

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Term structure of interest rates and monetary policy

### Permalink

<https://escholarship.org/uc/item/0sp2h654>

### Author

Wu, Jing Cynthia

### Publication Date

2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Term Structure of Interest Rates and Monetary Policy**

A dissertation submitted in partial satisfaction of the  
requirements for the degree  
Doctor of Philosophy

in

Economics

by

Jing Cynthia Wu

Committee in charge:

Professor James D. Hamilton, Chair  
Professor Davide Debortoli  
Professor Ivana Komunjer  
Professor Bruce N. Lehmann  
Professor Allan Timmermann

2011

Copyright  
Jing Cynthia Wu, 2011  
All rights reserved.

The dissertation of Jing Cynthia Wu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

---

---

Chair

University of California, San Diego

2011

DEDICATION

To my beloved family and esteemed mentor

## TABLE OF CONTENTS

Signature Page . . . . .		iii
Dedication . . . . .		iv
Table of Contents . . . . .		v
List of Figures . . . . .		vii
List of Tables . . . . .		viii
Acknowledgements . . . . .		ix
Vita and Publications . . . . .		x
Abstract of the Dissertation . . . . .		xi
Chapter 1	Numerical Challenges for Affine Term Structure Models: A Solution . . . . .	1
	1.1 Introduction . . . . .	1
	1.2 The Model . . . . .	5
	1.2.1 State and Short Rate Dynamics . . . . .	5
	1.2.2 Identifying Restrictions in the Literature . . . . .	5
	1.2.3 Bond Pricing and Difference Equations . . . . .	6
	1.2.4 Conventional Estimation with MLE . . . . .	7
	1.3 New Estimation Procedure . . . . .	8
	1.4 Empirical Results . . . . .	13
	1.4.1 Illustrative Results for Simulated Data . . . . .	13
	1.4.2 Robustness to Alternative Experiment Designs . . . . .	15
	1.4.3 The Consequences of Imposing Zeros Arbitrarily . . . . .	16
	1.5 Conclusion . . . . .	17
	1.6 Acknowledgements . . . . .	18
	1.7 Appendices . . . . .	19
Chapter 2	Identifying and Testing the Refutable Implications behind the Macro Finance Term Structure Models of Interest Rates . . . . .	25
	2.1 Introduction . . . . .	25
	2.2 The Baseline Macro Finance Term Structure Model . . . . .	27
	2.2.1 Factor Dynamics . . . . .	27
	2.2.2 Short Rate Process . . . . .	29
	2.2.3 Pricing Kernel . . . . .	30
	2.2.4 Bond Pricing . . . . .	30

2.3	Testable Restrictions Implied by the Macro Finance Structure . . . . .	31
2.3.1	Deriving the VAR representation of the Baseline Macro Finance Model . . . . .	32
2.3.2	Testable Restrictions . . . . .	33
2.3.3	Target Restrictions vs. Maintained restrictions . . . . .	34
2.4	Empirical Results . . . . .	35
2.4.1	A Special Case: Ang and Piazzesi (2003) . . . . .	37
2.4.2	Generalization: Alternative Specifications . . . . .	40
2.5	Robustness Check: Which Three Yields Are Priced Without Error . . . . .	44
2.5.1	A Special Case: Ang and Piazzesi (2003) . . . . .	44
2.5.2	Generalization: Alternative Specifications . . . . .	45
2.5.3	Robustness Check with $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$ . . . . .	46
2.6	Conclusion . . . . .	46
2.7	Acknowledgements . . . . .	47
2.8	Appendices . . . . .	48
Chapter 3	The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment . . . . .	61
3.1	Introduction. . . . .	62
3.2	Options for monetary stimulus at the zero lower bound. . . . .	65
3.3	Preferred-habitat investing and market arbitrage. . . . .	69
3.4	Estimation of Affine-Term-Structure Models. . . . .	71
3.5	Data. . . . .	73
3.6	The term structure of interest rates prior to the financial crisis. . . . .	74
3.7	The term structure of interest rates at the zero lower bound. . . . .	82
3.8	Discussion. . . . .	88
3.8.1	Comparison with other estimates. . . . .	88
3.8.2	Effects on non-Treasury securities. . . . .	90
3.8.3	Application: Evaluation of QE2. . . . .	93
3.9	Conclusion. . . . .	95
3.10	Acknowledgements . . . . .	96
3.11	Appendices . . . . .	97
Bibliography	. . . . .	118

## LIST OF FIGURES

Figure 3.1: Alternative measures of 5-year expected inflation . . . . .	107
Figure 3.2: Federal Reserve assets . . . . .	108
Figure 3.3: Maturity structure of U.S. federal debt . . . . .	109
Figure 3.4: Average maturity . . . . .	110
Figure 3.5: Yields and factors . . . . .	111
Figure 3.6: Actual and model fitted interest rates . . . . .	112
Figure 3.7: Factor loadings . . . . .	113
Figure 3.8: Predicted change . . . . .	114
Figure 3.9: Assorted long-term yields . . . . .	115
Figure 3.10: Assorted long-term yields . . . . .	116
Figure 3.11: Summaries for maturity structure . . . . .	117



## LIST OF TABLES

Table 1.1: parameters for simulation . . . . .	21
Table 1.2: Examples of local maxima . . . . .	22
Table 1.3: 2 “global” maxima . . . . .	22
Table 1.4: 6 global maxima . . . . .	23
Table 1.5: Parameter comparison . . . . .	24
Table 2.1: Likelihood ratio tests . . . . .	52
Table 2.2: AIC and BIC . . . . .	53
Table 2.3: RMSE . . . . .	53
Table 2.4: likelihood ratio tests . . . . .	54
Table 2.5: AIC ranking . . . . .	54
Table 2.6: BIC ranking . . . . .	55
Table 2.7: RMSE . . . . .	55
Table 2.8: LLF ranking . . . . .	56
Table 2.9: RMSE . . . . .	57
Table 2.10: log likelihood comparison . . . . .	58
Table 2.11: RMSE . . . . .	59
Table 2.12: AIC and BIC . . . . .	60
Table 2.13: RMSE . . . . .	60
Table 3.1: Parameter estimates . . . . .	103
Table 3.2: Holding-return forecasting regressions . . . . .	104
Table 3.3: Factor vector autoregression . . . . .	105
Table 3.4: $R^2$ for post-crisis sample . . . . .	105
Table 3.5: Comparison of different estimates . . . . .	105
Table 3.6: Risky securities . . . . .	106

## ACKNOWLEDGEMENTS

I am extremely fortunate to have Professor James Hamilton as the chair of my committee, and forever indebted to him for being incredibly inspiring, encouraging, generous, patient and always having faith in me, which ensure my academic, professional and moral well being. In every sense, none of this work would have been possible without him.

I would also like to dedicate my gratitude to my other committee members, Professor Davide Debortoli, Professor Ivana Komunjer, Professor Bruce Lehmann, and Professor Allan Timmermann.

Far too many people to mention individually have assisted in so many ways during my work. In particular, I would like to thank Michael Bauer, Christiane Baumeister, Bryan Brown, John Cochrane, Greg Duffee, Gauti Eggertsson, Jeff Hallman, Monika Piazzesi, Glenn Rudebusch, Ken Singleton, Eric Swanson, Dimitri Vayanos, Kenneth West, Michael Woodford, and seminar and conference participants at the University of California, San Diego, the University of Chicago Booth School of Business, the Federal Reserve Board, Federal Reserve Bank of Boston, Federal Reserve Bank of San Francisco, Federal Reserve Bank of New York, the Michigan State University, the Pennsylvania State University, Bank of Canada, the European Central Bank, the NBER Monetary Economics program Meeting and the Federal Reserve Bank of San Francisco's annual Macro/Monetary Economics conference for helpful comments.

Chapter 3 is coauthored with James Hamilton, and in full, has been submitted for publication of material. Later versions of Chapter 1 and 2 have been submitted for publication of material, and the submitted versions are coauthored with James Hamilton. I thank Jim for the permission to use these chapters in the dissertation.

## VITA AND PUBLICATIONS

- 2005 B. A. in Economics, Renmin (People's) University of China
- 2009 M. A. in Economics, University of California, San Diego
- 2011 Ph. D. in Economics, University of California, San Diego

James D. Hamilton and **Jing Cynthia Wu**, “The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment”, *revised and resubmitted*, 2011.

James D. Hamilton and **Jing Cynthia Wu**, “Testable Implications of Affine Term Structure Models”, *revised and resubmitted*, 2011.

ABSTRACT OF THE DISSERTATION

**Term Structure of Interest Rates and Monetary Policy**

by

Jing Cynthia Wu

Doctor of Philosophy in Economics

University of California, San Diego, 2011

Professor James D. Hamilton, Chair

My dissertation solves various difficulties of Affine-Term-Structure Models (ATSM) known or unknown in the literature, by providing an explicit mapping between the reduced-form and ATSM parameters for identification analysis, developing a more reliable, better-behaved and faster estimation procedure, and coming up with a framework to examine how well those models fit the data. I also apply ATSM to assess the effectiveness of unconventional monetary policy under the zero lower bound during the current financial crisis, and develop new measures for contribution of maturity structure of Treasury debt to the term structure of interest rates.

# Chapter 1

## Numerical Challenges for Affine Term Structure Models: A Solution

### Abstract

The conventional estimation procedure for affine term structure models with MLE is known to be problematic, plagued by singularity, local maxima and flat LLF surface along several directions. To solve the numerical challenges, I propose a new estimation procedure, which solves part of the problem with OLS after re-parameterization, and breaks the parameters down into three groups, to be estimated one group at a time with a combination of numerical and analytical methods. Empirical analysis establishes that the new method offers a substantial improvement over the traditional formulation. One of the problems that the procedure helps uncover is that the canonical representations proposed in the literature are not unique; one more restriction is needed to ensure the uniqueness.

### 1.1 Introduction

The conventional estimation procedure for affine term structure models with MLE (maximum likelihood estimation) is known to be problematic due to

highly non-linear and badly-behaved likelihood surface. For example, Kim and Orphanides (2005) observe:

In particular, the likelihood function seems to have multiple inequivalent local maxima which have similar likelihood values but substantially different implications for economic quantities of interest. Furthermore, the likelihood function around these maxima can be quite flat along many directions in the parameter space; the standard errors for the parameters and quantities of economic interest are often too large to be useful.

Duffee (2002) likewise cautions:

... general three-factor affine models are already computationally difficult to estimate owing to the number of parameters.

...

The QML functions for these models have a large number of local maxima. The most important reason for this is the lack of structure placed on the feedback matrix  $K$ . Similar QML values can be produced by very different interactions among the elements of the state vector.

Ang and Piazzesi (2003) have also encountered:

... difficulties associated with estimating a model with many factors using maximum likelihood when yields are highly persistent.

...

We need to find good starting values to achieve convergence in this highly non-linear system. In particular, since unconditional means of persistent series are difficult to estimate, the likelihood surface is very flat in  $\lambda_0$  which determines the mean of long yields.

...

While our particular procedure may be path dependent, we could not find a feasible alternative which implies unconditional means for long yields close to those in the data.

And Kim (2008) also notes:

Flexibly specified no-arbitrage models tend to entail much estimation difficulty due to a large number of parameters to be estimated and due to the nonlinear relationship between the parameters and yields that necessitates a nonlinear optimization. Even if one finds the parameter vector that corresponds to the global optimum of the criterion function, not all may be fine with the resulting estimate.

To solve the estimation problem of affine term structure models, I propose a new estimation procedure, which solves a linear regression with OLS, and breaks the parameters down into three groups and estimates one group at a time. I first rewrite the model in terms of its VAR (vector autoregression) representation. Next, I re-parameterize the VAR representation to make the error terms of two parts of the system orthogonal, then I estimate these two parts separately with OLS. Lastly, I break the underlying parameters into three groups, where such a grouping is supported by theory, and back out one group of parameters at a time from the estimated OLS coefficients. Two groups of parameters are solved numerically with systems of equations and one is solved analytically with closed-form solutions.

The advantages of the new method are as follows:

1. **Less non-linearity.** There are multiple levels of functions in the conventional method, and all the functions are highly non-linear. The new procedure eliminates one level of non-linear functions by solving the linear regression with OLS.
2. **Avoiding singularity.** The log likelihood function in the conventional procedure involves the inverse of matrices, which may cause a singularity problem. The new procedure totally avoids this problem by solving the re-parameterized system with OLS.
3. **Fewer parameters inferred at each step.** The conventional method solves all underlying parameters in the affine term structure models numerically at once. The new procedure breaks the parameter system down into three groups, and solves two of them numerically with systems of equations and one analytically with closed-form solutions. This grouping, which is supported by theory, simplifies the estimation tremendously.
4. **No trouble recognizing convergence.** The old procedure, maximizing the likelihood function, can employ no handy tool to tell local maxima from global ones except repeating with enough different starting values. The new procedure, solving systems of equations numerically or solving closed-form so-

lutions analytically, can recognize convergence with certainty if the equations are solved with the weighted sum of squared errors less than the tolerance.

Besides introducing the new estimation procedure to solve the recognized estimation problems of affine term structure models, my empirical findings also show that the canonical representations or identifying restrictions proposed in literature are not unique, and one more restriction, the ordering of the diagonal elements of the mean reversion matrix under the risk neutral measure, is needed to ensure the uniqueness.

The new approach is also superior to the common practice of imposing zero restrictions on some parameters arbitrarily in a first round of estimation (Duffee 2002, Dai and Singleton 2002, Duarte 2004, Ang and Piazzesi 2003, Kim and Orphanides 2005, Christensen, Diebold and Rudebusch forthcoming) discuss this approach and note its shortcomings. I demonstrate the potential consequences of that approach with a simulated example. Zero restrictions imposed arbitrarily may yield wrong parameter estimates and thus totally distinct economic implications, and can also cause multiple local maxima that don't even exist in the original system. Again, these problems are avoided completely with the new method.

The rest of the paper is organized as follows. Section 1.2 introduces a baseline affine term structure model with all state variables latent, and several versions of identifying restrictions imposed in the literature. Section 1.3 introduces the new estimation procedure in terms of five steps, and also discusses how the modified version of identifying restrictions facilitates the new estimation procedure. Section 1.4 illustrates the issues using simulated data. I use three-factor models to demonstrate estimation difficulties with the conventional MLE and how the new method solves them. Multiple global solutions found by the new procedure reveal that the "identifying" restrictions proposed in the literature are in fact not identifying and need one more restriction to ensure the uniqueness of the canonical representation. In that section I also illustrate the consequences of imposing zeros on some parameters arbitrarily.



## 1.2 The Model

### 1.2.1 State and Short Rate Dynamics

The baseline model is a discrete version of the standard affine term structure models. All the underlying factors are latent, and the  $N$ -dimensional vector  $f_t$  follows a vector autoregression:

$$f_t = \mu + \rho f_{t-1} + \Sigma u_t \quad (1.1)$$

where  $u_t \sim i.i.d.N(0, I_N)$ ,  $\mu$  is a  $N \times 1$  vector and  $\rho$  and  $\Sigma$  are  $N \times N$  matrices. The above factor process is specified under the physical probability measure  $P$ .

The no-arbitrage condition is equivalent to the existence of an equivalent martingale measure of  $P$ , or the risk-neutral measure  $Q$ , under which  $f_t$  follows the dynamics:

$$f_t = \mu^Q + \rho^Q f_{t-1} + \Sigma u_t^Q \quad (1.2)$$

The short rate is specified as an affine function of the underlying factors:

$$r_t = \delta_0 + \delta_1' f_t \quad (1.3)$$

### 1.2.2 Identifying Restrictions in the Literature

The latent state variables  $f_t$  may rotate and translate without changing the probability distribution of bond yields, hence not all parameters in the above system Eq.(1.1) – (1.3) can be identified. As the first effort, Dai and Singleton (2000) impose the identifying restrictions under the  $P$ -measure on Eq.(1.1) – (1.3), which they refer to as  $A_0(N)$  where  $N$  stands for the number of factors: the mean  $\mu = 0$ , the volatility matrix  $\Sigma = I$  and the mean reversion matrix  $\rho$  to be triangular. Singleton (2006) and Christensen et al. (forthcoming) impose the identifying restrictions under the  $Q$ -measure instead: the mean  $\mu^Q = 0$ , the volatility matrix  $\Sigma = I$  and the mean reversion matrix  $\rho^Q$  to be triangular. Besides the restrictions on the factor dynamics, the identifying restriction on the short rate process  $\delta_1 \geq 0$  is also standard in the literature.

However, Christensen et al. (forthcoming) point out:

from an economic point of view, these two identifications may not be equivalent, because the yield function being fit to the observed yields is determined solely by the dynamics under the  $Q$ -measure, so imposing restrictions on the  $Q$ -measure drift terms could limit the ability of the model to fit observed yields.

Taking that into consideration, I propose a set of equivalent identifying restrictions to restrict the mean reversion matrix under the  $Q$  measure  $\rho^Q$ , and the mean vector under the  $P$ -measure  $\mu$ :

- $\Sigma = I$
- $\mu = 0$
- $\rho^Q$  is lower triangular
- $\delta_1 \geq 0$ .

### 1.2.3 Bond Pricing and Difference Equations

Define the nominal pricing kernel  $m_{t+1}$  as

$$m_{t+1} = \exp \left( -r_t - \frac{1}{2} \lambda_t' \lambda_t - \lambda_t' u_{t+1} \right) \quad (1.4)$$

The variables  $\lambda_t$  are the time-varying market prices of risk associated with the sources of uncertainty  $u_{t+1}$ . As in the affine term structure literature, I parameterize  $\lambda_t$  as an affine function of the state variables:

$$\lambda_t = \lambda_0 + \lambda_1 f_t \quad (1.5)$$

Now we can link the parameters from  $P$  and  $Q$  dynamics together through  $\lambda_0$  and  $\lambda_1$ :

$$\begin{aligned} \mu^Q &= \mu - \lambda_0 = -\lambda_0 \\ \rho^Q &= \rho - \lambda_1 \end{aligned} \quad (1.6)$$

Hereafter, I will represent the parameter system in terms of 5 underlying parameters:  $\delta_0, \delta_1, \rho, \rho^Q$  and  $\lambda_0$ .

Having defined the nominal pricing kernel in Eq.(2.6), the prices of zero coupon bonds satisfy:

$$p_t^{n+1} = E_t (m_{t+1} p_{t+1}^n) \quad (1.7)$$

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

In the affine term structure models, the bond prices are exponential affine functions of the state variables:

$$p_t^n = \exp (\bar{a}_n + \bar{b}'_n f_t) \quad (1.8)$$

where  $\bar{a}_n$  and  $\bar{b}_n$  follow the difference equations:

$$\begin{aligned} \bar{a}_{n+1} &= -\delta_0 + \bar{a}_n - \bar{b}'_n \lambda_0 + \frac{1}{2} \bar{b}'_n \bar{b}_n \\ \bar{b}_{n+1} &= -\delta_1 + \rho^Q \bar{b}_n \end{aligned} \quad (1.9)$$

with  $\bar{a}_1 = -\delta_0$  and  $\bar{b}_1 = -\delta_1$ . The derivation of the difference equations can be found in Appendix A of Ang and Piazzesi (2003)

Then, the yields  $y_t^n$  are linear in the state variables

$$y_t^n = -\frac{\log p_t^n}{n} = a_n + b'_n f_t \quad (1.10)$$

with  $a_n = -\bar{a}_n/n$  and  $b_n = -\bar{b}_n/n$ .

The difference equations in Eq.(1.9) can be solved for  $\bar{a}_n$  and  $\bar{b}_n$  and therefore  $a_n$  and  $b_n$  as follows (see Appendix A for derivation):

$$\begin{aligned} \bar{b}_n &= - [(\rho^Q)^n - I] [\rho^Q - I]^{-1} \delta_1 \\ b_n &= \frac{1}{n} [(\rho^Q)^n - I] [\rho^Q - I]^{-1} \delta_1 \\ a_n &= \delta_0 + \frac{(\bar{b}'_1 + \bar{b}'_2 + \dots \bar{b}'_{n-1})}{n} \lambda_0 - \frac{(\bar{b}'_1 \bar{b}_1 + \bar{b}'_2 \bar{b}_2 + \dots + \bar{b}'_{n-1} \bar{b}_{n-1})}{2n} \end{aligned} \quad (1.11)$$

#### 1.2.4 Conventional Estimation with MLE

The conventional method estimates the affine term structure models through MLE (maximum likelihood estimation), and the likelihood function can be found, for example, in Appendix B of Ang and Piazzesi (2003).

### 1.3 New Estimation Procedure

This section introduces the new estimation procedure. I first write the affine term structure model in terms of a VAR representation, then re-parameterize the VAR representation to orthogonalize the error terms and estimate this re-parameterized system with OLS, and lastly back out the underlying parameters with a combination of numerical and analytical methods. I'll discuss why re-parameterization is necessary and how the new representation of identifying restrictions described in Section 2.2 facilitates the new estimation procedure.

#### Step1: VAR representation

The affine term structure model has been presented with latent state variables  $f_t$ , and this section will illustrate how to get rid of those latent variables and represent the system with only observables. The estimation procedure introduced by Chen and Scott (1993) assumes that the number of yields priced without error is equal to the number of unobserved factors  $N$ , so we can back out the unobserved factors given data and parameters.

The yields have  $N + M$  different maturities, collected in a vector  $Y_t$ . Partition this vector into an  $N \times 1$  vector  $Y_t^1$ , the yields for which the affine pricing model is presumed to hold without error, and an  $M \times 1$  vector  $Y_t^2$  for which measurement error is allowed. Stack Eq.(2.10) for  $N + M$  different maturities in the order of  $Y_t^1$  and  $Y_t^2$  into an  $(N + M) \times 1$  equation system, and add pricing errors for the elements in  $Y_t^2$ :

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} f_t + \begin{bmatrix} 0 \\ B^m \end{bmatrix} u_t^m \quad (1.12)$$

$$\text{with } B^m = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_M \end{pmatrix} \text{ and } u_t^m \sim i.i.d.N(0, I_M)$$

Solve the unobserved vector  $f_t$  from  $Y_t^1$  part of Eq.(2.11), and then substitute it into the factor dynamics Eq.(1.1) and  $Y_t^2$  part of Eq.(2.11). After some

derivation (see Appendix B), the VAR representation of the above macro finance system is given as follows:

$$\begin{aligned} Y_t^1 &= A_1^* + \phi_1^* Y_{t-1}^1 + u_{1t}^* \\ Y_t^2 &= A_2^* + \phi_2^* Y_{t-1}^1 + u_{2t}^* \end{aligned} \quad (1.13)$$

$$\begin{aligned} \text{with } A_1^* &= A_1 - B_1 \rho B_1^{-1} A_1 \\ A_2^* &= A_2 - B_2 \rho B_1^{-1} A_1 \\ \phi_1^* &= B_1 \rho B_1^{-1} \\ \phi_2^* &= B_2 \rho B_1^{-1} \\ \text{var} \begin{pmatrix} u_{1t}^* \\ u_{2t}^* \end{pmatrix} &= \begin{pmatrix} B_1 B_1' & B_1 B_2' \\ B_2 B_1' & B_2 B_2' + B^m B^{m'} \end{pmatrix} \end{aligned} \quad (1.14)$$

## Step 2: OLS with Re-parameterization

I have transformed the affine term structure model with latent factors into its VAR representation in Eq.(1.13), which can be solved by OLS (ordinary least squares). But before applying that technique, I will re-parameterize the VAR representation and make the error terms orthogonal. Such a re-parameterization simplifies the system and facilitates the following steps of estimation to solve out the underlying parameters.

Consider the population linear projection of  $u_{2t}^*$  in Eq.(1.13) on  $u_{1t}^*$ ,

$$u_{2t}^* = \eta u_{1t}^* + \tilde{u}_{2t}^* \quad (1.15)$$

The residual  $\tilde{u}_{2t}^*$  is uncorrelated with  $u_{1t}^*$ . The regression coefficient  $\eta$  turns out to be:

$$\eta = B_2 B_1^{-1} \quad (1.16)$$

where the derivation can be found in Appendix C.

To make the error terms orthogonal, subtract  $\eta Y_t^1$  from  $Y_t^2$ , where  $Y_t^1$  and  $Y_t^2$  are from Eq.(1.13):

$$Y_t^2 = \tilde{A}_2^* + \eta Y_t^1 + \tilde{u}_{2t}^* \quad (1.17)$$

with  $\tilde{A}_2^* = A_2^* - \eta A_1^*$ , and  $\tilde{u}_{2t}^*$  defined in Eq.(1.15). The derivation can be found in Appendix C.

If such a transformation were performed with the unrestricted version of Eq.(1.13), Eq.(1.17) would maintain the  $Y_{t-1}^1$  term. However, Eq.(1.17) doesn't include any lagged term. Therefore, re-parameterization doesn't only orthogonalize the error terms, but also practically imposes one more restriction on the linear regression, the coefficient for  $Y_{t-1}^1$  is zero in Eq.(1.17), reflected from the structural restrictions of Eq.(1.13).

Now the system in Eq.(1.13) has been written

$$\begin{aligned} Y_t^1 &= A_1^* + \phi_1^* Y_{t-1}^1 + u_{1t}^* \\ Y_t^2 &= \tilde{A}_2^* + \eta Y_t^1 + \tilde{u}_{2t}^* \end{aligned} \quad (1.18)$$

$$\text{with } E \begin{pmatrix} u_{1t}^* \\ \tilde{u}_{2t}^* \end{pmatrix} \begin{pmatrix} u_{1t}^* & \tilde{u}_{2t}^* \end{pmatrix} = \begin{pmatrix} B_1 B_1' & 0 \\ 0 & B^m B^{m'} \end{pmatrix} \quad (1.19)$$

and  $\tilde{A}_2^*$  and  $\eta$  defined in Eq.(1.17) and Eq.(1.16).

Having independent error terms, the  $Y_t^1$  and  $Y_t^2$  blocks can be estimated separately with OLS in this step. Regressing  $Y_t^1$  over a constant and its own lag  $Y_{t-1}^1$  gives us values for  $\widehat{A}_1^*$ ,  $\widehat{\phi}_1^*$  and  $\widehat{B}_1 B_1'$ . Regressing  $Y_t^2$  over a constant and  $Y_t^1$  yields  $\widehat{A}_2^*$ ,  $\widehat{\eta}$  and  $\widehat{B^m B^{m'}}$ .

### Step 3: Numerically Solve $\rho^Q, \delta_1, B_1$ and $B_2$ from $\widehat{B}_1 B_1'$ and $\widehat{\eta}$

The third step solves  $\rho^Q, \delta_1, B_1$  and  $B_2$  from  $\widehat{B}_1 B_1'$  and  $\widehat{\eta}$  numerically with system of equations. First, construct the estimate  $\widehat{B}_2 B_1' = \widehat{\eta} \times \widehat{B}_1 B_1'$ . We then use the following system of equations:

$$\begin{aligned} B_1 B_1'_{(N \times N)} &= \begin{bmatrix} b'_{i_1}(\rho^Q, \delta_1) \\ \vdots \\ b'_{i_N}(\rho^Q, \delta_1) \end{bmatrix} \begin{bmatrix} b_{i_1}(\rho^Q, \delta_1) & \cdots & b_{i_N}(\rho^Q, \delta_1) \end{bmatrix} \\ B_2 B_1'_{(M \times N)} &= \begin{bmatrix} b'_{j_1}(\rho^Q, \delta_1) \\ \vdots \\ b'_{j_M}(\rho^Q, \delta_1) \end{bmatrix} \begin{bmatrix} b_{i_1}(\rho^Q, \delta_1) & \cdots & b_{i_N}(\rho^Q, \delta_1) \end{bmatrix} \end{aligned} \quad (1.20)$$

where the expressions for  $b_n$ 's as functions of  $\rho^Q$  and  $\delta_1$  can be found in Eq.(1.11),  $i_1, \dots, i_N$  are the  $N$  maturities of the yields priced exactly and  $j_1, \dots, j_M$  are the  $M$  maturities of the yields priced with error.

The unique elements in  $\widehat{B_1 B_1'}$  and  $\widehat{B_2 B_2'}$  give us  $N \times (N + 1) / 2 + M \times N$  equations, and  $\rho^Q$  and  $\delta_1$  totally contain  $N \times (N + 1) / 2 + N$  unknowns with the identification restriction constraining  $\rho^Q$  to lower triangular. The system is “just identified” in the sense that the number of equations equals the number of unknowns if  $M = 1$ , and at this point, I will only analyze the “just-identified” case. Note we can use this system of equations to solve for  $\widehat{\rho^Q}$  and  $\widehat{\delta_1}$  numerically. Then, we can use the closed-form solutions in Eq.(1.11) to solve  $\widehat{b}_n$  for  $n = 1, 2, \dots, \max(i_1, \dots, i_N, j_1, \dots, j_M) - 1$ , and  $\widehat{b}_n$  for  $n = i_1, \dots, i_N, j_1, \dots, j_M$ . Finally,  $\widehat{B_1} = \left[ \widehat{b}_{i_1}(\rho^Q, \delta_1) \quad \dots \quad \widehat{b}_{i_N}(\rho^Q, \delta_1) \right]'$ ,  $\widehat{B_2} = \left[ \widehat{b}_{j_1}(\rho^Q, \delta_1) \quad \dots \quad \widehat{b}_{j_M}(\rho^Q, \delta_1) \right]'$ .

#### Step 4: Analytically Solve $A_1, A_2, B^m$ and $\rho$ from OLS Coefficients

In the second step, we have already solved  $\widehat{A_1^*}, \widehat{\phi_1^*}, \widehat{A_2^*}$  and  $\widehat{B^m B^{m'}}$ , which are direct functions of  $A_1, A_2, B_1, B_2, B^m$  and  $\rho$  in form of  $A_1^* = A_1 - B_1 \rho B_1^{-1} A_1$ ,  $\phi_1^* = B_1 \rho B_1^{-1}$  and  $\tilde{A}_2^* = A_2 - B_2 \rho B_1^{-1} A_1 - \eta A_1^*$ . Therefore, in this step, we want to back out  $A_1, A_2, B^m$  and  $\rho$  from the OLS coefficients obtained in the second step, and they can be solved with closed-form solutions in the following manner:

4.1 Solve  $\widehat{B^m}$  from  $\widehat{B^m B^{m'}}$  where  $B^m$  is a diagonal matrix.

4.2 Solve  $\widehat{\rho}$  from  $\widehat{\phi_1^*}$ , and  $\rho = B_1^{-1} \widehat{\phi_1^*} B_1$ .

4.3 Solve  $\widehat{A_1}$  from  $\widehat{A_1^*}$ , and  $A_1 = (I - \phi_1^*)^{-1} A_1^*$

4.4 Solve  $\widehat{A_2}$  from  $\widehat{A_2^*}$ , and  $A_2 = \tilde{A}_2^* + \eta A_1^* + B_2 \rho B_1^{-1} A_1$ .

#### Step 5: Numerically Solve $\delta_0$ and $\lambda_0$ from $\widehat{A_1}$ and $\widehat{A_2}$

Lastly, we can solve  $\delta_0$  and  $\lambda_0$  numerically from  $\widehat{A_1}$  and  $\widehat{A_2}$  with system of equations. We have solved  $\widehat{b}_n$  in Eq.(1.11) for different  $n$ 's in the third step, so we

can write  $A_1$  and  $A_2$  as functions of  $\delta_0$  and  $\lambda_0$  only:

$$A_1 = \begin{bmatrix} a_{i_1}(\delta_0, \lambda_0) \\ \vdots \\ a_{i_N}(\delta_0, \lambda_0) \end{bmatrix} \quad A_2 = \begin{bmatrix} a_{j_1}(\delta_0, \lambda_0) \\ \vdots \\ a_{j_M}(\delta_0, \lambda_0) \end{bmatrix} \quad (1.21)$$

where  $a_n$ 's as functions of  $\delta_0$  and  $\lambda_0$  can be found in Eq.(1.11),  $i_1, \dots, i_N$  are the  $N$  maturities of the yields priced exactly and  $j_1, \dots, j_M$  are the  $M$  maturities of the yields priced with error.

There are totally  $1 + N$  unknowns in  $\delta_0$  and  $\lambda_0$ , and the number of available equations is  $N + M$ . So with  $M = 1$ , the system is "just identified" in the sense that the number of equations equals the number of unknowns. I can solve  $\widehat{\delta}_0$  and  $\widehat{\lambda}_0$  numerically from  $\widehat{A}_1$  and  $\widehat{A}_2$  with this system of equations.

## Discussion

One might wonder why the re-parameterization is necessary. Without re-parameterization, the VAR representation in Eq.(1.13) as a whole can still be estimated with OLS. Given the restrictions in Eq.(1.14), we can still follow step 3 and numerically obtain values for  $\widehat{\rho}^Q, \widehat{\delta}_1, \widehat{B}_1$  and  $\widehat{B}_2$  from  $\widehat{B}_1 \widehat{B}'_1$  and  $\widehat{B}_1 \widehat{B}'_2$  in the covariance matrix. Next,  $\widehat{\rho}$  can be solved from  $\widehat{\phi}_1^*$  following step 4.2. However, having  $\widehat{\phi}_2^*, \widehat{\rho}, \widehat{B}_1$  and  $\widehat{B}_2$ , the equation  $\phi_2^* = B_2 \rho B_1^{-1}$  in Eq.(1.14) becomes redundant and doesn't hold for most of the time. One additional restriction on the linear regression imposed implicitly by re-parameterization solves this problem, as demonstrated with the algorithm in this section.

The modified version of identifying restrictions I have proposed with constraints on the mean reversion matrix under the  $Q$  measure  $\rho^Q$  and the mean vector under the  $P$ -measure  $\mu$  fits into the new estimation procedure better for the following reasons: 1)  $\rho$  can be solved in Step 4.2 with closed-form solution, so there is no need to constrain it to be triangular. 2) The mean reversion matrix  $\rho^Q$  is unidentifiable from  $B_1 B'_1$  and  $\eta = B_2 B_1^{-1}$  unless it's restricted to be triangular, as argued above. 3) Restricting  $\mu$  instead of  $\mu^Q$  is more desirable in terms of economic meaning.



## 1.4 Empirical Results

### 1.4.1 Illustrative Results for Simulated Data

This section conducts an experiment with 3 factors ( $N = 3$ ) and 4 maturities ( $M = 1$ ), of which 1, 12 and 60 months yields are priced exactly, and 36 months yield is priced with error. Both the conventional and new estimation methods will be performed, and results for both will be reported.

I simulate a set of 1000-month yield data with parameters shown in Table 1.1.

The experiment uses 100 randomly generated starting values for the key variable  $\rho^Q$  based on the following formula. Draw 3 random variables from the uniform distribution then divided by 2 and plus 1/2, and place them in the diagonal. Hence, the diagonal elements of  $\rho^Q$  are between 0.5 and 1, and off-diagonal elements are all zeros.

The initial values for other parameters are quite standard and non-stochastic. I set  $\lambda_1 = 0$  with the conventional procedure, which yields  $\rho = \rho^Q$ . The new procedure doesn't require an initial value for  $\rho$  because it solves  $\rho$  with a closed-form solution. All three elements in  $\delta_1$  start from 1e-4,  $\lambda_0$  from zeros, and  $\delta_0$  from 0.0046, the average short rate. This set of starting values is very close to the parameters used for simulation, and is among the best initial values we can start with.

I perform the conventional procedure with MLE first. For **26** out of 100 starting values, the conventional procedure stops because the initial values encounter the singularity problem. For another **69** starting values, it converges to some local maxima. Table 1.2 shows 4 examples of those, and all of which display similar log likelihood values but different parameter estimates or economic meanings.

Furthermore, there is an issue of multiple “global” maxima. For the remaining **5** starting values, the conventional method gets two “global” maxima in Table 1.3, which both imply the identical likelihood function with the highest LLF I can achieve of 28110.4. Thus, one problem with the canonical representation is people

haven't noticed that we need some more restrictions to ensure its uniqueness.

In addition, the so-called "global" maxima could turn out to be local if some other estimates produce higher LLF values.

Now, let's see how the new estimation method solves all those problems. With initial values generated by the same algorithm described above, the new procedure converges to one of the six configurations in Table 1.4 **100** out of 100 times, and all six estimates imply the identical LLF. Furthermore, with the new method, I can recognize the six estimates are the **global** solutions, because Eq.(1.20) and Eq.(1.21) in third and fifth steps are solved with approximately zero errors. To sum up, the new procedure eliminates all the local maxima and singularity problem, and converges to one of the global maxima 100%.

Compare the estimates in Table 1.3 and Table 1.4, I could confirm the "global" maxima found with the conventional estimation are global. "Global"1 in Table 1.3 is GLOBAL1 in Table 1.4 and "Global"2 is GLOBAL3 with only numerical approximation errors.

Both Table 1.3 and Table 1.4 confirm there are multiple global maxima with the identifying restrictions in the literature imposed. Comparing the 6 distinct global solutions in Table 1.4, I find the key parameter  $\rho^Q$  in all cases shares the same diagonal elements but in different orders. The parameters assessing the average short rate  $\delta_0$  and the measurement error  $B^m$  stay the same in different cases. For cases with the same entry for the (1,1) or (3,3) element of  $\rho^Q$ , the corresponding elements of other parameters are the same. For example, GLOBAL1 and GLOBAL2 share the same (1,1) element of  $\rho^Q$ , which causes the first elements of  $\delta_1$  and  $\lambda_0$  and the (1,1) element of  $\rho$  are the same.

Therefore, with the assistance of the new method, the empirical results show that the canonical representations proposed by Dai and Singleton (2000) and Singleton (2006) display multiple equivalent solutions. To make the canonical representations unique, one more restriction on the ordering of the diagonal elements of  $\rho^Q$  is needed, and I would make it the descending order.

**Proposition 1** *A canonical representation of system (1.1) – (1.3) satisfies the following restrictions:*

1.  $\Sigma = I$
2.  $\mu = 0$
3.  $\rho^Q$  is lower triangular
4.  $\delta_1 \geq 0$
5. **Diagonal elements of  $\rho^Q$  are in descending order**

### 1.4.2 Robustness to Alternative Experiment Designs

Using the same set of simulated data, the second experiment employs a different formula to generate 100 initial values. Set the diagonal elements of  $\rho^Q$  to 0.99, 0.93 and 0.7, add a random variable from  $N(0, 0.01)$  to each of them, and place them in the diagonal. This set of starting values for  $\rho^Q$  is somewhere near the “truth”, and unit root is allowed. All three elements in  $\delta_1$  start from 0.001, and other things remain the same. The conventional MLE stops with the singularity problem for **2** starting values, and converges to some local but not global maxima for another **90** starting values and only converges to the global maxima for **8** starting values. On the other hand, the new procedure converges to one of the six global maxima for **100** out of 100 times.

I simulate another set of 1000-month yield data with the same parameters shown in Table 1.1.

With the new set of simulated data and the same algorithm to produce the initial values as the first experiment in Section 1.4.1, the results of the third experiment are more dramatic. The conventional MLE stops with the singularity problem for **13** starting values, and converges to some local maxima for **86** starting values and only converges to the global maxima for **1** starting value. On the other hand, the new procedure again converges to one of the global maxima for **100** out 100 times.

### 1.4.3 The Consequences of Imposing Zeros Arbitrarily

This section will demonstrate the consequences of a common practice in literature, which imposes zeros on some parameters arbitrarily in order to solve the badly-behaved system citeDuffee,DaiSingleton2002, and show that the estimates in the original case and in the modified case with zero restrictions produce totally different economic implications. For simple demonstration, I use a single factor model and 1000 months of simulated yields with 2 different maturities of 1 and 60 months.

Table 1.5 shows the parameters used for simulation “SIMU” and compares the parameters estimated from the conventional MLE (“MLE”) and the new procedure (“NEW”) under 3 different scenarios. In the first scenario, the model is unrestricted and “identified” in the sense that the number of unknown parameters equals the number of equations. In this case, both procedures produce exactly the same parameters.

In literature, it’s commonly seen that most elements of  $\lambda_0$  are restricted to 0 due to large standard errors in the first round of estimation (Dai and Singleton 2002, Ang and Piazzesi 2003). The second scenario sets  $\lambda_0$  to 0 to demonstrate the consequences of such an exercise in literature. In order to achieve higher LLF, the conventional MLE gets very different values for  $\rho$  in physical dynamics and  $\delta_0$ , which is essentially the average short rate and should be somewhere close to 0.0043. Therefore, setting  $\lambda_0$ , which is commonly considered as a trivial parameter, to zero will result in significant changes in underlying economic meanings. On the other hand, the “NEW” procedure reproduces  $\rho^Q$ ,  $\rho$  and  $\delta_1$  exactly, because they could be solved without the knowledge of  $\lambda_0$ . Even though  $\delta_0$  does depend on  $\lambda_0$ , the value of  $\delta_0$  is not affected much by fixing  $\lambda_0$ .

The variable  $\delta_0$  is essentially the average short rate, and it shouldn’t be something like 0.0174. So I fix it at 0.0043 (similar to what Ang and Piazzesi (2003) did) besides setting  $\lambda_0$  to zero in the third scenario. Such an additional restriction doesn’t lead the MLE estimate to where it’s supposed to be, but on the contrary, the MLE leads us to two different local maxima. On the other hand, the “NEW” procedure reproduces  $\rho^Q$ ,  $\rho$  and  $\delta_1$  exactly the same as before.

Therefore, imposing zero restrictions arbitrarily in the first round of estimation may yield to very different parameter estimates and thus wrong economic implications, and it might also cause multiple local maxima, which don't even exist in the original "just-identified" system. On the other hand, there is no need for the new estimation procedure to impose extra zero restrictions besides the identifying ones because it always converges to the global maxima.

## 1.5 Conclusion

The conventional estimation procedure for affine term structure models with MLE is known to be problematic due to highly non-linear likelihood function and badly-behaved likelihood surface. To completely solve the numerical challenges, I propose a new estimation procedure. The new procedure solves part of the problem with OLS after re-parameterization, and by doing so, it reduces non-linearity of the system and solves the singularity problem. It also breaks the parameters down into three groups and estimates one group at a time with a combination of numerical and analytical methods. This grouping, which is supported by theory, simplifies the estimation tremendously. Moreover, by solving systems of equations numerically and closed-form solutions analytically, the new procedure has no difficulty recognizing the convergence, while the conventional way has trouble differentiating local maxima from global ones.

In three factor experiments, the conventional MLE encounters singularity problem with a good chance due to the inverse of matrices in the log likelihood function. Besides, it converges to some local but not global maxima for most of the times. The empirical analysis gives examples of those, and all of which display similar log likelihood values but different parameter estimates or economic meanings. Moreover, even though the conventional method gets to the "global" maxima by luck, there is a multiple "global" maxima issue. In addition, the "global" maxima obtained by the conventional method could turn out to be local if some other estimates produce higher LLF values.

In contrast, the new procedure proposed in this paper eliminates all the

local maxima and singularity problem, and converges to one of the global maxima 100%. Besides the fact that it doesn't run into local maxima, the new method can recognize the global solutions with certainty when the equation systems are solved with zero errors.

For a three factor model with identifying restrictions imposed in the literature, the new estimation procedure lands at 6 distinct global solutions with identical LLF's, which therefore shows that the canonical representations proposed by Dai and Singleton (2000) and Singleton (2006) display multiple equivalent solutions. To ensure the uniqueness of canonical representation, one more restriction on the ordering of the diagonal elements of  $\rho^Q$  is needed.

A simple experiment with one factor model and simulated data shows that the widely-used technique in the conventional estimation with MLE of imposing zero restrictions arbitrarily on some parameters may yield to very different parameter estimates and therefore different economic implications from the original system, and it might also cause multiple local maxima, which don't exist in the original system.

## 1.6 Acknowledgements

Later version of Chapter 1, has been submitted for publication of material, and the submitted version is coauthored with James Hamilton. I thank Jim for the permission to use this chapter in the dissertation.

## 1.7 Appendices

### A. Solve the Difference Equations.

Solve  $\bar{b}_{n+1}$  part in Eq.(1.9)

$$\begin{aligned}\bar{b}_{n+1} &= -[I + \rho^{Q'} + \dots + (\rho^{Q'})^n] \delta_1 \\ &= -[(\rho^{Q'})^{n+1} - I] [\rho^{Q'} - I]^{-1} \delta_1\end{aligned}$$

Then,  $\bar{a}_{n+1}$  part of the difference equations in Eq.(1.9) can be solved as well:

$$\begin{aligned}\bar{a}_{n+1} &= -\delta_0 + \bar{a}_n - \bar{b}'_n \lambda_0 + \frac{1}{2} \bar{b}'_n \bar{b}_n \\ &= \left(-\delta_0 - \bar{b}'_n \lambda_0 + \frac{1}{2} \bar{b}'_n \bar{b}_n\right) + \left(-\delta_0 - \bar{b}'_{n-1} \lambda_0 + \frac{1}{2} \bar{b}'_{n-1} \bar{b}_{n-1}\right) \\ &\quad \dots + \left(-\delta_0 - \bar{b}'_1 \lambda_0 + \frac{1}{2} \bar{b}'_1 \bar{b}_1\right) + a_1 \\ &= -(n+1) \delta_0 - (\bar{b}'_1 + \bar{b}'_2 + \dots + \bar{b}'_{n-1} + \bar{b}'_n) \lambda_0 \\ &\quad + \frac{1}{2} (\bar{b}'_1 \bar{b}_1 + \bar{b}'_2 \bar{b}_2 + \dots + \bar{b}'_{n-1} \bar{b}_{n-1} + \bar{b}'_n \bar{b}_n)\end{aligned}$$

So  $\bar{b}_n, b_n = -\bar{b}_n/n$  and  $a_n = -\bar{a}_n/n$  can be solved in Eq.(1.11).

### B: Derive the VAR Representation.

Solve the unobserved vector  $f_t$  from  $Y_t^1$  part of Eq.(2.11):

$$f_t = B_1^{-1} (Y_t^1 - A_1) \quad (1.22)$$

and then substitute it into the factor dynamics Eq.(1.1):

$$Y_t^1 = (A_1 - B_1 \rho B_1^{-1} A_1) + B_1 \rho B_1^{-1} Y_{t-1}^1 + B_1 u_t \quad (1.23)$$

The  $Y_t^2$  part of Eq.(2.11) now becomes:

$$Y_t^2 = (A_2 - B_2 \rho B_1^{-1} A_1) + B_2 \rho B_1^{-1} Y_{t-1}^1 + B_2 u_t + B^m u_t^m \quad (1.24)$$

Combining Eq.(1.23) and (1.24) results in the VAR representation of the affine term structure model in Eq.(1.13) with parameters defined in Eq.(1.14)

### C: Re-parameterization.

The regression coefficient  $\eta$  in Eq.(1.15) is

$$\begin{aligned}
 \eta' &= \left( \sum u_{1t}^* u_{1t}^{*'} \right)^{-1} \sum u_{1t}^* u_{2t}^{*'} \\
 &= [\text{var}(u_{1t}^*)]^{-1} \text{cov}(u_{1t}^*, u_{2t}^*) \\
 &= (B_1 B_1')^{-1} B_1 B_2' \\
 &= B_1'^{-1} B_1^{-1} B_1 B_2' \\
 &= B_1'^{-1} B_2' = (B_2 B_1^{-1})'
 \end{aligned}$$

Subtract  $\eta Y_t^1$  from  $Y_t^2$ , where  $Y_t^1$  and  $Y_t^2$  are from Eq.(1.13) with expressions for  $\phi^{*'}s$  plugged in from Eq.(1.14)

$$\begin{aligned}
 Y_t^2 - \eta Y_t^1 &= (A_2^* - \eta A_1^*) + (B_2 \rho B_1^{-1} - \eta B_1 \rho B_1^{-1}) Y_{t-1}^1 + (u_{2t}^* - \eta u_{1t}^*) \quad (1.25) \\
 Y_t^2 &= (A_2^* - \eta A_1^*) + \eta Y_t^1 + (B_2 \rho B_1^{-1} - B_2 B_1^{-1} B_1 \rho B_1^{-1}) Y_{t-1}^1 + (u_{2t}^* - \eta u_{1t}^*)
 \end{aligned}$$

It results in the expression for  $Y_t^2$  in Eq.(1.17).



# Tables

**Table 1.1:** parameters for simulation

$Y_t^1$	1mth	12mth	60mth
$Y_t^2$	36mth		
$\rho^Q$	0.9992	0	0
	0.0100	0.9320	0
	0.0294	0.2553	0.7034
$\delta_1$	1.714E-04	1.816E-04	4.418E-04
$\rho$	0.9812	0.0067	0.0614
	-0.0008	0.8629	0.1039
	0.0165	0.1849	0.6854
$\delta_0$	0.0046		
$\lambda_0$	-0.0401	-0.0154	-0.5501
$B^m$	9.128E-05		

This table shows the parameters used for simulation.

**Table 1.2:** Examples of local maxima

	LOCAL1			LOCAL2		
$\rho^Q$	0.9985	0	0	0.9336	0	0
	0.0113	0.9328	0	-0.0121	0.9982	0
	0.0224	0.2499	0.7200	0.2413	0.0670	0.7096
$\delta_1$	1.736E-04	1.729E-04	4.455E-04	1.455E-04	2.032E-04	4.465E-04
$\rho$	0.9807	0.0180	0.0683	0.8629	-0.0224	0.0978
	0.0086	0.8576	0.1182	-0.0126	0.9788	0.0983
	0.0152	0.2005	0.6994	0.1759	0.0638	0.7110
$\delta_0$	-0.0025			-0.0652		
$\lambda_0$	-0.0759	-0.1178	-0.0041	0.1553	-0.4352	-0.0139
<i>LLF</i>	28104.9			28099.2		
	LOCAL3			LOCAL4		
$\rho^Q$	0.9984	0	0	0.9976	0	0
	0.0122	0.9329	0	0.3113	-0.9988	0
	0.0153	0.2464	0.7212	0.1404	-0.7643	0.9433
$\delta_1$	1.785E-04	1.717E-04	4.502E-04	1.231E-04	1.928E-04	4.540E-04
$\rho$	0.9823	0.0085	0.0778	0.9634	0.0839	0.0708
	0.0004	0.8523	0.1208	0.0497	0.6374	-0.0413
	0.0268	0.1846	0.7179	0.0357	-0.1269	0.9367
$\delta_0$	0.2449			-0.0012		
$\lambda_0$	1.2496	-0.4875	-0.4293	-0.0620	0.0564	-0.1280
<i>LLF</i>	28097.9			28074.3		

This table shows 4 examples of local maxima achieved by the conventional estimation for affine term structure models.

**Table 1.3:** 2 “global” maxima

	“Global”1			“Global”2		
$\rho^Q$	0.9985	0	0	0.9328	0	0
	0.0116	0.9328	0	-0.0116	0.9985	0
	0.0219	0.2500	0.7202	0.2424	0.0652	0.7202
$\delta_1$	1.713E-04	1.710E-04	4.452E-04	1.386E-04	1.984E-04	4.452E-04
$\rho$	0.9696	0.0141	0.0671	0.8548	-0.0230	0.1040
	-0.0027	0.8533	0.1175	-0.0062	0.9681	0.0866
	0.0085	0.1985	0.6993	0.1940	0.0429	0.6993
$\delta_0$	0.0046			0.0046		
$\lambda_0$	-0.0416	-0.0085	-0.5315	-0.0012	-0.0425	-0.5315
<i>LLF</i>	28110.4			28110.4		

This table shows 2 “global” maxima achieved by the conventional estimation for affine term structure models.

**Table 1.4:** 6 global maxima

	<b>GLOBAL1</b>			<b>GLOBAL2</b>		
$\rho^Q$	0.9985	0	0	0.9985	0	0
	0.0116	0.9326	0	0.0054	0.7209	0
	0.0220	0.2500	0.7209	0.0243	-0.2500	0.9326
$\delta_1$	1.717E-04	1.704E-04	4.455E-04	1.717E-04	1.578E-04	4.502E-04
$\rho$	0.9697	0.0140	0.0672	0.9697	0.0327	0.0603
	-0.0028	0.8529	0.1177	0.0076	0.6329	-0.0611
	0.0086	0.1987	0.6997	0.0048	-0.1421	0.9196
$\delta_0$	0.0046			0.0046		
$\lambda_0$	-0.0416	-0.0084	-0.5305	-0.0416	-0.3364	-0.4103
$B^m$	9.109E-05			9.109E-05		
$LLF$	28110.4			28110.4		
	<b>GLOBAL3</b>			<b>GLOBAL4</b>		
$\rho^Q$	0.9326	0	0	0.9326	0	0
	-0.0116	0.9985	0	0.2387	0.7209	0
	0.2424	0.0649	0.7209	0.0440	-0.0649	0.9985
$\delta_1$	1.382E-04	1.985E-04	4.455E-04	1.382E-04	3.886E-04	2.948E-04
$\rho$	0.8544	-0.0230	0.1043	0.8544	0.1068	0.0014
	-0.0062	0.9681	0.0865	0.1905	0.6849	-0.0234
	0.1942	0.0428	0.6997	0.0382	0.0203	0.9828
$\delta_0$	0.0046			0.0046		
$\lambda_0$	-0.0011	-0.0424	-0.5305	-0.0011	-0.5069	-0.1621
$B^m$	9.109E-05			9.109E-05		
$LLF$	28110.4			28110.4		
	<b>GLOBAL5</b>			<b>GLOBAL6</b>		
$\rho^Q$	0.7209	0	0	0.7209	0	0
	-0.0054	0.9985	0	-0.2387	0.9326	0
	-0.2504	0.0194	0.9326	-0.0760	-0.0194	0.9985
$\delta_1$	1.545E-04	1.747E-04	4.502E-04	1.545E-04	3.825E-04	2.948E-04
$\rho$	0.6322	0.0010	-0.0623	0.6322	-0.0600	-0.0166
	0.0261	0.9703	0.0591	-0.1438	0.9071	-0.0166
	-0.1422	0.0020	0.9196	-0.0151	0.0405	0.9828
$\delta_0$	0.0046			0.0046		
$\lambda_0$	-0.3355	-0.0481	-0.4103	-0.3355	-0.3800	-0.1621
$B^m$	9.109E-05			9.109E-05		
$LLF$	28110.4			28110.4		

This table displays 6 distinct solutions for a 3 factor model with 4 simulated yields estimated from the new procedure.

**Table 1.5:** Parameter comparison

	SIMU	<i>Unrest</i> NEW	MLE	$\lambda_0 = 0$ NEW	MLE	$\lambda_0 = 0$ NEW	$\delta_0 =$ MLE	0.0043
$\rho^Q$	0.9961	0.9964	0.9964	0.9964	0.9964	0.9964	0.9710	1.0091
$\rho$	0.9712	0.9852	0.9852	0.9852	0.9996	0.9852	0.9851	0.9851
$\delta_1$	5.48E-04	5.47E-04	5.47E-04	5.47E-04	5.49E-04	5.47E-04	5.27E-04	4.80E-04
$\delta_0$	0.0043	0.0042	0.0042	0.0042	0.0187	0.0043	0.0043	0.0043
$\lambda_0$	-0.0954	-0.0960	-0.0961	0	0	0	0	0

This table compares the estimated parameters from the conventional MLE and the new procedure under 3 different scenarios. The column “SIMU” gives the parameters used for simulation. The first scenario “Unrestricted” means all parameters are free to estimate; the second case restricts  $\lambda_0$  to zero; the third case fixes  $\delta_0$  to 0.0043 besides  $\lambda_0 = 0$ .

## Chapter 2

# Identifying and Testing the Refutable Implications behind the Macro Finance Term Structure Models of Interest Rates

### Abstract

Macro finance term structure models of interest rates and their estimation procedure impose refutable restrictions on the VAR. I introduce the VAR representation of a class of macro finance models to illustrate why those restrictions are testable. Fitting the VAR and macro finance models into the same framework makes the comparison more tractable and better defined. The empirical results show that both the in-sample and out-of-sample criteria favor the VAR over the macro finance structure.

## 2.1 Introduction

Accurately describing and correctly modeling the behavior of the term structure of interest rates is essential for bond pricing in the financial market, and policy making for the central banks. Research on joint movement of the yield

curve and macro variables is becoming the trend in the term structure literature.

Empirical studies start to investigate the joint behavior of the yield curve and macro variables with VAR models. Representative research includes Estrella and Mishkin (1997) and Evans and Marshall (2007).

To rule out riskless arbitrage opportunities across maturities, Ang and Piazzesi (2003)'s seminal paper begins the macro finance literature by incorporating cross-equation restrictions into the VAR structure. Subsequent examples include Bernanke, Reinhart and Sack (2004), Ang, Piazzesi and Wei (2006), Dewachter and Lyrio (2006), Hördahl, Tristani and Vestin (2006), Rudebusch and Wu (2008) and Pericoli and Taboga (2008)

The macro finance structure together with an estimation technique assuming some interest rates are observed without error (Chen and Scott 1993) brings testable restrictions on the VAR. To my knowledge, all of the previous research simply adopts the "standard" model setup and estimation technique without first testing if they fit the data well. The main objective of this paper is to make the first effort to demonstrate in a systematic manner why those restrictions are testable, what they are and whether the models fit the data as well as people have assumed.

A related empirical literature has obtained mixed results for the data compatibility of the arbitrage-free models. On the one hand, Duffee (2002) among others notes that the associated arbitrage-free models demonstrate disappointing empirical performance, especially with regard to out-of-sample forecasting. On the other hand, Ang and Piazzesi (2003) claim a superior out-of-sample performance of one of their two macro finance models after comparing them with some VAR and other reduced form models. However, in addition to the conflicting results on the empirical performance of the arbitrage-free structure, the comparisons are limited to the forecasting dimension and the comparison models are chosen arbitrarily due to the lack of a unified framework of the reduced form models and the structural arbitrage-free models. To address those problems, I introduce the VAR representation of the macro finance models. Fitting the VAR and macro finance models into the same framework makes the multidimensional comparisons between them feasible, and makes the comparisons more tractable and better defined. This action

also addresses the specification uncertainty of the reduced form models mentioned by Cochrane and Piazzesi (2009)

The empirical tests find that with Ang and Piazzesi (2003)'s exact models and data, two out of three in-sample criteria (the likelihood ratio tests and AIC) prefer the reduced form VAR while BIC and out-of-sample forecasting results are ambiguous. After finding that those restrictions imposed by the macro finance structure are not as data compatible as people have assumed with their original setup, I examine some alternative macro finance model specifications. Unfortunately, none of the changes in the model specifications save the performance of the macro finance structure; on the contrary, all of the model selection criteria favor the unrestricted VAR setup, and this result is proved to be robust.

The paper is organized as follows. Section 2.2 introduces the baseline macro finance model setup. In section 2.3, I explain why the restrictions are testable by introducing the VAR representation of the baseline macro finance model, and then elaborate on what those restrictions are and where they are from. Section 2.4 performs various in-sample and out-of-sample empirical tests on those restrictions, starting with Ang and Piazzesi (2003)'s original setup, and then turning to some alternatives. A robustness check is conducted in Section 2.5, and Section 2.6 concludes the paper.

## **2.2 The Baseline Macro Finance Term Structure Model**

### **2.2.1 Factor Dynamics**

In the macro finance literature of the term structure models of interest rates, the underlying factors fall into two categories: the observable macro variables and the unobservable latent factors. The factor dynamics, with a minimal set of identification restrictions proposed by Pericoli and Taboga (2008), follow a VAR

process<sup>1</sup>:

$$\begin{aligned} f_t^o &= \rho_{oo}F_{t-1}^o + \rho_{ou}f_{t-1}^u + \sigma_o u_t^o \\ f_t^u &= \rho_{uo}F_{t-1}^o + \rho_{uu}f_{t-1}^u + u_t^u \end{aligned} \quad (2.1)$$

with  $(u_t^o, u_t^u)' \sim i.i.d.N(0, I_5)$ , where  $\sigma_o$  is a  $2 \times 2$  lower triangular matrix. The vector  $f_t^o$  contains two current macro variables: inflation and output gap; while  $f_t^u$  is a vector containing three unobserved latent factors, which are commonly referred to as "level", "slope" and "curvature" in the literature. The vector  $F_{t-1}^o = (f_{t-1}^o, \dots, f_{t-p}^o)'$ , for  $p = 1, 2, \dots, 12$ , has  $2p$  elements of past macro variables. For example, Ang and Piazzesi (2003) set  $p = 12$ , or  $F_{t-1}^o = (f_{t-1}^o, \dots, f_{t-12}^o)'$ , implying that the lagged macro variables up to  $t - 12$  are incorporated into the system to explain the current ones.

The unrestricted factor process (2.1) in the baseline model allows dependence between the macro and latent factors; while some variations of the baseline setup might impose an additional independence assumption (see, for example, Ang and Piazzesi (2003)), which would restrict the cross terms  $\rho_{ou}$  and  $\rho_{uo}$  to zeros.

Write all the macro variables in the factor process (2.1) in terms of  $F_t^o$  as opposed to  $f_t^o$ :

$$\begin{aligned} F_t^o &= \varrho_{oo}F_{t-1}^o + \varrho_{ou}f_{t-1}^u + \Sigma_o u_t^o \\ f_t^u &= \rho_{uo}F_{t-1}^o + \rho_{uu}f_{t-1}^u + u_t^u \end{aligned} \quad (2.2)$$

The first two rows of the macro dynamics of in (2.2) are identical to the macro dynamics in (2.1). That is, the first two rows of  $\varrho_{oo}$  are  $\rho_{oo}$ , those of  $\varrho_{ou}$  are  $\rho_{ou}$ , and those of  $\Sigma_o$  are  $\sigma_o$ . The rest of  $\varrho_{oo}$  contains an identity matrix on the left and a zero matrix on the right; while the other rows of  $\varrho_{ou}$  and  $\Sigma_o$  only contain zeros. Note that there are  $2p$  elements in  $F_t^o$  and only 2 in  $u_t^o$ , so  $\Sigma_o$  might not be a square matrix. In the special case with  $p = 1$ , Eq.(2.2) collapses to Eq.(2.1), and  $\varrho_{oo} = \rho_{oo}$ ,  $\varrho_{ou} = \rho_{ou}$ ,  $\Sigma_o = \sigma_o$ .

---

<sup>1</sup>I construct the macro series  $f_t^o$  to have means of zeros. As in the literature, I impose the identification restriction that the latent variables  $f_t^u$  also have zero means by assuming that the factor process be stationary, which is different from the restriction imposed by Pericoli and Taboga (2008) that  $f_0^u = 0$ . Due to the restrictions imposed above, the constant terms in the factor dynamics are close to but not identical to zeros. But for compatibility, I follow Ang and Piazzesi (2003) to drop the constant terms from the VAR.



The vector  $F_t = (F_t^{o'}, f_t^{u'})'$  follows a VAR(1) process in the companion form:

$$F_t = \varrho F_{t-1} + \Sigma u_t \quad (2.3)$$

where  $u_t = (u_t^{o'}, u_t^{u'})'$ . The matrix  $\varrho$  contains four blocks:  $\varrho_{oo}$ ,  $\varrho_{ou}$ ,  $\rho_{uo}$  and  $\rho_{uu}$ ; and  $\Sigma$  can be partitioned into four blocks as well:  $\Sigma_o$  and a zero matrix on the top, and a zero matrix and an identity matrix on the bottom.

## 2.2.2 Short Rate Process

The Taylor (1993) rule describes how the Fed adjusts the short term interest rate  $r_t$  in response to inflation and real activity fluctuations. In addition to the short rate being a linear function of macro variables in the standard Taylor rule, it is also a linear function of latent variables  $f_t^u$ . Following Ang and Piazzesi (2003)'s convention, I will employ two versions of the Taylor rules, and they differ in the definition of the macro factors  $\tilde{F}_t^o$  used to explain the short rate process:  $\tilde{F}_t^o$  contains only current macro factors in one version and both current and past macro factors in the other version

$$r_t = \delta_0 + \delta'_{1o} \tilde{F}_t^o + \delta'_{1u} f_t^u \quad (2.4)$$

The model is called "macro" if  $\tilde{F}_t^o = f_t^o$ , in which the short rate only depends on the contemporaneous variables; and called "macro lag" if  $\tilde{F}_t^o = F_t^o$ , in which the whole state vector  $F_t$  in the companion form of the factor dynamics (2.3) helps to form the short rate process. With  $p = 1$ , two cases collapse into one, and I will name it "macro". By Pericoli and Taboga (2008), the identification restriction requires  $\delta_{1u} \geq 0$ .

To unify the notation in the short rate process (2.4) with that in the factor dynamics (2.3), I write factors in terms of  $F_t^o$  or  $F_t$  as opposed to  $\tilde{F}_t^o$ , and the short rate equation becomes:

$$r_t = \delta_0 + \delta'_1 F_t \quad (2.5)$$

The "macro lag" model has  $\delta'_1 = (\delta'_{1o}, \delta'_{1u})$ ; while the "macro" model has  $\delta'_1 = (\delta'_{1o}, 0_{1 \times (2p-2)}, \delta'_{1u})$

### 2.2.3 Pricing Kernel

Define the nominal pricing kernel  $m_{t+1}$  as

$$m_{t+1} = \exp \left( -r_t - \frac{1}{2} \lambda_t' \lambda_t - \lambda_t' u_{t+1} \right) \quad (2.6)$$

This setup for the risk adjusted stochastic discount factor is very standard in the literature. The variables  $\lambda_t$  are the time-varying market prices of risk associated with the sources of uncertainty  $u_{t+1}$ . As in the affine term structure literature, I parameterize  $\lambda_t$  as an affine function of the state variables

$$\lambda_t = \lambda_0 + \lambda_1 F_t \quad (2.7)$$

In order not to over fit, the vector  $\lambda_t$  is constrained to only have 5 non-zero elements corresponding to the current variables, and is constrained to depend only on the state variables of the current period. Consequently, the time invariant vector  $\lambda_0$  only has 5 non-zero elements. The matrix  $\lambda_1$  has zeros in rows and columns corresponding to the lagged macro factors, and only has  $5 \times 5$  non-zero elements at four corners. If an independent model is adopted (see, for example, Ang and Piazzesi (2003)), the matrix  $\lambda_1$  is further restricted to be block diagonal with  $2 \times 2 + 3 \times 3$  non-zero elements; that is, the time variation in the compensation for shocks to latent variables is only driven by the latent variables themselves, and the same argument holds for macro variables.

### 2.2.4 Bond Pricing

Having defined the nominal pricing kernel in Eq.(2.6), the prices of zero coupon bonds satisfy

$$p_t^{n+1} = E_t (m_{t+1} p_{t+1}^n)$$

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

In the affine term structure models, the bond prices are exponential affine functions of the state vector,

$$p_t^n = \exp (\bar{a}_n + \bar{b}_n' F_t) \quad (2.8)$$

where  $\bar{a}_n$  and  $\bar{b}_n$  follow the difference equations:

$$\begin{aligned}\bar{a}_{n+1} &= -\delta_0 + \bar{a}_n - \bar{b}'_n \Sigma \lambda_0 + \frac{1}{2} \bar{b}'_n \Sigma \Sigma' \bar{b}_n \\ \bar{b}_{n+1} &= -\delta_1 + (\varrho - \Sigma \lambda_1)' \bar{b}_n\end{aligned}\tag{2.9}$$

with  $\bar{a}_1 = -\delta_0$  and  $\bar{b}_1 = -\delta_1$ . The derivation of the difference equations can be found in Appendix A of Ang and Piazzesi (2003).

Then, the yields  $y_t^n$  are linear in the state variables

$$y_t^n = -\frac{\log p_t^n}{n} = a_n + b'_n F_t\tag{2.10}$$

with  $a_n = -\bar{a}_n/n$  and  $b_n = -\bar{b}_n/n$ .

## 2.3 Testable Restrictions Implied by the Macro Finance Structure

In addition to those assumptions imposed by the macro finance structure to rule out opportunities for riskless arbitrage across maturities, practitioners usually assume some interest rates are observed without error (see Chen and Scott (1993)), and that combination will give us testable restrictions. To my knowledge, all of the previous researches simply adopt the "standard" model setup and estimation technique without testing if they fit the data well. Therefore, the main objective of this paper is to make the first effort to demonstrate why those restrictions are testable, what they are and whether they fit the data as well as people have assumed. I will address these questions in an orderly manner, and use the empirical results to demonstrate the model compatibility with the data.

In this section, I will first explain why those restrictions are testable by introducing the VAR representation of the baseline macro finance model, and then elaborate on what the testable restrictions are and where they come from. Finally, I'll explain how to separate those restrictions that the baseline macro finance structure imposes on the reduced form VAR from the restrictions that various versions of the macro finance models impose on the baseline macro finance structure, and I will focus on the former ones.

### 2.3.1 Deriving the VAR representation of the Baseline Macro Finance Model

The estimation procedure introduced by Chen and Scott (1993) assumes that the number of yields priced without error is equal to the number of unobserved factors, so we can back out the unobserved factors given data and parameters.

The yields in my data set have  $N = 5$  different maturities (1, 3, 12, 36 and 60 months), collected in a  $5 \times 1$  vector  $Y_t$ . Partition this vector into a  $3 \times 1$  vector  $Y_t^1$ , the yields for which the affine pricing model is presumed to hold without error (for example, in Ang and Piazzesi (2003)),  $Y_t^1 = (y_t^1, y_t^{12}, y_t^{60})'$ , and a  $2 \times 1$  vector  $Y_t^2$  for which measurement error is allowed (for example, in Ang and Piazzesi (2003),  $Y_t^2 = (y_t^3, y_t^{36})'$ ). Stack Eq.(2.10) for 5 different  $n$ 's in the order of  $Y_t^1$  and  $Y_t^2$  into a  $5 \times 1$  equation system, and add pricing errors for the elements in  $Y_t^2$ :

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} + \begin{bmatrix} B_{1o} \\ B_{2o} \end{bmatrix} F_t^o + \begin{bmatrix} B_{1u} \\ B_{2u} \end{bmatrix} f_t^u + \begin{bmatrix} 0 \\ B^m \end{bmatrix} u_t^m \quad (2.11)$$

$$\text{with } B^m = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \text{ and } u_t^m \sim i.i.d.N(0, I_2)$$

Solve the unobserved vector  $f_t^u$  from the  $Y_t^1$  part of the equation system (2.11), and then substitute it into the factor dynamics (2.2) and the  $Y_t^2$  part of (2.11). After some derivation (see Appendix A), the VAR representation of the above macro finance system is given as follows:

$$\begin{aligned} F_t^o &= A_o^* + \phi_{oo}^* F_{t-1}^o + \phi_{o1}^* Y_{t-1}^1 + u_{ot}^* \\ Y_t^1 &= A_1^* + \phi_{1o}^* F_{t-1}^o + \phi_{11}^* Y_{t-1}^1 + u_{1t}^* \\ Y_t^2 &= A_2^* + \phi_{2o}^* F_{t-1}^o + \phi_{21}^* Y_{t-1}^1 + u_{2t}^* \end{aligned} \quad (2.12)$$

The non-linear restrictions on the coefficients and the variance-covariance matrix of the error terms are given in Appendix A and B. For elements of  $F_t^o$  corresponding to lagged macro variables, if any, the corresponding rows of  $A_o^*$ ,  $\phi_{o1}^*$  and  $u_{ot}^*$  are zeros, and those of  $\phi_{oo}^*$  consist of an identity matrix and zeros. Note that  $Y_{t-1}^2$  doesn't enter into the right hand side of the VAR representation as an

explanatory variable, which means that only  $Y_{t-1}^1$  interest rates that are assumed to be accurately priced by the model should help to forecast both  $F_t^o$ ,  $Y_t^1$  or  $Y_t^2$ . Moreover, there are 7 random elements in the system from  $u_t^o$ ,  $u_t^u$  and  $u_t^m$ .

Deriving the VAR representation of the baseline macro finance model shows that the no-arbitrage affine macro finance structure implies testable restrictions. Moreover, fitting the structural and reduced form models into the same framework makes the comparison more tractable and better defined. For example, Ang and Piazzesi (2003) compare the out-of-sample forecasting performance of their macro finance models with  $p = 12$  in Eq.(2.1) with that of a VAR(12) for 5 yields and 2 macro variables. It can be seen from Eq.(2.12) that a VAR with  $p = 12$  lags of macro variables and only 1 lag of yields would be a better candidate to compare with.

### 2.3.2 Testable Restrictions

I've shown why the no-arbitrage affine macro finance structure implies testable restrictions by illustrating that the baseline macro finance model is a vector autoregression with linear and nonlinear restrictions. All of the previous applications of this framework simply adopt the "standard" model setup and estimation technique without noticing that they are testable.

Having seen why the restrictions are testable, we shall ask what those testable restrictions are and where they come from.

The unrestricted VAR version of (2.12) is:

$$\begin{aligned} F_t^o &= A_o^* + \phi_{oo}^* F_{t-1}^o + \phi_{o1}^* Y_{t-1}^1 + \phi_{o2}^* Y_{t-1}^2 + u_{ot}^* \\ Y_t^1 &= A_1^* + \phi_{1o}^* F_{t-1}^o + \phi_{11}^* Y_{t-1}^1 + \phi_{12}^* Y_{t-1}^2 + u_{1t}^* \\ Y_t^2 &= A_2^* + \phi_{2o}^* F_{t-1}^o + \phi_{21}^* Y_{t-1}^1 + \phi_{22}^* Y_{t-1}^2 + u_{2t}^* \end{aligned} \quad (2.13)$$

By comparing the restricted version in (2.12) with its unrestricted version in (2.13), the restrictions that the baseline macro finance model imposes on the VAR include the nonlinear ones shown in Appendix A and B and the linear ones:  $\phi_{o2}^* = 0, \phi_{12}^* = 0, \phi_{22}^* = 0$ . Those restrictions can be attributed to the interaction

between the affine macro finance structure ruling out arbitrage opportunities and the estimation procedure assuming some of yields are priced without error.

### 2.3.3 Target Restrictions vs. Maintained restrictions

I've already demonstrated what the restrictions imposed by the baseline macro finance model setup and estimation framework are. Those restrictions are of central interest of this paper and I will call them the "target" restrictions or "structural" restrictions thereafter. In addition, the variations in the macro finance model specifications may impose some extra restrictions on the baseline setup, and some of those restrictions will pass through to the VAR representation and add some additional non-structural restrictions on it. Those additional restrictions are not of central interest of this paper, so I will just take them as given without testing them, and refer them as the "maintained" restrictions. This section associates each macro finance model with a matching VAR model by adding these maintained restrictions to the unrestricted VAR. In other words, the difference between the VAR and macro finance models in the same pair should only consist of the structural restrictions, and the following shows how to construct the pairs from several dimensions.

The first dimension is the dependence between the macro and latent factors. The independent macro finance models impose  $\rho_{ou} = 0$  and  $\rho_{uo} = 0$  in the factor dynamics (2.1), implying additional restrictions on the VAR representation (2.12) in the form of  $\phi_{o1}^* = 0$  and  $A_o^* = 0^2$ . The detailed derivation can be worked through similarly to Appendix A, and the intuition is that the independence assumption on the factors results in one way dependence between the macro variables and the interest rates, which means only the yields depend on the information from the lagged macro variables and the reverse is not true. Therefore, the VAR models used to compare with the independent macro finance models impose  $\phi_{o1}^* = 0$ ,  $\phi_{o2}^* = 0$  and  $A_o^* = 0$  in Eq.(2.13), and with abuse of notation I will refer those VARs to be "independent" along the first dimension.

---

<sup>2</sup>I construct the macro series  $f_t^o$  to have means of zeros, so in the independent case with  $\phi_{o1}^* = 0$  in (2.12),  $A_o^*$  is close to but not identical to a vector of zeros. But for compatibility, I follow Ang and Piazzesi (2003) to impose  $A_o^* = 0$ .

Second, the number of macro lags in the factor dynamics (2.1)  $p$  may vary across macro finance models, so I choose to compare the VAR and macro finance models with the same  $p$ . For example, if  $p = 1$  in the macro finance model, then I compare it with a VAR for which also  $p = 1$ , meaning  $F_{t-1}^o = f_{t-1}^o$  in both cases.

On the other hand, changing to Taylor rule, that is, "macro" or "macro lag" in (2.4), won't bring any maintained restrictions. So each pair of the "macro" and "macro lag" models<sup>3</sup> with the same entries along the first two dimensions,  $p$  and dependence, has a reduced form VAR to nest both of them with only target restrictions; and this VAR model has the same traits as its macro finance counterparts along the other two dimensions.

## 2.4 Empirical Results

In this section, I will empirically test those restrictions implied by the macro finance structure imposing no arbitrage constraint and the estimation procedure assuming that some of the yields are priced with no error.

As a starting point, I will use Ang and Piazzesi (2003)'s exact models and data to see if those refutable implications derived in the above section could survive their original setup. After showing that those restrictions imposed by the macro finance structure are not as data compatible as people have assumed under their original setup, I'll turn to some alternative model specifications to see if those would help. Unfortunately, none of the various specifications can save the performance of the macro finance models; none of the macro finance models have better performance than their VAR counterparts.

I will use various in-sample and out-of-sample model selection criteria as follows.

First, I use likelihood ratio tests to compare nested models. The likelihood ratio is defined as:

$$Ratio = -2(\ln L_R - \ln L_U) \tag{2.14}$$

where  $L_R$  is the maximized likelihood of the restricted model, and  $L_U$  is the max-

---

<sup>3</sup>The pair will degenerate to the "macro" model if  $p = 1$ .

imized likelihood of the unrestricted model. The ratio asymptotically has a  $\chi^2$  distribution with degrees of freedom equal to the number of restrictions.

The traditional likelihood ratio tests may suffer from the small sample bias. As a robustness check, I will simulate the critical values of the ratios at 5% significance level for some of the applications. More specifically, I will simulate a set of data with the same time series length as the actual data given the estimated parameters and model structure of the restricted version, and then estimate with both restricted and unrestricted models to calculate the simulated likelihood ratio based on Eq.(2.14). I repeat the procedure 1000 times and sort the simulated ratios in an ascending order, so the simulation-based critical value is the 950th entry of the ratios. If the likelihood ratio calculated from the real data is greater than the simulation-based critical value, we will reject the restriction.

Second, the information criteria, the AIC (Akaike information criterion) and the BIC (Bayesian information criterion), are used to compare both nested and non-nested models. The two information criteria, penalizing on over parameterization to different degrees, are defined as follows:

$$\begin{aligned} AIC &= 2K - 2\ln(L) \\ BIC &= K \ln(T) - 2\ln(L) \end{aligned} \tag{2.15}$$

where  $K$  is the number of parameters,  $\ln(L)$  is the maximized value of the log likelihood function, and  $T$  is the length of time series. A smaller AIC or BIC means a better model. For time series data with  $\ln(T) > 2$ ,  $BIC$  penalizes the number of parameters more severely, so it will favor more parsimonious models.

Third, the out-of-sample RMSE (root mean squared error) employs a different philosophy from those in-sample likelihood-based tests in that it's not a direct function of the maximized likelihood value, which characterizes how good the in-sample fits are. In contrast, the out-of-sample RMSE reflects how well the models fit the data in terms of forecasting. The procedure for examining the out-of-sample forecasts is as follows. At each date  $t$ , I estimate the models using data up to and including time  $t$ , and then forecast the next month's yields at time  $t + 1$ , and I do so for all the five yields available in data.



### 2.4.1 A Special Case: Ang and Piazzesi (2003)

Ang and Piazzesi (2003) compare their models with some other models including one vector autoregression with both yields and macro factors from an out-of-sample forecasting perspective, and they conclude that one of their macro finance models has a better out-of-sample performance. I will say more about the post sample performance shortly, but first focus on the more basic question of whether the assumptions behind the macro finance structure are valid.

#### Data, Models and Estimation Methods

The monthly data I am using is the closest I can find to that used by Ang and Piazzesi (2003), and almost identically replicates the statistics and estimation results in their paper. I use zero-coupon bond yields with maturities of 1, 3, 12, 36 and 60 months from CRSP monthly treasury file. I obtain two groups of monthly US macroeconomic key indicators, seasonally adjusted if applicable, from Datastream. The first group consists of various inflation measures which are based on the CPI, the PPI of finished goods, and the CRB Spot Index for commodity prices. The second group contains variables that capture real activity: the Index of Help Wanted Advertising, Unemployment Rates, the growth rate of Total Civilian Employment and the growth rate of Industrial Production. All growth rates and inflation rates are measured as the difference in logs of index at time  $t$  and  $t - 12$ , for index's monthly data. To reduce the dimensionality of the system, I follow Ang and Piazzesi (2003)'s procedure: I first normalize each series separately to have zero mean and unit variance, then extract the first principal component of each group, and name them as the "inflation" and "real activity" indices, so each index has zero mean and unit variance by construction. The sample period for yields is from December 1952 to December 2000, and that for the macro indices is from January 1952 to December 2000.

The macro finance models in Ang and Piazzesi (2003) specify  $p = 12$  in the factor dynamics (2.1), which means 12 macro lags are used as explanatory variables. In addition, they assume independence between the macro and latent factors; in other words,  $\rho_{ou} = 0$  and  $\rho_{uo} = 0$  in the factor dynamics (2.1), and  $\lambda_1$  in (2.7) is

block diagonal with  $2 \times 2 + 3 \times 3$  non-zero elements. Given the independence, the macro dynamics in (2.2) and the macro part of the short rate process in (2.4) can be estimated by OLS. I will designate a macro finance model as "macro" if  $\tilde{F}_t^o = f_t^o$  in (2.4) and as "macro lag" if  $\tilde{F}_t^o = F_t^o$  in (2.4). During estimation, Ang and Piazzesi (2003) impose 10 additional zeros on those insignificant parameters in both models. Rather than following their path-dependent estimation procedure, I set those 10 parameters to zero before estimation. Lastly, they assume  $Y_t^1 = (y_t^1, y_t^{12}, y_t^{60})'$ , or 1, 12 and 60 months yields are priced exactly.

As I have described in Section 3, I will separate the maintained restrictions from the structural ones and just test those implied by the model setup and estimation procedure. Consequently, the VAR model I use to compare with the macro finance models in Ang and Piazzesi (2003) has the following properties:  $p = 12$  in (2.13) or  $F_{t-1}^o = (f_{t-1}^{o1}, \dots, f_{t-1}^{o12})'$ , and what I refer to as the "independent" assumption namely  $\phi_{o1}^* = 0, \phi_{o2}^* = 0, A_o^* = 0$  in (2.13).

The above described VAR is different from the VAR selected by Ang and Piazzesi (2003) in the following aspects. First, Ang and Piazzesi (2003) incorporate 12 lags for both macro variables and yields; while Eq.(2.12) shows that the VAR with one lag of yields is enough to nest the macro finance models, and 12 lags of yields together with 12 lags of macro variables make the model overparameterized. Second, the macro variables and yields are dependent in Ang and Piazzesi (2003)'s VAR model. Instead, I will use the "independent" version of the VAR models to nest those macro finance setups for parsimony.

The macro finance models are estimated with MLE (the maximum likelihood estimation), and the identical estimation results can be achieved two different ways. The first method is to solve the VAR in Eq.(2.12) with the structural restrictions provided in Appendix A and B and the maintained restrictions for independence mentioned above, and the second method is to use the likelihood function provided in Appendix B of Ang and Piazzesi (2003). The VAR with the independence assumption can be solved following the procedure described in Chapter 11 of Hamilton (1994).

## Results

Table 2.1 performs the large-sample likelihood ratio tests between the VAR and the macro finance models. The vector autoregression is the unrestricted model, while the two macro finance models, “macro” and “macro lag”, are the restricted ones. The conventional likelihood ratio tests from row 3 to row 5 reject the restrictions with both P values of 0.00%, and prefer the unrestricted VAR.

As a robustness check, I perform in the last row of the table the simulation-based likelihood ratio tests, which adjust for small sample bias. Both the likelihood ratios in the third row are greater than the simulated critical values at the 5% significance level, so the restrictions implied by the macro finance models in Ang and Piazzesi (2003) are rejected. Compared to the conventional likelihood ratio tests, the simulation-based likelihood ratio tests might change the results quantitatively, but not qualitatively. Therefore, both the large-sample and small-sample versions of the likelihood ratio tests agree with each other, and don’t support the structure imposed by the macro finance models.

Table 2.2 ranks the VAR and macro finance models based on the AIC, and the ranking preference is in a descending order throughout the table. It shows that the VAR model has a smaller AIC value than both macro finance models; thus it confirms the results from the likelihood ratio tests, and the key result of this paper is that the cross equation restrictions imposed by the macro finance structure do not fit the data well.

Compared to the AIC, the BIC penalizes over-parameterization more, and it tends to choose more parsimonious models. BIC prefers the “macro” model to both the VAR and “macro lag”. That might be because imposing the cross equation restrictions reduces the dimensionality of the system to certain degree. For both information criteria, the “macro lag” model is ranked at the bottom, which is saying that the short rate only depends on the contemporaneous variables.

Table 2.3 compares the VAR and macro finance models using the out-of-sample RMSE, and the best models for different maturities with the smallest RMSEs are indicated in bold. It shows that the VAR model does better in forecasting the shorter yields with the maturities of 1 month and 3 months; while the “macro”

model proposed by Ang and Piazzesi (2003) provides a better post sample fit for the longer yields with the maturities of 12, 36 and 60 months. On the other hand, the "macro lag" model is once again proved to be the worst with the highest RMSEs for most of the time. Thus, I find that the cross equation restrictions imposed by the macro finance models are desirable for forecasting the longer term yields, which is consistent with the findings in Ang and Piazzesi (2003). On the other hand, I find them not desirable for forecasting the shorter term yields, which is contrary to the results in Ang and Piazzesi (2003). The reason that I have obtained a different result from Ang and Piazzesi (2003) is that I am using a more restrictive VAR, which nests the two macro finance models with only "target" restrictions and thus is a more suitable candidate for the comparison.

Overall, the empirical results show that the "macro" model proposed by Ang and Piazzesi (2003) does have a lower BIC value and a better forecasting ability, but such ability is limited to the longer term yields as compared to the VAR model. On the other hand, the "macro lag" model performs poorly by all criteria. Moreover, both macro finance models have worse in-sample fits in terms of both likelihood ratio tests and AIC.

## 2.4.2 Generalization: Alternative Specifications

We have seen that under Ang and Piazzesi (2003)'s original setup, the restrictions imposed by macro finance models appear to be inconsistent with the data. This section examines an alternative data set and various model specifications, yet maintains the basic affine no-arbitrage macro finance structure and its estimation procedure, to see if such modifications can save the performance of the macro finance models.

### Data, Model Specifications and Estimation Methods

The data are the same as before, except that the time periods are shorter and more recent: the yields are from December 1970 to December 2007 and the

macro indices are from January 1970 to December 2007<sup>4,5</sup>.

Ang and Piazzesi (2003)'s original macro finance setups fail to pass various model selection tests, and I will try some alternative specifications for the macro finance models along several dimensions.

First, Ang and Piazzesi (2003) assume independence between macro and latent factors:  $\rho_{ou} = 0$  and  $\rho_{uo} = 0$  in (2.1), and  $\lambda_1$  in (2.7) is block diagonal with  $2 \times 2 + 3 \times 3$  non-zero elements. Such an assumption eliminates any potential feedbacks from the yields to the macro variables. The alternative is the dependent setup, which removes the constraints on  $\rho_{ou}$  and  $\rho_{uo}$ , and  $\lambda_1$  has  $5 \times 5$  non-zero elements on four corners.

The second dimension is how many lags of macro variables  $p$  are in the factor dynamics (2.1). In Ang and Piazzesi (2003)'s setup,  $p = 12$  or  $F_{t-1}^o = (f_{t-1}^o, \dots, f_{t-12}^o)'$ . However, the number of parameters explodes when so many lags are included, especially when dependence is introduced at the same time. The alternative is  $p = 1$  or  $F_{t-1}^o = f_{t-1}^o$  for its parsimony.

Third, even though the "macro lag" version of the models underperforms with the original setup, I'll keep it and see if the results are different with the new model setups. Still, the model is called "macro" if  $\tilde{F}_t^o = f_t^o$  in (2.4) and "macro lag" if  $\tilde{F}_t^o = F_t^o$  in (2.4). The "macro lag" version will only be assessed together with  $p = 12$ .

To sum up, the macro finance models have three different traits to be varied: dependence (dependent or independent),  $p$  (12 or 1) and the Taylor rule ("macro" or "macrolag"). By assuming that the 1, 12 and 60 months yields are priced exactly, totally 6 macro finance models will be compared among others, and I denote the models as  $MF(dep/ind, 1/12, macro/lag)$ , where "lag" only appears when the second entry is 12. As a special case, the models  $MF(ind, 12, macro)$  and

---

<sup>4</sup>Empirical term structure research has used different sample periods: some of them use longer periods, starting from 1950's or 1960's, for example, Ang and Piazzesi (2003), and Pericoli and Taboga (2008); some others use shorter periods, starting from 1980's and 1990's, for example, Kim and Wright (2005), Kim and Wright (2005), Diebold and Li (2006) and Rudebusch and Wu (2008). Combining these two strands, I follow Bikbov and Chernov (2008) among others to start from 1970 and extend to 2007.

<sup>5</sup>I also have done calculations not reported that use bigger panels, which are only available for this shorter period.

$MF(ind, 12, lag)$  are exactly the same as the models in Ang and Piazzesi (2003), and the only difference is I don't impose the extra 10 zeros on the parameters during the estimation.

As I have described in Section 3, each pair of the "macro" and "macro lag" models<sup>6</sup> with the same entries along the first two dimensions has a reduced form VAR to nest both of them with only target restrictions; and this VAR model should have the same traits as its macro finance counterparts along those two dimensions. So I will specify the first two dimensions of the VAR models the same as those of the macro finance models, and the four VAR models are named as  $VAR(dep/ind, 1/12)$ .<sup>7</sup> The model  $VAR(i, j)$  nests  $MF(i, j, \cdot)$  with only necessary structural restrictions, for  $i = ind$  or  $dep$ ,  $j = 1$  or  $12$ .

The likelihood ratio tests require nesting relationship, and therefore the VAR model and macro finance model in the same test should have the same model entries along the first two dimensions. In other words, I compare  $VAR(i, j)$  with  $MF(i, j, \cdot)$ . For other model selection criteria, nesting relationship is not necessary, so I will pool all the macro finance models and their matching VAR models together to make the comparisons.

For simplicity, I will solve the MLE of the macro finance models using the VAR representations in Eq.(2.12) with structural and maintained restrictions. The reduced form VARs can be solved following the procedure described in Chapter 11 of Hamilton (1994).

## Results

The likelihood ratio tests in Table 2.4 reject the restrictions imposed by the no-arbitrage macro finance structure under all model specifications at all significance levels with all of the p-values being 0.00%. The conclusion drawn from this set of likelihood ratio tests is perfectly in line with that from the likelihood ratio tests conducted with only Ang and Piazzesi (2003)'s models in Table 2.1, and the key result stands.

---

<sup>6</sup>The pair will degenerate to the "macro" model if  $p = 1$ .

<sup>7</sup>The "independent" VAR models have the restrictions:  $\phi_{o1}^* = 0, \phi_{o2}^* = 0, A_o^* = 0$  in (2.13).

Putting together the structural and reduced form models, the AIC will unanimously pick any VAR over any macro finance models (in Table 2.5), no matter which values the other two dimensions, "dep" and "p", take. The best model nominated by the AIC is VAR(ind,12,n). Once again, it confirms the point Table 2.2 has delivered and the key point of the paper that the macro finance models don't go well with the data.

Table 2.6 shows that the BIC chooses the parsimonious models with  $p = 1$  over the ones with  $p = 12$ . Comparing between the VAR and macro finance models, with the second entries being  $p = 1$ , the VAR models are preferred over the macro finance ones, which is in line with the key result. The new BIC results seem to overrule those old ones in Table 2.2, but they are not totally inconsistent. Putting the picture together, the macro finance models may have lower BICs, but only if  $p = 12$ ; and the VAR jointly with  $p = 1$  gives the best models as far as the BIC is concerned. Thus, incorporating more model specifications doesn't help the macro finance structure to perform but eliminates the deviation of the earlier BIC results from the key result.

The root mean squared errors compare the models from a forecasting perspective. A superior forecasting ability is claimed to be one of the key contributions of the macro finance models, and my experiment under Ang and Piazzesi (2003)'s exact setup in Table 2.3 confirms this point, but only for the longer end of the yield curve. However, with more macro finance and VAR models incorporated into the comparison, Table 2.7 indicates that the VAR model  $VAR(ind, 1)$  performs the best for almost all of the maturities. Accordingly, the reduced form VAR structure is supported even by the out-of-sample forecasting under the new setup, and up to this point, the key result is that the restrictions implied jointly by the macro finance structure and estimation procedure are not preferred by all of the model selection criteria.

Therefore, the changes in the macro finance model specifications do not save the performance of the structure introduced by Ang and Piazzesi (2003); on the contrary, all of the model selection criteria favor the unrestricted VAR setup.

## 2.5 Robustness Check: Which Three Yields Are Priced Without Error

The above comparisons between the macro finance models and VAR models are conducted by assuming  $Y_t^1 = (y_t^1, y_t^{12}, y_t^{60})'$ , or 1, 12 and 60 months yields are priced exactly. But it's actually testable which yields we want to treat as observable, and I will conduct this set of tests by comparing models assigning  $Y_t^1$  with different maturities. In particular, I will just compare within the VAR family, and the VAR models used for this comparison only impose maintained restrictions on the VAR representation (2.12). Incorporating different yields in  $Y_t^1$  implies different sets of the two yields which do not appear on the right hand side of the VAR models as the explanatory variables.

The VAR models used to compare with the macro finance models in the empirical section take the form of Eq.(2.13), and therefore they are symmetric; while those used to test which three yields we want to treat as observable in this section take the form of Eq.(2.12), and they are asymmetric.

### 2.5.1 A Special Case: Ang and Piazzesi (2003)

This section will test which three yields should be priced without error using Ang and Piazzesi (2003)'s data from 1952 to 2000, and adopting the first two dimensions from Ang and Piazzesi (2003) with  $p = 12$  and independence.

Table 2.8 compares the asymmetric VAR models assigning  $Y_t^1$  with different yields. The table ranks the models based on the log likelihoods, and it shows that  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  is the best choice. Since the numbers of the parameters are exactly the same for all those models, ranking the likelihoods is equivalent to ranking the AICs or BICs, and thus the likelihood-based tests suggest that 3, 12 and 60 months yields are better proxies for short, medium and long runs.

Table 2.9 compares the out-of-sample performance of those VAR models with different yields included in  $Y_t^1$ . No particular model is the best in terms of all the maturities, and the three best models highlighted in bold are  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  for 60 months yields,  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{36})'$  for 12 and 36 months and



$Y_t^1 = (y_t^1, y_t^3, y_t^{12})'$  for 1 and 3 months. Therefore, to forecast the yields with any particular maturity, the yields with this maturity should be incorporated in  $Y_t^1$ . For example, the model  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  has a lowest RMSE in predicting 60 months yields, and this model assumes that the 60 months yields are priced without error.

## 2.5.2 Generalization: Alternative Specifications

This section will further test which three yields should be priced without error by using the alternative data set from 1970 to 2007, and four specifications along the first two dimensions  $VAR(dep/ind, 1/12)$ .

Table 2.10 ranks the VAR models along the third dimension, which three yields are incorporated in  $Y_t^1$ , based on the order in Table 2.8. For all four different VAR specifications,  $VAR(dep/ind, 1/12)$ , the log likelihoods are almost in descending order, and the best models highlighted in bold always have  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$ . This implication that the 3, 12 and 60 months yields are good proxies for the short term, medium term and long term is consistent with Table 2.8.

Four different VAR specifications  $VAR(dep/ind, 1/12)$  nominate different sets of the best models, and each set contains 5 or less best models with the lowest RMSEs for the 5 yields. For example, in Table 2.11, the best models with  $VAR(dep, 12)$  are  $Y_t^1 = (y_t^1, y_t^3, y_t^{60})'$  for 1 month,  $Y_t^1 = (y_t^1, y_t^3, y_t^{12})'$  for 3 and 12 months,  $Y_t^1 = (y_t^1, y_t^3, y_t^{36})'$  for 36 months and  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{36})'$  for 60 months. Although different specifications favor different models, forecasting the yields with a specific maturity requires the yields with this maturity to be included in  $Y_t^1$ , for most of the cases. Surprisingly, the model chosen by Ang and Piazzesi (2003)  $Y_t^1 = (y_t^1, y_t^{12}, y_t^{60})'$  is never the best for any maturity.

To sum up,  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  is nominated as the best by in-sample likelihood-based tests consistently for all variations along the first two dimensions, and the results from the out-of-sample tests are far more ambiguous.

### 2.5.3 Robustness Check with $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$

We've seen that the in-sample tests overwhelmingly prefer the models with  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$ , so I will adopt this specification  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  and redo some tests in Section 4.1, which have been performed under Ang and Piazzesi (2003)'s setups.

Table 2.12 and 2.13 exactly reproduce the ordering in Table 2.2 and 2.3, and thus the key result stands.

## 2.6 Conclusion

Combined with its estimation procedure assuming some interest rates are observed without error, the macro finance structure imposes some testable restrictions in order to rule out opportunities for riskless arbitrage across maturities. This paper makes the first effort to demonstrate why those restrictions are testable, what they are and whether they fit the data as well as people have assumed. To address those questions, I introduce the VAR representation of the macro finance models and fit the VAR and macro finance models into the same framework, which makes the comparison more tractable and better defined.

Using Ang and Piazzesi (2003)'s exact models and data, the empirical results show that the "macro lag" model with the lagged macro variables incorporated in the short rate process performs poorly with all criteria. Both of the "macro" and "macro lag" models are seen to have worse in-sample fits in terms of both likelihood ratio tests and AIC, and the results from BIC and out-of-sample forecasting are ambiguous.

After finding that those restrictions imposed by the macro finance structure are not as data compatible as people have assumed with their original setup, I examine some alternative model specifications. Unfortunately, none of the changes in the macro finance model specifications save the performance of the structure introduced by Ang and Piazzesi (2003); on the contrary, all of the model selection criteria favor the unrestricted VAR setup.

Rather than assuming three particular yields are included in  $Y_t^1$ , the ro-

business section examines which yields we want to treat as observable. I find that  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  is nominated as the best by in-sample likelihood-based tests consistently for all variations along the first two dimensions, and the results from the out-of-sample tests are far more ambiguous. More interestingly, the one chosen by Ang and Piazzesi (2003)  $Y_t^1 = (y_t^1, y_t^{12}, y_t^{60})'$  is never the best for any criteria. Redoing some tests using  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  instead of  $Y_t^1 = (y_t^1, y_t^{12}, y_t^{60})'$  exactly reproduces the ordering of the models by various criteria, and the key result stands.

## 2.7 Acknowledgements

Later version of Chapter 2, has been submitted for publication of material, and the submitted version is coauthored with James Hamilton. I thank Jim for the permission to use this chapter in the dissertation.

## 2.8 Appendices

### A. Deriving the VAR representation of the Baseline Macro Finance Model

I follow Chen and Scott (1993) and others in assuming that the inverse of  $B_{1u}$  exists, so I can solve the unobserved vector  $f_t^u$  from the  $Y_t^1$  part of the equation system (2.11):

$$f_t^u = B_{1u}^{-1}Y_t^1 - B_{1u}^{-1}A_1 - B_{1u}^{-1}B_{1o}F_t^o \quad (2.16)$$

Note that the unobserved factors are an affine function of the 3 yields and the current and lagged macro variables.

Substitute (2.16) into the macro dynamics in (2.2):

$$\begin{aligned} F_t^o &= \varrho_{oo}F_{t-1}^o + \varrho_{ou}f_{t-1}^u + \Sigma_o u_t^o \\ &= \varrho_{oo}F_{t-1}^o + \varrho_{ou} \left( B_{1u}^{-1}Y_{t-1}^1 - B_{1u}^{-1}A_1 - B_{1u}^{-1}B_{1o}F_{t-1}^o \right) + \Sigma_o u_t^o \\ &= -\varrho_{ou}B_{1u}^{-1}A_1 + (\varrho_{oo} - \varrho_{ou}B_{1u}^{-1}B_{1o}) F_{t-1}^o + \varrho_{ou}B_{1u}^{-1}Y_{t-1}^1 + \Sigma_o u_t^o \end{aligned}$$

So the macro state variables are affine in their own lags and lagged yields which are priced without error:

$$F_t^o = A_o^* + \phi_{oo}^* F_{t-1}^o + \phi_{o1}^* Y_{t-1}^1 + u_{ot}^* \quad (2.17)$$

with

$$\begin{aligned} A_o^* &= -\varrho_{ou}B_{1u}^{-1}A_1 \\ \phi_{oo}^* &= \varrho_{oo} - \varrho_{ou}B_{1u}^{-1}B_{1o} \\ \phi_{o1}^* &= \varrho_{ou}B_{1u}^{-1} \\ u_{ot}^* &= \Sigma_o u_t^o \end{aligned} \quad (2.18)$$

Substitute the expression for  $f_t^u$  in (2.16) into the dynamics for the latent factors in (2.2):

$$\begin{aligned} B_{1u}^{-1}Y_t^1 - B_{1u}^{-1}A_1 - B_{1u}^{-1}B_{1o}F_t^o &= \rho_{uu} \left( B_{1u}^{-1}Y_{t-1}^1 - B_{1u}^{-1}A_1 - B_{1u}^{-1}B_{1o}F_{t-1}^o \right) \\ &\quad + \rho_{uo}F_{t-1}^o + u_t^u \end{aligned}$$

$$\begin{aligned}
Y_t^1 - A_1 - B_{1o}F_t^o &= B_{1u}\rho_{uo}F_{t-1}^o + B_{1u}\rho_{uu} (B_{1u}^{-1}Y_{t-1}^1 - B_{1u}^{-1}A_1 - B_{1u}^{-1}B_{1o}F_{t-1}^o) \\
&\quad + B_{1u}u_t^u \\
Y_t^1 &= A_1 + B_{1o} (A_o^* + \phi_{oo}^*F_{t-1}^o + \phi_{o1}^*Y_{t-1}^1 + u_{ot}^*) + B_{1u}\rho_{uo}F_{t-1}^o \\
&\quad + B_{1u}\rho_{uu} (B_{1u}^{-1}Y_{t-1}^1 - B_{1u}^{-1}A_1 - B_{1u}^{-1}B_{1o}F_{t-1}^o) + B_{1u}u_t^u \\
&= (I_3 - B_{1u}\rho_{uu}B_{1u}^{-1}) A_1 + B_{1o}A_o^* \\
&\quad + (B_{1o}\phi_{oo}^* + B_{1u}\rho_{uo} - B_{1u}\rho_{uu}B_{1u}^{-1}B_{1o}) F_{t-1}^o \\
&\quad + (B_{1o}\phi_{o1}^* + B_{1u}\rho_{uu}B_{1u}^{-1}) Y_{t-1}^1 \\
&\quad + B_{1o}u_{ot}^* + B_{1u}u_t^u
\end{aligned}$$

So  $Y_t^1$  can be written as a linear combination of  $F_{t-1}^o$  and  $Y_{t-1}^1$ :

$$Y_t^1 = A_1^* + \phi_{1o}^*F_{t-1}^o + \phi_{11}^*Y_{t-1}^1 + u_{1t}^* \quad (2.19)$$

with

$$A_1^* = (I_3 - B_{1u}\rho_{uu}B_{1u}^{-1}) A_1 + B_{1o}A_o^* \quad (2.20)$$

$$\phi_{1o}^* = B_{1o}\phi_{oo}^* + B_{1u}\rho_{uo} - B_{1u}\rho_{uu}B_{1u}^{-1}B_{1o} \quad (2.21)$$

$$\phi_{11}^* = B_{1o}\phi_{o1}^* + B_{1u}\rho_{uu}B_{1u}^{-1} \quad (2.22)$$

$$u_{1t}^* = B_{1o}u_{ot}^* + B_{1u}u_t^u \quad (2.23)$$

Substitute  $F_t^o$  in (2.17) and  $f_t^u$  in (2.16) into the  $Y_t^2$  part in (2.11):

$$\begin{aligned}
Y_t^2 &= A_2 + B_{2o}F_t^o + B_{2u}f_t^u + B^m u_t^m \\
&= A_2 + B_{2o}F_t^o + B_{2u} (B_{1u}^{-1}Y_t^1 - B_{1u}^{-1}A_1 - B_{1u}^{-1}B_{1o}F_t^o) + B^m u_t^m \\
&= A_2 - B_{2u}B_{1u}^{-1}A_1 + (B_{2o} - B_{2u}B_{1u}^{-1}B_{1o}) F_t^o + B_{2u}B_{1u}^{-1}Y_t^1 + B^m u_t^m \\
&= A_2 - B_{2u}B_{1u}^{-1}A_1 + (B_{2o} - B_{2u}B_{1u}^{-1}B_{1o}) (A_o^* + \phi_{oo}^*F_{t-1}^o + \phi_{o1}^*Y_{t-1}^1 + u_{ot}^*) \\
&\quad + B_{2u}B_{1u}^{-1} (A_1^* + \phi_{1o}^*F_{t-1}^o + \phi_{11}^*Y_{t-1}^1 + u_{1t}^*) + B^m u_t^m \\
&= A_2 - B_{2u}B_{1u}^{-1}A_1 + (B_{2o} - B_{2u}B_{1u}^{-1}B_{1o}) A_o^* + B_{2u}B_{1u}^{-1}A_1^* \\
&\quad + (B_{2o}\phi_{oo}^* - B_{2u}B_{1u}^{-1}B_{1o}\phi_{oo}^* + B_{2u}B_{1u}^{-1}\phi_{1o}^*) F_{t-1}^o \\
&\quad + (B_{2o}\phi_{o1}^* - B_{2u}B_{1u}^{-1}B_{1o}\phi_{o1}^* + B_{2u}B_{1u}^{-1}\phi_{11}^*) Y_{t-1}^1 \\
&\quad + (B_{2o} - B_{2u}B_{1u}^{-1}B_{1o}) u_{ot}^* + B_{2u}B_{1u}^{-1}u_{1t}^* + B^m u_t^m
\end{aligned}$$

Finally, the expression for  $Y_t^2$  in terms of lagged macro variables and yields is:

$$Y_t^2 = A_2^* + \phi_{2o}^* F_{t-1}^o + \phi_{21}^* Y_{t-1}^1 + u_{2t}^* \quad (2.24)$$

with

$$\begin{aligned} A_2^* &= A_2 - B_{2u} B_{1u}^{-1} A_1 + (B_{2o} - B_{2u} B_{1u}^{-1} B_{1o}) A_o^* + B_{2u} B_{1u}^{-1} A_1^* \quad (2.25) \\ \phi_{2o}^* &= B_{2o} \phi_{oo}^* - B_{2u} B_{1u}^{-1} B_{1o} \phi_{oo}^* + B_{2u} B_{1u}^{-1} \phi_{1o}^* \\ \phi_{21}^* &= B_{2o} \phi_{o1}^* - B_{2u} B_{1u}^{-1} B_{1o} \phi_{o1}^* + B_{2u} B_{1u}^{-1} \phi_{11}^* \\ u_{2t}^* &= (B_{2o} - B_{2u} B_{1u}^{-1} B_{1o}) u_{ot}^* + B_{2u} B_{1u}^{-1} u_{1t}^* + B^m u_t^m \end{aligned}$$

Putting (2.17), (2.19) and (2.24) together with their non-linear restrictions on the parameters in (2.18), (2.20) and (2.25), we obtain the VAR representation of the baseline macro finance model:

$$\begin{aligned} F_t^o &= A_o^* + \phi_{oo}^* F_{t-1}^o + \phi_{o1}^* Y_{t-1}^1 + u_{ot}^* \quad (2.26) \\ Y_t^1 &= A_1^* + \phi_{1o}^* F_{t-1}^o + \phi_{11}^* Y_{t-1}^1 + u_{1t}^* \\ Y_t^2 &= A_2^* + \phi_{2o}^* F_{t-1}^o + \phi_{21}^* Y_{t-1}^1 + u_{2t}^* \end{aligned}$$

## B. Non-linear restrictions on the variance-covariance matrix

As shown in Appendix A, the error terms for  $F_t^o$ ,  $Y_t^0$  and  $Y_t^1$  in the VAR representation of the baseline macro finance model (2.26) are:

$$\begin{aligned} u_{ot}^* &= \Sigma_o u_t^o \quad (2.27) \\ u_{1t}^* &= B_{1o} u_{ot}^* + B_{1u} u_t^u = B_{1o} \Sigma_o u_t^o + B_{1u} u_t^u \\ u_{2t}^* &= (B_{2o} - B_{2u} B_{1u}^{-1} B_{1o}) u_{ot}^* + B_{2u} B_{1u}^{-1} u_{1t}^* + B^m u_t^m \\ &= (B_{2o} - B_{2u} B_{1u}^{-1} B_{1o}) u_{ot}^* + B_{2u} B_{1u}^{-1} (B_{1o} u_{ot}^* + B_{1u} u_t^u) + B^m u_t^m \\ &= B_{2o} \Sigma_o u_t^o + B_{2u} u_t^u + B^m u_t^m \end{aligned}$$

The variance-covariance matrix of the error terms in (2.27) are as follows:

$$\text{var} \begin{pmatrix} u_{ot}^* \\ u_{1t}^* \\ u_{2t}^* \end{pmatrix} = \begin{pmatrix} E(u_{ot}^* u_{ot}^{*'}) & E(u_{ot}^* u_{1t}^{*'}) & E(u_{ot}^* u_{2t}^{*'}) \\ E(u_{1t}^* u_{ot}^{*'}) & E(u_{1t}^* u_{1t}^{*'}) & E(u_{1t}^* u_{2t}^{*'}) \\ E(u_{2t}^* u_{ot}^{*'}) & E(u_{2t}^* u_{1t}^{*'}) & E(u_{2t}^* u_{2t}^{*'}) \end{pmatrix} \quad (2.28)$$

Individual components of the variance-covariance matrix in (2.28) take the following forms:

$$E(u_{ot}^* u_{ot}^{*'}) = \Sigma_o \Sigma_o'$$

$$E(u_{ot}^* u_{1t}^{*'}) = \Sigma_o \Sigma_o' B'_{1o}$$

$$E(u_{ot}^* u_{2t}^{*'}) = \Sigma_o \Sigma_o' B'_{2o}$$

$$E(u_{1t}^* u_{1t}^{*'}) = B_{1o} \Sigma_o \Sigma_o' B'_{1o} + B_{1u} B'_{1u}$$

$$E(u_{1t}^* u_{2t}^{*'}) = B_{1o} \Sigma_o \Sigma_o' B'_{2o} + B_{1u} B'_{2u}$$

$$E(u_{2t}^* u_{2t}^{*'}) = B_{2o} \Sigma_o \Sigma_o' B'_{2o} + B_{2u} B'_{2u} + B^m B^{m'}$$

## Tables

**Table 2.1:** Likelihood ratio tests

Unrestricted	VAR	VAR
Restricted	macro	macrolag
Ratio	898	1305
d.f.	153	131
p-value	0.00%	0.00%
Simulated 5% critical value	481	512

This table conducts the likelihood ratio tests. The unrestricted model is the VAR and the restricted models are the macro finance models from Ang and Piazzesi (2003). Row 3 through row 5 performs the conventional likelihood ratio tests:  $Ratio = -2(\ln L_R - \ln L_U)$ , d.f. = the number of restrictions. The restriction can be rejected, if the p-value is smaller than or equal to the significance level. The last row performs the simulation-based likelihood ratio tests adjusted for the small sample bias, with the simulated critical values at 5% significance level shown in the last row of the table. If the actual ratio is greater than the critical value, I can reject the restriction. The VAR denotes that five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true.



**Table 2.2:** AIC and BIC

	# of para	LLF	AIC	BIC
VAR	226	21137	-41822	-40838
macro	73	20688	-41230	-40912
macrolag	95	20470	-40750	-40336

This table describes the AIC and BIC of one VAR and two macro finance models. The information criteria are defined as  $AIC = 2K - 2 \ln(L)$ ,  $BIC = K \ln(T) - 2 \ln(L)$ . A smaller AIC or BIC means a better model. The macro finance models are the exact ones from Ang and Piazzesi (2003). The VAR denotes that five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true.

**Table 2.3:** RMSE

	1mth	3mth	12mth	36mth	60mth
VAR	<b>0.2730</b>	<b>0.1802</b>	0.1999	0.2507	0.2628
macro	0.3074	0.2013	<b>0.1799</b>	<b>0.2405</b>	<b>0.2477</b>
macrolag	0.3293	0.2721	0.2188	0.2599	0.2519

This table compares the VAR and Ang and Piazzesi (2003)'s macro finance models using RMSE. A smaller RMSE indicates a better model, and the best models for different maturities are indicated in bold. The procedure for examining the out-of-sample forecasts over the last 60 months of the sample is as follows. At each date  $t$ , I estimate the models using data up to and including time  $t$ , and then forecast the next month's yields at time  $t+1$ , and I do so for all the five yields available in data. The VAR denotes that five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true.

**Table 2.4:** likelihood ratio tests

Unrestricted	VAR(dep,12)	VAR(dep,12)	VAR(dep,1)
Restricted	MF(dep,12,lag)	MF(dep,12,macro)	MF(dep,1,macro)
Ratio	367	376	392
d.f.	40	62	18
p-value	0.00%	0.00%	0.00%
Unrestricted	VAR(ind,12)	VAR(ind,12)	VAR(ind,1)
Restricted	MF(ind,12,lag)	MF(ind,12,macro)	MF(ind,1,macro)
Ratio	1159	701	532
d.f.	121	143	33
p-value	0.00%	0.00%	0.00

This table conducts the likelihood ratio tests between the macro finance and VAR models. The ratio is defined as  $Ratio = -2(\ln L_R - \ln L_U)$ , and the degree of freedom (d.f.) is the number of the restrictions. Accept the restriction if the p-value exceeds the significance level, otherwise, the unrestricted version is preferred.

**Table 2.5:** AIC ranking

Model		<i>dep</i>	<i>p</i>	<i>n<sub>para</sub></i>	LLF	<b>AIC</b>	BIC
VAR(ind,12)	<b>VAR</b>	ind	12	226	16286	<b>-32121</b>	-31195
VAR(dep,1)	<b>VAR</b>	dep	1	84	16137	<b>-32107</b>	-31762
VAR(dep,12)	<b>VAR</b>	dep	12	238	16291	<b>-32105</b>	-31131
VAR(ind,1)	<b>VAR</b>	ind	1	72	16124	<b>-32104</b>	-31809
MF(dep,12,macro)	MF	dep	12	176	16103	<b>-31854</b>	-31133
MF(dep,12,lag)	MF	dep	12	198	16107	<b>-31819</b>	-31008
MF(dep,1,macro)	MF	dep	1	66	15941	<b>-31751</b>	-31480
MF(ind,12,macro)	MF	ind	12	83	15936	<b>-31705</b>	-31365
MF(ind,1,macro)	MF	ind	1	39	15858	<b>-31638</b>	-31478
MF(ind,12,lag)	MF	ind	12	105	15707	<b>-31204</b>	-30774

This table ranks the VAR and macro finance models based on the AIC, and the ranking preference is in a descending order throughout the table. The information criteria are defined as  $AIC = 2K - 2\ln(L)$ ,  $BIC = K\ln(T) - 2\ln(L)$ . A smaller AIC or BIC means a better model.

**Table 2.6:** BIC ranking

Model		<i>dep</i>	<i>p</i>	<i>n<sub>para</sub></i>	LLF	AIC	<b>BIC</b>
VAR(ind,1)	<b>VAR</b>	ind	<b>1</b>	72	16124	-32104	<b>-31809</b>
VAR(dep,1)	<b>VAR</b>	dep	<b>1</b>	84	16137	-32107	<b>-31762</b>
MF(dep,1,macro)	MF	dep	<b>1</b>	66	15941	-31751	<b>-31480</b>
MF(ind,1,macro)	MF	ind	<b>1</b>	39	15858	-31638	<b>-31478</b>
MF(ind,12,macro)	MF	ind	12	83	15936	-31705	<b>-31365</b>
VAR(ind,12)	VAR	ind	12	226	16286	-32121	<b>-31195</b>
MF(dep,12,macro)	MF	dep	12	176	16103	-31854	<b>-31133</b>
VAR(dep,12)	VAR	dep	12	238	16291	-32105	<b>-31131</b>
MF(dep,12,lag)	MF	dep	12	198	16107	-31819	<b>-31008</b>
MF(ind,12,lag)	MF	ind	12	105	15707	-31204	<b>-30774</b>

This table ranks the VAR and macro finance models based on the BIC, and the ranking preference is in a descending order throughout the table. The information criteria are defined as  $AIC = 2K - 2 \ln(L)$ ,  $BIC = K \ln(T) - 2 \ln(L)$ . A smaller AIC or BIC means a better model.

**Table 2.7:** RMSE

Models	1mth	3mth	12mth	36mth	60mth
VAR(dep,12)	0.3361	0.3100	0.2955	0.2603	0.2424
VAR(dep,1)	0.2795	0.2367	0.2548	0.2490	0.2300
VAR(ind,12)	0.3338	0.3089	0.2944	0.2601	0.2420
VAR(ind,1)	<b>0.2640</b>	<b>0.2248</b>	0.2392	<b>0.2431</b>	<b>0.2262</b>
MF(dep,12,lag)	0.3251	0.2757	0.2807	0.2689	0.2385
MF(dep,12,macro)	0.3255	0.2798	0.2817	0.2667	0.2364
MF(dep,1,macro)	0.2814	0.2339	0.2451	0.2664	0.2358
MF(ind,12,lag)	0.5245	0.4009	0.2250	0.2817	0.2328
MF(ind,12,macro)	0.3454	0.2492	0.2142	0.2571	0.2343
MF(ind,1,macro)	0.2794	0.2363	<b>0.2016</b>	0.2559	0.2339

This table compares various VAR and macro finance models using the RMSE. A smaller RMSE indicates a better model, and the best models for different maturities are indicated in bold. The procedure for examining out-of-sample forecasts over the last 44 months of the sample is as follows. At each date  $t$ , I estimate the models using data up to and including time  $t$ , and then forecast the next month's yields at time  $t+1$ , and I do so for all the five yields available in data.

**Table 2.8:** LLF ranking

yields in $Y_t^1$	LLF
3,12,60	21035
3,36,60	20981
1,12,60	20959
3,12,36	20904
1,36,60	20895
12,36,60	20893
1,12,36	20838
1,3,60	20792
1,3,36	20697
1,3,12	20419

This table ranks the LLFs of the VAR models with different maturities included in  $Y_t^1$ , and the ranking preference is in a descending order throughout the table. The first column indicates that for each model, which three yields are priced exactly. For all the VAR models, five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true.

**Table 2.9:** RMSE

Yields in $Y_t^1$	1mth	3mth	12mth	36mth	60mth
3,12,60	0.2758	0.1802	0.2010	0.2572	<b>0.2601</b>
3,36,60	0.2855	0.1800	0.2075	0.2520	0.2643
1,12,60	0.3067	0.2203	0.2054	0.2600	0.2613
3,12,36	0.2739	0.1802	<b>0.1987</b>	<b>0.2499</b>	0.2846
1,36,60	0.3060	0.2530	0.2307	0.2522	0.2645
12,36,60	0.4447	0.2943	0.2072	0.2551	0.2664
1,12,36	0.3069	0.2194	0.2037	0.2519	0.2857
1,3,60	0.2767	0.1797	0.2074	0.2573	0.2663
1,3,36	0.2756	0.1800	0.2001	0.2565	0.2908
1,3,12	<b>0.2720</b>	<b>0.1796</b>	0.2114	0.3442	0.3931

This table compares the VAR models with different maturities included in  $Y_t^1$  using RMSE. A smaller RMSE indicates a better model, and the best models for different maturities are indicated in bold. The procedure for examining the out-of-sample forecasts over the last 60 months of the sample is as follows. At each date  $t$ , I estimate the models using data up to and including time  $t$ , and then forecast the next month's yields at time  $t+1$ , and I do so for all the five yields available in data. For all the VAR models, five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true.

**Table 2.10:** log likelihood comparison

Yields in $Y_t^1$	VAR(dep,12)	VAR(dep,1)	VAR(ind,12)	VAR(ind,1)
3,12,60	<b>16222</b>	<b>16051</b>	<b>16219</b>	<b>16042</b>
3,36,60	16189	16030	16187	16020
1,12,60	16151	15982	16149	15973
3,12,36	16130	15949	16127	15941
1,36,60	16125	15969	16124	15959
12,36,60	16103	15952	16100	15941
1,12,36	16066	15888	16063	15880
1,3,60	16015	15836	16013	15827
1,3,36	15952	15769	15951	15761
1,3,12	15779	15584	15778	15579

The VAR models with different maturities included in  $Y_t^1$ , and it ranks the models based on the order in Table 2.8. The first column indicates that for each model, which three yields are priced exactly. Four different VAR specifications along the first two dimensions,  $VAR(dep/ind, 1/12)$ , are considered.

**Table 2.11:** RMSE

Yields in $Y_t^1$	1mth	3mth	12mth	36mth	60mth
12,36,60	0.3636	0.3023	0.2770	0.2571	0.2465
3,36,60	0.3212	0.3056	0.3144	0.2609	0.2528
3,12,60	0.3302	0.3194	0.3160	0.2820	0.2500
3,12,36	0.3317	0.3202	0.3289	0.2609	<b>0.2426</b>
1,36,60	0.3789	0.3516	0.3371	0.2611	0.2524
1,12,60	0.3555	0.3145	0.2994	0.2780	0.2516
1,12,36	0.3543	0.3112	0.3065	0.2591	0.2549
1,3,60	<b>0.3200</b>	0.3014	0.4348	0.2896	0.2540
1,3,36	0.3211	0.2936	0.3872	<b>0.2569</b>	0.3787
1,3,12	0.3269	<b>0.2779</b>	<b>0.2460</b>	0.7329	1.0482

This table compares the VAR models with different maturities included in  $Y_t^1$  using RMSE. A smaller RMSE indicates a better model, and the best models for different maturities are indicated in bold. The procedure for examining the out-of-sample forecasts over the last 60 months of the sample is as follows. At each date  $t$ , I estimate the models using data up to and including time  $t$ , and then forecast the next month's yields at time  $t+1$ , and I do so for all the five yields available in data. For all the VAR models, five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true. The VAR specification along the first two dimensions is  $VAR(dep, 12)$ .

**Table 2.12:** AIC and BIC

	# of para	LLF	AIC	BIC
VAR	226	21137	-41822	-40838
macro	73	20684	-41221	-40903
macrolag	95	20511	-40831	-40418

This table describes the AIC and BIC of one VAR and two macro finance models. The information criteria are defined as  $AIC = 2K - 2\ln(L)$ ,  $BIC = K \ln(T) - 2\ln(L)$ . A smaller AIC or BIC means a better model. The macro finance models are the same as the ones from Ang and Piazzesi (2003) except  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$ . The VAR denotes that five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true.

**Table 2.13:** RMSE

	1mth	3mth	12mth	36mth	60mth
VAR	<b>0.2730</b>	<b>0.1802</b>	0.1999	0.2507	0.2628
macro	0.3074	0.1999	<b>0.1719</b>	<b>0.2401</b>	<b>0.2475</b>
macrolag	0.4755	0.2729	0.2118	0.2501	0.2512

This table compares the symmetric VAR and Ang and Piazzesi (2003)'s macro finance models with  $Y_t^1 = (y_t^3, y_t^{12}, y_t^{60})'$  using RMSE. A smaller RMSE indicates a better model, and the best models for different maturities are indicated in bold. The procedure for examining the out-of-sample forecasts over the last 60 months of the sample is as follows. At each date  $t$ , I estimate the models using data up to and including time  $t$ , and then forecast the next month's yields at time  $t+1$ , and I do so for all the five yields available in data. The VAR denotes that five yields and two macro variables regress over 12 lags of macro variables and 1 lag of yields, and the dependence is one way, which means only the yields depend on the lagged macro variables and the reverse is not true.



## Chapter 3

# The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment

### Abstract

This paper reviews alternative options for monetary policy when the short-term interest rate is at the zero lower bound and develops new empirical estimates of the effects of the maturity structure of publicly held debt on the term structure of interest rates. We use a model of risk-averse arbitrageurs to develop measures of how the maturity structure of debt held by the public might affect the pricing of level, slope and curvature term-structure risk. We find these Treasury factors historically were quite helpful for predicting both yields and excess returns over 1990-2007. The historical correlations are consistent with the claim that if in December of 2006, the Fed were to have sold off all its Treasury holdings of less than one-year maturity (about \$400 billion) and use the proceeds to retire Treasury debt from the long end, this might have resulted in a 14-basis-point drop in the 10-year rate and an 11-basis-point increase in the 6-month rate. We also develop a description of how the dynamic behavior of the term structure of interest rates changed after hitting the zero lower bound in 2009. Our estimates imply that at the zero lower bound, such a maturity swap would have the same effects as

buying \$400 billion in long-term maturities outright with newly created reserves, and could reduce the 10-year rate by 13 basis points without raising short-term yields.

### 3.1 Introduction.

The key instrument of monetary policy is the interest rate on overnight loans between banks, which in normal times is quite sensitive to the quantity of excess reserves. However, since December 2008, the Fed's target for the fed funds rate has been essentially zero. The level of reserves, which had typically been around \$10 billion prior to the financial crisis, has been maintained in the neighborhood of a trillion dollars. Trying to lower the short-term interest rate or increase the volume of reserves any further offers little promise of boosting aggregate demand. With the Fed's traditional tools incapable of providing further stimulus to the economy, it is of considerable interest to ask what other options might be available to the central bank.

Our study begins by briefly reviewing some of the available options and the Fed's experience with using them. That analysis leads us to focus on one strategy in particular, which is to try to influence the term structure of interest rates through the maturity structure of securities acquired by open-market purchases.

A number of previous studies have reported evidence that the relative supplies of Treasury securities of different maturities are correlated with yield spreads; see for example Roley (1982), Bernanke et al. (2004), Kuttner (2006), Gagnon, Raskin, Remache and Sack (2010), Doh (2010), Greenwood and Vayanos (2010), D'Amico and King (2010), and Swanson (2011).<sup>1</sup> But using those correlations to infer potential effects of nonstandard open-market operations raises questions from the perspective of both economic theory, in terms of the proposed mechanism whereby the effects could possibly be generated, as well as from the perspective of econometric methodology, in terms of whether it is reasonable to place a causal interpretation on the correlations. Our paper makes contributions in both areas.

---

<sup>1</sup>Other closely related research includes Krishnamurthy and Vissing-Jorgensen (2010), Baumeister and Benati (2010), Kitchen and Chinn (2010), and Hancock and Passmore (2011).

Our theoretical motivation follows Vayanos and Vila (2009), who developed a promising framework for understanding how the supplies of assets of different maturities might influence their respective yields. Vayanos and Vila postulate the existence of two groups of investors. The willingness of preferred-habitat investors to buy securities of maturity  $n$  is presumed to be an increasing function of the yield on that asset. A second group, known as arbitrageurs, is willing to hold any assets based on a simple tradeoff between expected return and risk. The behavior of the second group generates no-arbitrage conditions relating the yields on different securities.

Our empirical analysis follows Doh (2010) and Greenwood and Vayanos (2010) in using the Vayanos and Vila (2009) framework to try to quantify the ability of nonstandard open-market operations to change the yields on assets of different maturities. We differ from these earlier researchers in making more use of the details of the framework to inform the empirical estimates, developing a discrete-time version of the model and relating it directly to maximum-likelihood estimates of the dynamic behavior of the term structure of interest rates. We develop specific historical measures of how the maturity structure of debt issued to the public might be expected to affect the pricing of level, slope, and curvature risk according to this framework, and show that our inferred Treasury risk factors were historically quite helpful in predicting yields and excess returns. For example, we find that over 1990-2007, the excess one-year return from holding 2-year Treasuries over 1-year Treasuries can be predicted with an  $R^2$  of 71% on the basis of traditional term-structure factors along with our proposed Treasury risk factors.

One of the challenges for estimating potential policy effects on the basis of historical correlations is the problem of endogeneity, in that the correlation between bond supplies and interest rates may reflect the response of the Treasury or the Fed to interest rates. We try to minimize this endogeneity bias by looking at forecasting rather than contemporaneous regressions and including the current level, slope, and curvature as additional explanatory variables in the regression. Our impact estimates are based on the incremental contribution of the Treasury maturity structure to a one-month-ahead forecast of interest rates beyond the

information already contained in the current term structure, so that insofar as the maturities of debt issued by the Treasury or purchased by the Fed are responding to current interest rates, that response could not account for our estimated effects. Our dynamic formulation also avoids the potential spurious regression problem that could arise in simple contemporaneous regressions that make no allowance for near-unit-root dynamics.

We use our estimated forecasting relations to analyze the outcome of the following policy change. Suppose the Federal Reserve were to sell off all of its holdings of Treasury securities of less than one-year duration, and use the proceeds to buy up all the outstanding Treasury debt it could at the long end of the yield curve. For example, in 2006 this would have involved a \$400 billion asset swap that would have retired all Treasury debt of more than 10-years duration. Our estimates imply that, in an environment not affected by the zero lower bound, this would have decreased the 10-year yield by 14 basis points and increased the 6-month yield by 11 basis points.

We next develop a framework for analyzing the behavior of interest rates when the short-term interest rate hits the zero lower bound. Our basic approach is to postulate that movements in longer-term yields in such a setting are explained by arbitrageurs' assumption that the economy will eventually break out of the zero lower bound, and that, once it does, short-term interest rates would again fluctuate in response to the same kind of forces as they did historically. We propose a very parsimonious description in which arbitrageurs assume that, apart from a possible downward shift in the average level, the post-ZLB dynamics will be the same as those observed in the pre-ZLB experience. Given an exogenous probability of exiting the ZLB in any given period, we then develop a no-arbitrage theory of how the term structure evolves dynamically when at the ZLB. We find this model provides a reasonable empirical description of the behavior of the term structure during 2009 and 2010.

We then use this model to revisit the question analyzed for the pre-2007 data. We find that, at the ZLB, an asset swap could continue to depress long-term yields by the same amount that it would in normal times, without producing

any rise in short-term yields. Thus, whereas swapping short-term for long-term assets has no consequences for the overall level of interest rates in normal times, it is an available tool for lowering the overall level at the ZLB. Moreover, since at the ZLB newly created reserves are essentially equivalent to short-term T-bills, direct large-scale asset purchases are a feasible tool that the Fed could use to lower long-term interest rates when at the ZLB.

The plan of the paper is as follows. Section 3.2 reviews alternative mechanisms whereby monetary policy might still be able to influence interest rates for an economy at the ZLB, and explains our reason for focusing in particular on the possible effects arising through changes in the maturity composition of outstanding debt. Section 3.3 develops a discrete-time version of the Vayanos and Vila (2009) framework for analyzing the nature of preferred-habitat asset markets and the pricing of term-structure risk. Section 3.4 provides details of our method for obtaining maximum-likelihood estimates of parameters, while Section 3.5 reviews the data set assembled for this study. In Section 3.6 we analyze the effects of nonstandard open-market operations in an environment of fluctuating short-term interest rates, while Section 3.7 extends the analysis to an economy in which the short-term rate is temporarily stuck at some lower bound. We briefly compare our results with other recent estimates in Section 3.8. Section 3.9 concludes.

## **3.2 Options for monetary stimulus at the zero lower bound.**

When the short-term interest rate gets all the way to zero, an open-market purchase of a short-term Treasury security with newly created base money represents an exchange of essentially equivalent assets. Such an exchange is obviously incapable of lowering the short-term rate any further, and it's not clear how the exchange could affect any economic magnitude of interest. Eggertsson and Woodford (2003) describe this as a situation in which the demand for money is completely satiated. With over a trillion dollars in excess reserves, the United States presently appears to be well past the satiation point for Federal Reserve deposits.

Even if the demand for reserve balances is presently satiated, as long as the situation is not permanent, at some future date the Fed will regain its ability to influence overnight rates. Thus even at the zero lower bound, Krugman (1998) and Eggertsson and Woodford (2003) propose that the central bank could mitigate the current problems by successfully communicating its commitment to reverse any decreases in the price level, embracing the higher future inflation rates necessary to achieve that. Although such a strategy holds appeal in theory, in practice it appears to be quite hard to achieve. For example, the top panel of Figure 3.1 plots the 5-year expected inflation rate implied by the difference between nominal and inflation-indexed U.S. Treasuries. This plunged in the fall of 2008, and has yet to recover to its pre-crisis levels. Five-year expected inflation has also declined according to the average response to the Survey of Professional Forecasters (bottom panel). The failure of the Fed to follow the theoretical policy prescription of trying to increase inflationary expectations in response to the crisis is not so much an indictment of the Fed as it is a clear demonstration that these expectations are far more difficult to control in practice than simple theoretical treatments might sometimes suppose.

If buying T-bills with newly created reserves has no effect, the Fed could buy some other assets which clearly are not perfect substitutes for cash. One obvious class of assets to consider purchasing would be those denominated in foreign currencies. If the Fed announced a commitment to buy such assets without limit until the dollar depreciated, it is hard to imagine real-world market forces that could prevent the goal from being achieved. In terms of theoretical models, the ability of the Fed to make good on such a commitment could arise from a portfolio balance effect (McCallum (2000)), or the announcement could serve as an expectations coordinating mechanism (Svensson (2001)). In either case, it certainly seems one practical tool for preventing deflation even if no others are available.

If the private sector were indeed indifferent between holding freely created reserves and long-term Treasury debt, one wonders why the Federal Reserve wouldn't want to buy up the entire stock of outstanding public debt, thereby eliminating the need for future taxes to service that debt. A related question is why the

government would choose to use taxes rather than money creation as the means to pay for any of its current or projected future expenditures. Auerbach and Obstfeld (2005) explore the possible expansionary effects resulting from reducing the distortionary effect of taxes.

In the actual U.S. experience over 2008-2010, the Federal Reserve doubled the size of its balance sheet, buying two broad classes of assets (see Figure 3.2). In the first year of the crisis, the Fed was aggressively extending loans through a variety of new facilities such as the Term Auction Facility (essentially a term discount window open to all depository institutions on an auction basis), foreign currency swaps (used to assist foreign central banks in lending dollars), and the Commercial Paper Funding Facility (which helped provide loans for issuers of commercial paper). These measures could matter both in terms of making these markets more liquid (in the sense of reducing bid-ask spreads) as well as potentially absorbing some default risk onto the Fed's balance sheet. Christensen, Lopez and Rudebusch (2009), McAndrews, Sarkar and Wang (2008), Taylor and Williams (2009), Adrian, Kimbrough and Marchioni (2010) and Duygan-Bump, Parkinson, Rosengren, Suarez and Willen (2010) provide empirical assessments of the effectiveness of such measures.

Beginning in March 2009, these lending facilities began to be unwound and replaced by the gradual purchase of up to \$1.1 trillion in mortgage-backed securities, along with \$160B in agency debt and \$300B in new holdings of Treasury bonds with greater than one year maturity. Although rates on MBS and agency debt might be argued to include a default premium, with the de facto nationalization of Fannie and Freddie, it seems most natural to regard the effect of these purchases as coming from a change in the relative supply of longer-term assets.<sup>2</sup> As this has become the most important tool going forward, our analysis in this paper focuses on the potential of such operations to alter the term structure of interest rates.

The mechanism by which such asset purchases might have an effect is very different from that characterizing traditional open-market operations. The Federal

---

<sup>2</sup>Hancock and Passmore (2011) nevertheless found evidence that the MBS purchases did lower the premium on MBS relative to Treasury securities.

Reserve is the monopoly supplier of reserves held by depository institutions and currency held by the public, and the supply it creates of these assets unquestionably has consequences under normal economic conditions. However, when the demand for such assets is satiated, it is not clear that anything the Fed does could affect the pricing kernel determining other yields. For example, Eggertsson and Woodford (2003) elaborate conditions under which an open-market purchase of any asset whatever would have zero consequences for variables such as real output and the price level provided that it has no implications for the future conduct of monetary or fiscal policy. Woodford (2010) notes that if the operations have no effect on the asset's state-contingent income stream or on the state-contingent aggregate supply of goods available for consumption, they should have no effect on the price of the asset.

Certainly from the perspective of an individual investor, a 10-year Treasury bond has different risk characteristics from a 6-month T-bill, and these differences get priced by the market. If an individual investor changes her relative holdings of these assets, she perceives herself to have a different risk exposure, and perceives the U.S. Treasury to be the counterparty. By focusing on this aspect of bond pricing, as our paper does, our answer to the Eggertsson-Woodford critique is that, if the government changes the maturity structure of its outstanding debt, it is in fact committing to a different state-contingent path for spending, taxes, or inflation in order to maintain intertemporal budget balance under the altered debt structure.

Changing the risk exposure of the holders of government debt appears to be the key mechanism whereby changes in the maturity structure of government debt would be able to influence the term structure of interest rates in a class of formal descriptions of the portfolio-balance effect. We illustrate this point by developing in the following section a discrete-time version of the framework recently proposed by Vayanos and Vila (2009). This exercise both clarifies the mechanism whereby nonstandard open market operations could affect the term structure, and also suggests particular empirical measures that prove to be helpful for quantifying plausible sizes for these effects.



### 3.3 Preferred-habitat investing and market arbitrage.

Vayanos and Vila (2009) propose that the investors we will refer to as “arbitrageurs” care only about the mean and variance of  $r_{t,t+1}$ , the rate of return between  $t$  and  $t + 1$  on their total portfolio<sup>3</sup>:

$$E_t(r_{t,t+1}) - (\gamma/2)\text{Var}_t(r_{t,t+1}). \quad (3.1)$$

If  $y_{1t}$  denotes the return on a risk-free asset, arbitrageurs will choose portfolio weights such that for any asset with a risky yield  $r_{i,t,t+1}$ ,

$$y_{1t} = E_t(r_{i,t,t+1}) - \gamma\vartheta_{it} \quad (3.2)$$

where  $\vartheta_{it}$  is  $(1/2)$  the derivative of total portfolio variance with respect to holdings of asset  $i$ .

Consider a pure-discount  $n$ -period bond that is free of default risk, the log of whose price at date  $t$  (denoted  $p_{nt}$ ) is conjectured to be an affine function of a vector of  $J$  different macroeconomic factors (denoted  $f_t$ ),

$$p_{nt} = \bar{a}_n + \bar{b}'_n f_t. \quad (3.3)$$

The risk-free one-period rate is a function of the same factors,

$$y_{1t} = a_1 + b'_1 f_t,$$

where  $y_{1t} = -p_{1t}$ ,  $a_1 = -\bar{a}_1$ , and  $b_1 = -\bar{b}_1$ . Although these bonds have no default risk, the future pricing factors  $f_{t+s}$  are not known with certainty at date  $t$ , and so there is an uncertain one-period holding yield associated with buying the  $n$ -period bond at date  $t$  and selling the resulting  $(n - 1)$ -period bond at date  $t + 1$  given by

$$r_{n,t,t+1} = \exp\left(\bar{a}_{n-1} + \bar{b}'_{n-1} f_{t+1} - \bar{a}_n - \bar{b}'_n f_t\right) - 1. \quad (3.4)$$

---

<sup>3</sup>Vayanos and Vila (2009) assume that arbitrageurs maximize an objective function that is quadratic in the change in wealth rather than in the rate of return as here. Although their specification may have more theoretical appeal, their parameterization would be more difficult to bring to the data in the manner we propose here for an economy in which there is a trend in the level of wealth.

Suppose that the pricing factors follow a VAR(1) process,

$$f_{t+1} = c + \rho f_t + \Sigma u_{t+1} \quad (3.5)$$

with  $u_t \sim$  i.i.d.  $N(0, I_J)$ , and that the arbitrageurs hold a fraction  $z_{nt}$  of their portfolio in the bond of maturity  $n$ , so that the return on their portfolio is given by

$$r_{t,t+1} = \sum_{n=1}^N z_{nt} r_{n,t,t+1}.$$

Then, as we detail in Appendix A, an approximation to the portfolio optimization problem results in the following implication of (3.2) for each maturity  $n$ :

$$-\bar{a}_1 - \bar{b}'_1 f_t = \bar{a}_{n-1} + \bar{b}'_{n-1} (c + \rho f_t) + (1/2) \bar{b}'_{n-1} \Sigma \Sigma' \bar{b}_{n-1} - \bar{a}_n - \bar{b}'_n f_t - \bar{b}'_{n-1} \Sigma \lambda_t \quad (3.6)$$

$$\lambda_t = \gamma \Sigma' d_t \quad (3.7)$$

$$d_t = \sum_{n=2}^N z_{nt} \bar{b}_{n-1}. \quad (3.8)$$

If the number of maturities  $N$  is greater than the number of factors  $J$ , equation (3.6) implies a set of restrictions that bond prices must satisfy as a result of the actions of arbitrageurs, who will price factor  $j$  risk the same way no matter which bonds it may be reflected in.

Vayanos and Vila close the model by postulating that other credit market participants may have a particular preference for bonds of a given maturity. They present examples in which the borrowing demand from these participants for bonds of maturity  $n$ , denoted  $\xi_{nt}$ , is a decreasing affine function of the yield  $y_{nt}$ . In our application, we will express these demands relative to  $W_t$ , the net wealth of the arbitrageurs:

$$\xi_{nt}/W_t = \zeta_{nt} - \alpha_n y_{nt}.$$

Thus  $\zeta_{nt}$  reflects the overall level of preferred-habitat borrowing of bonds of maturity  $n$  and  $\alpha_n$  the sensitivity of this demand to the interest rate. Equilibrium

then requires that the net borrowing by the preferred-habitat sector equals the net lending from the arbitrage sector:

$$z_{nt} = \zeta_{nt} - \alpha_n y_{nt}. \quad (3.9)$$

Suppose that  $\zeta_{nt}$  is also an affine function of  $f_t$ . We show in Appendix B that in equilibrium,

$$\lambda_t = \lambda + \Lambda f_t. \quad (3.10)$$

Substituting (3.10) into (3.6), we see that

$$\bar{b}'_n = \bar{b}'_{n-1} \rho^Q - b'_1 \quad (3.11)$$

$$\rho^Q = \rho - \Sigma \Lambda \quad (3.12)$$

$$\bar{a}_n = \bar{a}_{n-1} + \bar{b}'_{n-1} c^Q + (1/2) \bar{b}'_{n-1} \Sigma \Sigma' \bar{b}_{n-1} - a_1 \quad (3.13)$$

$$c^Q = c - \Sigma \lambda. \quad (3.14)$$

### 3.4 Estimation of Affine-Term-Structure Models.

Equations (3.11) through (3.14) will be recognized as the no-arbitrage conditions for a standard affine-term-structure model (e.g., equations (17) in Ang and Piazzesi, 2003). Thus the Vayanos-Vila formulation can be viewed as one explanation for the origins of affine prices of risk. In this section we describe how we estimated parameters for this class of models; for further details see Appendix C.

Let  $y_{nt}$  denote the yield and  $p_{nt}$  the log price on an  $n$ -period pure discount bond, which are related by  $y_{nt} = -n^{-1} p_{nt}$ . From (3.3),

$$y_{nt} = a_n + b'_n f_t \quad (3.15)$$

with  $a_n = -\bar{a}_n/n$  and  $b_n = -\bar{b}_n/n$ . In the models we estimate, the factors  $f_t$  are represented by a  $(J \times 1)$  vector of observed variables, whose dynamic parameters  $c$  and  $\rho$  can be obtained from OLS estimation of (3.5). We suppose that we have available a set of  $M$  different observed yields  $Y_{2t} = (y_{n_1,t}, y_{n_2,t}, \dots, y_{n_M,t})'$  whose values differ from the theoretical prediction (3.15) by measurement error

$$Y_{2t} = A + Bf_t + \Sigma_e u_t^e \quad (3.16)$$

with  $u_t^e \sim N(0, I_M)$ . We assume that the measurement error  $u_t^e$  is independent of the factor innovation  $u_t$  in (3.5), but otherwise the structure of  $\Sigma_e$  does not affect the estimation procedure—full-information maximum-likelihood estimates of all parameters other than  $\Sigma_e$  will be numerically identical regardless of whether the matrix  $\Sigma_e$  is assumed to be diagonal.

Our estimates come from the minimum-chi-square estimation algorithm proposed by Hamilton and Wu (2010b) which allows OLS to do the work of maximizing the joint likelihood function and uses the theoretical model to translate those OLS estimates back into the asset-pricing parameters of interest. Note that the structure of (3.5) and (3.16) implies that OLS equation by equation is the most efficient procedure for estimation of these reduced-form parameters. In the special case of a just-identified model, in which the number of observed yields  $M$  is one more than the number of factors  $J$ , there is an exact solution for the parameters of interest in terms of these OLS coefficients, and the resulting estimates are numerically identical to those that would be obtained by maximization of the joint likelihood function  $f(Y_{2T}, f_T, Y_{2,T-1}, f_{T-1}, \dots, Y_{21}, f_1 | Y_{20}, f_0)$  with respect to the parameters of the affine-term-structure model, namely,  $c, \rho, \Sigma, c^Q, \rho^Q, b_1, a_1$  and  $\Sigma^e$ .

Among other advantages, this approach allows us to recognize instantly whether estimates represent a local rather than a global maximum to the likelihood function, and makes it feasible to calculate small-sample confidence intervals for any function of the parameters of interest, by simulating a thousand different samples for  $\{f_t, Y_{2t}\}_{t=1}^T$  from a postulated structure and calculating the estimates that result from the proposed procedure on each separate artificial sample.

### 3.5 Data.

Our baseline estimates use weekly observations for  $y_{nt}$ , based on constant-maturity Treasury yields as of Friday or the last business day of the week as reported in the FRED database of the Federal Reserve Bank of St. Louis.<sup>4</sup> We supplement this with monthly analysis of holding yields on securities of nonstandard maturities, for which we construct constant-maturity yields from the daily term-structure parameterization of Gürkaynak, Sack and Wright (2007) as of the last day of the month.<sup>5</sup>

We also constructed estimates of the face value of outstanding U.S. Treasury debt at each weekly maturity as of the end of each month between January 1990 and December 2009 as detailed in Appendix E. For purposes of the pure theory sketched above, we would want to interpret each semiannual coupon on a given bond as its own separate zero-coupon security (paying  $\$C$  at some time  $t + s$ ) and construct the market value of the bond as the sum of the market value of its individual components, each coupon viewed as a separate pure-discount bond. However, converting the face value into a market value by this device would be quite unsatisfactory for our larger purpose of identifying exogenous sources of variation in the supply of outstanding securities at different maturities. The true market value of a given security would be highly endogenous with respect to changes in interest rates, whereas the face value, by construction, is not.<sup>6</sup> Note moreover that, when issued, the face value of the original coupon bond should be close to the market value of the sum of its individual stripped components. For these reasons, we regard the face value as reported by the Treasury and the Fed to

---

<sup>4</sup>The 30-year yields are unavailable for 2002/2/19 to 2006/2/8. Over this interval we used instead the 20-year rate minus 0.21, which is the amount by which the 20-year rate exceeded the 30-year rate both immediately before and after the gap.

<sup>5</sup>Specifically, we calculated  $y_{nt}$  from their equations (6) and (9) as

$$y_{nt} = \beta_{0t} + n^{-1}\beta_{1t}\tau_{1t}[1 - \exp(-n/\tau_{1t})] + \beta_{2t}\tau_{1t}\{1 - [1 + (n/\tau_{1t})]\exp(-n/\tau_{1t})\} \\ + \beta_{3t}\tau_{2t}\{1 - [1 + (n/\tau_{2t})]\exp(-n/\tau_{2t})\}$$

using daily values for the parameters  $\{\beta_{0t}, \beta_{1t}, \beta_{2t}, \tau_{1t}, \tau_{2t}\}$  downloaded from <http://www.federalreserve.gov/econresdata/researchdata.htm>.

<sup>6</sup>Greenwood and Vayanos (2010) deal with this issue by stripping coupons off and converting from face value to present value using the historical average short rate.

be the better measures to use for our purposes, and simply use the number of remaining weeks to maturity on any given series as the value for  $n$ .

We separately constructed rough estimates of how much of the security of each maturity was held by the Federal Reserve, as detailed in Appendix E. The resulting data structures for outstanding Treasury debt and Fed holdings take the form of  $(240 \times 1577)$  matrices, with rows corresponding to months (ranging from January 31, 1990 to December 31, 2009) and columns corresponding to maturity in weeks up to 30 years. Figure 3.3 displays the information from the December 31, 2006 rows of these two matrices. Figure 3.4 provides a sense of some of the time-series variation, plotting the average maturity of debt held by the public for each month.<sup>7</sup> Average maturity dropped temporarily in the mid-1990s and began a more significant and sustained decrease after 2001. Average maturity dropped sharply between September 2007 and October 2008, but has since reverted back to September 2007 levels.

### 3.6 The term structure of interest rates prior to the financial crisis.

In our baseline specification, we took the  $J = 3$  observed factors to be the deviations from the sample mean of the level, slope, and curvature of the term structure implied by the 6-month, 2-year, and 10-year Treasuries<sup>8</sup>, sampled weekly from January 1990 through the end of July, 2007. These yields and the 3 implied factors are plotted in Figure 3.5. The level factor trended down over this period, with pronounced dips after the recessions of 1990-91 and 2001. During these episodes, the term structure also sloped up more than usual and the curvature increased as the 2-year yield fell away from the 10-year. The parameters  $c$ ,  $\rho$  and  $\Sigma$  reported in Table 3.1 were estimated by OLS regressions of each factor on a constant and lagged values of the other three factors. We chose  $M = 4$  other

<sup>7</sup>The graph plots  $\sum_{n=1}^N nz_{nt}$  for each  $t$ .

<sup>8</sup>That is, if maturities were measured in weeks, prior to demeaning we would have  $f_{1t} = (1/3)(y_{26,t} + y_{104,t} + y_{520,t})$ ,  $f_{2t} = y_{520,t} - y_{26,t}$ , and  $f_{3t} = y_{520,t} - 2y_{104,t} + y_{26,t}$ .

yields<sup>9</sup> (the 3-month, 1-year, 5-year, and 30-year) in the vector  $Y_{2t}$  in order to estimate the parameters  $c^Q, \rho^Q, a_1, b_1$  and  $\Sigma_e$  from equation (3.16). We measured  $f_t$  in annual percentage points to keep reporting units natural and measured  $y_{nt}$  in weekly discount units so that the asset-pricing recursions all hold as written; for example, a 5.2% continuously compounded annual rate would correspond to  $f_{1t} = 5.2$  and  $y_{nt} = 0.001$ .

The model described in Section 3.3 implies that an objective forecast (sometimes referred to as the  $P$ -measure expectation) of the 3 factors is given by

$$E_t^P(f_{t+1}) = c + \rho f_t.$$

However, as a result of risk aversion, arbitrageurs value assets the way a risk-neutral investor would if that investor believed that the forecast was instead characterized by the  $Q$ -measure expectation

$$E_t^Q(f_{t+1}) = c^Q + \rho^Q f_t.$$

The risk premium is the difference between these two forecasts,

$$\begin{aligned} E_t^P(f_{t+1}) - E_t^Q(f_{t+1}) &= \Sigma\lambda + \Sigma\Lambda f_t \\ &= \Sigma\lambda_t. \end{aligned}$$

---

<sup>9</sup>Note that this approach does not make full use of all the available information, in that we do not impose any connection between the model-implied value for

$$y_{520,t} - y_{26,t} = a_{520} - a_{26} + b'_{520}f_t - b'_{26}f_t$$

and the observed value of  $f_{2t}$  itself. However, the smooth structure of the ATSM causes these restrictions to be approximately satisfied even without imposing them, that is, the estimates reported below are characterized by

$$\begin{bmatrix} \hat{b}_{26} \\ \hat{b}_{104} \\ \hat{b}_{520} \end{bmatrix} \approx \begin{bmatrix} (1/3) & (1/3) & (1/3) \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -(1/2) & (1/6) \\ 1 & 0 & -(1/3) \\ 1 & (1/2) & (1/6) \end{bmatrix}.$$

It is possible instead to apply the minimum-chi-square algorithm to a system imposing restrictions such as the above equation directly. The effect of adding this restriction (along with the analogous expressions for level and curvature) is to fix the values of  $\rho^Q$  and  $b_1$  up to the eigenvalues of  $\rho^Q$ , which eigenvalues are then estimated from (3.16). We applied this approach to several of the systems examined below and obtained almost identical results to those from the simpler approach that ignores these restrictions. To minimize the computational and expositional burden, we only report here the estimates from the unrestricted version of the model.

At the beginning of the sample, investors behaved as if they expected next week's level of interest rates to be about 3 basis points higher (at an annual rate) than an objective forecast would imply, though this risk premium had mostly vanished by the end of the sample. Throughout the sample, arbitrageurs acted as if they expected the slope to be flatter than it usually turned out to be, and often expected the 10-year-2-year spread to move closer to the 2-year-6-month spread.

We next consider how the term-structure risk factors would be priced according to the Vayanos-Vila framework under the following special case. Suppose that (1) the preferred-habitat sector consisted solely of the U.S. Treasury and Federal Reserve, (2) the arbitrageurs comprise the entire private sector, and (3) U.S. Treasury debt is the sole asset held by arbitrageurs. These are obviously extreme assumptions, but they have the benefit of implying a clear answer to how changes in the maturity structure of outstanding Treasury debt would influence the price of risk in one highly stylized case. Under these conditions, the arbitrageurs' portfolio weights  $z_{nt}$  could be measured directly from the ratio of debt held by the public of maturity  $n$  to the total outstanding publicly held debt at that date. From equations (3.7) and (3.8), we would then predict that  $\Sigma\lambda_t = \gamma\Sigma\Sigma' \sum_{n=2}^N z_{nt}\bar{b}_{n-1}$ . Define

$$q_t = 100\Sigma\Sigma' \sum_{n=2}^N z_{nt}\bar{b}_{n-1} \quad (3.17)$$

where a value of  $\gamma = 100$  was assumed in order to bring the series roughly on the same scale as  $\Sigma\lambda_t$ . This series for  $q_t$  was calculated with the values  $\bar{b}_n$  calculated from equation (3.11) for  $\rho^Q$ , and  $b_1$  reported in Table 3.1. The values for the 3 elements of  $q_t$  are highly correlated, though as we shall see shortly, there is statistically useful information in the difference between them.

If the strong assumptions detailed above were literally true, then the vector  $q_t$  would be proportional to the corresponding series plotted in Figure 3.5, and indeed the level, slope, and curvature of the term structure could be described solely in terms of changes in the maturity composition of the public debt as summarized by these three factors. Obviously the assumptions do not hold, and the maturity composition of outstanding Treasury debt is just one of many factors potentially contributing to interest rate moves. However, it is interesting to look at what



connections there may be in the data between  $q_t$  and pricing of interest-rate risk. Before doing so, we emphasize that although the above theory suggests that  $q_t$  might be related to the behavior of interest rates, in terms of how the series is constructed mechanically from the data, the time-series variation in  $q_t$  is driven solely by changes in the composition of Treasury debt  $z_{nt}$  and not at all by changes in interest rates. We accordingly propose the vector  $q_t$  as a possible 3-dimensional summary statistic of how the maturity composition of Treasury debt changes over time, where the simple theory sketched above suggests that this might be a summary statistic of interest for purposes of analyzing changes over time in the term structure of interest rates.

We begin by examining the ability to predict excess holding yields for bonds of different maturities. Let  $p_{mt}$  denote the log price of a pure-discount  $m$ -month bond purchased on the last day of month  $t$ .<sup>10</sup> The  $k$ -month holding yield for the bond (quoted at an annual rate) is  $(12/k)(p_{m-k,t+k} - p_{mt})$ . This compares with the holding yield for a  $k$ -month bond of  $(12/k)(p_{0,t+k} - p_{kt}) = (12/k)(-p_{kt})$ . Let  $h_{mkt}$  denote the excess holding yield for an  $m$ -month relative to a  $k$ -month bond:

$$h_{mkt} = (12/k)(p_{m-k,t+k} - p_{mt} + p_{kt}).$$

We explored regressions to predict these holding yields on the basis of information available at date  $t$ :

$$h_{mkt} = c_{mk} + \beta'_{mk} f_t + \gamma'_{mk} x_t + u_{mkt}. \quad (3.18)$$

If investors were risk-neutral, all the coefficients in (3.18) would be zero. Our finding of nonzero elements for  $\lambda$  and  $\Lambda$  in Table 3.1 (and a huge literature before us) suggests nonzero values for  $c_{mk}$  and  $\beta_{mk}$ , though if the market pricing of risk were fully captured by the 3-factor affine-term-structure model, no other variables  $x_t$  should enter statistically significantly.<sup>11</sup>

<sup>10</sup>We inferred these prices from the daily term-structure summaries of Gürkaynak et al. (2007).

<sup>11</sup>Although  $u_{mkt}$  is uncorrelated with the regressors in (3.18), it is not independent of the regressors, and thus OLS is subject to the small-sample problems highlighted by Stambaugh (1999). Moreover, given that risk-neutrality does not hold, both the left-hand and right-hand variables in (3.18) are highly serially correlated, raising potential spurious regression concerns if these are near-unit-root processes.

Table 3.2 reports the results from OLS estimation of (3.18), giving the  $R^2$  of the regression and Newey-West tests of the hypothesis that  $\gamma_{mk}$  or subsets of  $\gamma_{mk}$  are zero for various specifications of  $x_t$ .<sup>12</sup> The first row reproduces the well-known result that the traditional level, slope, and curvature factors  $f_t$  can predict a significant amount of the excess holding yield on assets of assorted maturities, with for example an  $R^2$  of 0.33 in the case of predicting the excess returns from holding a 2-year bond for one year. The second row adds the average maturity of outstanding debt,

$$z_t^A = \sum_{n=1}^N n z_{nt},$$

which was one of the summary statistics examined by Greenwood and Vayanos (2010),<sup>13</sup> but which we find in our sample usually does not have statistically significant additional predictive power beyond that contained in  $f_t$ . On the other hand, the other measure they propose, the fraction of outstanding debt of more than 10-year maturity,

$$z_t^L = \sum_{n=521}^N z_{nt},$$

does statistically significantly predict excess returns.

One could consider various other linear combinations of  $\{z_{nt}\}_{n=1}^N$  as possible predictors, such as the first three principal components. We find in the fourth row of Table 3.2 that these are helpful for forecasting the holding returns on short-maturity assets, but are generally inferior to  $z_t^A$  or  $z_t^L$ .

The theory sketched above suggests three particular linear combinations of  $\{z_{nt}\}_{n=1}^N$  that should matter for term premia, namely the three elements of the vector  $q_t$  in (3.17). The sixth row of Table 3.2 shows that these turn out to be incredibly useful for predicting holding returns, with an  $R^2$  as high as 0.71 in the

---

<sup>12</sup>Note that even though the excess holding yield would follow an  $MA(k-1)$  process under the null hypothesis of risk neutrality, one would still need to let the Newey-West lag parameter go to infinity as the sample size grows in order to get a consistent estimate. The Newey-West approach is helpful under the alternative hypothesis of a possibly more complex serial correlation, and generates a positive-definite variance-covariance matrix by construction. We also performed these calculations using Hansen-Hodrick (1980) standard errors based on  $k-1$  lags. These produced the same results except for one case in which the Hansen-Hodrick standard error was negative.

<sup>13</sup>Greenwood and Vayanos (2010) use duration rather than maturity.

case of predicting the 2-over-1 excess return. The contribution of  $q_t$  is statistically significant for every maturity, even if the regression already includes both  $f_t$  and the first three principal components of  $\{z_{nt}\}_{n=1}^N$ .

Cochrane and Piazzesi (2005) propose a particular yield pricing factor that they have found very helpful for forecasting excess holding returns. In our application, we confirm that this factor<sup>14</sup> provides a statistically significant improvement over using just  $f_t$  alone (row 5 of Table 3.2). Nevertheless, our Treasury factors  $q_t$  still provide a very dramatic improvement in forecasting ability beyond that contained in  $f_t$  and the Cochrane-Piazzesi factor  $v_t$  (row 8).

We next examine the ability of the Treasury factors  $q_t$  to help predict the yields themselves, examining OLS regressions of the form

$$f_{t+1} = c + \rho f_t + \phi q_t + \varepsilon_{t+1} \quad (3.19)$$

for  $\phi$  a  $(3 \times 3)$  matrix. The first column of Table 3.3 reports that the vector  $q_t$  makes a useful contribution to predicting each of the term-structure factors, with the hypothesis that the  $i$ th row of  $\phi$  is zero being rejected for each  $i$ .

It is then tempting to use (3.19) to draw tentative conclusions about what the effects on yields of different maturities might be of a change in the composition of publicly held debt. Such calculations are subject to a well-understood endogeneity problem: historical variations in  $z_{nt}$  may have represented a response by the Treasury or the Fed to overall economic conditions or to term-structure developments in particular. Although this is also a potential concern for (3.19), our formulation has three advantages over traditional regressions which simply examine the contemporaneous correlations. First, any contemporaneous response of  $q_t$  to  $f_t$  could not account for a nonzero value of  $\phi$  in (3.19). We are explicitly asking about the ability of  $q_t$  to forecast future  $f_{t+1}$  over and above any information contained in  $f_t$  itself.<sup>15</sup> Second, because the statistics we report represent the answer to well-posed forecasting questions, the results have independent interest

<sup>14</sup>In our application, we constructed  $v_t$  from the fitted value of a regression of  $(1/4)(h_{24,12,t} + h_{36,12,t} + h_{48,12,t} + h_{60,12,t})$  on a constant and the 1- through 5-year forward rates at date  $t$ .

<sup>15</sup>On the other hand, if  $q_t$  only matters for  $f_{t+1}$  through its effect on  $f_t$ , we might understate the contribution of  $q_t$  using our approach.

as objective summaries of those forecasting relations, regardless of what the underlying dynamic structural relations may be. Third, because we include lags of the dependent variable in the regression, we avoid the potential spurious regression problem that could plague other popular approaches such as trying to use OLS to estimate a relation of the form  $f_t = \alpha + \beta z_t^A$ .

For purposes of focusing on a particular forecasting question that might be of interest to policy makers, we consider the following exercise. Suppose that at the end of month  $t$ , the Federal Reserve were to sell all its Treasury securities with maturity less than 1 year, and use the proceeds to buy up all of the outstanding nominal Treasury debt of maturity greater than  $n_{1t}$ , where  $n_{1t}$  would be determined by the size of the Fed's short-term holdings and outstanding long-term Treasury debt at time  $t$ . For example, if implemented in December of 2006, this would result in the Fed selling about \$400 B in short-term securities and buying about \$400 B in long-term securities, effectively retiring all the federal debt of ten-year and longer maturity. We then calculated what  $q_t^A$  would be under this counterfactual scenario, and calculated the average historical value of  $q_t^A - q_t$ , which turns out to be

$$\Delta = \begin{bmatrix} 0.026101 \\ 0.022712 \\ -0.00780 \end{bmatrix}. \quad (3.20)$$

We then asked, by how much would one expect  $f_{t+1}$  to change according to (3.19) if  $q_t$  were to change by  $\Delta$ ? As should be clear from the description of the exercise, we are talking about a quite dramatically counterfactual event. If one considers the analogous forecasting equations of the form  $q_{t+1} = c_q + \rho_q f_t + \phi_q q_t + \varepsilon_{q,t+1}$ , a change of  $q_t$  of the size of  $\Delta$  would represent a  $36\sigma$  event, obviously something so far removed from anything that was attempted during the historical sample as to raise doubts about interpreting the parameter estimates as telling policy makers what would happen if they literally implemented a change of this size.

The second column of Table 3.3 reports how a forecast of the traditional term-structure factors would be affected by this change. We find that changing  $q_t$  by this amount could flatten the slope of the yield curve by 25 basis points, with no effect on the level of interest rates themselves. If it reduces the slope

but has no effect on the level, that means it would reduce long-term yields and raise short-term yields. Indeed, our 3-factor ATSM has a prediction<sup>16</sup> as to how much any given interest rate would change if the factors were to change by the amount specified in Table 3.3, which predicted responses we plot in Figure 3.8. Yields on maturities longer than 2-1/2 years would fall, with those at the long end decreasing by up to 17 basis points. Yields on the shortest maturities would increase by almost as much.<sup>17</sup>

A separate question from the feasibility for the Federal Reserve to achieve such effects is the desirability of its attempting to do so. Although we have described this as a Fed operation, it is probably more natural to think of it as a Treasury operation, implemented by the Treasury doing more of its borrowing at the shorter end of the yield curve. According to the simple framework that motivated our definition of  $q_t$ , the average slope of the yield curve arises from the preference of the U.S. Treasury for doing much of its borrowing with longer-term debt. For reasons presumably having to do with management of fiscal risks, the Treasury is willing to pay a premium to arbitrageurs for the ability to lock in a long-term borrowing cost. If the Treasury has good reasons to avoid this kind of interest-rate risk, it is not clear why the Federal Reserve should want to absorb it.

Our conclusion is that, although it appears to be possible for the Fed to influence the slope of the yield curve in normal times through the maturity of the System Open Market Account holdings, very large operations are necessary to have an appreciable immediate impact. If there is no concern about a zero-lower-bound constraint, this potential tool should clearly be secondary to the traditional focus of open-market operations on the short end of the yield curve.

---

<sup>16</sup>The predicted change in  $y_{nt}$  is given by  $b'_n \hat{\phi} \Delta$  for  $b_n = -\bar{b}_n/n$ ,  $\bar{b}_n$  calculated from equation (3.11) using the values of  $\rho^Q$  and  $b_1$  reported in Table 3.1,  $\hat{\phi}$  the OLS estimates from equation (3.19), and  $\Delta$  given by (3.20).

<sup>17</sup>Our estimates would also allow us in principle to answer dynamic questions, though we are much less comfortable with using the framework for this purpose. One problem is that the standard errors for dynamic responses turn out to be quite large. Another challenge is trying to infer the permanent consequences of changes whose time-series variation has been transitory.

### 3.7 The term structure of interest rates at the zero lower bound.

The above analysis ended prior to the first stages of the financial crisis in August 2007. As discussed in Section 3.2, we divide subsequent developments into two phases. The first phase was characterized by high default premiums, failures of some leading financial institutions, and serious disruption of traditional lending patterns. Gürkaynak and Wright (2010) documented that under the financial strains, significant arbitrage opportunities between yields on different Treasury securities often persisted between October 2008 and February 2009. We will not attempt to address the many important issues having to do with monetary policy under those circumstances, but instead begin our analysis here with the second phase which began in March of 2009, and during which policy makers have confronted the longer-term issue of how to provide stimulus to aggregate demand when the short-term interest rate had essentially reached zero.

Figure 3.6 plots assorted yields over this period. The 3-month yield has remained stuck near zero over this period, and the 1-year, although higher, has also displayed little variability. Nonetheless, there has continued to be considerable fluctuation in longer-term yields. What is the nature of the developments driving long-term yields in this environment?

The natural answer is that investors do not believe the U.S. will remain at the zero lower bound forever. When the U.S. escapes from the ZLB, interest rates at all maturities will again respond as they always have to changes in economic fundamentals. Any news today that leads to revisions in the expectations of those future fundamentals shows up as changes in those longer-term yields.

We propose that one way to interpret current long-term yields is to postulate the existence of latent factors, denoted  $f_t$ , which would determine what interest rates would currently be doing if the ZLB were not binding, along with probabilities that arbitrageurs assign to escaping from the ZLB at various future dates. For the first task, what should we assume about the dynamic behavior of these latent factors? The most parsimonious hypothesis would obviously be that, when the

economy escapes from the ZLB, the factor dynamics would revert to their historic behavior as represented by equations like (3.5) or (3.19). The difference is that, when we originally introduced these equations, we were treating the factors  $f_t$  as directly observed from the level, slope, and curvature of the term structure, whereas we are proposing now to interpret them as latent factors characterizing what the level, slope, and curvature would be if we were not stuck at the ZLB. For the second task, we again adopt the simplest possible hypothesis, which is that arbitrageurs assign a constant  $Q$ -measure probability  $\pi^Q$  that the economy will remain at the ZLB next week.

To develop this idea in more detail, we postulate that, once the economy escapes from the ZLB, the short rate will return to being determined by the factors according to the structure

$$\begin{aligned}\tilde{y}_{1t} &= a_1 + b'_1 f_t \\ \tilde{p}_{nt} &= \bar{a}_n + \bar{b}'_n f_t\end{aligned}$$

where the sequences  $\{\bar{a}_n, \bar{b}_n\}_{n=1}^N$  can be calculated as before using the recursions (3.11) and (3.13). However, as long as the economy remains at the ZLB, we instead have

$$\begin{aligned}y_{1t}^* &= a_1^* \\ p_{nt}^* &= \bar{a}_n^* + \bar{b}'_n f_t.\end{aligned}$$

If the zero lower bound were interpreted literally, then  $a_1^*$  would be zero. We represent it instead with some number slightly above zero to match the U.S. experience in which an interest rate paid on reserves has prevented the rate from falling all the way to zero.

Let  $q_{n,t+1}$  denote the holding return on an  $n$ -period bond purchased at  $t$  and sold at  $t+1$ . Note that if  $t$  is characterized by the ZLB, the  $Q$ -measure expectation

of this return is given by

$$\begin{aligned}
E_t^Q(q_{n,t+1}) &= E_t^Q \left[ \frac{(P_{n-1,t+1} - P_{nt})}{P_{nt}} \right] \\
&= \pi^Q E_t^Q \left[ \frac{(P_{n-1,t+1}^* - P_{nt}^*)}{P_{nt}^*} \right] + (1 - \pi^Q) E_t^Q \left[ \frac{(\tilde{P}_{n-1,t+1} - P_{nt}^*)}{P_{nt}^*} \right] \\
&\approx \pi^Q \left[ \bar{a}_{n-1}^* + \bar{b}_{n-1}^{*'} (c^Q + \rho^Q f_t) \right] + (1 - \pi^Q) \left[ \bar{a}_{n-1} + \bar{b}_{n-1}' (c^Q + \rho^Q f_t) \right] \\
&\quad + (1/2) \pi^Q \bar{b}_{n-1}^{*'} \Sigma \Sigma' \bar{b}_{n-1}^* + (1/2) (1 - \pi^Q) \bar{b}_{n-1}' \Sigma \Sigma' \bar{b}_{n-1} - \bar{a}_n^* - \bar{b}_n^{*'} f_t.
\end{aligned}$$

No-arbitrage requires the  $Q$ -measure expected one-period holding yield for an  $n$ -period bond to equal  $y_{1t}$ ,

$$a_1^* = E_t^Q(q_{n,t+1}).$$

This requires

$$\bar{b}_n^{*'} = \pi^Q \bar{b}_{n-1}^{*'} \rho^Q + (1 - \pi^Q) \bar{b}_{n-1}' \rho^Q \quad (3.21)$$

$$\begin{aligned}
\bar{a}_n^* &= \pi^Q \bar{a}_{n-1}^* + (1 - \pi^Q) \bar{a}_{n-1} + \pi^Q \bar{b}_{n-1}^{*'} c^Q + (1 - \pi^Q) \bar{b}_{n-1}' c^Q \\
&\quad + (1/2) \pi^Q \bar{b}_{n-1}^{*'} \Sigma \Sigma' \bar{b}_{n-1}^* + (1/2) (1 - \pi^Q) \bar{b}_{n-1}' \Sigma \Sigma' \bar{b}_{n-1} - a_1^*. \quad (3.22)
\end{aligned}$$

Given  $c^Q, \rho^Q, a_1, b_1, \Sigma$  we can calculate  $\{\bar{a}_n, \bar{b}_n\}_{n=1}^N$  from (3.11) and (3.13). Given these and  $\bar{b}_1^* = 0$ , we can calculate  $\{\bar{a}_n^*, \bar{b}_n^*\}_{n=1}^N$  as functions of  $\pi^Q$  and  $a_1^*$ . Predicted bond yields under the ZLB are then given by

$$y_{nt}^* = a_n^* + b_n^{*'} f_t \quad (3.23)$$

where  $a_n^* = -\bar{a}_n^*/n$  and  $b_n^* = -\bar{b}_n^*/n$ .

As a first pass, we propose to use the same values for  $c^Q, \rho^Q, a_1, b_1, \Sigma$  as estimated from the earlier historical sample. Note that even though these parameters are the same as before, the implied mapping from factors  $f_t$  into observed yields has changed. Let  $Y_{1t} = (y_{26,t}, y_{104,t}, y_{520,t})'$  denote the 6-month, 2-year, and 10-year yields observed at time  $t$ . In our historical sample, these were related to the factors  $f_t$  according to

$$\begin{aligned}
Y_{1t} &= A_1 + B_1 f_t \quad (3.24) \\
A_1 &= \begin{bmatrix} a_{26} \\ a_{104} \\ a_{520} \end{bmatrix} \quad B_1 = \begin{bmatrix} b'_{26} \\ b'_{104} \\ b'_{520} \end{bmatrix}.
\end{aligned}$$



Because we treated the factors in normal times as directly observed from the 6-month, 2-year, and 10-year level, slope, and curvature, and because of the smoothness of the ATSM term structure, our estimates were characterized by

$$B_1 \approx \begin{bmatrix} (1/3) & (1/3) & (1/3) \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -(1/2) & (1/6) \\ 1 & 0 & -(1/3) \\ 1 & (1/2) & (1/6) \end{bmatrix}$$

where the approximation would have been exact if we had imposed the restriction that  $Y_{1t}$  is observed without error.

By contrast, under the ZLB, the relation is

$$Y_{1t} = A_1^* + B_1^* f_t \quad (3.25)$$

$$A_1^* = \begin{bmatrix} a_{26}^* \\ a_{104}^* \\ a_{520}^* \end{bmatrix} \quad B_1^* = \begin{bmatrix} b_{26}^{*'} \\ b_{104}^{*'} \\ b_{520}^{*'} \end{bmatrix}.$$

Let  $Y_{2t}$  denote the four other yields used in the estimation, namely the 3-month, 1-year, 5-year, and 30-year yields. The model implies that

$$Y_{2t} = A_2^* + B_2^* f_t + \varepsilon_t^e \quad (3.26)$$

$$A_2^* = \begin{bmatrix} a_{13}^* \\ a_{52}^* \\ a_{260}^* \\ a_{1560}^* \end{bmatrix} \quad B_2^* = \begin{bmatrix} b_{13}^{*'} \\ b_{52}^{*'} \\ b_{260}^{*'} \\ b_{1560}^{*'} \end{bmatrix}$$

where  $\varepsilon_t^e \sim N(0, \Omega_e)$  denotes measurement error. Substituting (3.25) into (3.26),

$$Y_{2t} = A_2^\dagger + B_2^\dagger Y_{1t} + \varepsilon_t^e \quad (3.27)$$

$$A_2^\dagger = A_2^* - B_2^*(B_1^*)^{-1}A_1^* \quad B_2^\dagger = B_2^*(B_1^*)^{-1}. \quad (3.28)$$

We applied the minimum-chi-square estimation approach developed by Hamilton and Wu (2010b) to weekly interest rate data from March 6, 2009 to August 4, 2010 to infer the values of  $\pi^Q$  and  $a_1^*$  from the OLS estimates of  $\hat{A}_2^\dagger$  and  $\hat{B}_2^\dagger$ ,

taking  $c^Q, \rho^Q, a_1, b_1, \Sigma$  as given by the pre-2007 parameter estimates, as detailed in Appendix D.

This procedure resulted in estimates  $5200\hat{a}_1^* = 0.037$  and  $\hat{\pi}^Q = 0.9834$ , implying that the ZLB is characterized by a one-week interest rate of 4 basis points (at an annual rate) and that arbitrageurs expect the ZLB to persist for  $1/(1 - \pi^Q) = 60$  weeks. We used these two parameters along with the pre-crisis values for  $c^Q, \rho^Q, a_1, b_1, \Sigma$  reported in Table 3.1 to calculate  $b_n^*$  and  $a_n^*$  from (3.21) and (3.22), and used these to infer a value for  $f_t$  on the basis of the observed 6-month, 2-year, and 10-year yield using (3.25). With this  $f_t$  we then have from (3.26) predicted values for each week's 3-month, 1-year, 5-year, and 30-year yields, which predictions are plotted as dashed lines of Figure 3.6. The  $R^2$  for each relation is reported in the first column of Table 3.4. We might compare these with the best possible fit as represented by an unrestricted OLS regression of each yield on a constant and the 6-month, 2-year, and 10-year yields, whose  $R^2$  is reported in the second column of Table 3.4. Particularly for the longer-term yields, the predictions from our simple restricted parameterization are not far from what is actually observed during the ZLB period.

A tougher test of the framework is whether it can successfully predict yields in advance. Here we used the  $f_t$  constructed as above, formed the one-week-ahead forecast  $E_t^P(f_{t+1}) = c + \rho f_t$  again on the basis of the pre-crisis parameters reported in Table 3.1, and calculated the implied yields  $y_{n,t+1}$  using (3.23). Again, particularly for the longer maturities, these forecasts are reasonably close to the best possible in-sample fit as represented by an unrestricted OLS regression of  $y_{n,t+1}$  on a constant and  $Y_{1t}$  (see columns 3 and 4 of Table 3.4).

Although the post-sample fit is good, the model could nevertheless still be improved. Hamilton and Wu (2010b) propose a test of the overidentifying restrictions, which is basically a test of the statistical significance of the difference in  $R^2$  between the first and second columns of Table 3.4. This leads to quite strong rejection, with a  $\chi^2(14)$  test statistic of 344.5.

We made one further simple adjustment to improve the fit further. We postulated that when the economy escapes from the ZLB, arbitrageurs anticipate a

different average level of interest rates (as governed by the parameter  $a_1$ ) compared to that observed in the pre-crisis episode. The estimated value of  $5200a_1$  is 2.19, meaning arbitrageurs expect the post-ZLB average short rate to be below the 4.12 level observed over 1990-2007. The new estimate of  $5200a_1^*$  is 0.068 and of  $\pi^Q$  is 0.9907, implying an expected ZLB duration of 108 weeks. These changes improve the fit relative to that of the model summarized in Figure 3.6 and Table 3.4, though the specification would still be rejected ( $\chi^2(13) = 176.0$ ).

Although one could relax other restrictions of the model until a perfect fit is achieved, we regard this as an attractive parsimonious framework that successfully captures the broad features of how interest rates have been observed to behave under the ZLB regime to date. Another benefit is that this framework gives us an immediate basis for drawing conclusions about how the effects of monetary policy differ under the ZLB from normal times.

Figure 3.7 plots the factor loadings, which summarize how the yield of any maturity  $n$  is predicted to respond to changes in any of the three factors. The main difference is that, under the ZLB, short-term yields are essentially unresponsive to any macroeconomic developments, with all three elements of  $b_n^*$  near zero for small  $n$ . This is because arbitrageurs see very little probability of escaping from the ZLB over most of the term of the security. As  $n$  increases, the response of the yield to macroeconomic factors becomes larger and approaches the response observed in normal conditions, because there is an increasing probability that the economy will be away from the ZLB for most of the security's duration.

This framework allows us to revisit the consequences of a shift in the maturity of the Fed's Treasury holdings. Given our assumption that the latent factors  $f_t$  are responding in the same way as they would when away from the ZLB, we can still use the prediction that a change in the maturity composition of publicly held debt that changes the Treasury risk factor vector by  $\Delta$  would change  $f_{t+1}$  by  $\phi\Delta$ . But whereas in normal times we premultiplied this vector by  $b'_n$  to see what the change  $\Delta$  implied for a yield of maturity  $n$ , at the ZLB we would instead premultiply  $\phi\Delta$  by  $b_n^{*t}$ . These predicted impacts are compared in Figure 3.8. The policy continues to depress long-term yields by the same amount as in normal times, but,

because of the ZLB, it has very little effect on short-term yields. Cumulative effects on short-term yields are also negligible, while the ability to bring long yields down is the same as without the ZLB, as seen in Figure 3.8.

We have analyzed here the effects of a swap by the Federal Reserve of short-term assets for longer-term assets. An alternative strategy, which might be characterized as quantitative easing, is for the Fed to buy longer-term assets outright with newly created reserves. At the ZLB, interest-bearing reserves are essentially indistinguishable from zero-risk 1-week bonds. The effect of quantitative easing is to reduce the available supply of longer-term securities without changing the private-sector's exposure to the risk associated with holding short-term securities. But at the ZLB, changes in the supply of short-term securities have essentially no effects. Thus, the economic consequences of quantitative easing would be identical to those of the maturity swap just described if the economy were at the ZLB.

## **3.8 Discussion.**

### **3.8.1 Comparison with other estimates.**

Here we compare our estimates with those obtained by other researchers. For this purpose, we standardize on the basis of the two scenarios analyzed above. The first scenario is a simultaneous sale by the Fed of \$400 B in securities at the short end and purchase of \$400 B in securities at the long end, implemented in December of 2006. The second scenario is an outright purchase of \$400 B in long-term securities, implemented at the zero lower bound.

Gagnon et al. (2010) used as an explanatory variable the face value of privately-held debt of more than one-year maturity as a percent of GDP, and as dependent variable the 10-year yield or 10-year term premium. They estimated the effect of debt supply on yields using regressions estimated 1986:M12 to 2008:M6 that included several other explanatory variables, and obtained a coefficient relating the 10-year yield to bond supply of 0.069. Since \$400 B would represent about 2.9% of U.S. GDP in 2006:Q4, their estimates imply a predicted decline in the

10-year yield under scenario 1 of  $(2.9)(0.069) = 20$  basis points. This is close to our estimate of a decline of 14 basis points, as reported in the first row of Table 3.5.<sup>18</sup>

In the analysis of Greenwood and Vayanos (2010), the right-hand variable was the fraction of privately-held debt with duration greater than 10 years, and the left-hand variable was assorted yield spreads. They found that a one-percentage-point increase in the share resulted in a 4-basis-point increase in the 5-year-1-year spread over the period 1952-2006. In the sample we studied (1990-2007), a maturity swap of the size contemplated in scenario 1 would have lowered the share of debt with maturity greater than 10 years by 9.8 percentage points. This gives an effect implied by the Greenwood-Vayanos estimates of  $(9.8)(4) = 39$  basis points. For comparison, our estimate of the size of the effect is 17 basis points for scenario 1, but only 9 basis points for scenario 2. The reason for the difference between the two scenarios is that, in our framework, part of the drop in the spread if the policy had been implemented over the period studied by Greenwood and Vayanos (2010) would have come from an increase in short-term yields, something that would not happen if the same purchase were implemented at the zero lower bound.

Another recent analysis comes from D'Amico and King (2010), who look at the change in yields of different maturities during the Fed's purchase of \$300 billion in long-term securities between March and October of 2009. They conclude that these purchases lowered the yield on 10-year Treasuries by about 50 basis points, which would translate into an effect of  $(4/3)(50) = 67$  basis points for the \$400 B purchase analyzed in Table 3.5, a somewhat larger effect than implied by our estimates. However, the 10-year yield was where these purchases were concentrated and where D'Amico and King found the biggest effects, and large standard errors are associated with any of these estimates.

Deutsche Bank (2010) attempted to synthesize the estimates of Gagnon et

---

<sup>18</sup>Gagnon et al. (2010)'s regressions in which the term premium rather than the yield is the left-hand variable would imply estimates as low as 12 basis points. However, these are harder to compare directly with those for our scenario. In our conception of the question being asked, we assume that the supply of securities with maturity less than one year increases by \$400 B, driving up the yield on those securities and making the decrease in the term premium larger than the decrease in the yield. This effect is not captured by the Gagnon et al. (2010) regressions.

al. (2010), Macroeconomic Advisers, and their own research staff, and estimated that \$1 trillion in long-term purchases in the current setting might produce a 50-basis-point decline in long-term yields, which we've translated as a 20-basis-point decline for the \$400 billion purchase reported in Table 3.5.

In 1961, the U.S. attempted to use Treasury and Fed operations to lower the fraction of publicly-held long-term debt in what was referred to as "Operation Twist." Swanson (2011) used a daily event study of announcements pertaining to the Operation Twist and found effects on bond yields that, when scaled by the change in size of outstanding Treasury debt, are broadly consistent with those summarized in our Table 3.5.

Although our estimates of the effects are the smallest in this group, they are generally in the same ballpark, which is somewhat surprising given the very different ways in which these estimates are derived. There is overall agreement that sufficiently large asset purchases could achieve a modest reduction in long-term yields. There is nevertheless considerable uncertainty, both in terms of the econometric standard errors and possible specification errors, in any of the estimates reported.

### 3.8.2 Effects on non-Treasury securities.

Here we sketch a generalization of the theoretical framework in Section 3.3 to allow arbitrageurs also to hold other securities with a nonzero probability of default.

Let  $P_{1t}^\dagger$  denote the price paid at  $t$  for a one-period bond whose value next period will be

$$P_{0,t+1}^\dagger = \begin{cases} 1 & \text{with probability } \exp(-\psi_t) \\ 0 & \text{with probability } 1 - \exp(-\psi_t) \end{cases} .$$

If the arbitrageurs hold a fraction  $z_{1t}^\dagger$  in the risky asset and if the probability of default  $\psi_t$  is independent of risk factors  $f_t$ , then using a similar approach to that in Appendix A, the contribution of the risky asset to the variance can be

approximated<sup>19</sup> by  $z_{1t}^{\ddagger 2} \psi_t$  and the no-arbitrage condition (3.2) becomes

$$y_{1t}^{\ddagger} = y_{1t} + \psi_t(1 + \gamma z_{1t}^{\ddagger}). \quad (3.29)$$

In the absence of risk aversion ( $\gamma = 0$ ), in equilibrium the risky security will offer the same expected return as the risk-free security, which requires a premium of  $\psi_t$  to compensate for the probability of default. With risk aversion ( $\gamma > 0$ ) and a positive exposure of arbitrageurs to this risk ( $z_{1t}^{\ddagger} > 0$ ), the risky asset will offer a higher expected return to compensate for the risk.

If the factors that govern  $\psi_t$  and determine equilibrium  $z_{1t}^{\ddagger}$  are independent of the factors  $f_t$  that determine the risk-free yield, the one-period risky rate would have identical loadings as  $y_{1t}$  on fluctuations in the level, slope, and curvature factors, as well as additional loadings on separate default-risk factors. A parallel result can be derived for risky assets of longer maturity, with  $p_{nt}^{\ddagger}$  loading on  $f_t$  with the same coefficients  $\bar{b}_n$  as for risk-free bonds, along with separate loadings on the default-risk factors.

Although the independence of Treasury and default risk factors is a highly stylized assumption, there is no question that risky yields of different maturities respond in a similar way to the factors driving Treasury yields. Figure 3.9 displays the comovement between the 10-year Treasury rate and that on 30-year mortgages and Aaa-rated and Baa-rated corporate debt<sup>20</sup>.

Rather than impose a particular loading of non-Treasury yields on the level, slope, and curvature factors, we can estimate the empirical loading directly by OLS

---

<sup>19</sup>If we conjecture that  $p_{1t}^{\ddagger} = h \left( \bar{a}_1^{\ddagger} + \bar{b}_1^{\ddagger} f_t + \bar{c}_{\psi}^{\ddagger} \psi_t + \bar{c}_{\zeta}^{\ddagger} \psi_t \zeta_{1t}^{\ddagger} \right)$  for  $\zeta_{1t}^{\ddagger}$  independent factors affecting the supply of risky assets,

$$\begin{aligned} E_t \left[ z_{1t}^{\ddagger} \left( \frac{P_{0,t+1}^{\ddagger}}{P_{0t}^{\ddagger}} - 1 \right) \right]^2 &= z_{1t}^{\ddagger 2} \{ \exp(-\psi_t h) \exp[-2h(\bar{a}_1^{\ddagger} + \bar{b}_1^{\ddagger} f_t + \bar{c}_{\psi}^{\ddagger} \psi_t + \bar{c}_{\zeta}^{\ddagger} \psi_t \zeta_{1t}^{\ddagger})] \\ &\quad - 2 \exp(-\psi_t h) \exp[-h(\bar{a}_1^{\ddagger} + \bar{b}_1^{\ddagger} f_t + \bar{c}_{\psi}^{\ddagger} \psi_t + \bar{c}_{\zeta}^{\ddagger} \psi_t \zeta_{1t}^{\ddagger})] + 1 \} \\ &= z_{1t}^{\ddagger 2} \psi_t h + o(h). \end{aligned}$$

<sup>20</sup>Aaa and Baa yield represent values as of the last day of the month, while the 30-year mortgage rate is for the last week of the month, from the FRED database of the Federal Reserve Bank of St. Louis.

estimation of

$$y_{jt}^{\ddagger} = a_j^{\ddagger} + b_j^{\ddagger\prime} f_t + u_t$$

over  $t = 1990:M1$  to  $2007:M7$  for assorted securities  $j$ . Note that if there is a correlation between the default risk factors and  $f_t$ , this will be incorporated in the estimated values of  $b_j^{\ddagger}$ . Table 3.6 reports the empirical factor loadings for these three risky yields, which, not surprisingly given Figure 3.9, turn out to be similar to those for 10-year Treasury bonds.

In the next-to-last column we use these estimated values of  $b_j^{\ddagger}$  to calculate the predicted effect in normal times of a shift in the maturity composition of Fed holdings.<sup>21</sup> Based on the historical correlations between bond yields, in the pre-crisis period, if the Fed were to sell \$400 billion of short-term Treasuries and buy \$400 billion in long-term Treasuries, the 10-year T-bond and the Aaa and Baa corporate yields would each be expected to decline by 14 basis points, and the 30-year fixed mortgage rate by 11 basis points.

We can also get a quick impression of what might be expected at the zero lower bound as follows. The predicted change in the 6-month, 2-year, and 10-year yields of this \$400 billion maturity swap when at the ZLB are given by the corresponding elements of the vector  $B_1^* \phi \Delta$ . If  $y_{jt}^{\ddagger}$  tracks these as estimated historically (namely, by  $b_j^{\ddagger\prime} B_1^{-1}$ ), then we get a predicted effect on  $y_{jt}^{\ddagger}$  at the ZLB of  $b_j^{\ddagger\prime} B_1^{-1} B_1^* \phi \Delta$ . These estimates are reported in the last column of Table 3.6. Interestingly, buying long-term Treasuries might if anything have an even bigger effect on risky yields when at the ZLB than it does in normal circumstances. Again, at the ZLB, in our framework the effects are the same whether the Fed finances the purchases with sales of short-term T-bills or with newly created reserves.

If the Fed were instead to purchase risky securities directly, the resulting reduction in arbitrageurs' holdings of these securities  $z_{nt}^{\ddagger}$  would both reduce the default risk premium (through equation (3.29)) as well as affect the pricing of Treasury level, slope or curvature risk (because by holding these risky securities an investor is also exposed to the conventional term structure factors). For example,

---

<sup>21</sup>These were calculated as  $b_j^{\ddagger\prime} \phi \Delta$  for  $\phi$  the matrix of OLS coefficients in (3.19) and  $\Delta$  given by (3.20).



the Fed's MBS purchases could both flatten the slope of the Treasury yield curve and narrow the spread between MBS and Treasury yields.

We should also comment on how arbitrageurs' holding of risky securities would influence our empirical estimates of the matrix  $\phi$  itself. If Treasuries represent only a subset of arbitrageurs' holdings, then Treasury holdings as a fraction of their total wealth  $z_{nt}$  would be a smaller number than we have assumed. If, for example, each  $z_{nt}$  were divided by 2, our vector  $q_t$  and therefore the magnitude  $\Delta$  would be divided by two, while the OLS estimates  $\hat{\phi}$  would be multiplied by two. Notice that a change in scale of this type would leave the estimated product  $\phi\Delta$  unchanged and have no effect on any of the estimates reported. This invariance results from the fact that ultimately our estimates are simply an empirical summary of the historical relations between observed yields and maturity shares  $z_{nt}$  defined as a percentage of total publicly held federal debt, and it is the historical covariation of yields with outstanding Treasury debt that determined these estimates. If we were trying to make an inference about structural coefficients such as the risk aversion parameter  $\gamma$ , getting the scale right would be important. But for the purposes for which the estimates are used here, the scale of  $z_{nt}$  does not matter for any of the reported results.

### 3.8.3 Application: Evaluation of QE2.

On November 3, 2010, the Federal Reserve announced its intention to implement additional measures to stimulate the economy, which was described in the financial press as a second round of quantitative easing (QE2). The plan was to purchase an additional \$600 billion of longer-term Treasury securities by the end of the second quarter of 2011, a pace of about \$75 billion per month. This differed in several details from the scenarios analyzed above.

The first difference is that, as implemented, the purchases were concentrated not on the longest-maturity securities, but instead focused primarily on securities between 2-1/2 and 10 years. Over the period 1990-2006, if the Fed had sold all its holdings of less than 1 year and used the proceeds to purchase outstanding Treasury debt evenly over the 2-1/2 to 10 year range, the resulting average change

in  $q_t$  would be not the value reported in expression (3.20) but instead

$$\Delta_2 = (0.006898, 0.004479, -0.004406)'. \quad (3.30)$$

Whereas the estimated effects of  $\Delta$  on the term structure are statistically significantly distinguishable from zero in our framework, those resulting from  $\Delta_2$  are not. Figure 3.10 compares the estimated effects on yields of  $\Delta$  and  $\Delta_2$  if implemented at the zero lower bound. The dashed curve summarizes the predicted effects if the Fed were to sell all its holdings of less than 1-year maturity, and use the proceeds to retire debt of the longest outstanding maturities. Note this is identical to the dashed curve in Figure 3.8. The solid curve summarizes the predicted effects if the Fed were to sell all its holdings of less than 1-year maturity, and spread the proceeds evenly to purchase outstanding Treasury debt in the 2-1/2 to 10 year range. The latter has a significantly smaller effect on long-term rates. Again we interpret an outright purchase of a comparable quantity of securities as having similar effects to a debt swap when the economy is at the zero lower bound.

A second important difference between QE2 as it's been implemented by the Fed and the scenarios analyzed here is in the timing, with the purchases associated with QE2 spread over a period of 8 months. Between November 2009 and November 2010, non-TIPS Treasury debt increased by \$127.3 billion per month, of which \$71.4 billion was in the 2-1/2 to 10 year maturity range. Hence, the proposed QE2 would barely absorb the newly issued medium-term debt, and debt of greater than 10 years would continue to increase rather than decline. The top panel of Figure 3.11 shows that the average maturity of publicly-held Treasury debt has been higher in each of the first three months of QE2 than it had been in any month over the preceding 2 years. The bottom panel shows that the fraction of publicly-held debt of more than 10 years maturity continued to increase even as the Fed was implementing its QE2 bond purchases. Our conclusion is that QE2 as implemented had little potential to lower long-term interest rates via the mechanism explored in this paper.

### 3.9 Conclusion.

We have found statistically significant forecasting relations over 1990-2007 between the maturity structure of Treasury debt held by the public and the behavior of U.S. interest rates. These relations suggest that in normal times, the Federal Reserve has some potential to flatten the yield curve, though not to reduce the overall level of interest rates, by selling short-term securities and buying long-term securities. Our estimates of the effect on impact suggest that quite massive operations would be necessary to have a measurable effect on interest rates.

We proposed that altering the maturity structure of publicly held Treasury debt would be equally effective at lowering long-term yields when the economy is at the zero lower bound. But because there are negligible consequences for short-term yields in such a setting, the policy of reducing public holdings of long-term bonds has the potential to bring the overall level of interest rates down for an economy at the ZLB, whereas it could not do so in a normal environment. Quantitative easing, defined as buying the long-term bonds with newly created reserves, has the identical potential in this model.

One might suppose that the potential small magnitude of the effect is not a concern as far as the latter policy is concerned— if hundreds of billions are not enough to make much difference, then perhaps purchases in the trillions, such as the Fed has embarked upon with its holdings of mortgage-backed securities, might do the trick. However, we would emphasize that, in the model of the ZLB proposed here, the entire ability to influence long-term yields comes from investors' perceptions of what fundamentals are going to be *after* normal conditions have returned. A policy that only kept the supplies off the market during the ZLB episode itself would have much more limited potential. In this sense, this particular form of nonstandard monetary policy could end up having limited effectiveness for the same reasons as policies that hope to influence the public's expectation of what the target will be for short-term interest rates once the economy escapes from the ZLB.

Our estimated effects are linear— twice as big a purchase is predicted to have twice as big an effect on yields. But this is simply an assumption of our

empirical estimation strategy and not a proposition we have tested directly in the data. Particularly since the magnitudes under discussion are so different from the observed historical variations from which our estimates were inferred, extrapolation of these effects to larger and larger policy measures is of necessity an uncertain exercise.

We also noted that, although we have framed the discussion here in terms of options available to the Federal Reserve, this policy tool could in many ways more naturally be implemented by the Treasury itself altering the term structure of debt that it issues. If the Treasury has sound reasons not to do so, it is unclear why the Federal Reserve should try to undo the Treasury's attempted hedging of the unified government's balance sheet with respect to interest rate risk. Conversely, if the Fed has good reasons to try to flatten the slope of the yield curve, it is unclear why the Treasury should resist being the agent to implement the plan.

### **3.10 Acknowledgements**

Chapter 3 is coauthored with James Hamilton, and in full, has been submitted for publication of material. I thank Jim for the permission to use these chapters in the dissertation.

### 3.11 Appendices

#### A. Details of the arbitrageurs' portfolio optimization problem.

Let  $P_{nt}$  denote the price of a pure-discount  $n$ -period bond (with  $P_{0t} = 1$ ),  $W_t$  the total wealth of the arbitrageurs, and  $z_{nt}$  the portion of their wealth allocated to each bond maturity. Then the arbitrageurs' wealth evolves according to

$$W_{t+1} = \sum_{n=1}^N z_{nt} \frac{P_{n-1,t+1}}{P_{nt}} W_t$$

with associated rate of return

$$r_{t,t+1} = \frac{W_{t+1} - W_t}{W_t} = \sum_{n=1}^N z_{nt} \left[ \frac{P_{n-1,t+1}}{P_{nt}} - 1 \right].$$

If the change in prices between  $t$  and  $t + 1$  is small,<sup>22</sup> the portfolio's mean return and variance can be approximated

$$E_t r_{t,t+1} \approx -z_{1t}(\bar{a}_1 + \bar{b}'_1 f_t) + \sum_{n=2}^N z_{nt} \left[ \bar{a}_{n-1} + \bar{b}'_{n-1}(c + \rho f_t) + (1/2)\bar{b}'_{n-1} \Sigma \Sigma' \bar{b}_{n-1} - \bar{a}_n - \bar{b}'_n f_t \right] \quad (3.31)$$

$$\text{Var}_t(r_{t,t+1}) \approx d'_t \Sigma \Sigma' d_t \quad (3.32)$$

where the  $(J \times 1)$  vector  $d_t$  summarizes exposures to each of the  $J$  factor risks associated with holding the  $(N \times 1)$  vector of bonds  $z_t$ . The arbitrageurs thus choose  $z_t$  so as to maximize (3.1) subject to (3.31), (3.32), (3.8), and  $\sum_{n=1}^N z_{nt} = 1$ , for which the first-order condition is given by (3.6).

## B. Arbitrage-free equilibrium.

Note that  $y_{nt} = -n^{-1}p_{nt} = -n^{-1}(\bar{a}_n + \bar{b}'_n f_t)$  and suppose that  $\zeta_{nt} = \zeta_n + \vartheta'_n f_t$ . If we multiply (3.9) by  $\bar{b}_{n-1}$  and sum over  $n = 2, \dots, N$ , we find using (3.8)

<sup>22</sup>Suppose that

$$q_{n,t+1} \equiv \frac{(P_{n-1,t+1} - P_{nt})}{P_{nt}} = \exp\left(\mu_n h + \sqrt{h} \varepsilon_{n,t+1}\right) - 1$$

where  $(\varepsilon_{1,t+1}, \dots, \varepsilon_{N,t+1})' \sim N(0, \Omega)$ . Our approximation is derived from the limiting behavior as  $h$  becomes small, analogous to those obtained when considering a continuous-time representation of a discrete-time process. Thus as in Merton (1969),

$$E_t \left( \sum_{n=1}^N z_{nt} q_{n,t+1} \right) = \sum_{n=1}^N z_{nt} [\mu_n h + \Omega_{nn} h/2 + o(h)]$$

$$\text{Var}_t \left( \sum_{n=1}^N z_{nt} q_{n,t+1} \right) = z'_t \Omega z_t h + o(h)$$

for  $\Omega_{nn}$  the row  $n$ , column  $n$  element of  $\Omega$  and  $z_t = (z_{1t}, \dots, z_{Nt})'$ . Equations (3.31) and (3.32) are obtained by setting  $h = 1$  and  $o(h) = 0$ . Specifically,

$$\frac{P_{n-1,t+1}}{P_{nt}} = \exp\left(\bar{a}_{n-1} + \bar{b}'_{n-1} f_{t+1} - \bar{a}_n - \bar{b}'_n f_t\right)$$

$$\mu_n = \bar{a}_{n-1} + \bar{b}'_{n-1}(c + \rho f_t) - \bar{a}_n - \bar{b}'_n f_t$$

$$\Omega_{nn} = \bar{b}'_{n-1} \Sigma \Sigma' \bar{b}_{n-1}.$$

that equilibrium requires

$$d_t = \sum_{n=2}^N \bar{b}_{n-1} \left[ \zeta_n + \vartheta'_n f_t + (\alpha_n/n)(\bar{a}_n + \bar{b}'_n f_t) \right].$$

Equation (3.10) is obtained from (3.7) with

$$\begin{aligned} \lambda &= \gamma \Sigma' \sum_{n=2}^N \bar{b}_{n-1} [\zeta_n + (\alpha_n/n) \bar{a}_n] \\ \Lambda &= \gamma \Sigma' \sum_{n=2}^N \bar{b}_{n-1} [\vartheta'_n + (\alpha_n/n) \bar{b}'_n]. \end{aligned}$$

### C. ATSM estimation for a just-identified model.

We first estimate the parameters of (3.5) and (3.16) by OLS:

$$\begin{aligned} \begin{bmatrix} \hat{c} & \hat{\rho} \end{bmatrix} &= \left( \sum_{t=2}^T f_t \begin{bmatrix} 1 & f'_{t-1} \end{bmatrix} \right) \left( \sum_{t=2}^T \begin{bmatrix} 1 \\ f_{t-1} \end{bmatrix} \begin{bmatrix} 1 & f'_{t-1} \end{bmatrix} \right)^{-1} \\ \hat{\Sigma} \hat{\Sigma}' &= (T-1)^{-1} \sum_{t=2}^T (f_t - \hat{c} - \hat{\rho} f_{t-1})(f_t - \hat{c} - \hat{\rho} f_{t-1})' \\ \begin{bmatrix} \hat{A} & \hat{B} \end{bmatrix} &= \left( \sum_{t=1}^T Y_{2t} \begin{bmatrix} 1 & f'_t \end{bmatrix} \right) \left( \sum_{t=1}^T \begin{bmatrix} 1 \\ f_t \end{bmatrix} \begin{bmatrix} 1 & f'_t \end{bmatrix} \right)^{-1} \\ \hat{\Sigma}_e \hat{\Sigma}'_e &= T^{-1} \sum_{t=1}^T (Y_{2t} - \hat{A} - \hat{B} f_t)(Y_{2t} - \hat{A} - \hat{B} f_t)'. \end{aligned}$$

The predicted value for row  $i$  of  $\hat{B}$  is given by

$$\hat{B}'_i = n_i^{-1} b'_1 \left[ I_J + \rho^Q + (\rho^Q)^2 + \dots + (\rho^Q)^{n_i-1} \right] \quad \text{for } i = 1, \dots, M.$$

For the just-identified case with  $M = J + 1$ , we solve this  $[(J + 1) \times J]$  system of equations for the  $J(J + 1)$  unknowns  $\rho^Q$  and  $b_1$  using numerical search. Taking these values for  $\rho^Q$  and  $b_1$  as given, we can then use (3.11) to solve for  $\bar{b}_n$  for any desired  $n$  along with

$$\bar{a}_n = n \bar{a}_1 + \sum_{\ell=1}^n \bar{b}'_{\ell-1} c^Q + (1/2) \sum_{\ell=1}^n \bar{b}'_{\ell-1} \Sigma \Sigma' \bar{b}_{\ell-1}.$$

The  $J + 1$  values for  $a_1$  and  $c^Q$  are then found by numerical solution of the  $J + 1$  equations

$$\hat{A}_i = -n_i^{-1} \bar{a}_{n_i} \quad \text{for } i = 1, \dots, M.$$

## D. ATSM estimation for an overidentified model.

We estimated (3.27) by unconstrained OLS,

$$\begin{bmatrix} \hat{A}_2^\dagger & \hat{B}_2^\dagger \end{bmatrix} = \left( \sum_{t=1}^T Y_{2t} \begin{bmatrix} 1 & Y'_{1t} \end{bmatrix} \right) \left( \sum_{t=1}^T \begin{bmatrix} 1 \\ Y_{1t} \end{bmatrix} \begin{bmatrix} 1 & Y'_{1t} \end{bmatrix} \right)^{-1}$$

for which the inverse of the usual variance matrix for the estimated coefficients is given by

$$\hat{R} = \hat{\Omega}_e^{-1} \otimes T^{-1} \sum_{t=1}^T \begin{bmatrix} 1 \\ Y_{1t} \end{bmatrix} \begin{bmatrix} 1 & Y'_{1t} \end{bmatrix}$$

with  $\hat{\Omega}_e$  given by diagonal elements of

$$T^{-1} \sum_{t=1}^T (Y_{2t} - \hat{A}_2^\dagger - \hat{B}_2^\dagger Y_{1t})(Y_{2t} - \hat{A}_2^\dagger - \hat{B}_2^\dagger Y_{1t})'.$$

The minimum-chi-square estimation procedure proposed by Hamilton and Wu (2010b) estimates the structural parameters of interest  $\theta = (\pi^Q, a_1^*)'$  or  $(\pi^Q, a_1^*, a_1)'$  by minimizing

$$T[\hat{\pi} - g(\theta)]' \hat{R}[\hat{\pi} - g(\theta)] \quad (3.33)$$

where  $\hat{\pi} = \text{vec} \left( \begin{bmatrix} \hat{A}_2^\dagger & \hat{B}_2^\dagger \end{bmatrix}' \right)$  and  $g(\theta)$  denotes the corresponding predicted value from (3.28). Under the null hypothesis that the model is correctly specified, the minimal value achieved for (3.33) should have an asymptotic  $\chi^2(k_1 - k_0)$  distribution, where  $k_1 = 14$  is the number of parameters in  $\hat{A}_2^\dagger$  and  $\hat{B}_2^\dagger$  and  $k_0 = 2$  or  $3$  is the number of elements in  $\theta$ .

## E. Details of data construction.

Following Greenwood and Vayanos (2010), we started with CRSP data for outstanding Treasury debt by individual CUSIP number to estimate outstanding nominal Treasury debt at the end of each month. We calculated  $n$  for each issue by calculating the number of days between maturity and the last Friday of the month, and converted to weeks by rounding up. The raw source for these data appears to be the *Monthly Statement of the Public Debt of the United States*. We checked these data by summing all the maturities and comparing this sum with



the sum of nominal bills, bonds, and notes recorded in the Haver database,<sup>23</sup> which also comes from the same *Monthly Statement*. We found numerous discrepancies, which came from such factors as the CRSP files on occasion missing individual CUSIP series and at other times having incorporated assorted data entry errors. We were able to correct CRSP data errors so as to reduce almost all discrepancies to less than \$200 M by hand comparison of the CRSP numbers with individual copies of the *Monthly Statement* itself.

Although the Federal Reserve currently reports outright Treasury holdings for the System Open Market Account by individual CUSIP, we were unable to secure access to historical archives of these, and settled for rough estimates constructed as follows. The Federal Reserve's weekly H41 release<sup>24</sup> reports SOMA each Wednesday by rough maturity breakdowns (less than 15 days, 16-90 days, 91 days to 1 year, over 1 year to 5 years, over 5 years to 10 years, and over 10 years), and we matched up the last Wednesday of each month for SOMA holdings with the last calendar day of the month for Treasury marketable debt. Unfortunately, the reported SOMA maturity categories include both nominal Treasuries as well as TIPS, which we exclude from our analysis. Our solution was to assume that Fed holdings of TIPS as a fraction of the Fed's total holdings of notes and bonds was the same across all maturity categories. Total Fed holdings of notes and bonds are reported on the H41, as are total TIPS holdings (though prior to December 2002, we had to read the latter by hand from the notes section of individual reports). We then multiplied each maturity category greater than 1 year by this ratio to get an estimate of total TIPS holdings in those categories. For maturity categories less than 1 year, we multiplied by the product of this ratio with the ratio of the Fed's notes and bonds of maturity less than 1 year to the Fed's total Treasury securities less than one year. We then subtracted the resulting estimates of TIPS holdings within each maturity category from the reported total holdings within each category to get our estimate of nominal Fed holdings for each maturity category. We

---

<sup>23</sup>We thank Christiane Baumeister for sharing these Haver data.

<sup>24</sup>Available in Table 2 of <http://www.federalreserve.gov/datadownload/Choose.aspx?rel=H41>. Prior to June 2003, we used the end-of-calendar month data compiled by Kuttner (2006) available at <http://econ.williams.edu/people/knk1/research>.

then allocated this ratio evenly across total outstanding Treasury securities of each weekly maturity falling within that category to arrive at our estimate of how much of those securities were held by the Federal Reserve's SOMA.

## Tables

Table 3.1: Parameter estimates

	Estimated parameters			Implied parameters			
$c^Q$	<b>0.0116</b> (0.0002)	<b>-0.0118</b> (0.0005)	<b>-0.0036</b> (0.0007)	$\lambda$	-0.1378 (0.0717)	<b>0.1604</b> (0.0727)	-0.0564 (0.0687)
$\rho^Q$	<b>0.9990</b> (0.0001)	<b>0.0094</b> (0.0002)	<b>-0.0140</b> (0.0005)	$\Lambda$	-0.0867 (0.0468)	-0.0480 (0.0594)	-0.0948 (0.1203)
	<b>0.0027</b> (0.0003)	<b>0.9870</b> (0.0004)	<b>0.0330</b> (0.0010)		0.0847 (0.0455)	-0.0266 (0.0825)	0.1773 (0.1200)
	<b>-0.0018</b> (0.0002)	<b>-0.0028</b> (0.0002)	<b>0.9867</b> (0.0008)		-0.0567 (0.0436)	0.0531 (0.0596)	-0.1862 (0.1594)
$c$	-0.0034 (0.0089)	-0.0003 (0.0074)	0.0006 (0.0066)				
$\rho$	<b>0.9895</b> (0.0072)	0.0042 (0.0081)	-0.0244 (0.0157)				
	0.0083 (0.0047)	<b>0.9826</b> (0.0081)	<b>0.0478</b> (0.0123)				
	-0.0013 (0.0041)	0.0055 (0.0058)	<b>0.9755</b> (0.0132)				
$a_1 \times 5200$	<b>4.1158</b> (0.0074)						
$b_1 \times 5200$	<b>1.0345</b> (0.0058)	<b>-0.6830</b> (0.0081)	<b>0.6311</b> (0.0189)				
$\Sigma$	<b>0.1094</b> (0.0236)	0	0				
	<b>0.0360</b> (0.0100)	<b>0.1027</b> (0.0045)	0				
	<b>-0.0670</b> (0.0188)	0.0025 (0.0130)	<b>0.0968</b> (0.0149)				
$\Sigma_c \times 5200$	<b>0.0978</b> (0.0023)	0	0	0			
	0	<b>0.0674</b> (0.0016)	0	0			
	0	0	<b>0.0531</b> (0.0013)	0			
	0	0	0	<b>0.1171</b> (0.0028)			

Parameter estimates for the weekly affine-term-structure model, Jan 5, 1990 to July 27, 2007. Small-sample standard errors in parentheses. Sample size:  $T = 917$ . Bold indicates statistically significantly different from zero at the 5% significance level.

**Table 3.2:** Holding-return forecasting regressions

Regressors	6m over 3m	1yr over 6m	2y over 1y	5y over 1y	10y over 1y
$c, f_t^*$	<b>0.357</b> (0.000)	<b>0.356</b> (0.000)	<b>0.331</b> (0.000)	<b>0.295</b> (0.000)	<b>0.331</b> (0.000)
$c, f_t, z_t^{A*}$	<b>0.410</b> (0.020)	0.420 (0.119)	0.373 (0.311)	0.300 (0.728)	0.336 (0.665)
$c, f_t, z_t^{L*}$	<b>0.428</b> (0.003)	<b>0.501</b> (0.008)	<b>0.524</b> (0.006)	<b>0.398</b> (0.035)	0.357 (0.196)
$c, f_t, z_t^{pc*}$	<b>0.368</b> (0.001)	<b>0.361</b> (0.007)	0.333 (0.062)	0.297 (0.098)	0.334 (0.051)
$c, f_t, v_t^*$	<b>0.385</b> (0.016)	<b>0.409</b> (0.001)	<b>0.388</b> (0.006)	<b>0.339</b> (0.008)	0.338 (0.227)
$c, f_t, q_t^*$	<b>0.444</b> (0.002)	<b>0.568</b> (0.000)	<b>0.714</b> (0.000)	<b>0.617</b> (0.000)	<b>0.549</b> (0.001)
$c, f_t, z_t^{pc}, q_t^*$	<b>0.452</b> (0.002)	<b>0.571</b> (0.000)	<b>0.717</b> (0.000)	<b>0.618</b> (0.000)	<b>0.550</b> (0.002)
$c, f_t, v_t, q_t^*$	<b>0.458</b> (0.001)	<b>0.595</b> (0.000)	<b>0.737</b> (0.000)	<b>0.640</b> (0.000)	<b>0.552</b> (0.002)
$c, f_t, z_t^A, z_t^L, q_t^*$	<b>0.476</b> (0.000)	<b>0.597</b> (0.001)	<b>0.741</b> (0.000)	<b>0.670</b> (0.002)	0.634 (0.054)

$R^2$  and hypothesis tests for holding-return forecasting regressions. Reported numbers are the  $R^2$  for the regressions, with  $p$ -values in parentheses, for tests of the null hypothesis that coefficients on starred variables are zero. All regressions also include a constant term (denoted by  $c$ ) and all hypothesis tests use Newey-West variance matrix with 20 lags. Bold indicates coefficients on starred variables are statistically significantly different from zero at the 5% significance level.

**Table 3.3:** Factor vector autoregression

	$F$ test	$\phi'_i \Delta$
level	<b>3.256</b> (0.023)	0.005 (0.112)
slope	<b>4.415</b> (0.005)	<b>-0.250</b> (0.116)
curvature	<b>2.672</b> (0.049)	-0.073 (0.116)

Granger-causality tests and scenario impact estimates for factor vector autoregression. First column reports  $F$  test ( $p$ -value in parentheses) of null hypothesis that  $\phi_i = 0$  in regression  $f_{it} = c_i + \rho'_i f_{t-1} + \phi'_i q_{t-1} + \varepsilon_{it}$ . Second column reports estimate of  $\phi'_i \Delta$  for that regression (with standard error) for  $\Delta$  the average change in  $q$  under the alternative scenario.

**Table 3.4:**  $R^2$  for post-crisis sample

	Contemporaneous		Forecast	
	restricted	unrestricted	restricted	unrestricted
3m	0.625	0.668	0.522	0.602
1y	0.891	0.924	0.652	0.767
5y	0.961	0.975	0.753	0.753
30y	0.965	0.972	0.735	0.787

$R^2$  for post-crisis sample (March 3, 2009 to Aug 10, 2010) for unrestricted OLS fit to post-crisis data and for prediction constructed from pre-crisis parameter estimates together with post-crisis estimates of  $\pi^Q$  and  $a_1^*$ . Contemporaneous: prediction of  $y_{nt}$  given current 6-month, 2-year and 10-year yields. Forecast: predictions of  $y_{nt}$  given lagged 6-month, 2-year and 10-year yields.

**Table 3.5:** Comparison of different estimates

Study	Measure	Original estimates		Hamilton-Wu estimates	
		Pre-crisis	ZLB	Pre-crisis	ZLB
Gagnon, et. al.	10 yr yield	-20		-14	-13
Greenwood-Vayanos	5yr-1yr spread	-39		-17	-9
	20yr-1yr spread	-74		-25	-18
D'Amico-King	10yr yield		-67	-14	-13
Deutsche Bank	10yr yield		-20	-14	-13

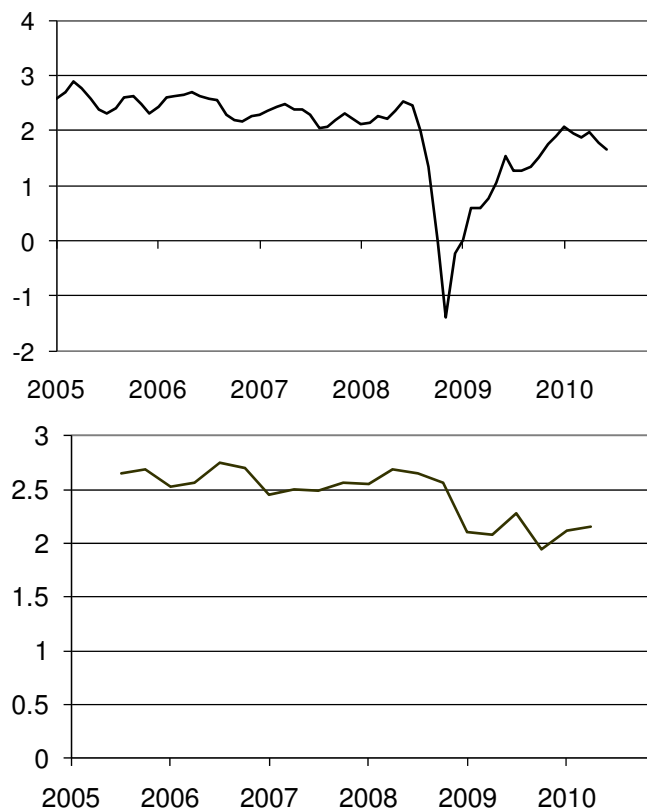
Comparison of different estimates of the effect of replacing \$400 billion in long-term debt with short-term debt.

**Table 3.6:** Risky securities

Yield	Factor loadings			Normal effect	ZLB effect
	level	slope	curvature		
10-year Treasury	1.000	0.500	0.167	-14	-13
Aaa Corporate	0.883	0.453	0.379	-14	-15
Baa Corporate	0.888	0.441	0.535	-14	-17
30-year Mortgage	0.933	0.363	0.325	-11	-13

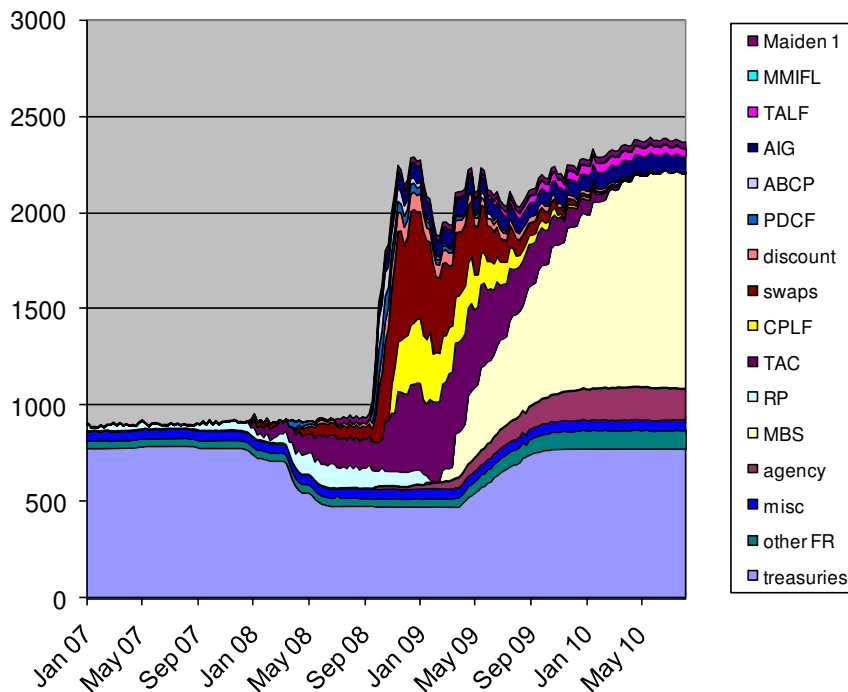
Empirical loadings of selected yields on Treasury level, slope and curvature factors, and predicted effect on yield (in basis points) of selling \$400 billion in short-term Treasury debt and buying \$400 billion in long-term Treasury debt.

## Figures



**Figure 3.1:** Alternative measures of 5-year expected inflation

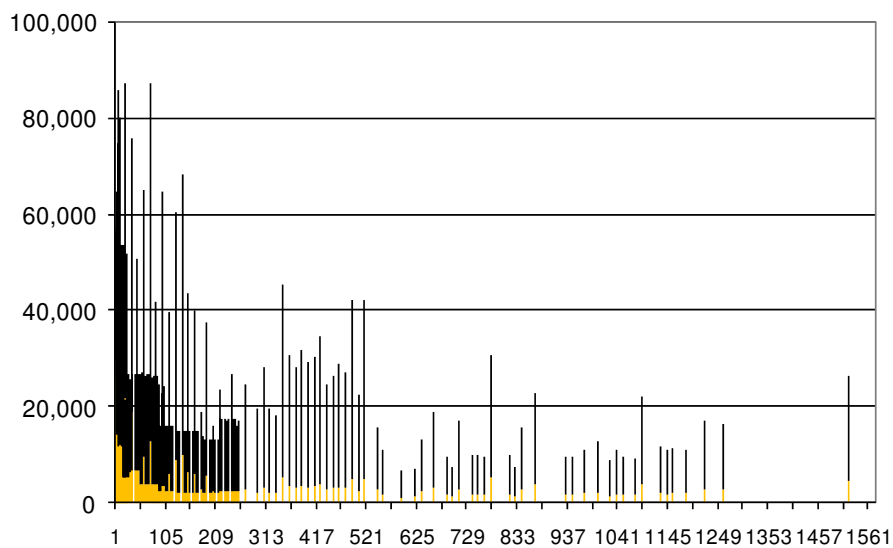
Top panel: 5-year break-even inflation rate, calculated as nominal yield of 5-year Treasury bond minus nominal yield on 5-year Treasury Inflation Protected Security, 2005:M1-2010:M6. Data source: FRED database of the Federal Reserve Bank of St. Louis. Bottom panel: 5-year expected CPI inflation, from the average response of the Survey of Professional Forecasters, 2005:Q3 to 2010:Q2. Data source: Federal Reserve Bank of Philadelphia.



**Figure 3.2:** Federal Reserve assets

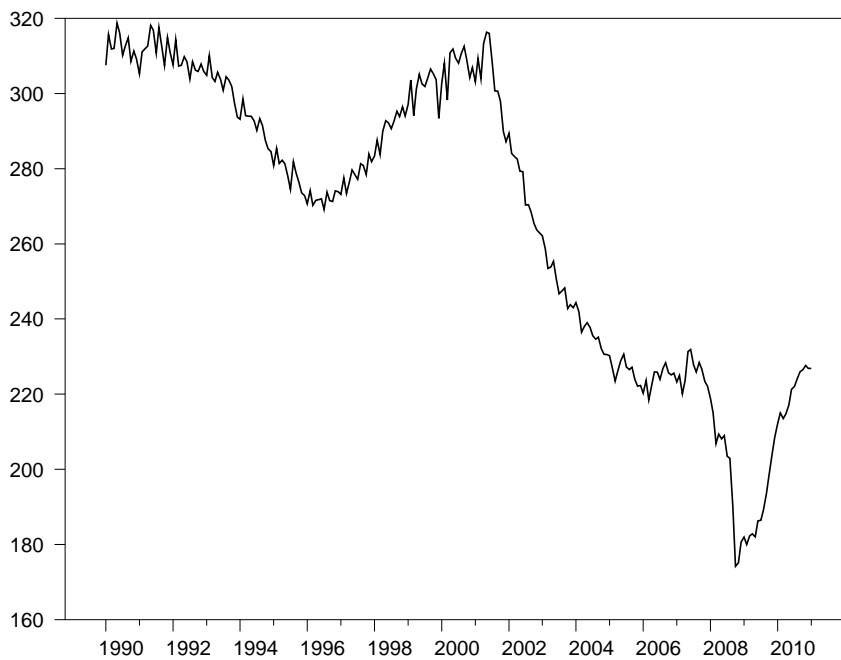
Federal Reserve assets, in billions of dollars, Jan 3, 2007 to Aug 4, 2010, Wednesday values, seasonally unadjusted, from Federal Reserve H41 release. Maiden 1: net portfolio holdings of Maiden Lane LLC; MMIFL: net portfolio holdings of LLCs funded through the Money Market Investor Funding Facility; TALF: loans extended through Term Asset-Backed Securities Loan Facility; AIG: sum of credit extended to American International Group, Inc. plus net portfolio holdings of Maiden Lane II and III; ABCP: loans extended to Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility; PDCF: loans extended to primary dealer and other broker-dealer credit; discount: sum of primary credit, secondary credit, and seasonal credit; swaps: central bank liquidity swaps; CPLF: net portfolio holdings of LLCs funded through the Commercial Paper Funding Facility; TAC: term auction credit; RP: repurchase agreements; MBS: mortgage-backed securities held outright; agency: federal agency debt securities held outright; misc: sum of float, gold stock, special drawing rights certificate account, and Treasury currency outstanding; other FR: Other Federal Reserve assets; treasuries: U.S. Treasury securities held outright.





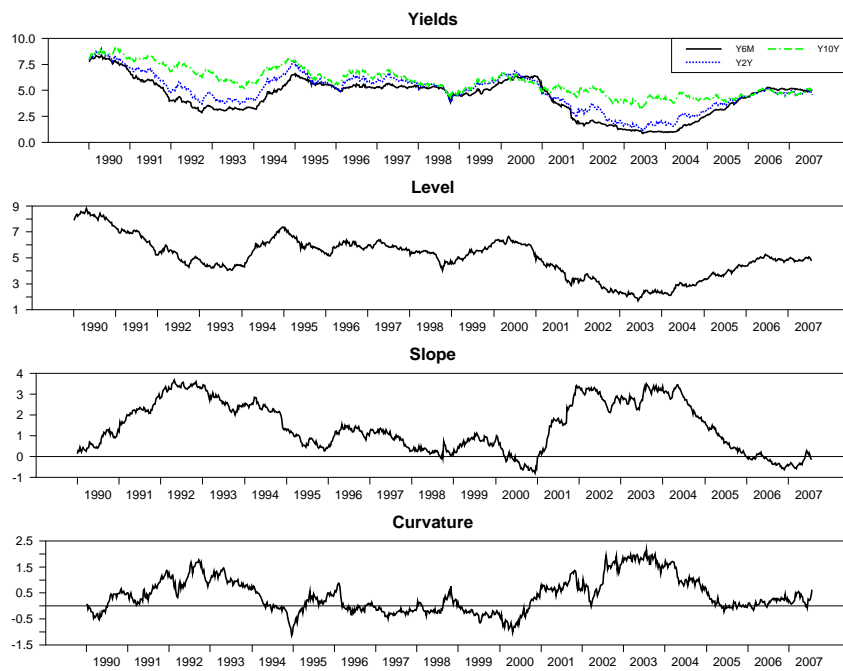
**Figure 3.3:** Maturity structure of U.S. federal debt

Maturity structure of U.S. federal debt as of December 31, 2006. Horizontal axis: maturity in weeks. Black bars: face value of marketable nominal Treasury securities of that maturity, in millions of dollars. Light bars: imputed holdings of the System Open Market Account of the U.S. Federal Reserve.



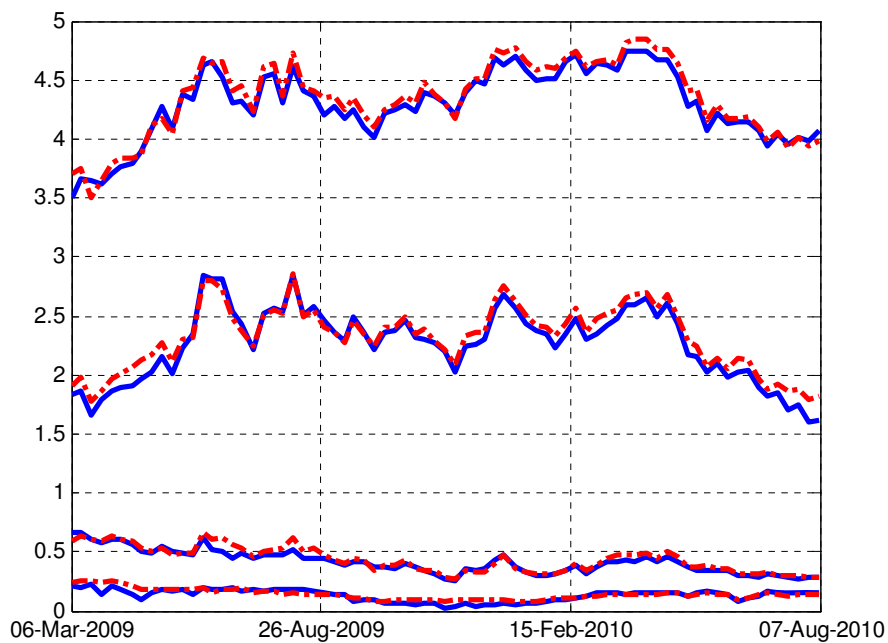
**Figure 3.4:** Average maturity

Average maturity in weeks of debt held by the public, plotted monthly from Jan 31, 1990 to Jan 31, 2011.



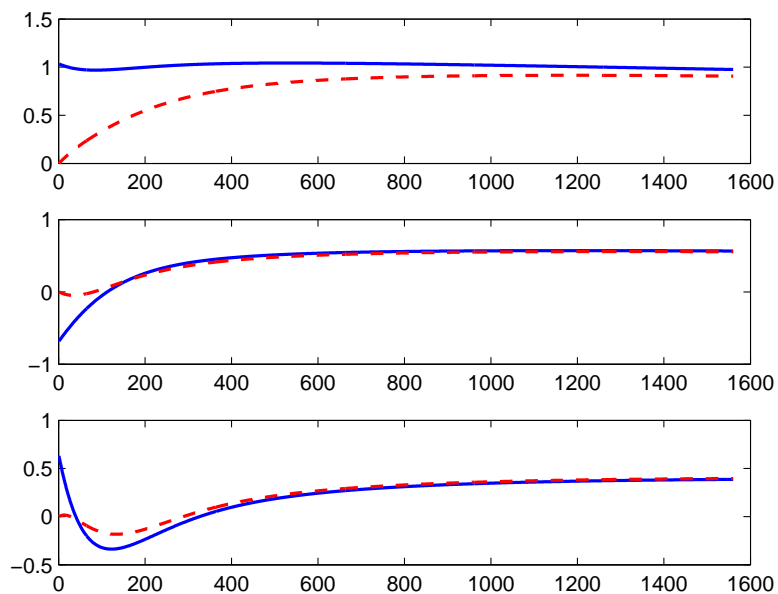
**Figure 3.5:** Yields and factors

Yields and factors used in baseline estimation, weekly from Jan 5, 1990 to July 27, 2007.



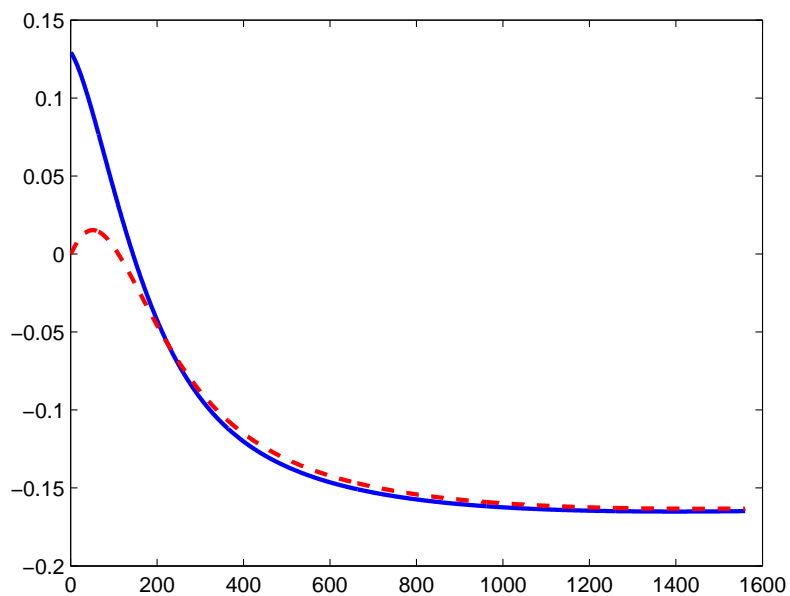
**Figure 3.6:** Actual and model fitted interest rates

Actual (solid) and predicted (dashed) behavior of selected interest rates, weekly from March 7, 2009 to August 10, 2010. Rates shown (in order from top to bottom) are the 30 year, 5 year, 1 year, and 3 month.



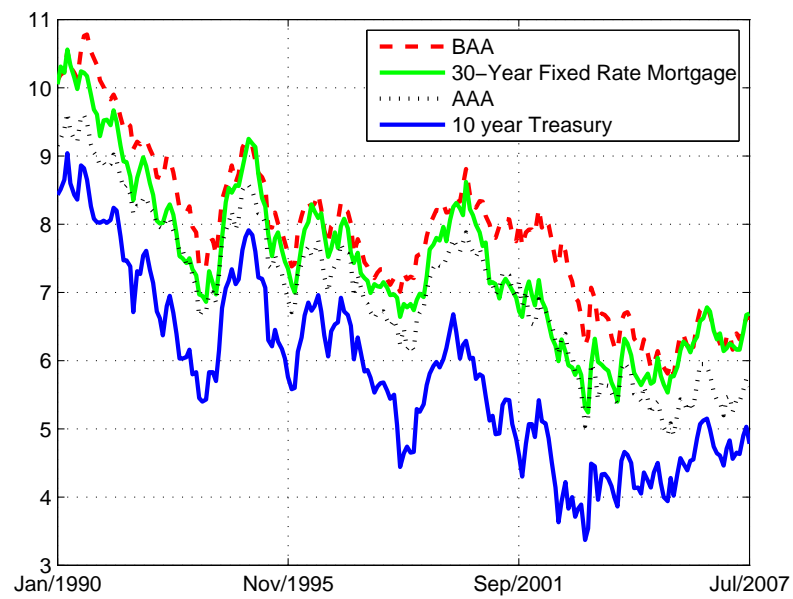
**Figure 3.7:** Factor loadings

Solid curves: normal loadings (plots of  $5200b_n$  as function of maturity  $n$  in weeks).  
Dashed curves: zero-lower-bound loadings ( $5200b_n^*$ ). Top panel: level loadings; middle panel: slope loadings; bottom panel: curvature loadings.



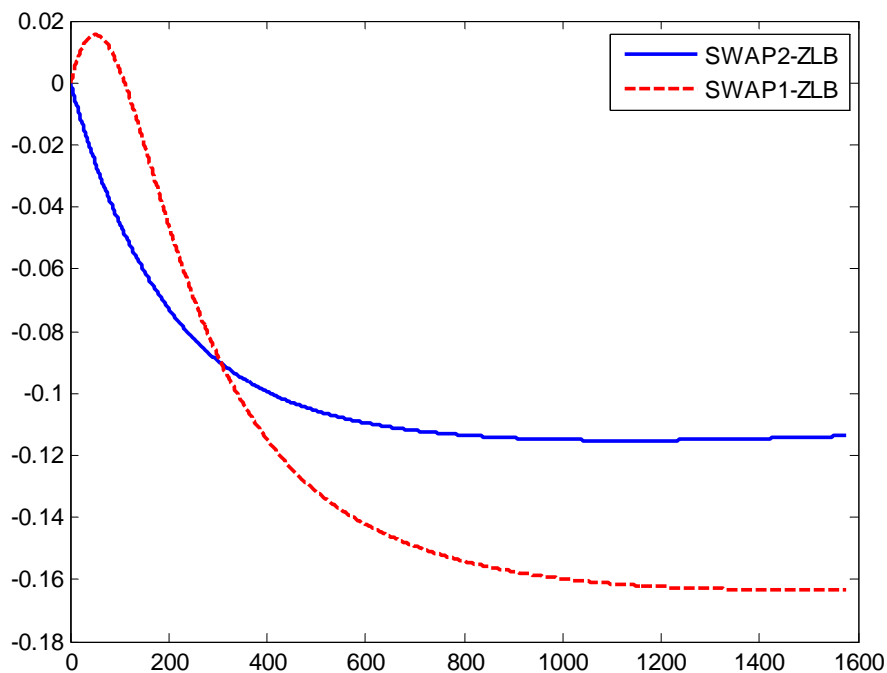
**Figure 3.8:** Predicted change

Predicted change in  $y_{n,t+1}$  (quoted in annual percentage points) as a function of weeks to maturity  $n$  in response to shift in  $q_t$  of size  $\Delta$ . Solid: effect in normal times (plot of  $5200b'_n\phi\Delta$  as a function of  $n$ ); dashed: effect at the zero lower bound (plot of  $5200b^*'_n\phi\Delta$ ).



**Figure 3.9:** Assorted long-term yields

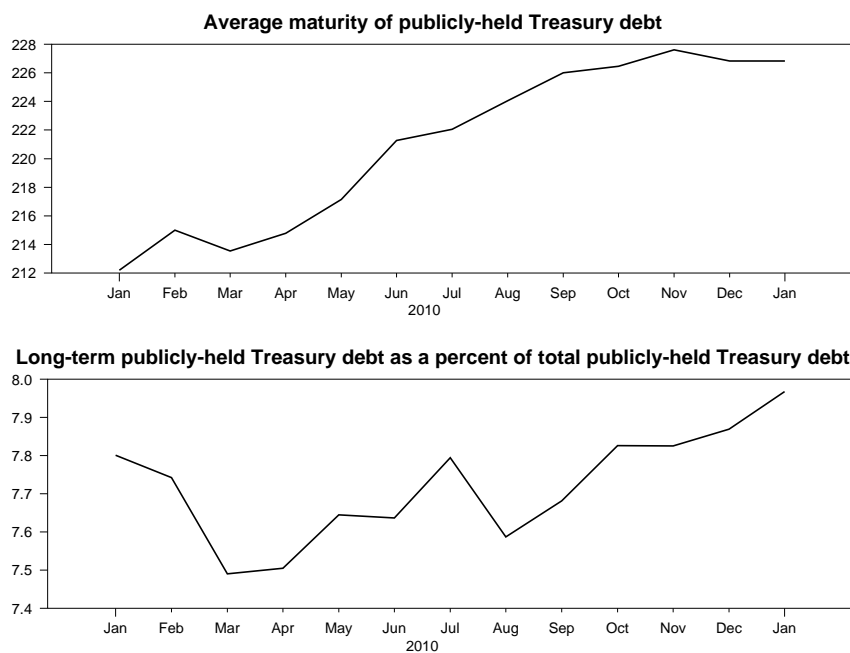
Assorted long-term yields, 1990:M1 to 2007:M7.



**Figure 3.10:** Assorted long-term yields

Effects of two different maturity swaps when implemented at the zero lower bound. Dashed curve: Fed sells all its holdings of less than 1-year maturity and retires debt at the longest end of the maturity structure (plot of  $5200b_n^* \phi \Delta$  as a function of  $n$ ). Solid curve: Fed sells all its holdings of less than 1-year maturity and retire debt evenly across 2-1/2 to 10 year maturities (plot of  $5200b_n^* \phi \Delta_2$ ).





**Figure 3.11:** Summaries for maturity structure

Top panel: average maturity of Treasury debt other than that held by the Federal Reserve ( $z_t^A$ ), 2010:M1-2011:M1. Bottom panel: fraction of outstanding Treasury debt not held by the Federal Reserve that is of 10 years or longer maturity ( $z_t^L$ ), 2010:M1-2011:M1.

# Bibliography

- Adrian, Tobias, Karin Kimbrough, and Dina Marchioni (2010) ‘The federal reserve’s commercial paper funding facility.’ *FRBNY Economic Policy Review*
- Ang, Andrew, and Monika Piazzesi (2003) ‘A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables.’ *Journal of Monetary Economics* 50, 745–787
- Ang, Andrew, Monika Piazzesi, and Min Wei (2006) ‘What does the yield curve tell us about gdp growth.’ *Journal of Econometrics* 131, 359–403
- Auerbach, Alan J., and Maurice Obstfeld (2005) ‘The case for open-market purchases in a liquidity trap.’ *American Economic Review* 95, 110–137
- Baumeister, Christiane, and Luca Benati (2010) ‘Unconventional monetary policy and the great recession: Estimating the impact of a compression in the yield spread at the zero lower bound.’ European Central Bank Working Paper
- Bernanke, Ben S., Vincent R. Reinhart, and Brian P. Sack (2004) ‘Monetary policy alternatives at the zero bound: An empirical assessment.’ *Brookings Papers on Economic Activity* 2, 1–100
- Bikbov, Ruslan, and Mikhail Chernov (2008) ‘Monetary Policy Regimes and the Term Structure of Interest Rates.’ *SSRN eLibrary*
- Chen, Ren-Row, and Louis Scott (1993) ‘Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates.’ *The Journal of Fixed Income* 3, 14–31
- Christensen, Jens H. E., Francis X. Diebold, and Glenn D. Rudebusch (forthcoming) ‘The affine arbitrage-free class of nelson-siegel term structure models.’ *Journal of Econometrics*
- Christensen, Jens H. E., Jose A. Lopez, and Glenn D. Rudebusch (2009) ‘Do central bank liquidity facilities affect interbank lending rates?’ Federal Reserve Bank of San Francisco, Working paper 2009-13

- Cochrane, John H., and Monika Piazzesi (2005) 'Bond risk premia.' *American Economic Review* 95, 138–160
- (2009) 'Decomposing the yield curve.' AFA 2010 Atlanta Meetings Paper
- Dai, Qiang, and Kenneth J. Singleton (2000) 'Specification analysis of affine term structure models.' *The Journal of Finance* 55, 1943–1978
- (2002) 'Expectation puzzles, time-varying risk premia, and affine models of the term structure.' *Journal of Financial Economics* 63, 415–441
- D'Amico, Stefania, and Thomas B. King (2010) 'Flow and stock effects of large-scale treasury purchases.' Working paper, Federal Reserve Board
- Dewachter, Hans, and Marco Lyrio (2006) 'Macro factors and the term structure of interest rates.' *Journal of Money, Credit & Banking (Ohio State University Press)* 38(1), 119 – 140
- Diebold, Francis X., and Canlin Li (2006) 'Forecasting the term structure of government bond yields.' *Journal of Econometrics* 130(2), 337 – 364
- Doh, Taeyoung (2010) 'The efficacy of large-scale asset purchases at the zero lower bound.' *Federal Reserve Bank of Kansas City Economic Review* Q2, 5–34
- Duarte, Jefferson (2004) 'Evaluating an alternative risk preference in affine term structure models.' *Review of Financial Studies* 17, 379–404
- Duffee, Gregory R. (2002) 'Term premia and interest rate forecasts in affine models.' *The Journal of Finance* 57, 405–443
- Duygan-Bump, Burcu, Patrick M. Parkinson, Eric S. Rosengren, Gustavo A. Suarez, and Paul S. Willen (2010) 'How effective were the federal reserve emergency liquidity facilities? evidence from the asset-backed commercial paper money market mutual fund liquidity facility.' Working paper, Federal Reserve Bank of Boston
- Eggertsson, Gauti, and Michael Woodford (2003) 'The zero bound on interest rates and optimal monetary policy.' *Brookings Papers on Economic Activity* 1, 139–211
- Estrella, Arturo, and Frederic S. Mishkin (1997) 'The predictive power of the term structure of interest rates in europe and the united states: Implications for the european central bank.' *European Economic Review* 41(7), 1375 – 1401
- Evans, Charles L., and David A. Marshall (2007) 'Economic determinants of the nominal treasury yield curve.' *Journal of Monetary Economics* 54(7), 1986 – 2003

- Gagnon, Joseph, Matthew Raskin, Julie Remache, and Brian Sack (2010) ‘Large-scale asset purchases by the federal reserve: Did they work?’ Federal Reserve Bank of New York Staff Reports
- Greenwood, Robin, and Dimitri Vayanos (2010) ‘Bond supply and excess bond returns.’ Working Paper, London School of Economics
- Gürkaynak, Refet S., and Jonathan H. Wright (2010) ‘Macroeconomics and the term structure.’ Working Paper, Johns Hopkins University
- Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007) ‘The u.s. treasury yield curve: 1961 to the present.’ *Journal of Monetary Economics* 54(8), 2291–2304
- Hamilton, James D. (1994) *Time Series Analysis* (Princeton, New Jersey: Princeton University Press)
- Hamilton, James D., and Jing Cynthia Wu (2010a) ‘The effectiveness of alternative monetary policy tools in a zero lower bound environment.’ Working paper, University of California, San Diego
- (2010b) ‘Identification and estimation of affine-term-structure models.’ Working paper, University of California, San Diego
- (2010c) ‘Testable implications of affine-term-structure models.’ Working paper, University of California, San Diego
- Hancock, Diana, and Wayne Passmore (2011) ‘Did the federal reserve’s mbs purchase program lower mortgage rates?’ Working paper, Federal Reserve Board
- Hördahl, Peter, Oreste Tristani, and David Vestin (2006) ‘A joint econometric model of macroeconomic and term-structure dynamics.’ *Journal of Econometrics* 131(1-2), 405 – 444
- Kim, Don H. (2008) ‘Challenges in macro-finance modeling.’ BIS Working Paper No. 240, FEDS Working Paper No. 2008-06
- Kim, Don H., and Athanasios Orphanides (2005) ‘Term structure estimation with survey data on interest rate forecasts.’ Federal Reserve Board, Finance and Economics Discussion Series 2005-48
- Kim, Don H., and Jonathan H. Wright (2005) ‘An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates.’ Federal Reserve Board, Finance and Economics Discussion Series 2005-33

- Kitchen, John, and Menzie D. Chinn (2010) 'Financing u.s. debt: Is there enough money in the world— and at what cost?' Working paper, University of Wisconsin-Madison
- Krishnamurthy, Arvind, and Annette Vissing-Jorgensen (2010) 'The aggregate demand for treasury debt.' Working paper, Northwestern University
- Krugman, Paul R. (1998) 'It's baaack: Japan's slump and the return of the liquidity trap.' *Brookings Papers on Economic Activity* 2, 137–205
- Kuttner, Kenneth N. (2006) 'Can central banks target bond prices?' NBER Working paper 12454
- McAndrews, James, Asani Sarkar, and Zhenyu Wang (2008) 'The effect of the term auction facility on the london inter-bank offered rate.' Working paper, Federal Reserve Bank of New York
- McCallum, Bennett T. (2000) 'Theoretical analysis regarding a zero lower bound on nominal interest rates.' *Journal of Money, Credit and Banking* 32, 870–904
- Merton, Robert C. (1969) 'Lifetime portfolio selection under uncertainty: The continuous time case.' *Review of Economics and Statistics* 51, 247–257
- Pericoli, Marcello, and Marco Taboga (2008) 'Canonical term-structure models with observable factors and the dynamics of bond risk premia.' *Journal of Money, Credit and Banking* 40, 1471–1488
- Roley, V. Vance (1982) 'The effect of federal debt-management policy on corporate bond and equity yields.' *Quarterly Journal of Economics* 97, 645–668
- Rudebusch, Glenn D., and Tao Wu (2008) 'A macro-finance model of the term structure, monetary policy and the economy.' *The Economic Journal* 118, 906–926
- Singleton, Kenneth J. (2006) *Empirical Dynamic Asset Pricing* (Princeton University Press)
- Stambaugh, Robert F. (1999) 'Predictive regressions.' *Journal of Financial Economics* 54, 375–421
- Svensson, Lars E.O. (2001) 'The zero bound in an open economy: A foolproof way of escaping from a liquidity trap.' *Monetary and Economic Studies* 19(S-1), 277–312
- Swanson, Eric T. (2011) 'Let's twist again: A high-frequency event-study analysis of operation twist and its implications for qe2.' Working paper, Federal Reserve Bank of San Francisco

- Taylor, John B. (1993) 'Discretion versus policy rules in practice.' *Carnegie-Rochester Conference Series on Public Policy* 39, 195 – 214
- Taylor, John B., and John C. Williams (2009) 'A black swan in the money market.' *American Economic Journal: Macroeconomics* 1, 58–83
- Vayanos, Dimitri, and Jean-Luc Vila (2009) 'A preferred-habitat model of the term structure of interest rates.' Working paper, London School of Economics
- Woodford, Michael (2010) 'Comments on ashcraft, garleanu, and pedersen, 'two monetary tools: interest rates and haircuts' .' In *NBER Macroeconomics Annual 2010*, ed. Daron Acemoglu and Michael Woodford, vol. 25 (Chicago, IL: University of Chicago Press)