however, there appears to be no theoretical link between the economy and the yields other than the three-month tenor (treated as $i_t$ in the NK system) and the 10-year tenor (since the UST 3m10y slope is treated as $q_t$, the 10-year yield is directly linked to the economy). Furthermore, we cannot resort to the ATS theory since Model 2 does not estimate the market prices of risk matrix $\lambda_t$. In the following section, I propose a parsimonious way to build a link between term premia and the economy in Model 2.

1.7.2 Treating $i_t$ and UST 3m10y slope as the level and slope factors of the entire yield curve

It has been widely known that most of the variation of the entire yield curve can be explained by three factors – level, slope and curvature. In the financial industry, practitioners usually use
principal component analysis (PCA) to construct these three factors (see, e.g., Soto 2004). Rudebusch and Wu (2008) find that the level and slope factors are sufficient to account for variation in the yield curve. The authors also give macroeconomic interpretations to the two factors, linking the level factor to inflation expectations and the slope factor to the cyclical stance of monetary policy.

I apply PCA to the Treasury yield curve historical data and confirm with Rudebusch and Wu (2008) that the first two principal components account for 99.95% of the total variation.

The level and slope factors generated by PCA are linear combinations of all the yields in the yield curve. Since only the three-month yield and the 10-year yield are incorporated in the NK system, we cannot use the PCA-produced level and slope factors to calculate term premia. However, I hypothesize that the three-month yield is a good proxy for the level factor and the UST 3m10y slope is a good proxy for the slope factor. A visual inspection, as shown in Figure 1.9, appears to verify that this is indeed the case. In fact, the correlation between the level factor and three-month yield is 0.96 and the correlation between the slope factor and UST 3m10y slope is 0.89.

To examine the explanatory power of the three-month yield and UST 3m10y slope on the yield curve, I run OLS regression of all the yields but the three-month and 10-year tenors on the three-month yield and UST 3m10y slope. As can be seen in Table 1.7, all the coefficients with respect to the three-month yield are close to 1, indicating the three-month yield’s role of yield curve level driver. Furthermore, coefficient of the UST 3m10y slope increases from nearly zero to nearly 1 as tenor increases, confirming the UST 3m10y slope’s role of yield curve slope driver.
Figure 1.9. Level factor vs. 3-month yield and slope factor vs. UST 3m10y slope

Table 1.7. OLS regression results of yields on 3-month yield and 10-year slope

<table>
<thead>
<tr>
<th>Yield tenor</th>
<th>$i_t$</th>
<th>$q_t$</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-month</td>
<td>1.029***</td>
<td>0.049***</td>
<td>0.998</td>
</tr>
<tr>
<td>1-year</td>
<td>1.064***</td>
<td>0.172***</td>
<td>0.994</td>
</tr>
<tr>
<td>2-year</td>
<td>1.103***</td>
<td>0.395***</td>
<td>0.991</td>
</tr>
<tr>
<td>3-year</td>
<td>1.108***</td>
<td>0.565***</td>
<td>0.991</td>
</tr>
<tr>
<td>4-year</td>
<td>1.099***</td>
<td>0.688***</td>
<td>0.993</td>
</tr>
<tr>
<td>5-year</td>
<td>1.083***</td>
<td>0.781***</td>
<td>0.995</td>
</tr>
<tr>
<td>6-year</td>
<td>1.066***</td>
<td>0.851***</td>
<td>0.997</td>
</tr>
<tr>
<td>7-year</td>
<td>1.048***</td>
<td>0.904***</td>
<td>0.998</td>
</tr>
<tr>
<td>8-year</td>
<td>1.031***</td>
<td>0.945***</td>
<td>0.999</td>
</tr>
<tr>
<td>9-year</td>
<td>1.014***</td>
<td>0.976***</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Note: * means significant at 5%, ** at 1% and *** at 0.1%

Next, I use the fitted yield curve (raw three-month and 10-year yields and OLS regression-fitted yields for other tenors) to calculate term premia. Since ATS theory is not used in Model 2, we cannot use Equation (1.18) to compute forward yields $f_t^n$. Rather, $f_t^n$ is now obtained as:
\[ f_t^n = y_t^{n+1}(n + 1) - y_t^n, \]  

(1.29)

where \( y_t^n \) and \( y_t^{n+1} \) come from the fitted yield curve. We can use Equation (1.29) together with Equations (1.17), (1.19) and (1.21) to calculate term premia.

Figure 1.10 exhibits the 10-year term premia produced by Model 1 and 2. We can see that they share almost the same pattern, though the levels are somewhat different.

Figure 1.10. 10-year term premia in Model 1 and 2 (%)

### 1.8 Conclusions and future research directions

In this paper I propose a macro-finance yield curve term premium model. My proposed model is unique in the literature in that it establishes a two-way feedback loop between the real economy and the yield curve under the NK framework in a micro-founded way. In doing so I
include a latent financial variable, which I term the Financial Risk Index, into the NK system. Furthermore, I model this latent variable as having a real balance effect on output gap and monetary policy.

Based on Affine Term Structure theory, I link the Treasury yield curve and the real economy by constructing the yield curve as an affine function of the NK system. I fit the NK-ATS model (Model 1) to macroeconomic variables and the yield curve using maximum likelihood method and obtains satisfactory goodness of fit. The term premia calculated in Model 1 match the empirical literature.

Given that the imputed Financial Risk Index in the NK-ATS model is highly correlated with the Treasury three-month vs. 10-year slope, I replace the Financial Risk Index with the Treasury 10-year slope as a state variable of the NK system and re-estimate a pure NK model (Model 2). To calculate term premia, I treat the three-month yield and the 10-year slope as the level and slope factors of the entire yield curve. The term premia calculated in Model 1 and 2 share almost the same historical pattern. In Model 2, since the three-month yield and the 10-year slope are part of the NK system and also part of the yield curve, a two-way feedback loop between the real economy and the yield curve is established. This setup also provides better interpretations of the impulse responses than Model 1.

I envisage a number of future research directions. First, a more fully-fledged forecasting model that can produce more plausible projections of policy rate in the long horizon and thus more plausible term premia can be explored. Second, the relationship between term premia and the real economy, whether in a structural or reduced form, can be investigated. Third, an analysis of the stability issues of the NK system can be conducted. Fourth, my proposed Model 1 and 2 can be extended to a small open economy.
Chapter 2

Correcting Estimation Biases in the New Keynesian Term Premium Model with Financial Risks

2.1 Introduction

2.1.1 Motivations and contributions

This paper contributes to the macro-finance literature by proposing an estimation method to correct estimation biases in the class of New Keynesian (NK) term premium models, and specifically, in a model presented by Fu (2017). These biases are caused by a full information maximum likelihood (FIML) estimation method typically employed by macro-finance term premium models. As pointed out by Bauer et al. (2012), an unbiased FIML estimator is difficult to find in practice because 1) the likelihood function has many local optima, 2) there is a high statistical uncertainty around the point estimates, and 3) the computationally intensive optimization procedure makes correcting a time series model’s small-sample bias infeasible.

Over the past 10 to 15 years, macro-finance term premium models have gained the attention of researchers and policy makers. In this literature, two classes of models are particularly popular: 1) vector autoregression-affine term structure (VAR-ATS) models and 2) NK-ATS models. Both classes of models treat the U.S. Treasury yield curve as an affine function of a system of state (macroeconomic and/or latent) variables, but the first approach models the state variables as a VAR system while the second approach uses an NK system to represent the state variables. The first approach’s seminal paper is Ang and Piazzesi (2003), while the second approach’s seminal papers are, among others, Hördahl et al. (2006), Rudebusch and Wu (2008) and Bekaert et al. (2010), with these papers offering different NK specifications. The main
drawback of these two classes of models is that there is no feedback from the yield curve to the economy in a micro-founded way (though there is feedback from the economy to the yield curve) since the yield curve is a function of the state variables but not the other way around.

Based on the NK-ATS framework, Fu (2017) provides a two-way feedback loop between the economy and the yield curve in a micro-founded way. Fu (2017) presents two models. Model 1 is an NK-ATS model that brings in a latent financial variable termed the Financial Risk Index (FRI) to form a four-equation (inflation, output gap, the Treasury three-month yield as a proxy of the policy rate, and the FRI) NK system. Model 2 uses the Treasury three-month vs. 10-year yield spread (henceforth the UST 3m10y slope) as a proxy for the FRI and thus drops the ATS part. Since the UST 3m10y slope is part of the yield curve and also part of the NK system, Model 2 provides a two-way bridge between the economy and the yield curve.

These two classes of models including Model 1 and 2 of Fu (2017) typically use an FIML method to estimate the ATS parameters and the VAR/NK parameters in a single step. This one-step estimation method is vulnerable to estimation biases, which tend to render expected future short rates less persistent than they should be and thus render the movements of estimated term premia too similar to those of underlying Treasury yields, as discovered by Bauer et al. (2012). Though some seminal papers of NK term premium models use some techniques to improve the robustness of the estimation (e.g., Hördahl et al. 2006 set some of the ATS parameters to zero to reduce the parameter space, Bekaert et al. 2010 use the Newey-West estimator to handle potentially heteroskedastic and/or autocorrelated residuals, etc.), these techniques do not change the fact that the one-step FIML method relies on an optimization procedure that could reach a local optimum or even a saddle point, which could generate a VAR coefficient matrix far away from the true value.
To correct estimation biases, Bauer et al. (2012) use a two-step estimation method to estimate a VAR-ATS model. In the first step, the paper estimates the reduced-form VAR process using ordinary least squares (OLS). Since the OLS-estimated VAR coefficient matrix is biased downward because the lagged state variables are included in the OLS procedure and thus strict exogeneity is violated, Bauer et al. (2012) adjust the coefficient matrix higher. The bias-corrected OLS VAR coefficient matrix thus produces unbiased expected future short rates and term premia. In the second step, Bauer et al. (2012) estimate the ATS parameters. The main advantage of this two-step method is that OLS can produce a stable estimated VAR coefficient matrix while the VAR coefficient matrix estimated by the one-step FIML method could vary for different initial values of the parameters and convergence criteria. Furthermore, there seems to be no way to correct the estimation bias to the coefficient matrix in the single-step FIML context.

While Bauer et al. (2012) address estimation biases in the class of VAR-ATS models, to my knowledge, no study to date has attempted to correct estimation biases in the class of NK-ATS term premium models. Such a task, in my understanding, has one more layer of complexity than that of Bauer et al. (2012) in that correcting estimation biases in an NK-ATS model requires recovering the NK structural parameters in addition to the ATS parameters from an already OLS-estimated reduced-form VAR model (Bauer et al. only recover the ATS parameters from an OLS-estimated VAR). This is the situation for Model 1 of Fu (2017). For Model 2 of Fu (2017), only the NK parameters need to be backed out from an OLS-estimated VAR. In a related vein, Keating (1990) proposes a two-step estimation method to identify rational-expectations (RE) models (not specifically NK models) within the structural VAR (SVAR) framework. Drawing on both Bauer et al. (2012) and Keating (1990), I develop a two-step estimation strategy to correct estimation biases seen in Model 2 of Fu (2017).
In the first step, I fit a VAR(1) model by OLS to the dataset used by Fu (2017) and obtain the coefficient matrix, which is shown to have larger diagonal elements than those estimated by the one-step FIML method in Model 2 of Fu (2017). This means that the state variables are less persistent and will revert to their means faster in Fu (2017) than in this study and explains the flatness of the 10-year average expected short rate over the past 30 years in Fu (2017). Note that the 10-year average expected short rate is the average of the short rates forecasted throughout a 10-year horizon by the estimated VAR(1) model. Therefore, the smaller the diagonal elements of the VAR(1) coefficient matrix, the closer the 10-year average expected short rate to the 30-year mean of the three-month yield. I use the OLS coefficient matrix to calculate the 10-year average expected short rate and the 10-year term premium. The 10-year average expected short rate series so calculated is shown to be much more cyclical (following economic fluctuations and monetary policy dynamics) and structural (on a downward trend due to structural factors such as demographics and thus descending natural rate of interest) than in Model 2 of Fu (2017).

Likewise, the OLS-estimated 10-year term premium is countercyclical while the Fu (2017) 10-year term premium has just moved in lockstep with the 10-year yield. Next, I correct the small-sample estimation bias in the coefficient matrix using an analytical approximation method (as opposed to a simulation method employed by Bauer et al. 2012) discussed in Engsted and Pedersen (2014). The bias-corrected OLS coefficient matrix preserves the cyclical and structural features of the 10-year average expected short rate and the countercyclical feature of the 10-year term premium, but slightly enlarges the variation ranges of the two variables.

In the second step, I use the bias-corrected OLS VAR(1) coefficient matrix and residuals to recover the NK parameters. I follow Keating (1990). The Keating method transforms a RE model’s RE terms into SVAR forms and also transforms structural disturbances into
representations in terms of reduced-form disturbances and finally maximizes the likelihoods of the transformed structural disturbances by an FIML procedure to recover RE structural parameters. Leu (2011) applies the same approach to an open economy NK model. Both Keating (1990) and Leu (2011) transform structural disturbance terms equation by equation. I simplify their approach by implementing the transformation in a matrix form. I obtain the NK parameters and find that none of the estimated parameter is statistically significant, that the estimated parameters of the IS equation are highly different from those in Fu (2017), and that the estimated NK parameters cannot recover the reduced-form VAR(1) coefficient matrix. I suspect that the parameter space may be too flat and the IS equation may be mis-specified so that the FIML optimization procedure cannot reach a satisfactory optimum.

To improve the SVAR-FIML method’s estimation results, I examine and identify the NK system equation by equation. My identification strategy is to transform each equation into a form that represents the dependent state variable as a linear combination of four lagged state variables. I then match the four loadings of the linear combination with the four coefficients of the corresponding equation of the reduced-form VAR(1) system. I minimize the residuals between the four loadings and four coefficients to recover the structural parameters of each NK equation.

I start from the Taylor rule since it can be exactly identified while other three equations are over-identified. I find that the exactly identified Taylor rule fails to satisfy the Taylor principle. I then convert the Taylor rule into a linear regression model and fits the model to the data. The fitted regression model shows the Taylor principle is held. I keep the regression-estimated coefficients for inflation and lagged short rate in the Taylor rule and apply the identification strategy again (this time it is over-identified) to recover the other two parameters for output gap and the FRI using a grid search method (assigning a presumed range to each structural
parameter). This calibration-aided identification strategy preserves the Taylor principle, though it also results in somewhat larger residuals than the exactly identification strategy.

I identify the Phillips curve and the FRI equation also by minimizing the residuals between each equation’s lagged variable linear combination’s loadings and the VAR(1) coefficients using a grid search method. In both cases the minimized residuals are satisfactorily small. For the Philipps curve, the parameter that governs the expected inflation and lagged inflation is shown to be dominant while for the FRI equation, the parameter that governs expected FRI and lagged FRI is dominant.

I apply the same identification strategy to the IS curve and find that the minimized residuals are unpleasantly large. I then revise the specification of the IS curve by adding a structural restriction. The revised specification yields reasonably small identification residuals. The parameter that governs the lagged output gap is shown to be dominant.

The equation-by-equation identification strategy outperforms the SVAR strategy and the estimation strategy in Model 2 of Fu (2017) because the former strategy recovers a reduced-form VAR(1) coefficient matrix much closer to the OLS-estimated coefficient matrix than the latter two strategies.

The contributions of this paper can be summarized as follows. First, the paper proposes a two-step estimation method to correct estimation biases in the class of NK term premium models, and specifically, in Model 2 of Fu (2017). The new estimation method produces a more cyclical and structural (hence fitting the reality better) 10-year average expected short rate and a more counter-cyclical 10-year term premium than Fu (2017). Second, by identifying the NK system equation by equation, the paper uncovers that the IS equation in Fu (2017) needs a structural restriction to better fit the data. Without a two-step estimation method, this finding would not
have been achievable. Third, the paper restores consistency between the structural NK system and the reduced-form VAR system as the OLS-estimated VAR coefficient matrix generates sensible NK parameters, which in turn recover the VAR coefficient matrix satisfactorily.

2.1.2 Organization of this paper

The rest of this paper is organized as follows. Section 2.2 summarizes Model 1 and Model 2 of Fu (2017) and their estimation procedures. It then discusses the estimation bias to the reduced-form VAR(1) coefficient matrix generated by the FIML method and demonstrates that the OLS method produces more realistic expected short rates and term premia than the FIML method. Moreover, it points out the small-sample bias generated by the OLS method. Section 2.3 presents a bias-correction two-step estimation strategy for Model 2 of Fu (2017). It uses Engsted and Pederson (2014)’s analytical approximation approach to address the OLS small-sample bias. It then details how to estimate the NK parameters using Keating’s SVAR method and how to simplify Keating’s method. Lastly, it shows the estimated NK parameters cannot recover the VAR(1) coefficient matrix and that the IS equation may be mis-specified. In Section 2.4, I explore the equation-by-equation identification strategy to recover the NK parameters. In particular, I present a new specification for the IS curve. Finally, I show that such a strategy outperforms the SVAR method and the single-step FIML method used by Model 2 of Fu (2012). Section 2.5 concludes and points to future research directions.

2.2 Model 2 of Fu (2017) and its estimation biases

2.2.1 The NK-ATS model of Fu (2017)

The Fu (2017) NK-ATS model starts from a modified three-equation NK system:
\[\pi_t = \beta E_t \pi_{t+1} + (1 - \beta) \pi_{t-1} + \kappa y_t + \epsilon_{\pi,t}, \quad (2.1)\]

\[y_t = \alpha E_t y_{t+1} + \psi y_{t-1} - \alpha (i_t - E_t \pi_{t+1}) - \theta q_t + \epsilon_{y,t}, \quad (2.2)\]

\[i_t = \mu_i + \rho i_{t-1} + (1 - \rho) (\delta_{\pi} \pi_t + \delta_{y} y_t - \delta_{q} q_t) + \epsilon_{i,t}. \quad (2.3)\]

Equation (2.1) is the Phillips curve, (2.2) the IS curve and (2.3) the Taylor rule. As usual, \(\pi_t\), \(y_t\) and \(i_t\) denote inflation, output gap and short rate, respectively. Note that this version of “modified” NK model is different than the “standard” NK model in that it adds lagged variables \(\pi_{t-1}\), \(y_{t-1}\) and \(i_{t-1}\) to capture the empirical evidence of adaptive inflation expectations, consumption habit formation and policy rate smoothing. Another difference is that in the IS equation, the expected output gap at time \(t+1\) and the current real interest rate share the same coefficient.

Fu (2017) adds to the NK system a latent variable – the FRI, which is denoted by \(q_t\) as shown in Equations (2.2) and (2.3). \(q_t\) is assumed to follow the process below:

\[q_t = \phi_q E_t q_{t+1} - \phi_y E_t y_{t+1} + \phi_r (i_t - E_t \pi_{t+1}) + (1 - \phi_q) q_{t-1} + \epsilon_{q,t}. \quad (2.4)\]

Note that \(q_t\) is not on the right-hand side of the Phillips curve.

Equations (2.1)-(2.4) are derived in Appendix 1.A of Fu (2017) in a micro-founded way. The FRI in Fu (2017) is similar to the aggregate stock market index in Nisticò (2012), which uses an overlapping generation NK model to study stock prices' wealth effect on the real economy and monetary policy.

Equations (2.1) – (2.4) can be written in a compact form:

\[BX_t = \mu + AE_t X_{t+1} + MX_{t-1} + \epsilon_t. \quad (2.5)\]

The four shocks in \(\epsilon_t\) are uncorrelated with one another, and \(\epsilon_t \sim N(0, D)\) where \(D\) is the diagonal variance matrix. This is a rational-expectation structural system. Using the undetermined coefficients method, a VAR(1) solution can be guessed as:
\[ X_t = c + \Omega X_{t-1} + \Gamma \epsilon_t, \]  

(2.6)

where \( \Gamma \) is a variance-covariance matrix of \( \epsilon_t \) with non-zero covariance (and correlation) coefficients.

Equation (2.6) is typically solved from Equation (2.5) by QZ method (Uhlig 1997).

Fu (2017) uses ATS finance theory (e.g., Duffie and Kan 1996, Dai and Singleton 2002) to connect the NK system and the yield curve. ATS models begin with a short rate that is an affine function of a state vector process \( X_t \):

\[ i_t = a_0 + a_1' X_t. \]  

(2.7)

\( X_t \) is as defined in Equation (2.6). Since \( i_t \) is the third state variable in \( X_t \), \( a_0 = 0 \) and \( a_1 = [0 \ 0 \ 1 \ 0]' \).

Next step is to make a zero-coupon Treasury bond (and yield) an affine function of \( X_t \). To price a Treasury bond, we need to transform the physical probability measure \( P \) on which the historical data of \( X_t \) are based to the risk-neutral probability measure \( Q \). The Radon-Nikodym derivative \( L_t \) serves this purpose: \( E^Q[W_{t+1}] = E^P[L_{t+1} W_{t+1}] / L_t \), where \( W_t \) is a random variable and \( L_t \) follows:

\[ L_{t+1} = L_t \exp(-\frac{1}{2} \lambda_t' \lambda_t - \lambda_t' \epsilon_{t+1}). \]  

(2.8)

\( \lambda_t \) measures market prices of risk and also is an affine function of \( X_t \):  

\[ \lambda_t = \lambda_0 + \lambda_1 X_t. \]  

(2.9)

The stochastic discount factor \( m_t \) is therefore:

\[ m_{t+1} = \exp(-i_t) \frac{L_{t+1}}{L_t} = \exp(-i_t - \frac{1}{2} \lambda_t' \lambda_t - \lambda_t' \epsilon_{t+1}). \]  

(2.10)

The stochastic discount factor is used to discount the bond price at time \( t+1 \) to time \( t \) as follows:
\[ p_t^{n+1} = E_t[m_{t+1} p_t^n]. \]  
(2.11)

where \( p_t^n \) is the price of an \( n \)-period zero-coupon bond at time \( t \).

Assume the price of \( p_t^n \) is an exponential affine function of \( X_t \):

\[ p_t^n = \exp(A_n + B'_n X_t). \]  
(2.12)

Then an \( n \)-period zero-coupon bond yield \( y_t^n \) is obtained as:

\[ y_t^n = -\frac{A_n}{n} - \frac{B'_n}{n} X_t. \]  
(2.13)

\( A_n \) and \( B_n \) can be solved iteratively. Their initial values are \( A_1 = -a_0 \) and \( B_1 = -a_1 \).

As such, the entire yield curve is an affine function of \( X_t \) with the parameters \( c, \Omega \) and \( \Gamma \) in Equation (2.6) and \( \lambda_0 \) and \( \lambda_1 \) in Equation (2.9). These parameters need to be estimated from data.

An \( n \)-period yield spread \( s_t^n \) is then the difference between the \( n \)-period yield and the three-month yield (also short rate in Fu 2017):

\[ s_t^n = y_t^n - i_t. \]  
(2.14)

To extract the term premium from an \( n \)-period yield, we need to first obtain the forward risk premium embedded in a one-period forward yield. The one-period forward premium, denoted as \( r_t^n \), is the difference between the one-period forward yield \( n \) periods ahead, \( f_t^n \), and the expected one-period short rate \( n \) periods ahead, \( E_t[i_{t+n}] \).

\[ r_t^n = f_t^n - E_t[i_{t+n}] = (A_n - A_{n+1}) + (B'_n - B'_{n+1} - a'_1 \Omega^n) X_t, \]  
(2.15)

where

\[ E_t[i_{t+n}] = E_t[a_0 + a'_1 X_{t+n}] = a'_1 \Omega^n X_t. \]  
(2.16)

The \( n \)-period term premium is the average of the forward premia up to time \( t+n-1 \):

\[ t_t^n = \frac{1}{n} \sum_{i=0}^{n-1} r_t^i. \]  
(2.17)
2.2.2 FIML estimation of the NK-ATS model

To estimate the NK-ATS model, Fu (2017) uses a dataset that contains quarterly Personal Consumption Expenditure (PCE) inflation, Congressional Budget Office (CBO) output gap, and Treasury three-month, six-month, one-year through 10-year zero-coupon bond yields. The sample period is from the first quarter of 1985 to the fourth quarter of 2016. The UST 3m10y slope is calculated as the difference between the 10-year yield and the three-month yield. This slope and other yields are affine functions of $X_t$. PCE inflation, CBO output gap, the three-month yield (as $i_t$), the UST 3m10y slope and some key yields (one-, two-, three-, five- and seven-year yields) are demeaned and then used in the estimation. The estimation problem is a state space model containing a measurement equation of yield-related variables and a transition equation of the four state variables.

Let $N = 6$ be the number of yield variables including the UST 3m10y slope. Following Chen and Scott (1993), Fu (2017) assumes the UST 3m10y slope is measured without error and those Treasury yields are measured with error. Let $K_2 = 1$ be the number of yields measured without error and $K = 4$ be the number of variables in $X_t$. The measurement equation is defined as follows:

$$Y_t = A_y + [B_y^o \ B_y^u] \begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix} + B_y^m u^m \tag{2.18}$$

where $Y_t$ contains all the yields and the UST 3m10y slope, $X_t^o$ consists of the observable output gap, inflation and short rate, $X_t^u$ is just the unobservable FRI, $[B_y^o \ B_y^u]$ includes the ATS loading coefficients for $Y_t$, and $u^m$ comprises the measurement errors for $Y_t$. Fu (2017) assumes the measurement errors are IID and are not cross-correlated. In this setup, suppose the last variable
of \( Y_t \) is the UST 3m10y slope, the corresponding component series of \( u_t^m \) will be zero and then \( X_t^u \) can be recovered by inverting the last equation of (2.18).

The transition equation is Equation (2.6). And the likelihood function is:

\[
L = -(T - 1) \ln|\det(J)| - \frac{T-1}{2} \ln|\det(\Gamma D')| - \frac{1}{2} \sum_{t=2}^T (X_t - \Omega X_{t-1})'(\Gamma D')^{-1} (X_t - \Omega X_{t-1}) - \frac{T-1}{2} \sum_{i=1}^{N-K_2} \frac{(u_i^m)^2}{\sigma_i^2},
\]

(2.19)

where \( T \) is the number of the observations, \( \sigma_i \) is the standard error of the \( i \)th measurement error series, and \( J \) is a Jacobian matrix such that:

\[
J = \begin{bmatrix}
I & 0 & 0 \\
B_y^0 & B_y^u & B_y^m
\end{bmatrix}.
\]

The FIML estimation method is an iterative optimization procedure that optimizes the parameter vector \([\beta \kappa \alpha \psi \theta \rho \delta_n \delta_y \delta_q \phi_r \phi_y \phi_q]'\), the standard errors of the four state variables \([\sigma_n \sigma_y \sigma_l \sigma_q]'\) and the market prices of risk matrix \( \lambda_1 \). The optimization procedure starts with a set of initial values of the parameters, constructs Equations (2.1)-(2.4) (equivalently Equation 2.5), and then uses QZ method to solve Equation (2.6) from Equation (2.5) for \( \Omega \) and \( \Gamma \). After that, Equation (2.18) is used to impute the FRI. The optimization procedure repeats these tasks with improved parameter estimates until certain pre-specified convergence criteria (e.g., the value of the likelihood function does not increase) are satisfied.

2.2.3 Model 2 and the estimation

Model 2 of Fu (2017) assumes the UST 3m10y is a good proxy of the FRI and thus the UST 3m10y slope directly enters Equations (2.2)-(2.4) as \( q_t \). The measurement equation (Equation 2.18) is not needed. Only Equations (2.5) and (2.6) need to be solved to recover the structural
parameters. FIML method is still used, but the parameter space does not contain the market prices of risk matrix \( \lambda_1 \) and the likelihood function is simplified to:

\[
L = -2(T - 1)\ln2\pi - \frac{T-1}{2} \ln| \Gamma' \Gamma | - \sum_{t=2}^{T} \left[ -\frac{1}{2} (X_t - \Omega X_{t-1})' (\Gamma' \Gamma)^{-1} (X_t - \Omega X_{t-1}) \right].
\] (20)

Fu (2017) finds that the imputed FRI in Model 1 resembles the UST 3m10y slope and confirms that the UST 3m10y slope is a good proxy of the FRI.

To calculate term premia in Model 2, Fu (2017) treats \( i_t \) and the UST 3m10y slope as the level and slope factors of the entire yield curve and then regresses the entire yield curve on these two factors to obtain the fitted yield curve (raw three-month and 10-year yields as well as OLS-fitted yields for other tenors). Forward yields \( f_t^n \) are now obtained as:

\[
f_t^n = y_t^{n+1}(n + 1) - y_t^nn,
\] (2.21)

where \( y_t^n \) and \( y_t^{n+1} \) come from the fitted yield curve. Equation (2.21) together with Equations (2.15)-(2.17) are used to calculate term premia.

Since \( i_t \) and the UST 3m10y slope are part of the NK system and also the two key drivers of the entire yield curve, Model 2 builds a two-way bridge between the economy and the yield curve.

### 2.2.4 FIML estimation bias to \( \Omega \)

In the FIML procedures in Model 1 and 2, parameters \( \Omega, \Gamma \) and \( \lambda_1 \) (\( \lambda_1 \) is not needed in Model 2) are estimated in a single step. Since it is an iterative procedure, the solution could settle at a local optimum or saddle point, which may be far away from the true value.

As shown in Equations (2.15)-(2.17), an \( n \)-period term premium is the sum of a series of differences between OLS-fitted forward yields and expected short rates in Model 2. Also, the expected short rate \( n \) periods ahead \( E_t[i_{t+n}] = \alpha_1' \Omega^n X_t = [0 \ 0 \ 1 \ 0] \Omega^n X_t \). In Model 2, therefore,
the $n$-period term premium is only determined by the estimated VAR(1) coefficient matrix $\Omega$. If the estimated $\Omega$ is biased (particularly the third row), estimated term premia also will be biased.

Below let us compare the OLS estimate of $\Omega$ with the FIML estimate. Modifying Equation (2.6) (changing the disturbance term and dropping the intercept), we can estimate $\Omega$ by OLS from the following equation:

$$X_t = \Omega X_{t-1} + e_t,$$  \hspace{1cm} (2.22)

where the reduced-form residual term $e_t$ is connected with the structural residual term $\epsilon_t$ as follows:

$$e_t = \Gamma \epsilon_t.$$  \hspace{1cm} (2.23)

Recall that the way in which the FIML procedure estimates $\Omega$ is to use a set of estimated structural parameters to construct Equation (2.5) and then solve for $\Omega$ from Equations (2.5) and (2.6). Therefore, a good estimate of $\Omega$ depends on a set of good structural parameter estimates, and more importantly, on a good specification of the structural NK system. The OLS method, on the other hand, estimates $\Omega$ directly from Equation (2.22) and does not involve the NK system.

The OLS and FIML estimates of $\Omega$ are as shown below, where the OLS estimate is denoted by $\hat{\Omega}_{OLS}$ and the FIML estimate by $\hat{\Omega}_{FIML}$.

$$\hat{\Omega}_{OLS} = \begin{bmatrix}
0.7670 & 0.0043 & 0.0614 & 0.0499 \\
-0.0795 & 0.9696 & 0.0192 & 0.0496 \\
-0.0057 & 0.1153 & 0.9666 & 0.1188 \\
-0.0207 & -0.1253 & 0.0087 & 0.8076
\end{bmatrix}$$

$$\hat{\Omega}_{FIML} = \begin{bmatrix}
0.8190 & 0.0262 & 0.1262 & 0.0814 \\
0.4469 & 0.1358 & -0.2928 & -0.0248 \\
0.1403 & 0.0148 & 0.8458 & 0.0752 \\
-0.3365 & -0.0219 & -0.2594 & 0.5910
\end{bmatrix}$$
We can see that most of the diagonal elements of $\hat{\Omega}_{OLS}$ are larger than those of $\hat{\Omega}_{FIML}$ (except for the first element), indicating that output gap, short rate and the UST 3m10y slope are more persistent (or less mean-reverting) based on the OLS estimate than based on the FIML estimate. In particular, the OLS estimate of the mean-reverting coefficient of the short rate is 0.967, much higher than 0.846 from the FIML estimate. This implies that the FIML-estimated expected short rates could adjust much faster to the long-term mean and could look much more stable than the OLS-estimated ones. As a result, the FIML-estimated term premia could unrealistically resemble the yield curve.

Figure 2.1 reveals that the above hypotheses are indeed true. The upper panel shows that the 10-year average short rate estimated in Model 2 of Fu (2017) is very stable around the short rate’s 30-year mean. The 10-year average short rate estimated by OLS, by contrast, is very volatile and exhibits a downward trend, implying that it has been driven by cyclical factors (economic fluctuations and monetary policy dynamics) and also by structural factors (demographic changes and decreasing productivity have brought down natural rate of interest). The lower panel shows that the 10-year term premium estimated in Model 2 of Fu (2017) has moved in lockstep with the 10-year yield as a result of the unrealistically stable estimated 10-year average short rate. The 10-year term premium estimated by OLS, by contrast, is countercyclical – it moves up in recessions and down in economic booms/low-volatility periods. In sum, the OLS-estimated 10-year average short rate and term premium are more realistic and intuitive.
From $\tilde{\Omega}_{\text{FIML}}$, we also can see that the mean-reverting coefficient of 0.136 for output gap looks too small, suggesting that Equation (2.2) may potentially be mis-specified.

### 2.2.5 Potential OLS estimation bias to $\Omega$

We have discovered that the optimization-based FIML estimation procedure can produce a highly biased estimate of $\Omega$ and thus unrealistic term premia. We also have found that an OLS estimation of the reduced-form VAR(1) process (Equation 2.22) performs better in this regard. Yet, the $\Omega$ coefficient matrix estimated by OLS also is biased in small sample as can be seen below:
\[
E[\hat{\Omega}_{\text{OLS}}|X_{t-1}] = \Omega + E[e_tX'_{t-1}(X_{t-1}X'_{t-1})^{-1}|X_{t-1}] .
\] (2.24)

The second term of the right-hand side of Equation (2.24) is not zero, i.e., it violates the strict exogeneity assumption of Gauss-Markov Theorem.

In the following sections, I will propose a method that addresses the estimation biases caused by the FIML and OLS procedures in the context of Model 2 of Fu (2017).

### 2.3 A bias-correction estimation method for Model 2 of Fu (2017)

My method is in essence similar to Bauer et al. (2012) in that my method also contains two steps.

In the first step, I estimate the VAR(1) process Equation (2.22) by OLS and correct the small-sample bias. In the second step, I then estimate the structural parameters in Equations (2.1)-(2.4) by FIML.

There are two differences between my method and Bauer et al. (2012): 1) in Step 1, while Bauer et al. (2012) use simulation (termed “inverse bootstrap” in their paper) to correct the small-sample bias, I use an analytical approach; and 2) in Step 2, Bauer et al. (2012) estimate the ATS parameters while my method estimates the structural parameters of the NK system.

#### 2.3.1 Correcting the small-sample bias in the OLS estimation of VAR(1)

I use an analytical procedure developed in Engsted and Pedersen (2014) to correct the small-sample bias in \(\hat{\Omega}_{\text{OLS}}\). For Equation (2.22), Engsted and Pedersen (2014) obtain the bias matrix \(B_{\text{OLS}}\) as:

\[
B_{\text{OLS}} = V_e\{\Omega'[I_N - \Omega^2]^{-1} + \sum_{i=1}^N \lambda_i \ (I_N - \lambda_i\Omega')^{-1}\}V'_X ,
\] (2.25)
where $I_N$ is an identity matrix, $\lambda_i$ is the $i$th eigenvalue of $\Omega$, and $V_e$ and $V_X$ are the covariance matrices of $e_t$ and $X_t$, respectively.

Therefore, the bias-adjusted OLS coefficient matrix $\hat{\beta}_{OLS}^c$ is:

$$\hat{\beta}_{OLS}^c = \hat{\beta}_{OLS} - B_{OLS}$$

Equation (2.25) is derived under the assumption that $X_t$ must be stationary, but $\hat{\beta}_{OLS}^c$ obtained from Equations (2.25) and (2.26) may make $X_t$ non-stationary. Kilian (1998) proposes a way to achieve stationarity: 1) calculate $\hat{\beta}_{OLS}^c$ using Equations (2.25) and (2.26); 2) if $\hat{\beta}_{OLS}^c$ does not have unit or explosive roots, then it is a good estimate; and 3) Otherwise, keep multiplying $B_{OLS}$ by a scaling factor $h \in \{0.99, 0.98, \ldots, 0\}$ and then subtracting the scaled $B_{OLS}$ from $\hat{\beta}_{OLS}$ to obtain $\hat{\beta}_{OLS}^c$ until $\hat{\beta}_{OLS}^c$ has no unit or explosive roots.

Note that Equation (2.25) requires knowledge of true values of $\Omega$ and $V_e$, which are unknown. Since the OLS estimator is a consistent estimator of $\Omega$, Engsted and Pedersen (2012) plug $\hat{\beta}_{OLS}$ into the right-hand side of Equation (2.25) to obtain $B_{OLS}$. Hence, it is called the “plug-in” approach. I also use this approach.

Once $\hat{\beta}_{OLS}^c$ is obtained, the residual series also needs to be re-estimated by plugging $\hat{\beta}_{OLS}^c$ into Equation (2.22). Let us denote it as $\hat{\epsilon}_{OLS,t}^c$.

The bias-corrected coefficient matrix $\hat{\beta}_{OLS}^c$ is as shown below. We can see that the diagonal elements are all larger than those of $\hat{\beta}_{OLS}$, making $X_t$ more persistent or less mean-reverting.

$$\hat{\beta}_{OLS}^c = \begin{bmatrix}
0.8116 & 0.0010 & 0.0506 & 0.0398 \\
-0.0725 & 0.9926 & 0.0188 & 0.0440 \\
-0.0024 & 0.1087 & 0.9727 & 0.1063 \\
-0.0210 & -0.1124 & 0.0108 & 0.8391
\end{bmatrix}$$
Figure 2.2 shows that the 10-year average short rate calculated by using $\hat{\Omega}_{OLS}^c$ is slightly less mean-reverting than the one calculated by using $\hat{\Omega}_{ols}$. This is mainly because the diagonal elements in $\hat{\Omega}_{OLS}^c$ are only slightly larger than those in $\hat{\Omega}_{OLS}$. As a result, the 10-year term premia estimated by using $\hat{\Omega}_{OLS}^c$ and $\hat{\Omega}_{ols}$ also are only slightly different.

2.3.2 Estimating the NK parameters

After obtaining $\hat{\Omega}_{OLS}^c$, we can use Equations (2.5) and (2.6) to estimate the NK parameters in Equations (2.1) – (2.4). In this section, I follow the estimation strategy used by Keating (1990), which transforms a system of RE equations into a SVAR system and then estimates the structural parameters from the SVAR system. Leu (2011) employs the same approach to estimate a four-
equation open economy NK model with the first three equations similar to the NK system in Clarida et al. (1999) and the last equation being the uncovered interest parity. Furthermore, I simplify part of Keating’s method in a matrix form.

2.3.2.1 The Keating (1990) SVAR estimation method

Let us use an example in Leu (2011) to illustrate Keating’s SVAR estimation method. The method first transforms each equation of the NK system into a structural disturbance representation in terms of reduced-form disturbances. Take the NK Phillips curve as an example:

\[ \pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \epsilon_{\pi,t}. \]

The transformation is done by subtracting from each variable the expectation at time \( t-1 \) of the same variable:

\[ \epsilon_{\pi,t} = \pi_t - E_{t-1} \pi_t - \beta (E_t \pi_{t+1} - E_{t-1} \pi_{t+1}) + \kappa (y_t - E_{t-1} y_t) \]
\[ = e_{\pi,t} - \beta (E_t \pi_{t+1} - E_{t-1} \pi_{t+1}) + \kappa e_{y,t}, \tag{2.27} \]

where \( e_{\pi,t} \) and \( e_{y,t} \) are the reduced-form disturbance terms for inflation and output gap.

Now, we need to represent \((E_t \pi_{t+1} - E_{t-1} \pi_{t+1})\) in terms of the reduced-form disturbances. We achieve this task by using the following equation:

\[ E_t \pi_{t+1} = d_1 \Omega X_t, \tag{2.28} \]

where \( d_i \) is a row vector of length 4 (Leu’s NK model has 4 equations) with the \( i \)th element being 1 and the other elements being 0. Therefore:

\[ E_t \pi_{t+1} - E_{t-1} \pi_{t+1} = d_1 \Omega (X_t - E_{t-1} X_t) = d_1 \Omega e_t. \tag{2.29} \]

Equation (2.27) can be rewritten as:

\[ \epsilon_{\pi,t} = e_{\pi,t} - \beta d_1 \Omega e_t + \kappa e_{y,t}. \tag{2.30} \]
Leu (2011) uses the procedure described above to convert $\epsilon_{yt}$ and $\epsilon_{it}$ into representations in terms of the reduced-form disturbances. Combining these structural disturbances, we have:

$$\epsilon_t = He_t.$$  \hspace{1cm} (2.31)

We can see that $H$ is a function of $\Omega$ and the structural parameters. Since $\Omega$ and $e_t$ have been estimated in the first step, the second step is to estimate the structural parameters by maximizing the log-likelihood function of $\epsilon_t$.

Let $D$ be the covariance matrix of $\epsilon_t$, the log-likelihood function is:

$$L = -2(T - 1)\ln 2\pi - \frac{T-1}{2} \ln |H^{-1}D(H^{-1})'| - \frac{1}{2} \sum_{t=2}^T \epsilon_t'D^{-1}\epsilon_t.$$ \hspace{1cm} (2.32)

### 2.3.2.2 Simplifying Keating’s disturbance transformation procedure

In a matrix form, the aforementioned procedure of transforming each structural disturbance into a representation of reduced-form disturbances used in Keating (1990) and Leu (2011) can actually be simplified and thus calculations can be reduced. Since $E_tX_{t+1} = \Omega X_t$, Equation (2.5) (dropping $\mu$) can be rewritten as a SVAR form:

$$(B - A\Omega)X_t = MX_{t-1} + \epsilon_t.$$ \hspace{1cm} (2.33)

Equation (2.33) can be transformed into a reduced-form VAR(1) form:

$$X_t = (B - A\Omega)^{-1}MX_{t-1} + (B - A\Omega)^{-1}\epsilon_t.$$ \hspace{1cm} (2.34)

Equating the terms of Equation (2.34) and those of Equation (2.6) (dropping $e$), we have:

$$\Gamma = (B - A\Omega)^{-1},$$ \hspace{1cm} (2.35)

$$\Omega = (B - A\Omega)^{-1}M,$$ \hspace{1cm} (2.36)

$$\epsilon_t = \Gamma^{-1}e_t.$$ \hspace{1cm} (2.37)

The representation of the structural disturbances in terms of the reduced-form disturbances as shown in Equation (2.37) is more straightforward to obtain than it is by Keating’s method.
Let $V$ be the covariance matrix of $e_t$. Since $V$ has been estimated in the first step of the estimation, the covariance matrix of $\varepsilon_t$, $D$, can be obtained as:

$$D = \Gamma V \Gamma^{-1}. \quad (2.38)$$

And the log-likelihood function becomes:

$$L = -2(T - 1) \ln 2\pi - \frac{T-1}{2} \ln |\Gamma \Gamma'| - \frac{1}{2} \sum_{t=2}^{T} \varepsilon_t' D^{-1} \varepsilon_t. \quad (2.39)$$

The estimation of the structural parameters is done in an FIML optimization procedure. Note that there are mainly two differences between this FIML procedure and the FIML procedure used in Model 2 of Fu (2017): 1) in this procedure, $\Omega$ and $e_t$ are already estimated by OLS and thus are inputs ($\hat{\Omega}_{OLS}$ and $\hat{e}_{OLS,t}$), while in Fu (2017) they are outputs; and 2) this procedure maximizes the log-likelihood function of $\varepsilon_t$ while Fu (2017) maximizes the log-likelihood function of $e_t$.

### 2.3.2.3 Estimation results

Table 2.1 displays the structural NK parameters estimated in Model 2 of Fu (2017) compared to those estimated by the SVAR strategy. There are three noticeable differences. First, none of the parameters estimated by SVAR is significant while most of the parameters estimated in Fu (2017) are significant, implying very flat parameter space and thus it is hard to reach an optimum in the SVAR estimation. Second, the estimates of $\alpha$ (in the IS equation), $\delta_\pi$ and $\delta_q$ (in the Taylor rule) in the SVAR estimation are very different from their counterparts in Fu (2017), suggesting a second look at the specifications of the IS curve and Taylor rule may be needed. Third, the number of the parameters estimated in the SVAR strategy is 12 compared to 16 in Fu (2017).

The reason is that, unlike in Fu (2017), in which the variances of the four structural disturbances
have to be estimated, in the SVAR strategy these variances can be directly computed from Equation (2.38).

Table 2.1. Structural NK parameters estimated by Model 2 of Fu (2017) and by SVAR

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 2 of Fu (2017)</th>
<th>SVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Z-value</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.596</td>
<td>8.09</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.083</td>
<td>0.97</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.515</td>
<td>5.73</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.115</td>
<td>1.26</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.132</td>
<td>1.52</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.874</td>
<td>10.91</td>
</tr>
<tr>
<td>( \delta_\pi )</td>
<td>1.378</td>
<td>15.19</td>
</tr>
<tr>
<td>( \delta_y )</td>
<td>0.803</td>
<td>9.05</td>
</tr>
<tr>
<td>( \delta_q )</td>
<td>0.916</td>
<td>10.45</td>
</tr>
<tr>
<td>( \phi_r )</td>
<td>0.136</td>
<td>1.58</td>
</tr>
<tr>
<td>( \phi_y )</td>
<td>0.235</td>
<td>2.41</td>
</tr>
<tr>
<td>( \phi_q )</td>
<td>0.514</td>
<td>6.46</td>
</tr>
</tbody>
</table>

To further examine the structural parameters estimated by the SVAR strategy, I use them to back out the VAR(1) coefficient matrix and compare this matrix with the bias-corrected OLS coefficient matrix. First, construct matrices B, A and M as in Equation (2.36). Next, solve Equation (2.36) for \( \Omega \) using QZ method (see a recipe of QZ method in Fu 2017). The solved \( \Omega \), denoted as \( \hat{\Omega}_{SVAR} \), is shown below:

\[
\hat{\Omega}_{SVAR} = \begin{bmatrix}
0.5242 & 0.0153 & 0.4643 & 0.6042 \\
0.0455 & 0.1812 & -0.0470 & 0.0697 \\
0.0546 & 0.0322 & 0.9341 & 0.5877 \\
-0.0585 & -0.0040 & 0.0668 & 0.5658
\end{bmatrix}
\]
We can see that $\hat{\Omega}_{SVAR}$ is rather different from $\hat{\Omega}_{OLS}$. The first two and last elements of the diagonal of $\hat{\Omega}_{SVAR}$ are much smaller than those of $\hat{\Omega}_{OLS}$. In particular, the second element of 0.181 is almost as small as 0.136 in $\hat{\Omega}_{FIML}$, suggesting a potential misspecification of the IS equation.

So far I have presented a bias-correction estimation method for Model 2 of Fu (2017). It can preserve the reduced-form mean-reverting coefficients of state variables and thus can produce more realistic term premia than the single-step FIML estimation method. However, this method runs into difficulties in estimating the structural NK parameters, which cannot recover the reduced-form mean-reverting coefficients. In the coming section, I will examine and identify the NK system equation by equation based on grid search method, economic theories and empirics.

### 2.4 Examining and identifying the NK system

Before starting the identification project, let us rewrite Equation (2.22) as follows:

\[
\begin{bmatrix}
\pi_t \\
y_t \\
i_t \\
q_t
\end{bmatrix} =
\begin{bmatrix}
\omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\
\omega_{21} & \omega_{22} & \omega_{23} & \omega_{24} \\
\omega_{31} & \omega_{32} & \omega_{33} & \omega_{34} \\
\omega_{41} & \omega_{42} & \omega_{43} & \omega_{44}
\end{bmatrix}
\begin{bmatrix}
\pi_{t-1} \\
y_{t-1} \\
i_{t-1} \\
q_{t-1}
\end{bmatrix} +
\begin{bmatrix}
\epsilon_{\pi,t} \\
\epsilon_{y,t} \\
\epsilon_{i,t} \\
\epsilon_{q,t}
\end{bmatrix}
\]  
(2.40)

In the above reduced-form system, each equation has four coefficients that have been estimated already. In the NK system (Equations 2.1-2.4), only the Taylor rule has four parameters, while the Phillips curve has two parameters, the IS curve and the FRI equation have three parameters. Hence, the Taylor rule can be exactly identified while other three NK equations are over-identified. It is natural to start the identification project with the Taylor rule.
2.4.1 Examining and identifying the Taylor rule

To identify Equation (2.3) (dropping $\mu_t$), let us write the right-hand side of the equation in terms of $X_{t-1}$:

$$i_t = \xi_3 d_1' \pi_{t-1} + \xi_3 d_2' y_{t-1} + (\rho + \xi_3 d_3') i_{t-1} + \xi_3 d_4' q_{t-1} + U_3 e_t + \epsilon_{i,t}. \quad (2.41)$$

where $d_i$ is a row vector of length 4 with the $i$th element being 1 and the other elements being 0. $U_3$ is a row vector whose representation is omitted here because it is not needed in identifying the Taylor rule. $\xi_3$ is defined as follows:

$$\xi_3 = (1 - \rho)(\delta_{\pi} d_1 + \delta_y d_2 - \delta_q d_4) \Omega. \quad (2.42)$$

To identify the Taylor rule, therefore, we need to solve the following system of equations:

$$
\begin{bmatrix}
\xi_3 d_1' \\
\xi_3 d_2' \\
\rho + \xi_3 d_3' \\
\xi_3 d_4'
\end{bmatrix}
= 
\begin{bmatrix}
\omega_{31} \\
\omega_{32} \\
\omega_{33} \\
\omega_{34}
\end{bmatrix}. \quad (2.43)
$$

In Equation (2.43), there are four equations and four unknowns, thus it is an exact identification problem, which can be solved by a direct inversion of Equation (2.43). The solution is $[\hat{\rho} \; \hat{\delta}_{\pi} \; \hat{\delta}_y \; \hat{\delta}_q]' = [0.969 \; 0.356 \; 3.911 \; 3.803]'$. The Euclidean norm of the residuals of the four equations is 2.31e-09. Note that $\hat{\delta}_{\pi}$ (0.356) is much smaller than 1, indicating that Taylor principle is not held in this case. Many previous studies have documented that this coefficient is greater than 1 empirically. For example, Bekaert et al. (2010) obtain 1.525 utilizing a specification including lagged short rate similar to Equation (2.3) (the main differences are that expected inflation rather than actual inflation is included and the FRI is excluded on the right-hand side of the Taylor rule in their paper) on a dataset from 1961 to 2003. Note that Equation (2.3) (demeaned) actually can be converted into a linear regression model that can be used to
check whether Taylor principle is held in this dataset:

\[ i_t = \rho_1 i_{t-1} + \rho_2 \pi_t + \rho_3 y_t + \rho_4 q_t + \epsilon_{i,t}. \]  

(2.44)

I run this regression and obtain \( \hat{\rho}_1 = 0.918 \) and \( \hat{\rho}_2 = 0.094 \). Since \( \hat{\rho}_2 = (1 - \hat{\rho}_1)\hat{\delta}_\pi \), \( \hat{\delta}_\pi = 1.142 \). In other words, Taylor principle is held. Note that Equation (2.44) also suffers from a small-sample estimation bias because \( i_{t-1} \) is included on the right-hand side. I use bootstrapping (by resampling \( \hat{\epsilon}_{i,t} \) with replacement) on Equation (2.44) to arrive at bias-corrected \( \hat{\rho}_1 = \hat{\rho} = 0.921 \) and \( \hat{\delta}_\pi = 1.168 \). Again, Taylor principle is held. These two calibrated parameters seem to fit empirics well, so I keep them fixed in Equation (2.43) to solve for the other two free parameters \( \hat{\delta}_y \) and \( \hat{\delta}_q \). The solution is \( [\hat{\rho} \ \hat{\delta}_\pi \ \hat{\delta}_y \ \hat{\delta}_q]' = [0.921 \ 1.168 \ 1.626 \ 1.507]' \). The norm of the residuals is 0.079, which shows that the VAR(1) coefficients for \( i_t \) to be backed out from this solution would be somewhat different from those in \( \hat{\Omega}_{OLS}^c \), but the differences should be much smaller than the differences between the coefficients in \( \hat{\Omega}_{SVAR}^c \) and those in \( \hat{\Omega}_{OLS}^c \).

Figure 2.3 shows that the norm of the residuals for the solution of Equation (2.43) is a convex function of \( \hat{\delta}_y \) and \( \hat{\delta}_q \) when \( \rho \) and \( \hat{\delta}_\pi \) are fixed at 0.921 and 1.168, respectively, and confirms that the minimum function value lies at a point where \( \hat{\delta}_y = 1.626 \) and \( \hat{\delta}_q = 1.507 \).
2.4.2 Examining and identifying the Phillips curve

The Phillips curve (Equation 2.1) can be rewritten as follows:

\[ \pi_t = (\xi_1 d_1' + 1 - \beta) \pi_{t-1} + \xi_1 d_2' y_{t-1} + \xi_1 d_3' i_{t-1} + \xi_1 d_4' q_{t-1} + U_1 e_t + \epsilon_{\pi,t}, \] (2.45)

where \( \xi_1 = (\beta d_1 \Omega + \kappa d_2)\Omega \). \( U_1 \) is a row vector whose representation is omitted here.

The Phillips curve can be identified by solving the following system of equations:

\[ \begin{bmatrix} \xi_1 d_1' + 1 - \beta \\ \xi_1 d_2' \\ \xi_1 d_3' \\ \xi_1 d_4' \end{bmatrix} = \begin{bmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{13} \\ \omega_{14} \end{bmatrix} \] (2.46)

Equation (2.46) is over-identified because only two parameters – \( \beta \) and \( \kappa \) – need to be solved for. I employ a grid search approach to find the optimum solution. Bekaert et al. (2010) use a Phillips curve specification almost the same as Equation (2.1) and obtain 0.611 and 0.064 for \( \beta \)
and κ with both estimates statistically significant. Bekaert et al. (2010) also point out that previous studies failed to obtain reasonably large (much smaller than their estimate of 0.064) and statistically significant estimates of κ. Based on these studies, I start from relatively large parameter ranges of [0.1, 0.9] for β and [0.001, 0.3] for κ and then gradually narrow down the ranges to [0.52, 0.58] and [0.001, 0.05]. It is discovered that the norm of the residuals for the solution of Equation (2.46) is a monotonically increasing function of κ when β is fixed and a convex function of β when κ is small (the convexity flattens out as κ increases (see Figure 2.4).

The minimum norm of the solution residuals is 0.002, which corresponds to \[\hat{\beta}, \hat{\kappa} = \begin{bmatrix} 0.551 \\ 0.001 \end{bmatrix} \]. Note that \(\hat{\kappa} = 0.001\) means that inflation is almost insensitive to output gap (though expected inflation contains information about output gap). However, since this is a finding shared by previous studies, I decide to let the data speak for themselves and keep these two parameter estimates.

Figure 2.4. The norm of solution residuals of Equation (2.46) as a function of β and κ
2.4.3 Examining and identifying the FRI equation

The FRI equation (Equation 2.4) can be rewritten as follows:

\[ q_t = \xi_4 d'_1 \pi_{t-1} + \xi_4 d'_2 y_{t-1} + \xi_4 d'_3 l_{t-1} + (\xi_4 d'_4 + 1 - \phi_q) q_{t-1} + U_4 e_t + \epsilon_{q,t}, \]  

(2.47)

where \( \xi_4 = \left( (\phi_q d_4 - \phi_y d_2 - \phi_r d_1) \Omega + \phi_r d_3 \right) \Omega. \) \( U_4 \) is a row vector whose representation is omitted here.

The FRI equation can be identified by solving the following system of equations:

\[
\begin{bmatrix}
\xi_4 d'_1 \\
\xi_4 d'_2 \\
\xi_4 d'_3 \\
\xi_4 d'_4 + 1 - \phi_q
\end{bmatrix} =
\begin{bmatrix}
\omega_{41} \\
\omega_{42} \\
\omega_{43} \\
\omega_{44}
\end{bmatrix}.
\]  

(2.48)

Equation (2.48) is over-identified because only three parameters \( [\phi_q \phi_y \phi_r] \) need to be solved for. Since the estimates of \( [\phi_q \phi_y \phi_r] \) by Model 2 of Fu (2017) and by the SVAR method have similar quantities (\([0.514 0.235 0.136]\) by Fu (2017) and \([0.584 0.121 0.096]\) by SVAR), I start the grid search from parameter ranges for \( [\phi_q \phi_y \phi_r] \) that comfortably encompasses those estimates: \([0.3, 0.7], [0.01, 0.4]\) and \([0.01, 0.4]\). Figure 2.5 shows the relationship between the norm of the four residuals in Equation (2.48) and \( \phi_q \). Note that there are wild oscillations for a specific value of \( \phi_q \) and different values of \( \phi_y \) and \( \phi_r \). The minimum norm of 0.008 corresponds to a parameter point of \( [\hat{\phi}_q \hat{\phi}_y \hat{\phi}_r] = [0.525 \ 0.01 \ 0.01] \).
Figure 2.5. The relationship between the norm of residuals and $\phi_q$

By fixing $\phi_q$ at 0.525, we can see in Figure 2.6 that the norm of residuals is monotonically increasing functions of $\phi_r$ and $\phi_y$. This means that in Equation (2.4), the FRI (the UST 3m10y slope) is rather insensitive to real interest rate and expected output gap.

We can examine the above implication by running a regression converted from Equation (2.4):

$$q_t = \phi_1 q_{t+1} + \phi_2 y_{t+1} + \phi_3 (i_t - \pi_{t+1}) + \phi_4 q_{t-1} + \epsilon_{q,t}.$$  \hfill (2.49)
In other words, by removing the expectation operators from Equation (2.4), I assume a variable’s expected value at \( t+1 \) is just its actual value at \( t+1 \). This is a loose assumption, but just for the purpose of examination rather than validation.

I obtain the coefficient vector \([\hat{\phi}_1 \hat{\phi}_2 \hat{\phi}_3 \hat{\phi}_4]\) = \([0.532 0.019 -0.004 0.517]\) with \( \hat{\phi}_1 \) and \( \hat{\phi}_4 \) statistically significant but \( \hat{\phi}_2 \) and \( \hat{\phi}_3 \) insignificant. Note that \( \hat{\phi}_1 \) is roughly the same as \( \hat{\phi}_q \) and, like \( \hat{\phi}_y \) and \( \hat{\phi}_r \), \( \hat{\phi}_2 \) and \( \hat{\phi}_3 \) also are close to zero.

Though the statement that the UST 3m10y slope is insensitive to real interest rate and expected output gap appears true, it is only true when the expected UST 3m10y slope is included as an explanatory variable – the expected UST 3m10y slope contains information about current real interest rate and current output gap. If we remove \( \phi_1 q_{t+1} \) from Equation (2.49), we obtain \([\hat{\phi}_2 \hat{\phi}_3 \hat{\phi}_4]\) = \([-0.117 -0.022 0.810]\) with only \( \hat{\phi}_3 \) statistically insignificant, i.e., \( \hat{\phi}_2 \) is now meaningfully different from zero.
If we remove $\phi_1 q_{t+1}$ from Equation (2.49) and also change $y_{t+1}$ to $y_t$ and $\pi_{t+1}$ to $\pi_t$, the coefficient vector becomes: $[\hat{\phi}_2 \ \hat{\phi}_3 \ \hat{\phi}_4] = [-0.117 \ -0.023 \ 0.796]$, which is almost the same as the coefficient vector estimated when $y_{t+1}$ and $\pi_{t+1}$ are retained in Equation (2.49). Again, only $\hat{\phi}_3$ is statistically insignificant.

These exercises show that the UST 3m10y slope is not so sensitive to real interest rate, but it is sensitive to output gap. It is just that the sensitivity to output gap is absorbed into the expected UST 3m10y slope in Equation (2.4). As such, it appears that we can drop the two terms of $-\phi_y E_t y_{t+1}$ and $\phi_r (i_t - E_t \pi_{t+1})$ from Equation (2.4). However, since Equation (2.4) is derived from micro-foundations and since only the dataset used by Fu (2017) shows that $\hat{\phi}_y$ and $\hat{\phi}_r$ are nearly zero, I decide to keep those two terms to avoid the Lucas critique.

### 2.4.4 Examining and identifying the IS equation

The IS equation (Equation 2.2) can be rewritten as follows:

$$y_t = \xi_2 d_1' \pi_{t-1} + (\xi_2 d_2' + \psi) y_{t-1} + \xi_2 d_3' i_{t-1} + \xi_2 d_4' q_{t-1} + U_2 e_t + \epsilon_{y,t},$$

(2.50)

where $\xi_2 = [\alpha((d_1 + d_2)\Omega - d_3) - \theta d_4] \Omega$. $U_2$ is a row vector whose representation is omitted here.

The IS equation can be identified by solving the following system of equations:

$$\begin{bmatrix} \xi_2 d_1' \\ \xi_2 d_2' + \psi \\ \xi_2 d_3' \\ \xi_2 d_4' \end{bmatrix} = \begin{bmatrix} \omega_{21} \\ \omega_{22} \\ \omega_{23} \\ \omega_{24} \end{bmatrix}.$$  

(2.51)

Again, Equation (2.51) is over-identified because only three parameters $[\alpha \ \psi \ \theta]$ need to be solved for. I still employ grid search for the optimum solution. Since the estimates of $[\alpha \ \psi \ \theta]$ by Model 2 of Fu (2017) and by the SVAR method have rather different quantities $([0.515 \ 0.115$
0.132] by Fu (2017) and [0.190 0.177 0.010] by SVAR), I start from large parameter ranges of
[0.05, 0.8], [0.001, 0.4] and [0.001, 0.4].

The results are not ideal. Figure 2.7 depicts the relationship between the norm of the four
residuals in Equation (2.51) and $\alpha$. The minimum norm of 0.488 at the parameter point of
$[\hat{\alpha} \hat{\psi} \hat{\theta}] = [0.270 0.396 0.011]$ is very large. In particular, the second residual of -0.359
$(\xi_2 d_2' + \psi - \omega_{22})$ at the minimum norm shows that output gap’s mean-reverting coefficient
estimated by minimizing the residuals of Equation (2.51) is 0.359 smaller than the corresponding
coefficient of 0.993 in $\hat{\theta}_{OLS}$. This again suggests that the IS equation (Equation 2.2) may be mis-
specified.

Figure 2.7. The relationship between the norm of residuals and $\alpha
2.4.4.1 A new specification of the IS equation

In Fu (2017), the IS equation is derived by incorporating consumption habit formation into the IS equation of Nisticò (2012). The IS equation of Nisticò (2012) looks as follows:

\[ y_t = \alpha E_t y_{t+1} - \alpha (i_t - E_t \pi_{t+1}) + \xi s_t + \epsilon_{y,t}. \] (2.52)

where \( s_t \) is the aggregate stock market index. Fu (2017) treats \( s_t \) as a representative risky asset and the FRI \( q_t = -\rho s_t \) as the measure of the riskiness of \( s_t \), which means when the risky asset price falls, the FRI rises proportionally. Consequently, Equations (2.52) becomes:

\[ y_t = \alpha E_t y_{t+1} - \alpha (i_t - E_t \pi_{t+1}) - \theta q_t + \epsilon_{y,t}. \] (2.53)

By incorporating consumption habit formation into the consumer utility function, Fu (2017) adds \( \psi y_{t-1} \) into the right-hand side of Equation (2.53) to arrive at Equation (2.2) in this paper.

I suspect that the large residuals of Equation (2.51) may come from the fact that there are no restrictions on \( \alpha \) and \( \psi \) in Equation (2.2). Bekaert et al. (2010), for example, require \( \alpha + \psi = 1 \).

Their IS equation is also derived by incorporating consumption habit formation and looks like:

\[ y_t = \alpha E_t y_{t+1} + (1 - \alpha) y_{t-1} - \theta (i_t - E_t \pi_{t+1}) + \epsilon_{y,t}. \]

To apply such a restriction to Equation (2.2), I rewrite \( y_t \) as a weighted average of \( y_{t-1} \) and the right-hand side of Equation (2.53):

\[ y_t = \psi y_{t-1} + (1 - \psi)\{\alpha [E_t y_{t+1} - (i_t - E_t \pi_{t+1})] - \theta q_t\} + \epsilon_{y,t}. \] (2.54)

2.4.4.2 Identifying the new IS equation

Based on this specification, \( \xi_2 \) in Equation (2.50) now becomes:

\[ \xi_2 = (1 - \psi)\{\alpha ((d_1 + d_2) \Omega - d_3) - \theta d_4\} \Omega. \]

The IS equation is still identified by solving Equation (2.51).
Equation (2.54) now resembles the specification of the Taylor rule (Equation 2.3). Recall that in the initial estimation of Equation (2.3) by a direct inversion of Equation (2.43), $\hat{\beta} = 0.969$, which is similar to the mean-reverting coefficient of 0.973 for short rate in $\hat{\Omega}_{OLS}^c$. Therefore, I suspect $\hat{\psi}$ also should be similar to the mean-reverting coefficient of 0.993 for output gap in $\hat{\Omega}_{OLS}^c$. I then start the grid search algorithm from parameter ranges of [0.7, 0.999], [0.05, 1.5] and [0.001, 0.4] for $\psi$, $\alpha$ and $\theta$. The estimation results reveal much smaller residuals than the old specification of the IS equation. Figure 2.8 shows that the norm of residuals falls as $\psi$ rises, but the norm appears to flatten out when $\psi$ approaches the upper limit of 0.999. Indeed, the minimum norm of 0.087 is reached at a parameter point of $[\hat{\psi} \hat{\alpha} \hat{\theta}] = [0.995 \ 0.05 \ 0.001]$, i.e., $\hat{\psi}$ is slightly smaller than its upper limit, $\hat{\alpha}$ and $\hat{\theta}$ are at their lower limits. Note that $\hat{\psi} = 0.995$ is very close to the mean-reverting coefficient of 0.993 for output gap in $\hat{\Omega}_{OLS}^c$.

Figure 2.8. The relationship between the norm of residuals and $\psi
Figure 2.9. The norm of residuals as increasing functions of $\alpha$ and $\theta$ when $\psi = 0.995$

Recall that in the estimation of the FRI equation, only $\hat{\phi}_q$ is large while the other two parameters $\hat{\phi}_r$ and $\hat{\phi}_y$ are very small. We have the same situation in the estimation of the new IS equation – $\hat{\alpha}$ and $\hat{\theta}$ are very small. Unsurprisingly, Figure 2.9 shows that the norm of residuals is increasing functions of $\alpha$ and $\theta$ when $\psi = 0.995$. However, the surface of the norm is very flat, especially along the edge of $\theta$ – when $\alpha = 0.05$ and $\theta$ increases from 0.001 to 0.4, the norm only increases from 0.087 to 0.088. In fact, even if both $\alpha$ and $\theta$ increase from their lower limits to their upper limits, the norm only increases from 0.087 to 0.093. This flat norm space is different from the FRI equation whose norm space is very steep with respect to $\phi_r$ and $\phi_y$ (see Figure 2.6). This creates a problem for a Newton-type optimization algorithm because the algorithm could settle at any values of $\alpha$ and $\theta$ when $\psi$ approaches 1.
2.4.5 Comparing the NK parameters estimated by Fu (2017), SVAR and grid search

Table 2.2 exhibits the NK parameter estimates produced by the FIML strategy in Model 2 of Fu (2017), the SVAR strategy and the equation-by-equation identification strategy using grid search method combined with calibration. We can see that the estimates of $\beta$ and $\phi_q$ and $\rho$ by the three methods are similar ($\rho$ by the third method is the largest), implying that these three parameters, which are the dominant drivers of the Phillips curve, FRI equation and Taylor rule, are rather model- and estimation method-invariant. Note that the meanings of $\alpha$, $\psi$ and $\theta$ in the third method are different from the first two methods and thus these three parameters’ estimates by the three methods are incomparable.

Table 2.2. Structural NK parameters estimated by Model 2 of Fu (2017), SVAR and grid search (plus calibration)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FIML Estimate</th>
<th>SVAR Estimate</th>
<th>Grid Search Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.596</td>
<td>0.648</td>
<td>0.551</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.083</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.515</td>
<td>0.190</td>
<td>0.050</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.115</td>
<td>0.177</td>
<td>0.995</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.132</td>
<td>0.010</td>
<td>0.001</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.874</td>
<td>0.804</td>
<td>0.921</td>
</tr>
<tr>
<td>$\delta_\pi$</td>
<td>1.378</td>
<td>0.919</td>
<td>1.168</td>
</tr>
<tr>
<td>$\delta_y$</td>
<td>0.803</td>
<td>0.922</td>
<td>1.626</td>
</tr>
<tr>
<td>$\phi_q$</td>
<td>0.916</td>
<td>4.200</td>
<td>1.507</td>
</tr>
<tr>
<td>$\phi_r$</td>
<td>0.136</td>
<td>0.096</td>
<td>0.010</td>
</tr>
<tr>
<td>$\phi_y$</td>
<td>0.235</td>
<td>0.121</td>
<td>0.010</td>
</tr>
<tr>
<td>$\phi_\tau$</td>
<td>0.514</td>
<td>0.584</td>
<td>0.525</td>
</tr>
</tbody>
</table>
2.4.6 Backing out the reduced-form VAR(1) coefficient matrix

With the NK parameters estimated by the grid search method, I now proceed to back out the reduced-form VAR(1) coefficient matrix and compare this matrix with the one estimated by OLS. Again, I populate matrices B, A and M in Equation (2.36) and use QZ method to solve the equation for $\Omega$. The solved $\Omega$, denoted as $\hat{\Omega}_{GS}$, is shown below:

$$\hat{\Omega}_{GS} = \begin{bmatrix} 0.8438 & -0.0854 & 0.1751 & 0.1697 \\ 0.0002 & 0.9952 & -0.0002 & 0.0000 \\ 0.0722 & 0.1070 & 0.9409 & 0.1169 \\ -0.0478 & -0.1089 & 0.0311 & 0.8505 \end{bmatrix}$$

We can see that the diagonal vector of $\hat{\Omega}_{GS}$ is now similar to the diagonal vector $[0.812 \ 0.993 \ 0.973 \ 0.839]'$ of $\hat{\Omega}_{OLS}$, suggesting a large improvement over the SVAR method (recall that the diagonal vector of $\hat{\Omega}_{SVAR}$ is $[0.524 \ 0.181 \ 0.934 \ 0.566]'$). Thus, the grid search method preserves state variables’ persistency.

One thing worth mentioning is that in the second row of $\hat{\Omega}_{GS}$, because the mean-reverting coefficient (0.995) is very large, the other three elements are very small (the second row of $\hat{\Omega}_{OLS}$ is $[-0.073 \ 0.993 \ 0.019 \ 0.044]'$), implying that the $t-1$ values of inflation, short rate and the UST 3m10y slope almost have no effect on output gap. Given the highly flat residual norm space with respect to $\alpha$ and $\theta$ when $\psi$ is large, if we raise the quantities of $\alpha$ and $\theta$, the three small elements of the second row of $\hat{\Omega}_{GS}$ should become closer to their counterparts of $\hat{\Omega}_{OLS}$.

2.5 Conclusions and future research directions

In this paper I propose an estimation method that corrects estimation biases in the class of NK term premium models, and specifically, Model 2 of Fu (2017). Because this model estimates the
structural NK parameters and the reduced-form VAR(1) coefficient matrix in a single step by an FIML optimization procedure, which often settles at a local optimum or saddle point instead of the global optimum, the estimated VAR(1) coefficient matrix is often biased. As a result, the estimated expected short rates and term premia are often biased as well. In fact, the 10-year average expected short rate estimated by Model 2 of Fu (2017) using 30 years of historical data appears too stable – it just hovers around the short rate’s 30-year mean, and the movements of the estimated 10-year term premium mimic those of the 10-year yield. The estimation method I propose proceeds in two steps. In the first step, I estimate the reduced-form VAR(1) coefficient matrix by OLS. The OLS-estimated 10-year average short rate is more realistic – it reflects economic cycles and structural economic factors, and the OLS-estimated 10-year term premium is countercyclical. This first step up to this point is in essence the same as the first step of Bauer et al. (2012), which estimate a VAR-ATS model. Next, I correct the small-sample bias in the OLS VAR estimation using an analytical approximation approach. This is different from Bauer et al. (2012), which use a simulation approach. In the second step, I recover the NK parameters following a SVAR strategy developed by Keating (1990). This is again different from Bauer et al. (2012), which back out the ATS parameters using an FIML procedure. In this second step, I simplify Keating’s structural disturbance transformation procedure using a matrix form. Because the NK parameters estimated by the SVAR strategy cannot recover the reduced-form VAR(1) coefficient matrix, I use an equation-by-equation identification strategy to recover the NK system. In the implementation of such a strategy, I propose a different specification of the IS curve than the one in Model 2 of Fu (2017) and obtain a better fit. The NK parameters so identified satisfactorily back out the reduced-form VAR(1) coefficient matrix. Thus, the two main contributions of my proposed method are: 1) it generates more realistic expected short rates
and term premia; and 2) it builds consistency between the structural NK system and the reduced VAR system.

I envision two venues for future research. First, since the equation-by-equation identification strategy shows that only a few NK parameters are dominant while other NK parameters are almost zero, more estimation exercises using datasets of different time periods can be conducted to check whether this is still the case. If so, a further examination of the specification of the NK system may be warranted. Second, though the 10-year average short rate estimated by my method in this paper is more cyclical and structural than the one estimated in Model 2 of Fu (2017), it is still too stable in the period of 2009 to date (see Figure 2.1). In other words, since the Fed immediately cut the policy rate to almost zero in December 2008, the 10-year average short rate has been too stable – it has tracked the nearly zero policy rate too closely, while in reality the 10-year average short rate should have closely tracked the natural interest rate, which arguably has trended downward since the subprime crisis. Whether this is due to the Fed’s monetary policy zero lower bound (ZLB) constraint is uncertain. If yes, further research can explore how to address the ZLB constraint in an NK framework.
Chapter 3
A Shadow Rate New Keynesian Term Premium Model with Financial Risks

3.1 Introduction

The concept of “shadow rate” is first introduced by Black (1995), which treats the short rate as a call option on an underlying stochastic process that can take positive or negative values. When the underlying process has a negative value, the short rate is zero. When the underlying is positive, however, the value of the short rate is the same as that of the underlying. Black (1995) terms this underlying process the “shadow short rate,” or “shadow rate” in short.

Faced with a severe sub-prime mortgage crisis, which resulted in the failures of many large financial institutions such as Bear Stearns, Countrywide Financial and Lehman Brothers, the Federal Reserve (Fed) slashed its policy rate (short rate in finance terms) – the federal funds rate – to a range of 0-0.25% in December 2008. The Fed had held the policy rate at this level until December 2015, when the central bank started its rate hiking cycle that took the policy rate to 2.25-2.5% as of April 17, 2019. Economics and finance researchers and professionals call the period of December 2008 to December 2015 the zero lower bound (ZLB) period. During this period, since the short rate was at the ZLB and the Taylor rule was inactive (the short rate could not respond to changes in inflation and output gap), the Fed had used a number of unconventional monetary policy tools such as large-scale asset purchases (commonly known as quantitative easing or QE) and forward guidance in order to provide additional stimulus to the economy. How to address the ZLB constraint and the resulting inactive Taylor rule have invited a large volume of research efforts, of which I find the Wu-Xia shadow rate (Wu and Xia 2016)
and the Wu-Zhang Shadow Rate New Keynesian (SR-NK) model (Wu and Zhang 2016) most relevant to my ongoing research project on the relationship between U.S. Treasury (UST) yield curve term premia and the real economy (see Fu 2017 and 2018).

3.1.1 Motivations and contributions

This paper’s main contribution is proposing a new version of SR-NK term premium model that addresses the ZLB constraint by explaining the effects of the Fed’s forward guidance and QE separately. It is an extension to Fu (2017), which proposes an NK term premium model that builds a two-way feedback loop between the economy and the yield curve in a micro-founded way. The class of NK term premium models has been popular over the past 10 to 15 years, perhaps due to its well-established simple NK structure that describes the economy and its usage of equally well-established affine term structure (ATS) finance theory to specify the yield curve as an affine (constant plus linear term) function of the economy. Seminal papers in this literature include Hördahl et al. (2006), Rudebusch and Wu (2008) and Bekaert et al. (2010), with these papers offering different NK specifications. These papers provide no micro-founded feedback from the yield curve to the economy. Fu (2017) addresses this one-way feedback issue in the second of the two models presented. Model 1 is an NK-ATS model that incorporates a latent financial variable termed the Financial Risk Index (FRI) into the NK system. Model 2 uses the UST three-month vs. 10-year yield spread (henceforth the UST 3m10y slope) as a proxy for the FRI and thus drops the ATS part. Since the UST 3m10y slope is part of the yield curve and also part of the NK system, Model 2 provides a two-way bridge between the economy and the yield curve. However, both of the two models in Fu (2017) are unable to address the ZLB constraint because the Taylor rule equation in the NK system uses the short rate as the dependent variable.
Also, the short rate is an independent variable in the IS equation and cannot stimulate the output gap. As a result, Fu (2017)’s estimated 10-year average expected short rate appears unintuitively high and stable and the 10-year term premium’s variations appear to resemble those of the long-end yields too much during the ZLB period.

This paper borrows the building blocks of the Wu-Xia shadow rate and Wu-Zhang SR-NK model to address the ZLB constraint. Wu and Xia (2016) estimate the shadow rate using a three-factor ATS model. Like Black (2015), they define the short rate as the maximum of a lower bound and the shadow rate. They then define the shadow rate as the sum of a constant and the first two latent factors of the yield curve. The entire shadow rate term structure (the Wu-Xia shadow rate is the shadow rate with the shortest tenor in the term structure) is treated as an affine function of the latent factors. They derive a nonlinear function of the latent factors for observable UST forward yields. They use extended Kalman filter to estimate the shadow rate with an assumed lower bound value of 0.25%. The estimated shadow rate quickly dipped below zero in early 2009 and touched as low as -3% in summer 2014. Since the shadow rate is considered a combination of the first two latent factors, which are widely known as the level and slope drivers of the yield curve, the shadow rate is set up to capture the joint variations of short yields (through the level factor) and long yields (through the slope factor). And since forward guidance is found to mainly affect short yields while QE to mainly affect long yields, the shadow rate is considered to capture the effects of both forward guidance and QE. This could explain why the shadow rate reached such a negative level of -3% during the ZLB period. This is convenient, but it leaves the problem of decomposing the effects of the two unconventional monetary policy tools unsolved.
Wu and Zhang (2016) propose a three-equation NK model containing the Phillips curve, IS curve and Taylor rule, similar to the standard NK model. They provide microfoundations for QE by replacing the short rate with a benchmark interest rate for the private sector. They define the shadow rate as a linear function of the Fed’s bond purchases with a negative coefficient on the grounds that the Wu-Xia shadow rate and some private sector interest rates were seen to be negatively correlated with the Fed’s balance sheet at the ZLB. They set the benchmark interest rate as the sum of the Wu-Xia shadow rate and a constant risk premium during the ZLB period. During normal times, however, the benchmark rate is the sum of the policy rate and the constant risk premium. As such, the benchmark rate can be swapped with a splined series that fuses together the policy rate during normal times and the Wu-Xia shadow rate at the ZLB. This splined series activates the Taylor rule at the ZLB, but it causes two issues: 1) it creates a tenor mismatch between the ZLB period and normal times because the shadow rate captures QE’s strong effect on long-tenor yields while the short rate has the shortest tenor; and 2) it makes difficult, if not impossible, to build a connection between the Wu-Zhang SR-NK model and the yield curve. The reason is that, if the NK system drives the yield curve, then the Wu-Xia shadow rate must be one of the drivers of the yield curve but cannot be a combination of two drivers. In other words, the Taylor rule cannot drive both the level and slope of the yield curve.

I propose a new version of SR-NK model that can remove the above two issues. My model inherits the four-equation NK system of Fu (2017), which contains the Phillips curve, IS curve, Taylor rule and the FRI pricing equation. Like Wu and Zhang (2016), I replace the short rate with the shadow rate, which is different from the Wu-Xia shadow rate in that my shadow rate is considered to determine only the level of the yield curve. This treatment leaves the role of determining the yield curve slope to the FRI. As such, I dedicate the shadow rate to capture the
effect of forward guidance and the FRI to capture the effect of QE during the ZLB period. In this setup, there is no tenor mismatch between the short rate and shadow rate. Also, a connection between the NK system and the yield curve is naturally established with the shadow rate as the level driver and the FRI as the slope driver.

Like Wu and Xia (2016), I use the ATS theory to connect observable forward yields with the NK system. There are two latent variables in my model – the shadow rate and the FRI.

Also like Wu and Xia (2016), I use extended Kalman filter to estimate my model. My measurement equation is different from that of Wu and Xia (2016) in that mine includes forward yields, inflation and output gap while theirs contains forward yields only. I then modify their algorithm to filter out only the values of the shadow rate and FRI at each time point. To reduce the parameter space, I use the values of the structural parameters related to inflation and output gap estimated by the grid search method of Fu (2018) and leave the remaining structural parameters to be estimated. These remaining structural parameters’ estimates are found to be similar to those obtained by the grid search method of Fu (2018).

After the model estimation, I find that the level of the imputed shadow rate is quite lower than the three-month yield, which is used as the proxy for the short rate. To be exact, the mean difference between the three-month yield and the imputed shadow rate before the ZLB period was 1.04%. By contrast, the level of the imputed FRI is much higher than the UST 3m10y slope. This could be because in the optimization problem there are many local optima or saddle points, each of which has a unique set of levels for the shadow rate and the FRI – the optimization procedure just reaches one local optimum or saddle point. The good thing is that, after being added the pre-ZLB mean difference of 1.04% compared to the three-month yield, the imputed shadow rate is shown to have tracked the three-month yield very well before the ZLB period and
to have dipped below zero and touched a minimum of -0.53% during the ZLB period. In fact, the imputed shadow rate is found to be most correlated with the one-year yield (with a correlation of 0.908) during the ZLB period. I also find that the imputed FRI generally shares a same trend with the UST 3m10y slope, though their correlation of 0.692 is not high. This could be because inflation and output gap enter the estimation and weaken the FRI’s role of yield curve slope driver, as documented by Ang and Piazzesi (2003), which add that “level factor survives largely intact when macro variables are incorporated.”

The fact that the imputed FRI does not tightly track the UST 3m10y slope, coupled with Krippner (2017)’s critiques that the Wu-Xia shadow rate is highly sensitive to the lower bound parameter and is inconsistent with some of the Fed’s monetary policy events, calls for a proxy for the shadow rate during the ZLB period. I choose a combination of the one- and two-year yields as the proxy for the shadow rate at the ZLB given Zhang (2017)’s finding that forward guidance has the largest effect on the yield with a tenor of 16 months and also given that three out of five forward guidance announcements promised to hold the ZLB for one and a half years to three years. I apply the principal component analysis (PCA)-weighted variations of the one- and two-year yields during the ZLB period to the three-month yield as if the three-month yield had reacted to forward guidance announcements as the one- and two-year yields. The adjusted three-month yield was below -0.3% for three years during the ZLB period with a minimum level of -0.38% reached in September 2011. These negative levels are in line with other countries’ negative policy rates, e.g., the deposit rate of -0.4% for the European Central Bank (ECB), -0.75% in Switzerland, -0.35% in Sweden, etc.. Such negative levels are supported by the predictions of a modified Taylor rule (regressing the thee-month yield on inflation, output gap and the UST 3m10y slope), which is fitted to an in-sample period between the first quarter of
1997 and the third quarter of 2008 and which predicts a minimum of -0.69% during the ZLB period, vs. -2.21% by a standard Taylor rule without the UST 3m10y slope.

I construct Model 2 with the adjusted three-month yield replacing the shadow rate and the UST 3m10y slope replacing the FRI in the NK system. Term premia can be readily calculated without using the ATS part. I follow Fu (2018) to estimate Model 2 in three steps: 1) fit a VAR(1) process to the data; 2) correct the small-sample bias in the coefficient matrix; and 3) apply a structural VAR (SVAR) procedure to estimate the structural NK parameters. The parameter estimates are very similar to those obtained by Fu (2018).

I compute the 10-year average expected short rate and term premium using Model 2. The 10-year average short rate during the ZLB period was on average 21 basis points lower than the one obtained by Fu (2018). I regress output growth on differenced 10-year average short rate and term premium and confirm with Rudebusch et al. (2007) that the two predictors have intuitive effects on output growth, but their effects seem to be dominated by lagged output growth.

The contributions of this paper can be summarized below. First, the paper points out two issues in the Wu-Zhang SR-NK model: 1) a tenor mismatch between the policy rate during normal times and the shadow rate during the ZLB period and 2) the difficulty (or impossibility) of building a link between their SR-NK model and the yield curve. Second, it addresses these two issues by proposing a SR-NK model that assigns the shadow rate the role of yield curve level driver and the FRI the role of slope driver. Likewise, the shadow rate is dedicated to capture the effect of the Fed’s forward guidance and the FRI to capture the effect of QE. In this setup, the imputed shadow rate is shown to have tracked the short rate very well before the ZLB period and to have dipped to -0.53% during the ZLB period, a negative level that is in line with the existing negative policy rates adopted by other central banks and that is much higher than as
much as -3% seen in the estimated Wu-Xia shadow rate series. This provides policy makers valuable reference information regarding what level of negative interest rate is feasible if the ZLB constraint is removed and if quantitative easing is restarted to complement the negative rate policy, allowing policy makers to strike a balance between the two unconventional policies. Third, the paper proposes a simple yet plausible way to replace the latent shadow rate by adjusting the short rate during the ZLB period using the variations of the one- and two-year yields. The adjusted short rate is shown to touch -0.38% during the ZLB period, also consistent with other central banks. This adjustment to the short rate constructs a ZLB constraint-free NK model without latent variables. It greatly simplifies model implementation, avoids the imputed latent variables’ sensitivity to parameter values, and provides policy makers and financial market practitioners observable market variables to monitor and to communicate with.

3.1.2 Organization of this paper

The rest of this paper is organized as follows. Section 3.2 introduces the Wu-Xia shadow rate model and the Wu-Zhang SR-NK model. Section 3.3 presents my proposed SR-NK-ATS term premium model and discusses the rationale for assigning the shadow rate and the FRI different roles in driving the yield curve and in capturing the Fed’s forward guidance and QE. Section 3.4 briefly describes data. Section 3.5 discusses the extended Kalman filter estimation procedure and show that the imputed shadow rate and FRI have done a fairly good job in their assigned roles. Section 3.6 covers how to find a proxy for the shadow rate and presents Model 2’s estimation results. Section 3.7 exhibits the calculated 10-year average short rate and term premium and assesses whether these two variables can predict economic growth. Section 3.8 concludes.
3.2 The Wu-Xia SR-ATS model and Wu-Zhang SR-NK model

3.2.1 The Wu-Xia SR-ATS model

Like Black (1995), Wu and Xia (2016) model the short rate as the maximum of the shadow rate \( s_t \) and a lower bound \( i_L \):

\[
i_t = \max(i_L, s_t).
\]

(3.1)

\( i_L \) is set at 0.25\% in the estimation procedure of Wu and Xia (2016).

3.2.1.1 Modeling the shadow rate term structure using the ATS theory

Wu and Xia (2016) then use the ATS finance theory (e.g., Duffie and Kan 1996, Dai and Singleton 2002) to model the shadow rate term structure. The first shadow rate in the term structure, \( s_t \), is treated as an affine function of a state vector process \( X_t \):

\[
s_t = a_0 + a_1' X_t,
\]

(3.2)

where \( X_t \) follows a VAR(1) process under the physical probability measure \( P \):

\[
X_t = c + \Omega X_{t-1} + \Gamma \varepsilon_t.
\]

(3.3)

Next step is to transform the \( P \)-measured VAR(1) process of \( X_t \) to a corresponding VAR(1) process under the risk-neutral probability measure \( Q \). The Radon-Nikodym derivative \( L_t \) serves this purpose:

\[
E_Q[\varepsilon_{t+1}] = E_P[L_{t+1}\varepsilon_{t+1}]/L_t,
\]

where \( L_t \) follows:

\[
L_{t+1} = L_t \exp(-\frac{1}{2} \lambda_t' \lambda_t - \lambda_t' \varepsilon_{t+1}),
\]

(3.4)

\( \lambda_t \) measures market prices of risk and also is an affine function of \( X_t \):

\[
\lambda_t = \lambda_0 + \lambda_1 X_t.
\]

(3.5)

Therefore, the \( Q \)-measured VAR(1) process of \( X_t \) is as follows:

\[
X_t = c^Q + \Omega^Q X_{t-1} + \Gamma^Q \varepsilon_t.
\]

(3.6)
The VAR(1) parameters under the P and Q measures have the following relations:

\[ c^Q = c - \Gamma \lambda_0, \]  
\[ \Omega^Q = \Omega - \Gamma \lambda_1. \]  

(3.7)  
(3.8)

Let \( s_{t+n} \) be the shadow rate \( n \) periods ahead. \( s_{t+n} \) also is an affine function of \( X_t \):

\[ s_{t+n} = \bar{A}_n + B'_n X_t, \]  

(3.9)

where

\[ \bar{A}_n = a_0 + a'_1 \{ \sum_{j=0}^{n-1} (\Omega^Q)^j \} c^Q, \]  
\[ B'_n = a'_1 \{ (\Omega^Q)^n \}. \]  

(3.10)  
(3.11)

Intuitively, we can regard \( s_{t+n} \) as a latent intrinsic future short rate, which can be either positive or negative, while the real-world future short rate \( i_{t+n} \) that can be extracted from today’s yield curve cannot go below the lower bound \( i_L \).

3.2.1.2 Modeling the forward yield term structure

Wu and Xia (2016) derive a formula for the one-period forward yield \( n \) periods ahead, \( f_t^n \):

\[ f_t^n = i_L + \sigma_n^Q g(\frac{A_n + B'_n X_t - i_L}{\sigma_n^Q}), \]  

(3.12)

where \( \sigma_n^Q \) is the standard deviation of \( s_{t+n} \) under the Q measure and is defined below:

\[ (\sigma_n^Q)^2 = \sum_{j=0}^{n-1} a'_1 (\Omega^Q)^j \Gamma \Gamma' (\Omega^Q)^j a_1. \]  

(3.13)

The function \( g(z) = z \Phi(z) + \emptyset(z) \) contains a normal cumulative distribution function \( \Phi(.) \) and a normal probability density function \( \emptyset(.) \). \( A_n \) is defined as:

\[ A_n = \bar{A}_n - \frac{1}{2} a'_1 \{ \sum_{j=0}^{n-1} (\Omega^Q)^j \} \Gamma \Gamma' \{ \sum_{j=0}^{n-1} (\Omega^Q)^j \} a_1. \]  

(3.14)

Therefore, forward yields are no longer an affine function but rather a nonlinear function of \( X_t \).
A detailed derivation of Equations (3.12)-(3.14) is in Appendix A of Wu and Xia (2016).

### 3.2.1.3 Estimation of the shadow rate

Since the shadow rate $s_t$ is unobservable, it has to be estimated from observable market variables. Wu and Xia (2016) estimate $s_t$ from the forward yield term structure using extended Kalman filter. The measurement equation is:

$$f_t^n = i_L + \sigma_n^Q g\left(\frac{A_n + B'_n X_t - i_L}{\sigma_n^Q}\right) + \eta_{n,t}. \quad (3.15)$$

And the transition equation is Equation (3.3), the $P$-measured VAR(1) process for $X_t$. $X_t$ contains three latent factors, which, together with the parameters $(c, c^Q, \Omega, \Omega^Q, \Gamma, a_0, a_1)$, have to be backed out in the estimation procedure. Note that one also can estimate $\lambda_0$ and $\lambda_1$ instead of $c^Q$ and $\Omega^Q$, but Wu and Xia (2016) choose to estimate $c^Q$ and $\Omega^Q$. For identification, the authors employ the following normalizing restrictions, which are proposed by Joslin, Singleton, and Zhu (2011): 1) $a_1 = [1 \ 1 \ 0]'$, 2) $c^Q = 0$, 3) $\Omega^Q$ is a real Jordan form, and 4) $\Gamma$ is lower triangular. In fact, $\Omega^Q$ is assumed to be in the following form:

$$\Omega^Q = \begin{bmatrix}
\omega^Q_1 & 0 & 0 \\
0 & \omega^Q_2 & 1 \\
0 & 0 & \omega^Q_2
\end{bmatrix}. \quad (3.16)$$

In other words, only two diagonal elements $\Omega^Q$, $\omega^Q_1$ and $\omega^Q_2$, have to be estimated. Therefore, estimating $c^Q$ and $\Omega^Q$ can greatly reduce the parameter space compared to estimating $\lambda_0$ and $\lambda_1$.

Once the latent factor vector $X_t$ is recovered, $s_t$ is just the sum of $a_0$ and the first two factors:

$$s_t = a_0 + X_{1,t} + X_{2,t}. \quad (3.17)$$

And the rest of the shadow rate term structure can be computed from Equation (3.9).
Figure 3.1 shows the imputed Wu-Xia shadow rate compared to the effective federal funds rate (EFFR), UST one- and two-year yields. From 1998 to the end of 2008, the shadow rate tracked the EFFR rather well (the former was smoother than the latter), yet the shadow rate deviated from the EFFR during the ZLB period. In fact, the shadow rate once dipped to -3% in summer 2014 vs. a minimum level of 0.04% for the EFFR during the ZLB period.

Figure 3.1. Wu-Xia shadow rate vs. EFFR, UST 1-year and 2-year yields (%)

3.2.2 The Wu-Zhang SR-NK model

Wu and Zhang (2016) propose a three-equation NK model that incorporates the shadow rate $s_t$ as one of the three state variables:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa y_t, \quad (3.18)$$

$$y_t = E_t y_{t+1} - \frac{1}{\sigma}(s_t - E_t \pi_{t+1} - s), \quad (3.19)$$

$$s_t = \delta_s s_{t-1} + (1 - \delta_s)(\delta_\pi \pi_t + \delta_y y_t + s), \quad (3.20)$$
where $\pi_t$ is inflation, $y_t$ is output gap and $s$ is the steady-state level of the shadow rate. Equation (3.18) is the standard Phillips curve, Equation (3.19) is a shadow rate IS curve and Equation (3.20) is a shadow rate Taylor rule.

In this model, $s_t$ can be considered the private sector’s benchmark interest rate (minus a constant risk premium) through which the Fed’s conventional and unconventional monetary policy tools take effect on the economy. This model is almost the same as the standard NK model except that the policy rate $i_t$ in the latter model is replaced with $s_t$. Since $s_t$ can be positive or negative, this adjustment removes the ZLB constraint faced by the standard NK model.

### 3.2.2.1 Micro-foundations of the SR-NK model for QE

Assume households maximize their utility function:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left( \frac{C_t^{1-\sigma}S_t^{1-\gamma}}{1-\sigma} - \frac{N_t^{1+\tau}}{1+\tau} \right),$$

(3.21)

with the budget constraint:

$$C_t + \frac{B_t^H}{P_t} = \frac{B_{t-1}^H}{P_{t-1}} + W_t N_t + T_t,$$

(3.22)

where $C_t$ is consumption, $N_t$ is labor supply, $P_t$ is the price level, $W_t$ is real wage, $T_t$ contains transfers and $B_t^H$ represents households’ nominal bond holdings from time $t-1$ to $t$ with a gross return of $R_{t-1}^B$.

The first-order condition for $B_t^H/P_t$ can be derived as:

$$C_t^{-\sigma} = \beta R_t^B E_t \left( \frac{C_{t+1}^{\sigma}}{\pi_{t+1}} \right),$$

(3.23)

where $\pi_{t+1} = P_{t+1}/P_t$ is inflation.
Replacing $C_t$ with output $Y_t$ and log-linearizing the Euler equation leads to the following IS curve:

$$y_t = E_t y_{t+1} - \frac{1}{\sigma} (r^B_t - E_t \pi_{t+1} - r^B),$$

(3.24)

The only difference between Equations (3.24) and (3.19) is that in the former, the bond return $r^B_t$ and its steady-state level $r^B$ replace the shadow rate and its steady-state level in the latter. $r^B_t$ can be regarded as a benchmark interest rate for households.

Define $r^B_t$ as the Fed’s policy rate plus a time-varying risk premium $rp_t$:

$$r^B_t = i_t + rp_t.$$  

(3.25)

Denote the central bank’s holdings of nominal bonds as $B^C_t$ and its steady-state level as $B^C$. Then, $rp_t$ is negatively correlated with changes in the central bank’s bond holdings (QE):

$$rp_t = rp - \phi(B^C_t - B^C),$$

(3.26)

where $rp$ is the steady-state level of risk premium. Therefore,

$$r^B_t = i_t + rp - \phi(B^C_t - B^C).$$

(3.27)

When the policy rate $i_t$ has not hit the ZLB, $B^C_t = B^C$ and $r^B_t = i_t + rp$. At the ZLB, however, $i_t = 0$ and the Fed chooses to launch QE programs by increasing $B^C_t$ to stimulate the economy. The shadow rate $s_t$ can serve as a transmission vehicle of QE to the real economy through:

$$s_t = -\phi(B^C_t - B^C).$$

(3.28)

At the ZLB, therefore,

$$r^B_t = s_t + rp.$$  

(3.29)

The shadow rate is just the private sector’s benchmark interest rate minus risk premium’s steady-state level. As such, $r^B_t$ and $s_t$ can be interchangeable in the IS curve and Taylor rule.
Wu and Zhang (2016) provide no discussion about the estimation of their NK model and no connection of the model with the Treasury yield term structure. The authors try to estimate the shadow rate Taylor rule (Equation 3.20) using a splined series of the EFFR during normal times and the Wu-Xia shadow rate at the ZLB as \( s_t \), however.

### 3.3 My proposed SR-NK-ATS term premium model

#### 3.3.1 The NK structural system

Similar to Fu (2018), my proposed model starts from a modified four-equation NK system:

\[
\pi_t = \beta E_t \pi_{t+1} + (1 - \beta) \pi_{t-1} + \kappa y_t + \epsilon_{\pi,t}, \tag{3.30}
\]

\[
y_t = \psi y_{t-1} + (1 - \psi) \{\alpha [E_t y_{t+1} - (s_t - E_t \pi_{t+1})] - \theta q_t\} + \epsilon_{y,t}, \tag{3.31}
\]

\[
s_t = \mu_t + \rho s_{t-1} + (1 - \rho) (\delta_{\pi} \pi_t + \delta_{y} y_t - \delta_{q} q_t) + \epsilon_{i,t}, \tag{3.32}
\]

\[
q_t = \phi_q E_t q_{t+1} - \phi_y E_t y_{t+1} + \phi_r (s_t - E_t \pi_{t+1}) + (1 - \phi_q) q_{t-1} + \epsilon_{q,t}, \tag{3.33}
\]

where \( q_t \) represents the FRI. The only difference between this NK system and the one in Fu (2018) is that policy rate \( i_t \) in the latter is replaced with the shadow rate \( s_t \).

The micro foundations for bringing in \( q_t \) and for including lagged state variables into the NK system are provided in Appendix 1.A of Fu (2017). The FRI is similar to the aggregate stock market index in Nisticò (2012), which uses an overlapping generation NK model to study stock prices' wealth effect on the real economy and monetary policy.

Equations (3.30) – (3.33) can be written in a compact form:

\[
BX_t = \mu + AE_t X_{t+1} + MX_{t-1} + \epsilon_t. \tag{3.34}
\]

The four shocks in \( \epsilon_t \) are uncorrelated with one another, and \( \epsilon_t \sim N(0, D) \) where \( D \) is the diagonal
variance matrix. This is a rational-expectation structural system. Using the undetermined coefficients method, the solution is Equation (3.3).

3.3.2 Adding the ATS part

Since the solution to my proposed NK system is the same as the $P$-measured VAR(1) process of the state variables in Wu and Xia (2016), we can reuse their ATS setup. This is also the approach employed in Model 1 of Fu (2017). In other words, we can reuse Equations (3.4)-(3.14). Note that $X_t$ in Wu and Xia (2016) is composed of three latent variables, while $X_t$ in my proposed model contains only two latent variables – the shadow rate and the FRI – together with observable inflation and output gap.

I make a minor adjustment: in Equation (3.2), $a_0 = 0$ and $a_1 = [0 0 1 0]'$. The necessity of this adjustment is obvious in that $s_t$ is the third variable of $X_t$.

3.3.3 The roles of $s_t$ and $q_t$ in driving the yield curve

This minor adjustment in the ATS part makes my proposed model rather different than the Wu-Xia SR-ATS model. In their model, $s_t$ is an intercept plus the first two latent factors. Since these two factors are widely considered to dynamically determine the level and slope of the entire yield curve in voluminous studies of ATS theory (e.g., Ang and Piazzesi 2003; Christensen, Diebold and Rudebusch 2011; Hamilton and Wu 2012), $s_t$ in Wu and Xia (2016) is the composite driver of the level and slope of the entire yield curve (this is clear in Figure 3.2). By contrast, $s_t$ in my proposed model is the first latent factor and thus should determine only the level of the yield curve, potentially leaving $q_t$ a role of driving the slope of the yield curve.
Figure 3.2 shows the Wu-Xia shadow rate and the first two latent factors between December 2008 and December 2013 produced by the data and MATLAB code of Wu and Xia (2016). Until spring 2012, the decreasing first factor had dragged the shadow rate down, partially offset by somewhat upward moving second factor. After that, the first factor increased quickly, but more than offset by the more quickly decreasing second factor. As a result, the shadow rate moved even lower.

Figure 3.2. Wu-Xia shadow rate vs. two latent factors during the ZLB period (%)

My model setup is similar to Ang and Piazzesi (2003). In their model, $X_t$ contains two macro variables and three latent factors, which the authors pin down to “level”, “slope” and “curvature”. The first macro variable is the first principal component (PC) of a group of inflation-related variables and the second macro variable is the first PC of a group of real
activity-related variables. Therefore, the two macro variables in Ang and Piazzesi (2003) are essentially $\pi_t$ and $y_t$ in my model. Likewise, $s_t$ and $q_t$ in my model can be related to the first two, or level and slope, latent factors in Ang and Piazzesi (2003). The main difference is that in their model, there is no structural NK system for $X_t$. However, this difference does not affect the same roles of $[s_t, q_t]$’ in my model in driving the yield curve as those of the first two latent factors in their model. The reason is the NK system only governs the structural relationships within $X_t$ rather than the relationship between $X_t$ and the yield curve.

3.3.4 The roles of $s_t$ and $q_t$ in explaining the Fed’s unconventional policies

Having constructed a hypothesis that $s_t$ and $q_t$ can serve as the level and slope drivers of the yield curve, we can proceed to explore the roles of $s_t$ and $q_t$ in explaining the Fed’s unconventional policies at the ZLB.

Swanson (2016) compiles a table of the Fed’s unconventional monetary policy announcements between 2009 and 2014 as shown in Table 3.1. These announcements can be categorized into two types: forward guidance and QE. These two types of announcements often came together, e.g., on March 18, 2009, September 13, 2012 and December 12, 2012. Three out of five forward guidance announcements specified a length of extending the ZLB with a range of one and a half years to three years.
Zhang (2017) studies the effects of the Fed’s forward guidance and QE announcements on the yield curve. The study finds that forward guidance has the largest effect on the yield with a tenor of 16 months and that QE’s effect on yields increases as tenor increases, i.e., QE affects the longest-tenor yield the most.

Based on the above discussion, it appears safe to assume that, in my model, $s_t$ should explain the effect of forward guidance and $q_t$ should explain the effect of QE. Therefore, my model
should provide a structural approach to studying the effects of forward guidance and QE separately. This is a salient feature that the Wu-Xia shadow rate cannot provide. Since the Wu-Xia shadow rate combines the first two latent factors, it should capture the effects of both forward guidance and QE. This is convenient, but it leaves the problem of decomposing the effects of the two unconventional monetary policy tools unsolved. This also could explain why Wu and Zhang (2016) provide no connection of their SR-NK model with the yield curve because if the NK system drives the yield curve, then the Taylor rule featuring the Wu-Xia shadow rate $s_t$ must be one of the drivers of the yield curve but cannot be a combination of two drivers. Unless the shadow rate is set to be one of the latent factors, there seems to be no connection that can be established between the Wu-Zhang SR-NK model and the yield curve.

Furthermore, the Wu-Xia shadow rate used in the Wu-Zhang SR-NK model has a tenor mismatch problem. In their model, the private sector’s benchmark interest rate $r_t^B$ is the sum of $s_t$ and a constant risk premium at the ZLB (see Equation 3.29). During normal times, however, $r_t^B$ is the sum of policy rate $i_t$ and the constant risk premium. Since $s_t$ captures QE’s strongest effect on the longest-tenor yield while $i_t$ has the shortest tenor, it creates a tenor mismatch between the ZLB period and normal times. By contrast, $s_t$ in my model is assumed to strongly affect only short-tenor yields and thus $s_t$ and $i_t$ have no tenor mismatch issue.

As a summary of Sections 3.3.3 and 3.3.4, $s_t$ in my model setup should capture the variations of short-tenor yields, e.g., one- to three-year yields, and $q_t$ should capture the variations of the yield curve slope during the ZLB period.
3.3.5 Extracting the term premium

To extract an $n$-period term premium, we need to first obtain the forward risk premium embedded in a one-period forward yield. The one-period forward premium, denoted as $r_t^n$, is the difference between the one-period forward yield $n$ periods ahead, $f_t^n$, and the expected one-period short rate $n$ periods ahead under the $Q$ measure, $E_t^Q[i_{t+n}]$:

$$r_t^n = f_t^n - E_t^Q[i_{t+n}]. \quad (3.35)$$

$f_t^n$ is calculated by Equation (3.12). $E_t^Q[i_{t+n}]$, as derived by Wu and Xia (2016), is obtained as:

$$E_t^Q[i_{t+n}] = i_L + \sigma_n^Q g \left( \frac{\bar{\eta}_t + \eta_t' X_t - i_t}{\sigma_n^Q} \right). \quad (3.36)$$

The term premium is the average of the forward premia up to time $t+n-1$:

$$t_t^n = \frac{1}{n} \sum_{i=0}^{n-1} r_t^i. \quad (3.37)$$

3.4 Data

I use the dataset of Fu (2017) and Fu (2018). It contains quarterly PCE inflation, CBO output gap, and Treasury three-month, six-month, one-year through 10-year zero-coupon bond yields. Quarterly average yields are computed and used. According to CBO’s white paper (CBO 2004), the output gap series is constructed using a growth model. The sample period is from the first quarter of 1985 to the fourth quarter of 2016. Data are obtained from Bloomberg. Forward yields are directly obtained from the yield curve using the following equation:

$$f_t^n = y_t^{n+1}(n+1) - y_t^n n \quad (3.38)$$

Hence, Equation (3.12) is actually the fitted value of $f_t^n$ with respect to the state vector $X_t$. 

100
3.5 Estimation of my proposed SR-NK-ATS model

3.5.1 Estimation by extended Kalman filter

Like Wu and Xia (2016), I use extended Kalman filter to estimate the model. The transition equation is Equation (3.3), same as in Wu and Xia (2016). The measurement equation, however, is slightly different than Equation (3.15) used by Wu and Xia (2016). The reason is that $X_t$ contains only three unobservable variables in their model but it contains two unobservable variables plus two observable variables $\pi_t$ and $y_t$ in my model. Hence, $\pi_t$ and $y_t$ need to be stacked to the left-hand side of my measurement equation. Let Equation (3.15) be rewritten as:

$$F_t = F(X_t, n, \theta) + u_t,$$

where $F_t$ contains all the observable forward yields $f_t^n$, $\theta$ is the parameter vector and $F(X_t, n, \theta)$ is the nonlinear function for $f_t^n$. Then, my measurement equation becomes:

$$\begin{bmatrix} \pi_t \\ y_t \\ F_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} X_t + F(X_t, n, \theta) + \eta_t. \quad (3.39)$$

The parameters to be estimated are the structural parameter vector $[\beta \kappa \alpha \psi \theta \rho \delta \pi \delta_y \delta_q \phi_r \phi_y \phi_q]'$, the standard errors of the four state variables $[\sigma_\pi \sigma_y \sigma_i \sigma_q]'$ and $c^Q, \Omega^Q$. Since the number of parameters is large, I decide to reduce the parameter space by setting the structural parameters related to $\pi_t$ and $y_t$ to the values obtained by the grid search method of Fu (2018), i.e., I set $[\beta \kappa \alpha \psi \rho \delta \pi \delta_y]' = 0.55 0.001 0.05 0.995 0.92 1.168 1.626]'$ and only leave $[\theta \delta_q \phi_r \phi_y \phi_q]'$ to be estimated.

Like Wu and Xia (2016), I set $c^Q = 0$. But unlike Wu and Xia (2016), I set $\Omega^Q$ to a lower triangular form (following Singleton 2006) rather than a real Jordan form, considering that the Jordan matrix’s repeated diagonal elements are too restrictive. I also require that the diagonal
elements of $\Omega^Q$ should be greater than 0.5 but smaller than 1 to ensure reasonable persistence (or reasonably slow mean-reversion) but to avoid nonstationarity.

In addition, I set $i_L = 0.125\%$ rather than 0.25% in Wu and Xia (2016), considering that the Fed’s policy rate was a range of 0-0.25% rather than 0.25% during the ZLB period. A detailed algorithm of the extended Kalman filter including the updating and prediction schemes can be found in Appendix B of Wu and Xia (2016). Since my measurement equation is different from that of Wu and Xia (2016), I present my different algorithm in Appendix 3.A of this paper.

### 3.5.2 Estimation results

The optimization algorithm converges quickly. Table 3.2 shows the estimated NK structural parameters (shaded) are very similar to those by the grid search method of Fu (2018).

Table 3.2. NK structural parameter estimates vs. those by the grid search of Fu (2018)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fu (2018) grid search</th>
<th>Kalman filter in this paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.551</td>
<td>0.551</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.995</td>
<td>0.995</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.001</td>
<td>0.011</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.921</td>
<td>0.921</td>
</tr>
<tr>
<td>$\delta_\pi$</td>
<td>1.168</td>
<td>1.168</td>
</tr>
<tr>
<td>$\delta_y$</td>
<td>1.626</td>
<td>1.626</td>
</tr>
<tr>
<td>$\delta_q$</td>
<td>1.507</td>
<td>1.466</td>
</tr>
<tr>
<td>$\phi_r$</td>
<td>0.010</td>
<td>0.012</td>
</tr>
<tr>
<td>$\phi_y$</td>
<td>0.010</td>
<td>0.014</td>
</tr>
<tr>
<td>$\phi_q$</td>
<td>0.525</td>
<td>0.496</td>
</tr>
</tbody>
</table>
Figure 3.3. $B_n$ coefficients of forward yields w.r.t. the four state variables of $X_t$.

Figure 3.3 depicts the $B_n$ (see Equation 3.15) coefficient term structure of shadow rates and forward yields with respect to the four state variables in $X_t$. We can see that the $B_n$ term structure with respect to $s_t$ dominates, which means the latent shadow rate $s_t$ has the largest effect on the shadow rate and forward yield term structures. This implies that $s_t$ drives the level of the term structure. Furthermore, the effect of $s_t$ declines as tenor increases, which is intuitive. The hump-shaped $B_n$ term structure with respect to $q_t$ shows that $q_t$ has the largest effect in the tenor area of three to four years. The effect of $y_t$ peaks around the one-year tenor while the effect of $\pi_t$ seems more long-lasting. These results are somewhat different than those of Ang and Piazzesi (2003), which show that $B_n$ is upward-sloping for the latent level and slope factors with the latter slope much steeper. This should be understandable because the $B_n$ coefficients of Ang and Piazzesi (2003) are calculated for spot yields rather than for forward yields in this study.
Figure 3.4 exhibits the imputed shadow rate series $s_t$ compared to some short-tenor yields. We can see that $s_t$ has shared the same trend as those short-term yields, but the general level of $s_t$ is much lower than those of the short-tenor yields. This can be attributed to the fact that two latent variables have to be imputed in this model. In the optimization problem there may be many local optima or saddle points, each of which is associated with one unique set of levels for the latent shadow rate and the FRI. The optimization procedure just reaches one local optimum or saddle point and throws out one set of shadow rate and FRI. It is likely that the low level of $s_t$ is offset by the high level of $q_t$, which is indeed the case when we look at the imputed $q_t$ in Figure 3.6 later.

Figure 3.4. Imputed shadow rate vs. three-month, one-year and two-year yields (%, 1998-2016)
To make Figure 3.4 more meaningful, I calculate the mean difference between the three-month yield and the imputed $s_t$ prior to the ZLB. The mean difference is 1.043%, which is then added to the imputed $s_t$ series (on an assumption that $s_t$ should closely track the three-month yield before the ZLB) and plotted in Figure 3.5.

We can see in Figure 3.5 that the adjusted $s_t$ had tracked those short-tenor yields very well before the ZLB. Between the beginning of 2009 and early 2014, however, $s_t$ had trended downward while the three-month yield had been stuck at nearly zero. During the entire ZLB period, the adjusted $s_t$ actually appeared to be most correlated with the one-year yield as they both touched the lowest levels of -0.53% and 0.08%, respectively, in 2014. In fact, the correlation coefficients of $s_t$ with the one-year, three-month and two-year yields are 0.908, 0.779 and 0.873 during the ZLB period. For the entire sample period, these correlation coefficients are 0.99, 0.988 and 0.984.

Figure 3.5. Adjusted shadow rate vs. three-month, one-year and two-year yields (%, 1998-2016)
Figure 3.6 plots the imputed $q_t$ together with the UST 3m10y slope in the upper panel. In the lower panel, the imputed $q_t$ is adjusted by subtracting a constant from it. We can see that $q_t$ generally moves in tandem with the UST 3m10y slope, though the correlation coefficient of 0.692 is not so high. Note that the level of the imputed $q_t$ is much higher than the level of the UST 3m 10y slope as can be seen in the upper panel. This offsets the low level of the imputed $s_t$. We also can see in the lower panel that the adjusted $q_t$ does not track the UST 3m10y slope very closely.

Figure 3.6. Imputed FRI vs. UST 3m10y slope (%; 1998-2016)

Figures 3.4-3.6 confirm that $s_t$ and $q_t$ indeed serve their roles in driving the yield curve’s level and slope and in explaining the effects of the Fed’s forward guidance and QE, as
hypothesized in Sections 3.3.3 and 3.3.4. The not-so-high correlation between $q_t$ and the UST 3m10y slope reveals that $q_t$’s role of yield curve slope driver is weaker than $s_t$’s role of yield curve level driver. This is probably because my proposed model incorporates macro variables, which dilute $q_t$’s slope driver effect on the yield curve, as documented in Ang and Piazzesi (2003), which also state that “level factor survives largely intact when macro variables are incorporated.”

The facts that the levels of the imputed $s_t$ and $q_t$ are quite different from short-tenor yields and the UST 3m10y slope and that the imputed $q_t$ does not track the UST 3m10y slope tightly show a model-based imputed latent variable’s difficulty in closely tracking the variations of a real-world financial market variable. This has many reasons, which, among others, could include: 1) a model-based latent variable may be too smooth; 2) the model may be mis-specified; 3) the latent variable may be too sensitive to changes in parameter values; and 4) if the model has to be estimated by an optimization procedure, the procedure may reach a local optimum or saddle point and thus the imputed variable may not behave as expected. The apparently different levels of $s_t$ and $q_t$ from short-tenor yields and the UST 3m10y slope may have been caused by one or several of the above reasons.

We also can consider the Wu-Xia shadow rate as an example. Though it has received wide recognition among researchers and monetary policy makers so that the Federal Reserve Bank of Atlanta (Atlanta Fed)’s Website hosts the shadow rate’s live data, it also has invited many critiques. Krippner (2017), for example, point out that the Wu-Xia shadow rate has, among others, the following shortcomings: 1) the estimated shadow rate is sensitive to different values of the lower bound $i_L$ and to different estimation sample lengths; and 2) the estimated shadow rate is inconsistent with some of the Fed’s unconventional monetary policy events.
I use Wu and Xia (2016)’s data (sample period: January 1990 – December 2013) and MATLAB code and I am able to verify Krippner (2017)’s findings. As depicted in Figure 3.7, though the estimated shadow rate with a lower bound of 0.25% (the benchmark shadow rate series presented in Wu and Xia 2016 and provided by the Atlanta Fed) displayed a clear downward trend between 2009 and 2013, the other three estimated series with lower bounds of 0.125%, 0.05% and 0% all exhibited an upward trend between mid-2009 and late 2010, when the QE1 and QE2 programs had been running. Arguably the QE1 and QE2 programs should have brought down long-term yields, as evidenced by the declining UST 3m10y slope during that period. The shadow rate should have declined as well since it is considered to capture the effects of both forward guidance and QE. Furthermore, arguably a lower bound of zero or nearly zero should be more plausible than 0.25% because 1) the Fed’s policy rate was 0-0.25% and 2) the EFFR averaged 0.14% and once touched 0.04% during that period. However, it was the shadow rate series with the highest lower bound of 0.25% that made sense during the period. This shows how sensitive the Wu-Xia shadow rate model is to different parameter values, and the sensitivity is unintuitive. Wu and Xia (2016) actually attempt to address this parameter sensitivity issue. Their paper tests a lower bound of 0.19% to produce a shadow rate series that has higher values than the benchmark series but that does not show an upward trend. The paper then claims: “the dynamics of the two series exhibit a strong comovement,” and “the comovement rather than difference in levels between the shadow rates is what drives the key results.” Such a conclusion is inadequate because the paper does not show the estimated shadow rates with a lower bound of zero or nearly zero.
Another period during which all the four shadow rate series were counterintuitive was the so-called “taper tantrum” period between May 2013 and December 2013, a short period during which the yield curve had steepened sharply on expectations for the Fed to reduce purchases of long-term bonds. The taper tantrum ended in December 2013, when the Fed announced its QE tapering plan. During that period, the UST 3m10y slope had moved up sharply, but all the shadow rates had moved down even more sharply. It appears that the estimated shadow rates interpreted the Fed’s tapering intention as dovish rather than hawkish.

It appears appropriate to summarize this section using a comment of Krippner (2017): “One response to the sensitivity of shadow rate estimates and their associated applications is to avoid using them altogether, necessitating an alternative proxy for unconventional monetary policy accommodation.”
3.6 Model 2: finding a proxy for the shadow rate at the ZLB

In search for a proxy for \( s_t \), I keep the following three criteria in mind:

1. It must assume the role of yield curve level driver at all times.
2. It must be able to explain the effect of the Fed’s forward guidance at the ZLB.
3. It would be better to be an observable variable or a combination of observables than a latent variable to avoid or mitigate the issues of model mis-specification and sensitivity to parameter values.

In Fu (2017), I show that the UST three-month yield can do a good job in driving the level of the yield curve. Hence, we can keep using the three-month yield as \( s_t \) for the sample period prior to the ZLB starting date, which was December 16, 2008 when the Fed slashed policy rate from 1% to a range of 0-0.25%. During the period of December 2008 to December 2015, however, the three-month yield had been stuck in a range of 0-0.25% and thus could not capture the effect of forward guidance. Hence, we need to either find a new variable to substitute the three-month yield or adjust the three-month yield for this period.

3.6.1 Using PCA-weighted variations of the one- and two-year yields as the proxy

Recall Zhang (2017)’s finding that forward guidance has the largest effect on the yield with a tenor of 16 months, and also consider that the Fed’s three out of five forward guidance announcements promised to hold the ZLB for one and a half years to three years. It seems that the one-, two- or three-year yield or a combination of them can be a good proxy for \( s_t \) during the ZLB period. After all, unlike the three-month yield, which touched zero during the ZLB period, these three yields had only seen their lowest levels of 0.08%, 0.16% and 0.3%, respectively. It implies that these three yields had not been constrained by the ZLB. Furthermore, the estimated
in my SR-NK-ATS model is most correlated with the one-year yield. As such, I choose a combination of one- and two-year yields as the proxy. I construct the proxy as follows:

(1) Extract the first principal component (PC) of the monthly variations of the one- and two-year yields (with an eigenvector of [0.675 0.738]’) and obtain the three-month yield monthly variation’s OLS regression coefficient (0.481) with respect to the first PC using the data from January 1985 to November 2008.

(2) Calculate the averages of the three-month, one- and two-year yields in November 2008 as their pre-ZLB level (in fact, the three-month yield already dropped to below 25 basis points on November 13, 2008 in anticipation of an aggressive cut by the Fed).

(3) Subtract the levels of the one- and two-year yields between December 2008 and December 2015 from their pre-ZLB levels to obtain their deviations during the ZLB period and then apply the PCA weights of [0.675 0.738]’ to the two deviation series.

(4) Apply this PCA-weighted deviation series (multiplied by 0.481) to the three-month yield for the ZLB period to obtain an adjusted three-month yield (henceforth “Adj 3m yield”).

This is in essence similar to Wu and Xia (2016)’s and Wu and Zhang (2016)’s approach, which fuse the EFFR during normal times and the Wu-Xia shadow rate at the ZLB together.

A comparison of the Adj 3m yield with other short-tenor yields and the Wu-Xia shadow rate is depicted in Figure 3.8. We can see that the Adj 3m yield had stayed below zero during most of the ZLB period, and remained below -0.3% for three years with a minimum level of -0.38% reached in September 2011. Such negative levels are in line with other countries’ negative policy rates, e.g., the deposit rate of -0.4% for the European Central Bank (ECB), -0.75% in Switzerland, -0.35% in Sweden, etc. It further confirms that the Wu-Xia shadow rate, which reached as negative as -3% (though it could be positive during almost the entire ZLB period if
the lower bound parameter is set to zero), captures the variations of short-tenor and long-tenor rates. It also echoes the need to decompose the effects of forward guidance and QE in an NK modeling framework. It is hard to imagine how it makes sense to plug in a -3.3% interest rate into the Fed’s reaction function in the real world, at least for now. If the Fed’s policy rate is between -1% and 0%, most, if not all, of the interest rates for the private sector still will be above zero. If the policy rate is -3%, however, many types of interest rates for the private sector, e.g., the interest rate for home equity line of credit, may become negative. This would be a highly challenging adventure for the Fed and for commercial banks, at least for now.

Figure 3.8. Adjusted 3m yield vs. the Wu-Xia shadow rate and actual 3m, 1y, 2y yields (%)

We can use the Taylor rule to investigate how negative the policy rate could go during the ZLB period. Let us run two linear regression models. The first one regresses the three-month
yield on inflation and output gap prior to the ZLB (from the first quarter of 1985 to the third quarter of 2008) and then uses the estimated coefficients to forecast the three-month yield during the ZLB period. The second model adds the UST 3m10y slope to the explanatory variable list. Let us compare the three-month yield’s fitted and forecast values in the first and second models. As shown in Figure 3.9, the second model fits better to the in-sample data than the first model with a smaller residual standard error. The first model, i.e., the standard Taylor rule, predicts that the three-month yield could reach as negative as -1.63% during the ZLB period, while the UST 3m10y slope-augmented Taylor rule predicts the three-month yield could only go as low as 0.05%. This implies that the declining UST 3m10y slope during the ZLB period served to add stimulus to the economy and thus reduced the need for a very negative policy rate.

Figure 3.9. UST 3m yield’s fitted and forecast values in 2 Taylor rules (1Q 1985 to 3Q 2008)

Note that in general, both of the two Taylor rules did not explain the in-sample variations very well, i.e., they both underestimated the three-month yield between 1985 and 1990 and
significantly overestimated the three-month yield between 2002 and 2007. This could be because the Fed’s tradeoff between the volatility of inflation and output gap has varied over time. I reckon that shrinking the in-sample period may reduce the variability of the Fed’s preference. I then fit the above two linear regression models to an in-sample period between the first quarter of 1997 and the third quarter of 2008. The results are depicted in Figure 3.10. We can see that both of the two models fit the shorter in-sample data much better and the second model still outperforms the first one. For the ZLB period, the standard Taylor rule predicts the three-month yield to reach as negative as -2.21%, while the UST 3m10y slope-augmented Taylor rule predicts a minimum of -0.69%, which is now in line with the Adj 3m yield and the negative policy rates of other countries. This again shows that the declining UST 3m10y slope during the ZLB period has reduced the need for a very negative policy rate and echoes the necessity of decomposing the effects of forward guidance and QE in an NK modeling framework.

Figure 3.10. UST 3m yield’s fitted and forecast values in 2 Taylor rules (1Q 1997 to 3Q 2008)
Besides having negative values in line with the negative policy rates of other countries, the Adj 3m yield is also seen to have been barely impacted by the taper tantrum (see Figure 3.8). The reason, I believe, was that the short-term yields were anchored by forward guidance (recall from Table 3.1 that on September 13, 2012, the Fed said it expected to keep rates unchanged “at least through mid-2015.”) Note that the two-year yield, whose maturity was slightly beyond mid-2015, increased more than the one-year yield and thus also increased more than the Adj 3m yield during the taper tantrum.

To examine whether the Adj 3m yield can serve the role of the yield curve level driver, I run OLS regression of all the yields but the three-month and 10-year tenors (note that the adjusted six-month yield tightly tracks the original series during the ZLB period and thus no adjustment for the six-month yield) on the Adj 3m yield and UST 3m10y slope. As can be seen in Table 3.3, yields’ coefficients with respect to the Adj 3m yield all are close to one. Furthermore, yields become more sensitive to the UST 3m10y slope as tenor increases, which is intuitive.

Table 3.3. OLS regression results of yields on Adj 3m yield and 3m10y slope

<table>
<thead>
<tr>
<th>Yield tenor</th>
<th>$s_\tau$</th>
<th>$q_\tau$</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-month</td>
<td>1.029***</td>
<td>0.049***</td>
<td>0.998</td>
</tr>
<tr>
<td>1-year</td>
<td>1.054***</td>
<td>0.172***</td>
<td>0.994</td>
</tr>
<tr>
<td>2-year</td>
<td>1.103***</td>
<td>0.399***</td>
<td>0.991</td>
</tr>
<tr>
<td>3-year</td>
<td>1.108***</td>
<td>0.565***</td>
<td>0.991</td>
</tr>
<tr>
<td>4-year</td>
<td>1.099***</td>
<td>0.688***</td>
<td>0.993</td>
</tr>
<tr>
<td>5-year</td>
<td>1.083***</td>
<td>0.781***</td>
<td>0.995</td>
</tr>
<tr>
<td>6-year</td>
<td>1.066***</td>
<td>0.851***</td>
<td>0.997</td>
</tr>
<tr>
<td>7-year</td>
<td>1.048***</td>
<td>0.904***</td>
<td>0.998</td>
</tr>
<tr>
<td>8-year</td>
<td>1.033***</td>
<td>0.945***</td>
<td>0.999</td>
</tr>
<tr>
<td>9-year</td>
<td>1.014***</td>
<td>0.976***</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Note: * means significant at 5%, ** at 1% and *** at 0.1%
At this point, it appears that the Adj 3m yield meet the three selection criteria mentioned at the beginning of Section 3.6. There may be a potential concern that if the Fed’s monetary policy hits the ZLB again in the future, whether this type of proxy will still be valid for $s_t$. My answer is that this concern may be even unnecessary. The reasoning is: after so many central banks have entertained exercises of negative policy rates and all of them have declared success (though I believe they have just succeeded in kicking the can, i.e., next crisis, down the road and have actually caused lots of side problems such as ever increasing wealth inequality and thus more and more acceptance of socialism in the Western countries), the Fed may well be ready to take this adventure in the future. Indeed, on March 6, 2019, the Federal Reserve Bank of New York President John Williams said the Fed could consider negative policy rate in the event of an economic downturn in the future.

Having selected the Adj 3m yield as the proxy for $s_t$, I proceed to use the UST 3m10y slope as the proxy for $q_t$ (as is done by Fu 2017) to construct Model 2. Note that the UST 3m10y slope series needs to be recalculated as the difference between the 10-year yield and the Adj 3m yield.

### 3.6.2 Estimation of Model 2

Since the two latent variables $s_t$ and $q_t$ have been replaced by the Adj 3m yield and UST 3m10y slope, Model 2 becomes much simpler and more parsimonious than the SR-NK-ATS model (Model 1) presented in Section 3.3. In fact, the ATS part is no longer needed because Model 2 can be directly fitted to the sample data of $X_t$. Thus, Model 2 is simplified into an NK model because $s_t$ is now the adjusted $i_t$ and the ZLB no long binds. This is similar to Fu (2017), in which Model 1 is an NK-ATS model while Model 2 is just an NK model. This Model 2 retains the two-way bridge between the economy and the yield curve, a bridge initially built by Model 2.
of Fu (2017). I apply the structural VAR (SVAR) estimation procedure employed by Fu (2018), which includes the following three steps:

**Step 1**: Fit the following VAR(1) model to the demeaned data of $X_t$ using OLS:

$$X_t = \Omega X_{t-1} + e_t.$$  \hspace{1cm} (3.40)

Obtain the estimated coefficient matrix $\hat{\Omega}_{OLS}$:

$$\hat{\Omega}_{OLS} = \begin{bmatrix} 0.7673 & 0.0044 & 0.0610 & 0.0502 \\ -0.0795 & 0.9734 & 0.0199 & 0.0554 \\ -0.0122 & 0.1272 & 0.9712 & 0.1309 \\ -0.0138 & -0.1409 & 0.0026 & 0.7901 \end{bmatrix}$$

**Step 2**: correct the small-sample bias in $\hat{\Omega}_{OLS}$ (see, e.g., Bauer 2012) using the method of Engsted and Pedersen (2014) to arrive at $\hat{\Omega}_{OLS}^c$:

$$\hat{\Omega}_{OLS}^c = \begin{bmatrix} 0.8117 & 0.0016 & 0.0501 & 0.0408 \\ -0.0724 & 0.9959 & 0.0194 & 0.0500 \\ -0.0083 & 0.1209 & 0.9767 & 0.1193 \\ -0.0148 & -0.1279 & 0.0056 & 0.8207 \end{bmatrix}$$

and the bias-corrected reduced-form residuals $\hat{e}_{OLS,t}^c$. Note that $\hat{\Omega}_{OLS}^c$ is only slightly different (less mean-reverting) than $\hat{\Omega}_{OLS}$ because the sample size of 128 is not too small.

**Step 3**: Use $\hat{\Omega}_{OLS}^c$ and $\hat{e}_{OLS,t}^c$ to recover the NK structural parameter vector $[\beta \kappa \alpha \psi \theta \rho \delta_{1} \delta_{2} \delta_{3} \delta_{4} \delta_{5} \phi_{r} \phi_{y} \phi_{q}]'$. This is done by a full information maximum likelihood (FIML) method. The log-likelihood function is:

$$L = -2(T - 1)\log 2\pi - \frac{T-1}{2} \log |\Gamma \Gamma'| - \frac{1}{2} \sum_{t=2}^{T} e_t' D^{-1} e_t,$$  \hspace{1cm} (3.41)

where $e_t$ is the structural residual term as defined in Equation (3.3), and is related to the reduced-form residual term $e_t$ as follows:
\[ \epsilon_t = \Gamma^{-1} e_t. \]

\( \epsilon_t \) also is in the following SVAR representation re-written from Equation (3.34) with \( \mu = 0 \):

\[ (B - A\Omega)X_t = MX_{t-1} + \epsilon_t. \]

The NK structural parameters are contained in the matrices \( B, A \) and \( M \).

Let \( V \) be the covariance matrix of \( e_t \) already estimated by OLS. The covariance matrix of \( \epsilon_t \), \( D \), can be obtained as:

\[ D = \Gamma V \Gamma^{-1}. \]

The FIML method maximizes the log-likelihoods of \( \epsilon_t \).

Table 3.4 compares the structural parameter estimates of Fu (2018) by the grid search method with the SVAR estimates just obtained. The two sets of parameter values are similar, considering that the two data sets are slightly different (\( s_t \) and \( q_t \) are adjusted in this paper) and that the grid search optimizes each equation of the NK system while the SVAR method in this study optimizes the entire NK system.

Table 3.4. Structural parameter estimates by SVAR and by Fu (2018)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fu (2018) grid search</th>
<th>SVAR in this paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>0.551</td>
<td>0.568</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.050</td>
<td>0.039</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.995</td>
<td>0.921</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.001</td>
<td>0.016</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.921</td>
<td>0.908</td>
</tr>
<tr>
<td>( \delta_{\pi} )</td>
<td>1.168</td>
<td>1.467</td>
</tr>
<tr>
<td>( \delta_{y} )</td>
<td>1.626</td>
<td>1.883</td>
</tr>
<tr>
<td>( \delta_{q} )</td>
<td>1.507</td>
<td>1.792</td>
</tr>
<tr>
<td>( \phi_{r} )</td>
<td>0.010</td>
<td>0.051</td>
</tr>
<tr>
<td>( \phi_{y} )</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>( \phi_{q} )</td>
<td>0.525</td>
<td>0.452</td>
</tr>
</tbody>
</table>
3.7 Computing the 10-year term premium and assessing its impact on growth

3.7.1 Computing the 10-year term premium using Model 2

An $n$-period term premium is still calculated by Equation (3.37). Equation (3.35) for the one-period forward premium still holds, though $E_t^Q[i_{t+n}]$ is just $E_t[i_{t+n}]$ since Model 2 no long uses the ATS theory. And $E_t[i_{t+n}]$ is obtained as:

$$E_t[i_{t+n}] = [0 \ 0 \ 1 \ 0] (\hat{\beta}_{OLS}^c)^n X_t.$$  \hspace{1cm} (3.42)

Forward yields are calculated from the OLS regression-fitted yield curve except for the Adj 3m yield and the 10-year yield, for which the observed yields are used.

Figure 3.11 exhibits the calculated 10-year term premium and the 10-year average expected short rate as well as their comparison during the ZLB period to those calculated by Fu (2018). Note that Fu (2018) also uses a bias-corrected OLS-estimated VAR(1) coefficient matrix, but it uses the 3m yield rather than the Adj 3m yield. We can see that the 10-year average expected short rate from Fu (2018) had notably higher values during the ZLB period than the one calculated using the Adj 3m yield (the bottom panel of Figure 3.11). The average difference at the ZLB is 21 basis points. Both Fu (2018) and this study reveal that the Fed’s unconventional monetary policy tools had succeeded in bringing down interest rates and term premia across the term structure. But the top panel of Figure 3.11 also shows that the 10-year term premium (generated in this study) since the 2008-2009 financial crisis has not reached as low as it had during the 2004-2006 “Great Moderation” period (though the two lows are rather close). It was the declining 10-year average expected short rate that has mainly contributed to drag down the 10-year yield since the Great Moderation. Note that some other models, e.g., Adrian et al. (2013), produce negative values for the 10-year term premium for some days over the past
couple of years. It is worth exploring the differences between those models and my model in future research.

Figure 3.11. Computed 10-year term premium and average expected short rate (\%) 

3.7.2 Assessing whether the 10-year term premium can predict economic growth

Whether term premia can predict economic growth has been a popular topic in many previous studies. Hamilton and Kim (2002), for example, decompose the UST 3m10y slope into the expectation component and term premium component using an instrumental variable approach and then use the two components to forecast future economic growth. They use the following linear regression model:
\[ y_{t+4} - y_t = \beta_0 + \beta_1(y_t - y_{t-4}) + \beta_2 E_t[i_t^{10y}] + \beta_3 t_p^{10y} + \epsilon_t, \]  

(3.43)

where \( y_t \) is log output and thus \((y_t - y_{t-4})\) is this quarter’s economic growth from one year ago, \( E_t[i_t^{10y}] \) is the 10-year average expected short rate and \( t_p^{10y} \) is the 10-year term premium. The authors find that \( \beta_1 \) and \( \beta_2 \) are statistically significant, though both of them are positive, which are counter-intuitive since declining term premium is supposed to be associated with faster future growth.

Rudebusch et al. (2007) modify Equation (3.43) by differencing \( E_t[i_t^{10y}] \) and \( t_p^{10y} \) with a four-quarter lag. They argue that since in the IS curve, output gap is a function of real interest rate, output growth should be a function of interest rate difference and of term premium difference. Hence, Equation (3.43) becomes:

\[ L^4 y_{t+4} = \beta_0 + \beta_1 L^4 y_t + \beta_2 L^4 E_t[i_t^{10y}] + \beta_3 L^4 t_p^{10y} + \epsilon_t, \]  

(3.44)

where \( L^4 \) is the lag operator with a lag of four quarters such that \( L^4 y_{t+4} = y_{t+4} - y_t \). They use the \( E_t[i_t^{10y}] \) and \( t_p^{10y} \) data produced by Kim and Wright (2005)’s three-factor ATS model. They use two sample periods – 1962 to 2005 and 1985 to 2005. The second sample period is a subset of my dataset (mine is from 1985 to 2016). For that sample period, they obtain \( \hat{\beta}_1 = 0.36 \) with a \( t \)-statistic of 2.68, \( \hat{\beta}_2 = 0.3 \) with a \( t \)-statistic of 1.37 and \( \hat{\beta}_3 = -0.59 \) with a \( t \)-statistic of -1.93. In their model setup, therefore, both the 10-year average expected short rate and term premium show intuitive effects on economic growth, though \( \hat{\beta}_2 \) and \( \hat{\beta}_3 \) are barely statistically significant while the coefficient of the lagged growth is statistically significant.

I proceed to use Rudebusch et al. (2007)’s approach to assess whether the 10-year term premium produced by my model can predict economic growth. Before applying Equation (3.44), I examine the four-quarter forecasting horizon’s validity by calculating output growth’s auto-
correlation and its cross-correlation coefficients with the four-quarter differences in $E_t[t_{10y}]$ and in $tp_t^{10y}$.

Table 3.5. Output growth’s cross-correlation with itself, differenced $E_t[t_{10y}]$ and $tp_t^{10y}$

<table>
<thead>
<tr>
<th></th>
<th>t-4</th>
<th>t-3</th>
<th>t-2</th>
<th>t-1</th>
<th>t</th>
<th>t+1</th>
<th>t+2</th>
<th>t+3</th>
<th>t+4</th>
</tr>
</thead>
<tbody>
<tr>
<td>10y short rate</td>
<td>0.321</td>
<td>0.396</td>
<td>0.457</td>
<td>0.505</td>
<td>0.504</td>
<td>0.457</td>
<td>0.393</td>
<td>0.299</td>
<td>0.194</td>
</tr>
<tr>
<td>10y term premium</td>
<td>-0.285</td>
<td>-0.362</td>
<td>-0.391</td>
<td>-0.377</td>
<td>-0.341</td>
<td>-0.297</td>
<td>-0.265</td>
<td>-0.228</td>
<td>-0.185</td>
</tr>
<tr>
<td>Output growth</td>
<td>0.338</td>
<td>0.523</td>
<td>0.726</td>
<td>0.896</td>
<td>1</td>
<td>0.896</td>
<td>0.726</td>
<td>0.523</td>
<td>0.338</td>
</tr>
</tbody>
</table>

Table 3.5 actually shows that $L^4y_t$ leads $L^4E_t[t_{10y}]$ by one quarter (though the cross-correlations at $t$ and $t-1$ are almost the same), and $L^4tp_t^{10y}$ by two quarters. It seems that it is the output growth that can predict the other two variables, especially the term premium. Also, output growth’s auto-correlation with a lag of four quarters is only 0.338 vs. 0.896 for a lag of one quarter, but 0.338 is still higher than the cross-correlations of $L^4y_{t+4}$ with $L^4E_t[t_{10y}]$ and with $L^4tp_t^{10y}$ (only 0.194 and -0.185, respectively). These imply: 1) using $L^4y_{t+1}$ rather than $L^4y_{t+4}$ as the dependent variable may be more appropriate; and 2) on the right-hand side of the regression, the effect of the lagged output growth may dominate those of the 10-year short rate and term premium.

The results, as shown in Table 3.6, indeed confirm with the above implications. When $L^4y_t$ is taken as one predictor, $L^4E_t[t_{10y}]$ and $L^4tp_t^{10y}$ are statistically insignificant (p-values in parentheses are large) whether $L^4y_{t+4}$ or $L^4y_{t+1}$ is the response. Furthermore, $\hat{\beta}_2$ is even negative when $L^4y_{t+4}$ is the response. When $L^4y_t$ is not an explanatory variable and $L^4y_{t+1}$ is the response, $\hat{\beta}_2$ is highly positive and significant, which seems to indicate that $L^4E_t[t_{10y}]$
assumes the predictive power of $L^4 y_t$. Lastly, in three out of the four regressions, $\hat{\beta}_3$ is negative. It is positive but small (0.028) when $L^4 y_{t+1}$ is the response. This probably should not be interpreted too much as all the predictive power seems to be dominated by $L^4 y_t$ ($\hat{\beta}_1 = 0.897$).

Table 3.6. Regression results of output growth on differences in short rate and term premium

<table>
<thead>
<tr>
<th>Response</th>
<th>$L^4 y_t$</th>
<th>$L^4 E_t[\bar{i}_{t}^{10y}]$</th>
<th>$L^4 t_{p_{t}^{10y}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L^4 y_{t+4}$</td>
<td>0.321 (0.002)</td>
<td>-0.031 (0.887)</td>
<td>-0.15 (0.436)</td>
</tr>
<tr>
<td>$L^4 y_{t+4}$</td>
<td>0.252 (0.219)</td>
<td>-0.209 (0.295)</td>
<td></td>
</tr>
<tr>
<td>$L^4 y_{t+1}$</td>
<td>0.897 (0.000)</td>
<td>0.023 (0.82)</td>
<td>0.028 (0.752)</td>
</tr>
<tr>
<td>$L^4 y_{t+1}$</td>
<td></td>
<td>0.803 (0.000)</td>
<td>-0.119 (0.506)</td>
</tr>
</tbody>
</table>

As such, it seems safe to draw the following conclusions: 1) $L^4 E_t[\bar{i}_{t}^{10y}]$ has a higher predictive power for output growth than $L^4 t_{p_{t}^{10y}}$, though both become meaningless when $L^4 y_t$ is a predictor; and 2) $L^4 E_t[\bar{i}_{t}^{10y}]$ and $L^4 t_{p_{t}^{10y}}$ both have intuitive (positive and negative) effects on output growth. This can be seen from Figure 3.12, which confirms with Table 3.5 to show output growth’s positive and negative correlations with the 10-year average expected short rate and term premium; and 3) there should be many other variables that can better serve as predictors of economic growth than term premia, e.g., the ISM manufacturing and non-manufacturing indices, weekly initial unemployment claims, etc.. That said, term premia have their importance. For example, they can be used as an indicator on whether investors are too risk-taking or risk-averse. When term premia are close to zero or even negative, policy makers may need to consider preventing too many future financial risks from building. On the other hand, when term premia are high while short-term rates are low, it may indicate private agents are risk-averse. This was
the case in mid-2010. Policy makers launched QE2, which helped depress term premia at the long end of the term structure.

Figure 3.12. GDP growth, differenced 10-year term premium & average expected short rate (%)

3.8 Conclusions and future research directions

In this paper, I present two unique shadow rate NK models that address the ZLB constraint. The first model uses the shadow rate and a financial risk indicator to capture the effects of the Fed’s forward guidance and QE separately. It then links the NK system to observable forward yields using the affine term structure theory. This is different than the Wu-Xia shadow rate model and the Wu-Zhang shadow rate NK model, which treat the shadow rate as to explain the joint effect of forward guidance and QE and thus create a maturity mismatch between the short rate and shadow rate. My second model adjusts the short rate during the ZLB period using the PCA-weighted variations of the one- and two-year yields and replaces the latent shadow rate
with the adjusted short rate. My Model 2 retains the two-way bridge between the economy and the yield curve, a bridge initially built by Fu (2017). Model 2 also is more parsimonious than Model 1 and can produce more robust term premium estimates.

I envision two venues for future research. First, an empirical comparison of the term premia generated by my model with those generated by other models, e.g., Adrian et al. (2013), Kim and Wright (2005), etc., can be explored. Second, it may be worth applying my model to other countries that have adopted negative policy rates.
Appendices

Appendix 1.A  The derivation of the extended NK model

1.A.1  Consumer’s utility maximization

My proposed extended NK model is in essence similar to Nisticò (2012), which uses an overlapping generation NK model to study stock prices’ wealth effect on the real economy and monetary policy. In Nisticò (2012), there are an infinite number of cohorts of consumers with a probability $\gamma$ of dying each period. The cohort of consumers born in period $j$ maximize their lifetime utility as below:

$$E_0 \sum_{t=0}^{\infty} \beta^t (1 - \gamma)^t \left[ \nu_t \ln C_{j,t} + \eta_t \ln \left( 1 - N_{j,t} \right) \right],$$  \hspace{1cm} (1.A.1)

where $\nu_t$ and $\eta_t$ are time-varying shocks to consumption and labor. The $j$-cohort’s budget constraints are:

$$P_t C_{j,t} + E_t \left[ F_{t,t+1} B_{j,t+1} \right] + P_t \int_0^{1} S_t(i) Z_{j,t+1}(i) di \leq W_t N_{j,t} - P_t T_{j,t} + \omega_{j,t}.$$  \hspace{1cm} (1.A.2)

where $F_{t,t+1} = 1/(1+i_t)$ is the discount factor, $B_{j,t+1}$ is the risk-free bond holdings, $T_{j,t}$ is the transfer, $Z_{j,t+1}(i)$ is the equity shares of the $i$th intermediate goods-producing firm and $S_t(i)$ is the stock price. $\omega_{j,t}$ is the $j$-cohort’s nominal financial wealth carried over from the previous period:

$$\omega_{j,t} = \frac{1}{1-\gamma} \left\{ B_{j,t} + P_t \int_0^{1} [S_t(i) + D_t(i)] Z_{j,t}(i) di \right\},$$  \hspace{1cm} (1.A.3)

where $D_t(i)$ is the $i$th firm’s dividend.

Equation (1.A.2) can be rewritten as:

$$P_t C_{j,t} + E_t \left[ F_{t,t+1} (1 - \gamma) \omega_{j,t+1} \right] = W_t N_{j,t} - P_t T_{j,t} + \omega_{j,t}.$$  \hspace{1cm} (1.A.4)
Define the human wealth for Cohort $j$ as:

$$h_{j,t} = E_t \left[ \sum_{k=0}^{\infty} F_{t,t+k} (1 - \gamma)^k (W_{t+k} N_{j,t+k} - P_{t+k} T_{j,t+k}) \right].$$  \hfill (1.A.5)

Solving Equation (1.A.4) forward to obtain:

$$\omega_{j,t} = E_t \left[ \sum_{k=0}^{\infty} F_{t,t+k} (1 - \gamma)^k P_{t+k} C_{j,t+k} \right] - h_{j,t}.$$ \hfill (1.A.6)

1.A.2 The IS curve and the stock pricing equation

The first-order condition for consumption is:

$$P_{t+1} C_{j,t+1} = \frac{\beta}{f_{t,t+1}} \exp(\nu_{t+1} - \nu_t) P_t C_{j,t}.$$ \hfill (1.A.7)

And the first-order condition for the $i$th stock price is:

$$P_t S_t(i) = E_t \left[ F_{t,t+1} P_{t,t+1} (S_{t+1}(i) + D_{t+1}(i)) \right].$$  \hfill (1.A.8)

Plugging Equation (1.A.7) into Equation (1.A.6), we can see that Cohort $j$’s current consumption is a linear function of its nominal financial and human wealth (see Appendix A of Nisticò 2005):

$$P_t C_{j,t} = \frac{1}{\Sigma_t} (\omega_{j,t} + h_{j,t}),$$ \hfill (1.A.9)

where $\Sigma_t = E_t \sum_{k=0}^{\infty} [\beta (1 - \gamma)]^k \exp(\nu_{t+k} - \nu_t)$, a complicated parameter that collapses into a constant in steady state.

The aggregate value of a state variable $x$ is the weighted average of the cohort-specific counterparts:

$$x_t = \sum_{j=-\infty}^{\infty} (1 - \gamma)^{t-j} x_{j,t}.$$  \hfill (1.A.10)

Aggregating across cohorts, the economy’s current aggregate consumption is also a weighted average of its aggregate financial and human wealth:

$$P_t C_t = \frac{1}{\Sigma_t} (\omega_t + h_t).$$ \hfill (1.A.11)
Aggregating Equation (1.A.4) leads to:

\[ P_t C_t + E_t[F_{t,t+1} \omega_{t+1}] = W_t N_t - P_t T_t + \omega_t. \]  \hspace{1cm} (1.A.12)

Plugging (1.A.12) into (1.A.11) to replace \( \omega_t \):

\[ P_t C_t = \frac{1}{\Sigma_t} \{ P_t C_t + E_t[F_{t,t+1} \omega_{t+1}] + h_t - (W_t N_t - P_t T_t) \}. \]  \hspace{1cm} (1.A.13)

And the aggregate financial and human wealth can be represented as:

\[ \omega_t = B_t + P_t \int_0^1 [S_t(i) + D_t(i)] Z_t(i) di, \]  \hspace{1cm} (1.A.14)

\[ h_t = E_t[\sum_{k=0}^{\infty} F_{t,t+k}(1 - \gamma)^k(W_{t+k} N_{t+k} - P_{t+k} T_{t+k})] \]
\[ = W_t N_t - P_t T_t + E_t[F_{t,t+1}(1 - \gamma) h_{t+1}]. \]  \hspace{1cm} (1.A.15)

Forwarding Equation (1.A.11) by one period and multiplying both sides by \( \Sigma_{t+1} F_{t,t+1}(1 - \gamma) \):

\[ \Sigma_{t+1} F_{t,t+1}(1 - \gamma) P_{t+1} C_{t+1} = F_{t,t+1}(1 - \gamma)(\omega_{t+1} + h_{t+1}). \]  \hspace{1cm} (1.A.16)

Taking conditional expectations of and rearranging Equation (1.A.16):

\[ E_t[F_{t,t+1}(1 - \gamma) h_{t+1}] = E_t[\Sigma_{t+1} F_{t,t+1}(1 - \gamma) P_{t+1} C_{t+1}] - E_t[F_{t,t+1}(1 - \gamma) \omega_{t+1}]. \]  \hspace{1cm} (1.A.17)

Plugging (1.A.15) into (1.A.17):

\[ h_t - (W_t N_t - P_t T_t) = E_t[\Sigma_{t+1} F_{t,t+1}(1 - \gamma) P_{t+1} C_{t+1}] - E_t[F_{t,t+1}(1 - \gamma) \omega_{t+1}]. \]  \hspace{1cm} (1.A.18)

Now, plugging (1.A.18) into (1.A.13) to have consumption at time \( t \) represented as the weighted average of the financial wealth and consumption at time \( t+1 \):

\[ P_t C_t = \frac{1}{\Sigma_t} \{ \gamma E_t[F_{t,t+1} \omega_{t+1}] + (1 - \gamma) E_t[F_{t,t+1} \Sigma_{t+1} P_{t+1} C_{t+1}] \}. \]  \hspace{1cm} (1.A.19)

Since \( E_t[F_{t,t+1} \omega_{t+1}] = P_t S_t \), current consumption is a function of current aggregate stock market price and future consumption:

\[ C_t = \frac{1}{\Sigma_t} \{ \gamma S_t + (1 - \gamma) E_t[\Sigma_{t+1} \frac{1+\pi_{t+1}}{1+\pi_t} C_{t+1}] \}. \]  \hspace{1cm} (1.A.20)
We can see that if \( \gamma = 0 \), i.e., if consumers live indefinitely, future financial wealth and thus current stock price will not impact current consumption. In this case, the Nisticò model converges to the standard NK model.

Equation (1.A.8) for the stock market price can be rewritten as:

\[
S_t = E_t \left[ \frac{1 + \pi_{t+1}}{1 + i_t} (S_{t+1} + D_{t+1}) \right]. \tag{1.A.21}
\]

Log-linearizing Equations (1.A.20) and (1.A.21), we have the following New Keynesian-style IS equation and stock pricing equation:

\[
c_t = (1 - \gamma) \Sigma^{ss} [E_t c_{t+1} - (E_t \pi_{t+1})] + \frac{\gamma}{\Sigma^{ss-1} C^{ss}} S_t + \epsilon_{c,t} \\
= \alpha E_t c_{t+1} - \alpha (E_t \pi_{t+1}) + \xi s_t + \epsilon_{c,t}, \tag{1.A.22}
\]

\[
s_t = \frac{S^{ss}}{S^{ss} + D^{ss}} E_t s_{t+1} + \frac{D^{ss}}{S^{ss} + D^{ss}} E_t d_{t+1} - (E_t \pi_{t+1}) + \epsilon_{s,t} \\
= \phi_s E_t s_{t+1} + (1 - \phi_s) E_t d_{t+1} - (E_t \pi_{t+1}) + \epsilon_{s,t}, \tag{1.A.23}
\]

where the lower-case letters are their upper-case counterparts’ log derivations from steady-state values and \( \Sigma^{ss}, S^{ss}, C^{ss}, D^{ss} \) are the steady-state values of \( \Sigma, S, C \) and \( D \), respectively.

Replacing \( c_t \) with \( y_t \) and assuming \( d_t \) is a proportion of \( y_t \), we have:

\[
y_t = \alpha E_t y_{t+1} - \alpha (E_t \pi_{t+1}) + \xi s_t + \epsilon_{y,t}, \tag{1.A.24}
\]

\[
s_t = \zeta_s E_t s_{t+1} + \zeta_y E_t y_{t+1} - (E_t \pi_{t+1}) + \epsilon_{s,t}. \tag{1.A.25}
\]

Note that \( S_t \) is the stock market price in Nisticò (2012) while my proposed model calls for the FRI \( Q_t \). Let us suppose \( S_t \) is a representative risky asset instead of the stock market index. It should be reasonable to assume that the FRI \( q_t = -\rho_s s_t \), which means when the risky asset price falls, the FRI rises proportionally. Consequently, Equations (1.A.24) and (1.A.25) become:
\[ y_t = \alpha E_t y_{t+1} - \alpha(i_t - E_t \pi_{t+1}) - \theta q_t + \epsilon_{y,t}, \quad (1.A.26) \]

\[ q_t = \phi_q E_t q_{t+1} - \phi_y E_t y_{t+1} + \phi_r (i_t - E_t \pi_{t+1}) + \epsilon_{q,t}. \quad (1.A.27) \]

### 1.A.3 Adding lagged output gap to the IS curve

At the moment, there are no lagged output gap and FRI in (1.A.26) and (1.A.27). However, as discussed earlier, historical data of the UST 3m10y slope and AAA 10y spread, which are considered good candidate components of the FRI, exhibit strong serial correlation. To add \( y_{t-1} \) into Equation (1.A.26), I incorporate consumption habit formation into the utility function, which now has a CRRA preference:

\[
E_0 \sum_{t=0}^{\infty} \beta^t (1 - \gamma)^t \left[ v_t \frac{(C_{j,t-\delta} - \delta C_{j,t-1})^{1-\sigma}}{1-\sigma} - \eta_t \frac{N_{j,t}^{1+\tau}}{1+\tau} \right], \quad (1.A.28)
\]

where \( \delta \) is a habit parameter. This is an additive habit form (see, e.g., Dennis 2009).

The first-order condition for consumption is:

\[
\left( \frac{C_{j,t+1} - \delta C_{j,t}}{C_{j,t-\delta} C_{j,t-1}} \right)^\sigma = \beta (1 + i_t) \frac{P_t}{P_{t+1}} \frac{v_{t+1}}{v_t}. \quad (1.A.29)
\]

Rewrite (1.A.29) as:

\[
\frac{(C_{j,t+1} - \delta C_{j,t}) P_{t+1}}{(C_{j,t-\delta} C_{j,t-1}) P_t} = \alpha_t, \quad (1.A.30)
\]

where \( \alpha_t = \frac{v_{t+1}}{v_t} \left[ \beta (1 + i_t) \right]^{1/\sigma} (1 + \pi_{t+1})^{1+\sigma}. \)

Expand (1.A.30) as:

\[
P_{t+1} C_{j,t+1} = \delta (1 + \pi_{t+1}) P_t C_{j,t} + \alpha_t P_t C_{j,t} - \delta \alpha_t (1 + \pi_t) P_t C_{j,t-1}. \quad (1.A.31)
\]

Letting \( a_t = \delta (1 + \pi_{t+1}) + \alpha_t \) and \( b_t = \delta \alpha_t (1 + \pi_t) \), we have:

\[
P_{t+1} C_{j,t+1} = a_t P_t C_{j,t} - b_t P_t C_{j,t-1}. \quad (1.A.32)
\]

Forwarding (1.A.32) by one period to arrive at:
\[ P_{t+2}C_{j,t+2} = a_{t+1}P_{t+1}C_{j,t+1} - b_{t+1}P_{t+1}C_{j,t} \]
\[ = [a_{t+1}a_t - b_{t+1}(1 + \pi_{t+1})]P_tC_{j,t} - a_{t+1}b_tP_tC_{j,t-1}. \]  
(1.A.33)

Continue to forward \( P_{t+k}C_{j,t+k} \) where \( k = 3, \ldots, \infty \). We can see that \( P_{t+k}C_{j,t+k} \) always can be represented as a linear function of \( P_tC_{j,t} \) and \( P_tC_{j,t-1} \). Therefore, we have the cumulative lifetime consumption also as a linear function of \( P_tC_{j,t} \) and \( P_tC_{j,t-1} \):

\[ \sum_{k=0}^{\infty} P_{t+k}C_{j,t+k} = U_tP_tC_{j,t} - V_tP_tC_{j,t-1}, \]  
(1.A.34)

where \( U_t \) and \( V_t \) are time-varying coefficients.

Plugging (1.A.34) into (1.A.6), we obtain Cohort \( j \)'s consumption at time \( t \) as a linear function of financial wealth and human wealth at time \( t \) and of consumption at time \( t-1 \):

\[ P_tC_{j,t} = \frac{1}{M_tU_t}[(\omega_{j,t} + h_{j,t} + V_tP_tC_{j,t-1}], \]  
(1.A.35)

where \( M_t = E_t[\sum_{k=0}^{\infty} F_{t,t+k} (1 - \gamma)^k] \).

We can see that (1.A.35) has an additional term involving past nominal consumption. This term is not in (1.A.9).

Therefore, aggregate consumption is:

\[ P_tC_t = \frac{1}{M_tU_t}[(\omega_t + h_t + V_tP_tC_{t-1}], \]  
(1.A.36)

Plugging (1.A.12) into (1.A.36) yields:

\[ P_tC_t = \frac{1}{M_tU_t}\{P_tC_t + E_t[F_{t,t+1}\omega_{t+1}] + h_t - (W_tN_t - P_tT_t) + V_tP_tC_{j,t-1}\}. \]  
(1.A.37)

With (1.A.36), (1.A.18) becomes:

\[ h_t - (W_tN_t - P_tT_t) = E_t[M_{t+1}U_{t+1}F_{t,t+1}(1 - \gamma)P_{t+1}C_{t+1}] \]
\[ -E_t[F_{t,t+1}(1 - \gamma)\omega_{t+1}] - E_t[F_{t,t+1}(1 - \gamma)V_{t+1}P_{t+1}C_{t+1}]. \]  
(1.A.38)

Plugging (1.A.38) into (1.A.37), we have consumption at time \( t \) represented as the weighted average of the stock market price at time \( t \), consumption at time \( t+1 \) and consumption at \( t-1 \):
\[ C_t = \frac{1}{W_t} \{ \gamma S_t + (1 - \gamma) E_t \left[ M_{t+1} U_{t+1} \left[ 1 + \frac{\pi_{t+1}}{1 + \lambda} C_{t+1} \right] \right] + V_t C_{t-1} \}, \quad (1.A.39) \]

where \( W_t = M_t U_t - 1 + (1 - \gamma) \frac{1 + \pi_{t+1}}{1 + \lambda} E_t [V_{t+1}] \).

Log-linearizing (1.A.39) to arrive at the IS equation for consumption that includes the stock market price:

\[ c_t = \alpha E_t c_{t+1} + \psi c_{t-1} - \alpha (i_t - E_t \pi_{t+1}) + \xi S_t + \epsilon_{c,t}. \quad (1.A.40) \]

Therefore, the IS equation for output gap that includes the FRI is:

\[ y_t = \alpha E_t y_{t+1} + \psi y_{t-1} - \alpha (i_t - E_t \pi_{t+1}) - \theta q_t + \epsilon_{y,t}. \quad (1.A.41) \]

As for Equation (1.A.27) for \( q_t \), using the overlapping NK setup, there seems to be no way to add \( q_{t-1} \). However, considering the FRI data’s autocorrelation, I insert \( q_{t-1} \) to arrive at Equation (4).

1.A.4 The Phillips curve

In Nisticò (2012), the NK Phillips curve is rather standard:

\[ \pi_t = \beta E_t \pi_{t+1} + \kappa mc_t, \quad (1.A.42) \]

where \( mc_t \) represents marginal cost. To add \( \pi_{t-1} \), I borrow the approach of Galí and Gertler (2000), which extends Calvo’s model (Calvo 1983) by assuming a fraction \( w \) of the firms set prices according to a backward-looking rule while the remaining fraction \( 1 - w \) of the firms follow a forward-looking rule. Galí and Gertler (2000) derive a micro-founded hybrid NK Phillips curve:

\[ \pi_t = \beta_f E_t \pi_{t+1} + \beta_b \pi_{t-1} + \kappa mc_t, \quad (1.A.43) \]

where \( \beta_f \) and \( \beta_b \) are functions of the intertemporal discount factor \( \beta \), \( w \) and the Calvo price.
setting probability $\theta$. I apply a restriction that $\beta_f + \beta_b = 1$, which is common in the literature, e.g., in Bekaert et al. (2010). As such, I obtain Equation (1.1).

## 1.A.5 The Taylor rule

In the standard NK model, the Taylor rule posits that the Fed's target rate linearly responds to output gap and inflation. Since the FRI is introduced into my proposed model, it is natural to add the FRI to the right-hand side of the Taylor rule.

$$i_t^* = \bar{i} + \delta_\pi \pi_t + \delta_y y_t - \delta_q q_t.$$  \hspace{1cm} (1.A.44)

where $i_t^*$ is an implicit policy rate target. Considering the Fed's tendency to smooth interest rate adjustments, Clarida, Galí and Gertler (1999) allow for partial adjustments to policy rate:

$$i_t = \rho i_{t-1} + (1 - \rho) i_t^* + \epsilon_{i_t}$$  \hspace{1cm} (1.A.45)

Combining Equations (1.A.44) and (1.A.45) leads to Equation (1.3).
Appendix 1.B  The derivation of the IS equation using non-separable utility preference and consumption habit formation

Assume an infinitely-lived representative consumer maximizes the following utility function:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left( \frac{v_t c_{t-1}^{\rho} c_t^{1-\sigma} s_t^{1-\gamma}}{1-\sigma} - \lambda \frac{N_t^{1+\tau}}{1+\tau} \right), \quad (1.B.1)$$

where $S_t$ is the real value of a representative risky asset, $v_t$ is a demand shock and $\delta$ is a habit parameter. This is a multiplicative habit form (see, e.g., Dennis 2009).

The consumer’s budget constraint is:

$$C_t + B_t + S_t \leq W_t N_t + \frac{1+i_{t-1}}{1+\pi_t} B_{t-1} + \frac{s_{t-1}}{1+\pi_t} + \frac{T_t}{P_t}, \quad (1.B.2)$$

where $B_t$ is the real value of the bond holding and $T_t$ is the nominal value of the transfer.

From the first-order conditions for consumption at $t$ and $t+1$ and for bond, we have:

$$\left( \frac{c_{t+1}}{c_t} \right)^{\sigma} = \beta \frac{v_{t+1}}{v_t} \left( \frac{s_{t+1}}{s_t} \right)^{1-\gamma} \left( \frac{c_t}{c_{t-1}} \right)^{\delta} \frac{1+i_t}{1+\pi_{t+1}}. \quad (1.B.3)$$

Log-linearizing Equation (1.B.3) to yield:

$$c_t = \frac{\sigma}{\sigma+\delta} E_t c_{t+1} + \frac{\delta}{\sigma+\delta} c_{t-1} - \frac{1-\gamma}{\sigma+\delta} (E_t s_{t+1} - s_t) - \frac{1}{\sigma+\delta} (i_t - E_t \pi_{t+1}) + \epsilon_{v,t}, \quad (1.B.4)$$

where $c_t$ and $s_t$ are the percentage derivations of $C_t$ and $S_t$ from their corresponding steady states.

Replacing consumption with output and letting the FRI $q_t = -\rho s_t$ in Equation (1.B.4), we obtain the following IS equation:

$$y_t = \alpha E_t y_{t+1} + (1-\alpha)y_{t-1} + \zeta (E_t q_{t+1} - q_t) - \xi (i_t - E_t \pi_{t+1}) + \epsilon_{v,t}. \quad (1.B.5)$$

Output gap at $t$ depends on its expected value at $t+1$, its past value at $t-1$, the difference between the expected one-period ahead FRI and current FRI, and real interest rate. It is hard to justify a positive effect of the FRI difference on output gap.
Appendix 3.A  The extended Kalman filter algorithm

We have the transition equation:

\[ X_t = c + \Omega X_{t-1} + \Gamma \epsilon_t, \]  \hspace{1cm} (3.A.1)

and the measurement equation:

\[
\begin{bmatrix}
\pi_t \\
y_t \\
F_t
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
X_t \\
F(X_t, n, \theta) + \eta_t
\end{bmatrix},
\]  \hspace{1cm} (3.A.2)

where \( F(X_t, n, \theta) \) is the function of \( X_t \) for an \( n \)-period ahead forward yield as defined in Equation (3.12). We need to use extended Kalman filter to estimate and structural NK parameters and filter out the latent vector of variables \( W_t = [s_t \ q_t]' \) given the observed \( F_t \) and \( M_t = [\pi_t \ y_t]' \). The algorithm is divided into the following steps:

**Step 1.** Compute the expected value and variance of \( X_t \) given the information flow up to time \( t-1, \phi_{t-1} \):

\[
E[X_t|\phi_{t-1}] = X_{t|t-1} = c + \Omega X_{t-1},
\]  \hspace{1cm} (3.A.3)

\[
Var[X_t|\phi_{t-1}] = \Sigma^X_{t|t-1} = \Omega \Sigma^X_{t-1|t-1} \Omega' + \Gamma \Gamma'.
\]  \hspace{1cm} (3.A.4)

**Step 2.** Compute the expected value and variance of \( F_t \) given \( X_{t|t-1} \) and \( \Sigma^X_{t|t-1} \):

\[
E[F_t|X_{t|t-1}] = F_{t|t-1} = F(X_{t|t-1}, n, \theta).
\]  \hspace{1cm} (3.A.5)

Define

\[
H_t = \frac{\partial}{\partial X_{t|t-1}} F(X_{t|t-1}, n, \theta).
\]  \hspace{1cm} (3.A.6)

Then, we have the variance of \( F_{t|t-1} \) as:

\[
Var[F_t|X_{t|t-1}] = \Sigma^F_{t|t-1} = H_t \Sigma^X_{t|t-1} H_t' + \Sigma^\eta,
\]  \hspace{1cm} (3.A.7)

where \( \Sigma^\eta \) is the variance-covariance matrix of the measurement error term \( \eta_t \).

Step 1 and 2 are the same as in Wu and Xia (2016).
**Step 3.** Update $X_{t|t}$ and $\Sigma^X_{t|t}$ given $f_t$. We just need to update $W_{t|t}$ and $\Sigma^W_{t|t}$ because $M_{t|t} = M_t$ and $\Sigma^M_{t|t} = 0$. First, calculate the Kalman gain $K_t$:

$$K_t = H_t \Sigma^X_{t|t-1}(\Sigma^F_{t|t-1})^{-1}. \quad (3. A. 8)$$

Let $K^W_t$ be the part of $K_t$ associated with the latent vector of variables $W_t$ and let $H^W_t$ be the part of $H_t$ associated with $W_t$. Then, $W_{t|t}$ and $\Sigma^W_{t|t}$ can be updated as:

$$W_{t|t} = W_{t|t-1} + K^W_t (F_t - F_{t|t-1}), \quad (3. A. 9)$$

$$\Sigma^W_{t|t} = (I - K^W_t H^W_t) \Sigma^W_{t|t-1}. \quad (3. A. 10)$$

where $I$ is an identify matrix.

Therefore,

$$X_{t|t} = [M_t \ W_{t|t}]', \quad (3. A. 11)$$

and

$$\Sigma^X_{t|t} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma^W_{t|t} \end{bmatrix}. \quad (3. A. 12)$$

Step 3 is different from Wu and Xia (2016) because in their model all the variables in $X_t$ are latent while in my model $X_t$ contains the observed $M_t$ and latent $W_t$.

The log-likelihood function is calculated as:

$$L(F_t|f_{t-1}) = - \frac{1}{2} \left\{ n_F \log(2\pi) + \log |\Sigma^F_{t|t-1}| + (F_t - F_{t|t-1})' (\Sigma^F_{t|t-1})^{-1} (F_t - F_{t|t-1}) \right\}, \quad (3. A. 13)$$

where $n_F$ is the number of the forward yields used in the estimation.

This completes one iteration of the extended Kalman filter and then the two updated values by Equations (3. A. 10) and (3. A. 11) are reused in Step 1 to start a new iteration for time $t+1$.

The algorithm minimizes the negative of the sum of the log-likelihood function values for all the time periods in the dataset.
Bibliography


