UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Forecasting and the Price of Risk in Commodity and Bond Markets

Permalink
https://escholarship.org/uc/item/9vm860fn

Author
Zhecheva, Irina Yurieva

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Forecasting and the Price of Risk in Commodity and Bond Markets

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

in

Economics

by

Irina Yurieva Zhecheva

Committee in charge:

>   Professor James Hamilton, Chair
>   Professor Ivana Komunjer
>   Professor Allan Timmermann
>   Professor Rossen Valkanov
>   Professor Ruth Williams

2017

The Dissertation of Irina Yurieva Zhecheva is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California, San Diego

2017

DEDICATION

To my mother Ivanka Hristova

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Professor James Hamilton for his invaluable guidance and endless support. He has been incredibly inspiring and helpful, and I am extremely indebted to him. I could not have asked for a better advisor. It was a privilege to learn from him and have him as a mentor.

I am very grateful to my other committee members, Professor Allan Timmermann, Professor Rossen Valkanov, Professor Ivana Komunjer, and Professor Ruth Williams. I am very grateful for having them on my dissertation committee.

I would also like to acknowledge Johannes Wieland, David Lagakos, Julie Cullen, Mark Machina, Yixiao Sun, Brendan Beare, Asad Dossani, Shihan Xie, Leland Farmer, Dora Xia, Lawrence Schmidt, Sung Je Byun, Xuan Liang, Hie Joo Ahn, and seminar participants at UCSD for their helpful comments.

Finally, I would like to thank my mother Ivanka Hristova for always being there for me and supporting me unconditionally.

Chapter 1, in full, is currently being prepared for submission for publication of the material. Zhecheva, Irina Y. The dissertation author was the primary investigator and author of this material.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Zhecheva, Irina Y. The dissertation author was the primary investigator and author of this material.

VITA

2008 B.A. in Mathematics and Economics, Williams College

2012 M.A. in Economics, Univerisity of California, San Diego

2017 Ph.D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Forecasting and the Price of Risk in Commodity and Bond Markets

by

Irina Yurieva Zhecheva

Doctor of Philosophy in Economics

University of California, San Diego, 2017

Professor James Hamilton, Chair

In the first two chapters of my dissertation, I study the pricing of risk in commodities futures and bond markets. In the first chapter, I provide a new way to characterize risk in commodities futures markets. I apply my framework to the natural gas futures market and study the consequences of changes in regime on the risk premium. In the second chapter, I study risk pricing in bond yields and investigate whether regime shifts are important for our understanding of bond risk premia and the term structure. I produce novel empirical estimates to characterize risk premia and the term structure of bond yields and natural gas

futures contracts. I also propose a new method for estimating Gaussian affine term structure models subject to regime switching, which solves the serious numerical difficulties encountered by other methods in the literature. The third chapter of my dissertation investigates whether forecast aggregation helps in forecasting commodity prices.

Chapter 1

Pricing of Risk in Natural Gas Futures Contracts: A New Approach to Affine Term Structure Models Subject to Changes in Regime

Abstract

Commodities alternate between periods of contango, when the near month futures trade at a discount to back month futures, and backwardation, when near month futures are priced higher than back month futures. Market participants (hedgers and speculators) are exposed to substantial risk due to the possibility of the market switching from one state to the other. I provide a new way to characterize these risks in commodities futures markets. I apply this framework to the natural gas futures market where these risks are particularly substantial, and study the consequences of changes in regime on the risk premium. Motivated by the historically observed switches of the natural gas market between states of contango and backwardation, I propose and estimate a Markov regime-switching Gaussian affine term structure model and estimate the states of the market from

the data. I find very strong evidence that there are regimes in the data, using a dataset of prices from 1995 to 2014. Moreover, the regimes correspond precisely to historically observed periods of contango and backwardation. I find that the market acts as if regimes are more persistent than they really are, which could be a result of hedging pressure. Moreover, I find that regime switching risk is priced. Calculating expected returns conditional on each regime confirms the claim that agents face significant risks from the possibility of a change in the regime. The maximum likelihood based methods common in the literature for estimating Gaussian affine term structure models with regime switching pose significant numerical challenges. A separate contribution of this paper is to propose a new approach to estimating this type of models. My approach avoids these numerical difficulties and allows for computationally fast estimation.

1.1   Introduction

Commodities alternate between periods when the spread between the longer term and shorter term futures contracts is positive and periods when it is negative. When the spread between the longer term and the shorter term contracts is positive, commodities futures are said to be in contango, whereas when that spread is negative, they are said to be in backwardation. Participants in commodities futures markets are exposed to substantial risks due to the possibility of the market switching from one state to the other. This paper provides a new way to characterize these risks in commodities futures markets using a regime switching approach.

As futures contracts approach expiry, investors who want to stay long in a given contract roll their position, i.e. sell their position and buy the next month contract. In backwardation, investors can often earn a positive return by rolling their position, i.e. selling out of the contract at a higher price than what they

initially paid[1], and reinvesting at a lower price. That is why a common trading strategy in backwardation is to be long in a futures contract and roll your position each month. In contango, investors sell out of the contract at a lower price than what they initially paid.[2] Hence, a common trading strategy in contango is to be short in a futures contract and roll your position each month. As long as the market continues to be in backwardation, an investor may be getting a positive return by being long, but if the market unexpectedly switches to contango, the investor would likely experience a significant loss. Therefore, investors in commodities futures face significant risk due to the possibility of the market switching from one state to the other, and it is crucial to be able to determine the turning points.

Natural gas futures prices are very volatile, and the risks faced by market participants due to the possibility of switching from one state to the other are particularly substantial. Thus, the natural gas market is a useful setting for studying this type of risk. Natural gas is one of the most heavily traded commodity futures contracts in the United States. The natural gas futures market is highly liquid with daily open interest of up to about 300,000 contracts for the one month contract and total open interest of up to 1,400,000 contracts. The second contribution of this paper is to characterize these risks in the natural gas futures market. The method I develop is particularly well suited for identifying these risks in this setting. Motivated by the historically observed switches of natural gas futures prices between states of contango and backwardation, I propose and estimate a Markov regime-switching Gaussian affine term structure model and estimate the

---

[1]This is the case if the term structure of futures prices has approximately the same level and slope at time t when the contract was bought and time $t + 1$ when investors sell out of the contract.

[2]This is the case if the term structure of futures prices has approximately the same level and slope at time t when the contract was bought and time $t + 1$ when investors sell out of the contract.

states of the market from the data. In my model, the level and volatility of natural gas futures are allowed to switch with the regime. The risk premium is also regime dependent. I choose a model with two regimes in order to be able to see how the regimes estimated from the data are related to the two states of contango and backwardation observed in the market. I find that the estimated regimes correspond exactly to periods of contango and backwardation observed in the data. Moreover, I find very strong statistical evidence that there are regimes in the data. I also find that regime shift risk is priced.

I provide empirical estimates to characterize risk premia and the term structure of natural gas futures. I find that expected futures returns change sign depending on which regime we are conditioning on.[3] For example, I find that in the backwardation regime, an investor with a long position in the 9-month contract would on average earn a positive expected monthly holding return of 2.5%. However, in the contango regime, the investor would earn a negative expected monthly holding return of about -0.5% on average. Thus, as long as the market is in backwardation, the investor would on average profit from being long in the 9-month contract. However, if the regime switches to contango, the investor would experience a substantial loss. This is consistent with the claim that agents face significant risks from the possibility of a change in the regime.

The third contribution of this paper is methodological. I propose a new approach to the estimation of Markov regime-switching Gaussian affine term structure models, which solves the numerical issues encountered by other estimation methods in the literature. Most methods for estimating Gaussian affine term structure models in the literature are based on maximum likelihood and exploit both the distributional assumptions and the no-arbitrage restrictions. These methods

---

[3]This holds for all contracts except for the 6-month contract.

estimate latent pricing factors jointly with the model parameters by explicitly imposing the cross-equation no-arbitrage restrictions. As a result, these methods rely on numerical optimization, are computationally intensive, and pose significant numerical challenges, which become especially severe when there is the possibility of changes in regime. Estimation difficulties commonly arise due to highly non-linear and badly behaved likelihood surfaces, which are flat along many directions of the parameter space, and it is hard to achieve convergence. These problems can make estimation by MLE very difficult or infeasible. These difficulties have been documented by multiple researchers such as Kim and Orphanides (2005), Duffee (2002), Ang and Piazzesi (2003), Kim (2008), Duffee and Stanton (2008), Duffee (2009), and Ang and Bekaert (2002). To facilitate estimation of my model, I propose a new estimation method for Markov regime-switching affine term structure models. I use a regression based method to estimate the reduced-form parameters in the first stage, and then estimate the prices of risk and risk-neutral transition probabilities via minimum-chi-square estimation in the second stage. The no-arbitrage restrictions are not used or imposed in the first-stage reduced-form estimation, but are exploited in the second stage of the estimation. The minimum-chi-square procedure chooses the values of the structural parameters so that the values for the recursive pricing parameters implied by the no-arbitrage restrictions most closely fit the unrestricted first-stage estimates. This approach is asymptotically equivalent to full information maximum likelihood. The numerical component in the second stage is far simpler to implement than the one associated with other maximum likelihood based methods in the literature. In this way, I bypass the numerical difficulties encountered by other methods and have no problem achieving convergence. Another advantage of the minimum-chi-square approach is that the value of the objective function provides a way to test whether the no-arbitrage restrictions are consistent with the

data. I show how to price the time series and cross-section of the term structure of commodities futures prices in the case of regime switching, and apply my approach to estimate my model.

The class of Gaussian affine term structure models was originally developed by Vasicek (1977), Duffie and Kan (1996), Dai and Singleton (2000), Duffee (2002), and Piazzesi (2010) to characterize the relation between yields on bonds of different maturities. The Gaussian affine term structure framework is based on the assumptions that the pricing kernel is exponentially affine, prices of risk are affine in the state variables, and innovations to the state variables are conditionally Gaussian. Under these assumptions, the price process is affine in the state variables, and no-arbitrage restrictions constrain the coefficients on the state variables.

Hamilton and Wu (2014) adapt this class of models to commodities. They show that an affine factor structure of commodity futures prices can result from the interaction between arbitrageurs and commercial producers seeking hedges or financial investors seeking diversification. Schwartz (1997), Schwartz and Smith (2000), and Casassus and Collin-Dufresne (2006), among others, also develop related models to describe commodity futures prices.

Regime-switching models for the term structure of interest rates have been proposed and estimated by Dai, Singleton, and Yang (2007), Bansal and Zhou (2002), and Ang and Bekaert (2002) among others. Dai, Singleton, and Yang (2007) and Ang and Bekaert (2002) use maximum-likelihood based methods based on an iterative procedure developed by Hamilton (1989). Bansal and Zhou (2002) use a two-step efficient method of moments estimator. These methods are subject to the numerical issues mentioned earlier, which this paper resolves.

Almansour (2016) models the futures term structure of crude oil and nat-

ural gas using a convenience yield model.[4] He extends the two-factor stochastic convenience yield model of Gibson and Schwartz (1990) to allow the convenience yield level as well as other parameters to be regime dependent. He uses a two-factor model with the log of the spot price and the convenience yield as latent factors. The main advantages of my framework over his is that my framework gives a clear way to test the underlying model assumptions, while Almansour's does not. I provide results of a test of the underlying assumptions, and show that they are satisfied. Moreover, I provide tests of a number of hypotheses about the differences in regime, which he does not. Furthermore, Almansour does not allow the average values of the historic regime-switching transition probabilities (which he assumes to be time-varying) to differ from the risk-neutral transition probabilities, and I show that is key for risk pricing in the natural gas futures market. Another difference is that I discuss the implications for investment strategy and hedging, whereas he does not.

Adrian, Crump and Moench (2013) and Diez de los Rios (2015) have recently proposed regression based algorithms for estimation of single-regime affine term structure models that avoid the numerical difficulties associated with maximum likelihood estimation. The approach I propose in this paper allows for regime-switching and combines regression based and numerical calculations.

1.1.1 Background on the natural gas market

Natural gas accounts for 30% of U.S. electricity generation and is predicted to account for an even larger proportion in the future. This commodity has gained a lot of attention in recent years with the discovery of shale gas and the advancement of drilling technology.

---

[4]Convenience yield is the benefit derived from holding the underlying physical commodity rather than the futures contract.

Some of the potential reasons for the existence of the two market states of contango and backwardation are shortages, weather concerns, or geopolitical events. Contango can result from trader perceptions of future shortage or short-term supply glut, whereas backwardation can result from near term shortage or future supply glut. For example, if there is a hurricane that is expected to disrupt supply only for a short time, the near term futures may spike while the longer term futures can be relatively unaffected. This unanticipated weather event would induce a period of backwardation.

The transitions from one market state to another are in part due to seasonal changes in supply and demand conditions. Natural gas production is relatively constant throughout the year, but consumption tends to peak during the winter heating season (November through March) as home heating use rises and tends to be moderate in other seasons. In general, there is often a shortage of natural gas in the winter as supply is often unable to react quickly to short-term increases in demand, so winter months are often characterized by backwardation. At the same time, the inventory stored for the winter can help to meet demand, and if there is no current shortage, the market may not be in backwardation in the winter. In the remaining seasons when supply and demand are in balance or there is short-term oversupply, the market can often be in contango. Other factors affecting supply and demand can also affect whether the market is in contango or backwardation. On the supply side, amount of gas in storage, pipeline capacity, and imperfect information about storage can affect the state of the market.

Most of U.S. natural gas consumption is from domestic production. U.S. dry production has been steadily increasing since 2006. It reached its highest recorded annual total in 2015 and is still rising during 2016. The increases in production were due to more efficient, cost-effective drilling techniques which have allowed for

horizontal drilling in shale formations. This has led to an unprecedented surge in supply. As a result, since around 2009, the natural gas market has been in contango. This observation is captured by my model, which classifies the data sample post 2009 as being in the contango regime.

The fluctuations in the amount of natural gas in storage and the expectations of future oversupply or shortage can cause the market to switch between contango and backwardation. Although the fluctuations of prices between periods of contango and periods of backwardation have a seasonal component, they are far from deterministic. In some years there is virtually no shortage of natural gas in the winter, while in others there could be shortage well beyond the winter months. Thus, backwardation does not always occur in the winter, and other seasons are not always characterized by contango. The duration of episodes of contango and backwardation also varies from year to year. The type of nondeterministic seasonality observed in the natural gas market cannot be fully captured by seasonal dummy models or other methods for modeling deterministic seasonality. To capture these historically observed fluctuations, I model the level and volatility of natural gas futures using a Markov regime-switching affine term structure model and estimate the states of the market from the data.

The uncertainty about gas prices introduces the possibility that commercial producers or commercial users may at times make significant use of natural gas futures contracts for purposes of hedging. If commercial producers believe that the price of natural gas may fall in the future, they can take a short position in futures to secure a higher selling price for the natural gas they produce. Conversely, if commercial users think that the price of natural gas may rise in the future, they can take a long position in futures to lock in a lower price for natural gas. Natural gas futures are also traded by speculators, who are willing to assume the price risk that

hedgers try to avoid in return for a risk premium. I find evidence that commercial users and commercial producers use natural gas futures contracts for purposes of hedging. I find that in the contango regime, commercial producers may be trying to hedge their short positions in natural gas by selling 4-6 month contracts, while in the backwardation regime, commercial producers may be trying to hedge their short positions by selling 6-9 month contracts. I obtain analogous implications for commercial users. Hamilton and Wu (2014) show how variation in hedging pressure could influence the term structure of commodity futures prices. I study the consequences of changes in regime on the risk premium by generalizing the futures-pricing model in Hamilton and Wu (2014) to allow for changes in regime.

The rest of the paper is organized as follows. Sections 1.2 and 1.3 present the model framework, Section 1.4 describes the estimation approach, Section 1.5 gives empirical results for a one factor model of natural gas futures prices from 1995 to 2014, and Section 1.6 concludes.

1.2   Model

Let $F_t^{(n)}$ be the price of a futures contract with maturity n at time t. I assume that the log of the price is a function of a factor X_t that follows a Gaussian autoregression:

$$X_{t+1} = \mu_{s_t} + \Phi X_t + v_{t+1} \ , \ v_{t+1}|s_t \sim N(0, \sigma_{s_t}^2) \tag{1.1}$$

I find that a one factor model using the first principal component of the log of the futures prices with maturities 3 months to 9 months as a factor describes the data well.[5] Thus, in my empirical application X_t will be a scalar. I find that shorter

---

[5]In section 1.5.2 I provide some evidence on why a one factor model is appropriate.

term contracts have a systematically different behavior and do not fit well in my framework together with the long term contracts.

The intercept parameter is regime switching, and s_t denotes the regime at time t, $s_t \in \{1, 2\}$. I assume that the slope parameter Φ is regime independent.[6]

The no-arbitrage assumption implies the existence of a pricing kernel $M_{t,t+1}$ such that

$$F_t^{(n)} = E_t^{\mathbb{P}} \left[M_{t,t+1} F_{t+1}^{(n-1)} | s_t = j \right] \tag{1.2}$$

if the regime at time t is j. Following Dai et al. (2007), I assume that the pricing kernel is exponentially affine and takes the following form:

$$M_{t,t+1} = \exp \left[-\Gamma_{s_t, s_{t+1}} - \frac{1}{2} \lambda_{t,s_t}^2 - \lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right] \tag{1.3}$$

where $\Gamma_{s_t, s_{t+1}}$ is the market price of regime shift risk, λ_{t,s_t} is the market price of factor risk, and v_{t+1} is the innovation from the factor equation (1.1). $\Gamma_{s_t, s_{t+1}}$ is referred to as the market price of regime shift risk because it can be interpreted as the log expected return per unit of regime shift risk exposure as I show in equation (1.29) in section (1.5.2). I also assume that the market price of factor risk λ_{t,s_t} is an affine function of the state variable:

$$\lambda_{t,s_t} = \sigma_{s_t}^{-1} \left(\lambda_{0,s_t} + \lambda_1 X_t \right) \tag{1.4}$$

Here the market price of factor risk λ_{t,s_t} is time-varying and regime-dependent. I assume that λ_1 does not depend on the regime.[7] No arbitrage implies the existence

---

[6]I also estimated a version of the model with Φ allowed to vary with regime, but found that this led to only a trivial increase in the log likelihood for the system in equations (1.14)-(1.17) that I estimate. Since the equations are simpler and more intuitive with Φ constant, I only discuss the simpler case in this paper.

[7]When I estimated a version of the model in which Φ varies with the regime, I also allowed λ_1 to vary with the regime, denoting it λ_{1,s_t}. I failed to reject the hypothesis that $\lambda_{1,1} - \lambda_{1,2} = 0$.

of an equivalent martingale measure - the risk neutral measure $\mathbb{Q}$. The historic measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$ are related through the pricing kernel $M_{t,t+1}$. The price $P(X_t)$ of an asset with payoff $g(X_{t+1})$ in regime j can be computed as

$$P(X_t) = E_t^{\mathbb{P}}[M_{t,t+1}g(X_{t+1})|s_t = j] = E_t^{\mathbb{Q}}[g(X_{t+1})|s_t = j] \tag{1.5}$$

Under the $\mathbb{Q}$-measure, the factor X_t follows a Gaussian autoregression:

$$X_{t+1} = \mu_{s_t}^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t + v_{t+1}^{\mathbb{Q}} \tag{1.6}$$

where

$$\mu_j^{\mathbb{Q}} = \mu_j - \lambda_{0,j} \tag{1.7}$$

and

$$\Phi^{\mathbb{Q}} = \Phi - \lambda_1 \tag{1.8}$$

and $v_{t+1}^{\mathbb{Q}}|s_t = j \sim^{\mathbb{Q}} N(0, \sigma_j^2)$. The above relations are derived in Appendix 1.7.1.

Let $f_t^{(n)} \equiv \ln F_t^{(n)}$. Equations (1.1), (1.2), and (1.3) together imply that the log of the futures price is affine in the state variable:

$$f_t^{(n)} = A_{s_t}^{(n)} + B^{(n)} X_t + u_t^{(n)} \tag{1.9}$$

From equation (1.8) and equation (1.39), it follows that the factor loadings $B^{(n)}$ are regime independent. This ensures exact closed-form solutions for the futures prices, and is consistent with Dai et al. (2007). The intercept term is allowed to change with the regime.

Let $rx_{t+1}^{(n-1)}$ denote the one-month log holding return of a futures contract

---

Hence, I consider the simpler model in which λ_1 is regime independent.

maturing in n months:

$$rx_{t+1}^{(n-1)} = f_{t+1}^{(n-1)} - f_t^{(n)} \tag{1.10}$$

The holding return is the return on buying an n-month futures contract in month t and then selling it as an $(n-1)$-month futures contract in month $t+1$.

In my application, I assume that there are 2 regimes that govern the dynamic properties of the factor X_t. The unobserved regime variable s_t is presumed to follow a 2-state Markov chain, with the risk-neutral probability of switching from regime $s_t = j$ to regime $s_{t+1} = k$ given by $\pi^{\mathbb{Q}jk}, 1 \leq j, k \leq 2$, with $\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} = 1$, for $j = 1, 2$. I assume that the risk-neutral transition probabilities $\pi^{\mathbb{Q}jk}$ and the real-world transition probabilities $\pi^{\mathbb{P}jk}$ are regime independent. I allow $\pi^{\mathbb{P}jk} \neq \pi^{\mathbb{Q}jk}$. Agents are presumed to know the regime they are currently in, as well as the history of the factor X_t and of the regime. The econometrician does not observe the regime. The Markov process governing regime changes is assumed to be conditionally independent of the X_t process for tractability.

1.3　No arbitrage conditions for futures contract prices

Under my assumptions, the log of the futures price is affine in the factor X_t:

$$f_t^{(n)} = A_{s_t}^{(n)} + B^{(n)} X_t + u_t^{(n)} \tag{1.11}$$

My model implies the following cross-equation non-arbitrage restrictions on the parameters $A_j^{(n)}$, $j = 1, 2$ and $B^{(n)}$ characterizing the futures contract price:

$$A_j^{(n)} = \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}} \right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \tag{1.12}$$

for $j = 1, 2$ and

$$B^{(n)} = B^{(n-1)}(\Phi - \lambda_1) \qquad (1.13)$$

Equations (1.12) and (1.13) are derived in Appendix 1.7.2 in equations (1.38) and (1.40). They are very similar to the standard recursions for affine term structure models in the bond pricing literature (see for example Ang and Piazzesi (2003)). In bond pricing, the recursion for the intercept adds a term δ_0 corresponding to the interest earned each period. For commodities, such a term does not appear since there is no initial capital investment. The recursions above represent non-linear cross-equation no arbitrage restrictions. These restrictions are not used or imposed in the initial reduced-form estimation, but are exploited in the second stage of inference described below.

1.4 Estimation procedure

I assume that the factor X_t is observed, and it is the first principal component extracted from the demeaned log prices of the futures contracts with maturities from 3 months to 9 months.[8] Based on equations (1.9) and (1.1), I propose the following two-step method for estimating the parameters of the model.

---

[8]I treat the first principal component extracted from futures prices as observed. Internal consistency between actual and model-implied principal components requires $h'A_j = -k$ for $j = 1, 2$ and $h'B = 1$ where h is the 7×1 vector of principal component coefficients, A_j is the 7×1 vector of model-implied values of the coefficients $A_j^{(n)}$ from equation (1.12) (for $n = 3, \ldots, 9$), B is the 7×1 vector of model-implied coefficients $B^{(n)}$ from equation (1.13) (for $n = 3, \ldots, 9$), and $k = \frac{-h'\sum_{t=1}^{T} f_t}{T}$, where $f_t = \left(f_t^{(3)}, f_t^{(4)}, f_t^{(5)}, f_t^{(6)}, f_t^{(7)}, f_t^{(8)}, f_t^{(9)} \right)'$. I do not impose these internal consistency conditions in the estimation. However, I find that they are satisfied to a very large degree of accuracy, and any inconsistency is negligible. Specifically, I find that $h'A_1 = 3.7861$, $h'A_2 = 3.7875$, $k = -3.7868$, $h'B = 1.0021$, so that $h'A_1 \approx -k$, $h'A_2 \approx -k$, and $h'B \approx 1$.

1.4.1 Estimation of reduced-form parameters via regime-switching VAR's

First, I estimate the following regime-switching regressions:

$$f_t^{(n)} = A_{s_t}^{(n)} + B^{(n)} X_t + u_t^{(n)}, n = 3, \ldots, 9 \tag{1.14}$$

where

$$\begin{pmatrix} u_t^{(3)} \\ u_t^{(4)} \\ u_t^{(5)} \\ u_t^{(6)} \\ u_t^{(7)} \\ u_t^{(8)} \\ u_t^{(9)} \end{pmatrix} | s_t \sim N(0, \Omega) \tag{1.15}$$

and

$$\Omega \equiv \begin{pmatrix} \Omega_{(3)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \Omega_{(4)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Omega_{(5)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Omega_{(6)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Omega_{(7)} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Omega_{(8)} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Omega_{(9)} \end{pmatrix} \tag{1.16}$$

jointly with the regime-switching regression for X_t:

$$X_{t+1} = \mu_{s_t} + \Phi X_t + v_{t+1} \,, \ v_{t+1} | s_t \sim N(0, \sigma_{s_t}^2) \tag{1.17}$$

as a vector system of regime-switching equations. The time-series regressions in equation (1.14) estimate exposures of the futures prices with respect to the contemporaneous pricing factor. The regime-switching regression in equation (1.17) serves to decompose the pricing factor into a predictable component and a factor innovation by regressing the factor on its lagged level.

The estimation is done via the EM algorithm and is explained in detail in Appendix 1.7.3. The general vector version of the EM algorithm is found in Hamilton (2016).

1.4.2 Minimum-chi-square estimation of structural parameters

I use a minimum-chi-square approach to estimate the price of risk parameters $\lambda_{0,1}$, $\lambda_{0,2}$, and λ_1 and the risk-neutral probabilities $\pi^{\mathbb{Q}_{11}}$ and $\pi^{\mathbb{Q}_{22}}$. The procedure chooses the values of $\lambda_{0,1}$, $\lambda_{0,2}$, λ_1, $\pi^{\mathbb{Q}_{11}}$, and $\pi^{\mathbb{Q}_{22}}$ so that the values for the recursive pricing parameters $A_j^{(n)}$ and $B^{(n)}$ implied by the no-arbitrage restrictions in equations (1.12) and (1.13) most closely fit the unrestricted first-stage estimates from equation (1.14). Minimum-chi-square estimation in the setting of single regime Gaussian affine term structure models is described in Hamilton and Wu (2012).

Let π denote the vector of reduced-form parameters (VAR coefficients, variance of the factor, measurement error variances, and $\mathbb{P}$-measure regime-switching probabilities). Let $\mathcal{L}(\pi; Y)$ denote the log likelihood for the entire sample, and let $\hat{\pi} = \arg\max \mathcal{L}(\pi; Y)$ denote the full-information maximum likelihood estimate. If $\hat{R}$ is a consistent estimate of the information matrix,

$$R = -T^{-1} E \left[\frac{\partial^2 \mathcal{L}(\pi; Y)}{\partial \pi \partial \pi'} \right] \tag{1.18}$$

then θ can be estimated by minimizing the chi-square statistic

$$T\left[\hat{\pi} - g(\theta)\right]' \hat{R}\left[\hat{\pi} - g(\theta)\right] \tag{1.19}$$

As noted by Hamilton and Wu (2012),the variance of $\hat{\theta}$ can be approximated with $T^{-1}(\hat{\Gamma}'\hat{R}\hat{\Gamma})^{-1}$ for $\hat{\Gamma} = \frac{\partial g(\theta)}{\partial \theta'}\big|_{\theta=\hat{\theta}}$.

In my case, I want to minimize the distance between the unrestricted maximum likelihood estimates of the coefficients $A_j^{(n)}$ and $B^{(n)}$ (from the regime-switching regressions) and the values of $A_j^{(n)}$ and $B^{(n)}$ implied by the no arbitrage restrictions. According to equations (1.12) and (1.13), these are predicted to be functions of θ, a vector of structural parameters summarized in equation (1.22) below. Let $\hat{\pi}$ be the vector of the unrestricted maximum likelihood estimates from the regime-switching VAR:

$$\hat{\pi} = \big(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\Phi}, vec(\tilde{A}'), vec(\tilde{B}), \tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \tilde{\Omega}_{(3)}, \tilde{\Omega}_{(4)}, \tilde{\Omega}_{(5)}, \tilde{\Omega}_{(6)}, \tilde{\Omega}_{(7)}, \tilde{\Omega}_{(8)}, \tilde{\Omega}_{(9)}, \tilde{\pi}^{\mathbb{P}_{11}},$$
$$\tilde{\pi}^{\mathbb{P}_{22}}\big)' \tag{1.20}$$

where $\tilde{\mu}_1$, $\tilde{\mu}_2$, $\tilde{\Phi}$, $\tilde{\sigma}_1^2$, and $\tilde{\sigma}_2^2$ are the unrestricted maximum likelihood estimates of the parameters μ_1, μ_2, Φ, σ_1^2, and σ_2^2 from the regime-switching autoregression for

the factor in equation (1.17),

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1^{(3)} & \tilde{A}_2^{(3)} \\ \tilde{A}_1^{(4)} & \tilde{A}_2^{(4)} \\ \tilde{A}_1^{(5)} & \tilde{A}_2^{(5)} \\ \tilde{A}_1^{(6)} & \tilde{A}_2^{(6)} \\ \tilde{A}_1^{(7)} & \tilde{A}_2^{(7)} \\ \tilde{A}_1^{(8)} & \tilde{A}_2^{(8)} \\ \tilde{A}_1^{(9)} & \tilde{A}_2^{(9)} \end{bmatrix} \tag{1.21}$$

and $\tilde{B} = \left(\tilde{B}^{(3)}, \tilde{B}^{(4)}, \tilde{B}^{(5)}, \tilde{B}^{(6)}, \tilde{B}^{(7)}, \tilde{B}^{(8)}, \tilde{B}^{(9)} \right)'$ are the unrestricted maximum likelihood estimates of the coefficients of the regime-switching regressions for the futures prices in equation (1.14), $\tilde{\Omega}_{(n)}$, $n = 3 \ldots 9$ are the unrestricted maximum likelihood estimates of the measurement error variances $\Omega_{(n)}$, $n = 3 \ldots 9$ in equation (1.16), and $\tilde{\pi}^{\mathbb{P}_{11}}$ and $\tilde{\pi}^{\mathbb{P}_{22}}$ are the unrestricted maximum likelihood estimates of the regime switching probabilities $\pi^{\mathbb{P}_{11}}$ and $\pi^{\mathbb{P}_{22}}$ from the regime-switching VAR. Let

$$\begin{aligned} \theta = & \left(\mu_1, \mu_2, \Phi, A_1^{(3)}, A_2^{(3)}, B^{(3)}, \sigma_1^2, \sigma_2^2, \Omega_{(3)}, \Omega_{(4)}, \Omega_{(5)}, \Omega_{(6)}, \Omega_{(7)}, \Omega_{(8)}, \Omega_{(9)}, \pi^{\mathbb{P}_{11}}, \pi^{\mathbb{P}_{22}}, \right. \\ & \left. \lambda_{0,1}, \lambda_{0,2}, \lambda_1, \pi^{\mathbb{Q}_{11}}, \pi^{\mathbb{Q}_{22}} \right)' \end{aligned} \tag{1.22}$$

and

$$\begin{aligned} g(\theta) = & \left(\mu_1, \mu_2, \Phi, vec(A'(\theta)), vec(B(\theta)), \sigma_1^2, \sigma_2^2, \Omega_{(3)}, \Omega_{(4)}, \Omega_{(5)}, \Omega_{(6)}, \Omega_{(7)}, \Omega_{(8)}, \Omega_{(9)}, \right. \\ & \left. \pi^{\mathbb{P}_{11}}, \pi^{\mathbb{P}_{22}} \right)' \end{aligned} \tag{1.23}$$

Here

$$A(\theta) = \begin{bmatrix} A_1^{(3)} & A_2^{(3)} \\ A_1^{(4)} & A_2^{(4)} \\ A_1^{(5)} & A_2^{(5)} \\ A_1^{(6)} & A_2^{(6)} \\ A_1^{(7)} & A_2^{(7)} \\ A_1^{(8)} & A_2^{(8)} \\ A_1^{(9)} & A_2^{(9)} \end{bmatrix} \tag{1.24}$$

and $B(\theta) = \left(B^{(3)}, B^{(4)}, B^{(5)}, B^{(6)}, B^{(7)}, B^{(8)}, B^{(9)} \right)'$. $A_j^{(n)}$ in $vec\left(A'(\theta)\right)$ and $B^{(n)}$ in $vec\left(B(\theta)\right)$ for $n = 4, \ldots, 9$ are defined by the no arbitrage restrictions from equations (1.12) and (1.13):

$$A_j^{(n)} = \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}_{jk}} e^{A_k^{(n-1)}} \right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \tag{1.25}$$

for $j = 1, 2$ and

$$B^{(n)} = B^{(n-1)}(\Phi - \lambda_1) \tag{1.26}$$

For $n = 3$, $g(A_1^{(3)}) = A_1^{(3)}$, $g(A_2^{(3)}) = A_2^{(3)}$, and $g(B^{(3)}) = B^{(3)}$. Then $\hat{\theta}$ is obtained as

$$\hat{\theta} \equiv \operatorname{argmin}_\theta \{ T \left[\hat{\pi} - g(\theta)\right]' \hat{R} \left[\hat{\pi} - g(\theta)\right] \} \tag{1.27}$$

In this way I obtain estimates of the prices of risk $\lambda_{0,1}$, $\lambda_{0,2}$, and λ_1 and of the risk-neutral transition probabilities $\pi^{\mathbb{Q}_{11}}$ and $\pi^{\mathbb{Q}_{22}}$ as part of the vector $\hat{\theta}$. I also obtain second-stage estimates of μ_1, μ_2, Φ, σ_1^2, σ_2^2, $\pi^{\mathbb{P}_{11}}$, $\pi^{\mathbb{P}_{22}}$, $A_1^{(3)}$, $A_2^{(3)}$, $B^{(3)}$, and $\Omega_{(n)}$, $n = 3 \ldots 9$. By using the estimators for $\lambda_{0,1}$, $\lambda_{0,2}$, λ_1, $\pi^{\mathbb{Q}_{11}}$ and $\pi^{\mathbb{Q}_{22}}$ outlined above, I obtain the price of risk parameters and risk-neutral transition probabilities so that the values of $A_j^{(n)}$ and $B^{(n)}$ implied by equations (1.25) and (1.26) most

closely fit the unrestricted estimates from equation (1.14).

1.5 Empirical results

1.5.1 Data

I estimate a one factor model using data on prices of natural gas futures contracts traded on NYMEX with maturities 3 months to 9 months for the period from January, 1995 to June, 2014. The factor is constructed as the first principal component extracted from the demeaned log prices of these contracts. The data is obtained from Datastream. Natural gas contracts expire three business days prior to the first calendar day of the delivery month. Figure 1.1 shows the log of the observed 3-month futures price. I use a cross-section of $N = 7$ maturities in my estimation. I estimate the reduced-form parameters $\{A_j, \mu_j, \sigma_j^2\}$ for $j = 1, 2$, B, Ω, and Φ, and the probabilities $\pi^{\mathbb{P}_{11}}$, $\pi^{\mathbb{P}_{22}}$, and ρ_1 in the first step of the estimation procedure, and then estimate the prices of risk $\lambda_{0,j}$ and λ_1 and the risk-neutral probabilities $\pi^{\mathbb{Q}_{11}}$ and $\pi^{\mathbb{Q}_{22}}$ in the second step. Here ρ_1 is the probability that the initial state is regime 1. Since the longest maturity contract I use in my estimation is the 9-month contract, whereas the shortest term contract I use is the 3-month contract, I define the market as being in contango if the price of the 9-month contract is higher than the price of the 3-month contract. Conversely, I define the market as being in backwardation if the price of the 9-month contract is lower than the price of the 3-month contract.

1.5.2 Estimation results

The first principal component used as factor in my model captures 98.73% of the variation in futures prices. I use the Eigenvalue Ratio and Growth Ratio tests proposed in Ahn and Horenstein (2013) (described in Appendix 1.7.6) in order

to estimate the number of factors in my model. Both of these tests yield 1 as the number of factors that need to be used. This justifies my use of a one factor model.

Table 1.1 shows the estimates of the reduced-form parameters and the historical transition probabilities from the first stage of the estimation. Table 1.2 shows second stage estimates, including the estimates of the market prices of factor risk and the risk-neutral transition probabilities. The value of the chi-squared objective function from the second stage estimation has an asymptotic $\chi^2(q)$ distribution under the null hypothesis that the model-implied no-arbitrage restrictions are satisfied by the data. Here q is the number of reduced-form parameters in the first stage, which is equal to 35. In my estimation, the value of the objective function is 14.5270, which indicates that we fail to reject the null hypothesis. Hence, I conclude that the no-arbitrage restrictions are satisfied by the data.

I find that the factor is very persistent, with $\hat{\Phi} = 0.9785$. The factor is a stationary stochastic process under the $\mathbb{P}$-measure. The loadings of the futures prices on the first principal component are basically constant across maturities. Thus, the factor essentially represents a parallel change in prices. Because of its effect on price levels, I refer to this factor as the level factor. This is commonly done in the literature using principal component factor models.

Table 1.3 shows Wald t-statistics for the hypothesis tests testing whether there is regime switching in the various parameters. I find very strong evidence that there are regimes in the data. The null hypotheses that the level of the factor μ and the levels $A^{(3)}$, $A^{(4)}$, $A^{(5)}$, $A^{(7)}$, $A^{(8)}$, $A^{(9)}$ of the contracts are not regime switching are strongly rejected. The level μ of the factor is higher in regime 1 than in regime 2, and $\mu_1 - \mu_2$ is statistically significantly different from 0. The levels of the contracts with maturity 3-5 months are statistically significantly lower in

regime 1, while the levels of the contracts with maturity 7-9 months are statistically significantly higher in regime 1. The estimated variances in the two regimes are not statistically significantly different. The coefficients $\hat{B}$ and the variance of the measurement errors $\hat{\Omega}$ are regime independent by assumption. Figure 1.2 shows the spread between the 9-month contract and the 3-month contract plotted against the smoothed probability of regime 2. It can be seen that in regime 1 the spread is higher and is almost always positive, whereas in regime 2 the spread is lower and is almost always negative. Thus, I find that the estimated regimes correspond to the previously defined market states of contango and backwardation. Regime 1 (the positive spread regime) represents contango, while regime 2 (the negative spread regime) represents backwardation. Figures 1.3 and 1.4 show the log of the observed 3-month futures price and the factor, respectively, with shaded areas representing the backwardation regime.

Figure 1.5 shows the expected monthly holding returns for each contract conditional on the contango regime and the backwardation regime, respectively, averaged over each regime. Here I define the expected monthly holding return on the n-month contract conditional on being in regime j at time t as $\frac{E_t^{\mathbb{P}}[F_{t+1}^{(n-1)}-F_t^{(n)}|s_t=j]}{F_t^{(n)j}}$.[9] The fact that the expected holding returns for all contracts (except the 6-month contract) change sign depending on the regime and are large in magnitude in each regime shows that agents face very substantial risk due to the possibility of changes in regime. For instance, in the backwardation regime, an investor holding a long position in the 9-month contract would on average earn a positive expected monthly holding return of 2.5%. On the other hand, in the contango regime, the investor

---

[9]The formula for $\frac{E_t^{\mathbb{P}}[F_{t+1}^{(n-1)}-F_t^{(n)}|s_t=j]}{F_t^{(n)j}}$ is derived in equation (1.45) in Appendix 1.7.2. In Figure 1.5, I am not reporting the expected monthly holding return on the 3-month contract conditional on the regime, since calculating that requires using values for $A_j^{(2)}$ and $B^{(2)}$, which I am not obtaining from my model.

would earn a negative expected monthly holding return of about -0.5% on average. Thus, as long as the market is in backwardation, the investor would on average earn a profit, but if the regime switches, he would experience a significant loss.

From Figure 1.5, we can also see that in the contango regime, buying the 4-6 month contracts today and selling them next month on average yields a profit. So does shorting the 7-9 month contracts today and closing out the position next month. In the backwardation regime, shorting the 4-5 contracts today and closing out the position next month on average yields a profit. So does going long on the 6-9 month contracts today and selling them next month. Thus, on average it is profitable to be long in the 4-6 month contracts in the contango regime, whereas it is profitable to be long in the 6-9 month contracts in the backwardation regime.

The fact that there is a positive return to a long position in the 4-6 month contracts in the contango regime suggests that there is demand for short positions in futures. This could mean that in the contango regime, commercial producers are trying to hedge their short positions in natural gas by selling 4-6 month futures contracts. Moreover, in the contango regime there is a positive return to a short position in the 7-9 month contracts, which could indicate demand for long positions in these contracts. In turn, this could indicate that commercial users are trying to hedge their long positions in natural gas by buying 7-9 month contracts. Similarly, in the backwardation regime there is a positive return to a long position in the 6-9 month contracts, which could be a result of commercial producers trying to hedge their short positions by selling 6-9 month contracts. Moreover, in the backwardation regime there is a positive return to a short position in the 4-5 month contracts, which could be a result of commercial users trying to hedge their long positions by buying 4-5 month contracts.

By allowing for $\pi^{\mathbb{P}} \neq \pi^{\mathbb{Q}}$ in my model, I have another channel through which

risk preferences can affect expected returns. I test the restrictions $\pi^{\mathbb{P}11} = \pi^{\mathbb{Q}11}$ and $\pi^{\mathbb{P}22} = \pi^{\mathbb{Q}22}$, and find that they are rejected. Thus, my results suggest that $\pi^{\mathbb{P}} \neq \pi^{\mathbb{Q}}$. I find that $\pi_{11}^{\mathbb{Q}} > \pi_{11}^{\mathbb{P}}$ and $\pi_{22}^{\mathbb{Q}} > \pi_{22}^{\mathbb{P}}$. This suggests that investors act as though regimes are more persistent than they really are. This is consistent with what we see in reality, as market participants often just act as if the current state of the market will continue next period. This overestimation of the persistence of the regimes by investors could be a result of hedging pressure, which I discuss later in this section.

The expected log holding return $E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t = j\right]$ is related to the risk premium investors demand in regime j for holding a futures contract maturing in n months for 1 month. The expression for the expected log holding return of the n-month contract conditional on the regime $E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t = j\right]$, derived in equation (1.41) in Appendix 1.7.2, is

$$E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t = j\right] = \sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right) + B^{(n-1)}\lambda_{0,j} + $$
$$B^{(n-1)}\lambda_1 X_t - \frac{1}{2}B^{(n-1)^2}\sigma_j^2 \tag{1.28}$$

From the formula, we can see that nonzero expected returns arise in part due to the difference between the historic probabilities $\pi^{\mathbb{P}jk}$ and the risk-neutral probabilities $\pi^{\mathbb{Q}jk}$. The difference between $\pi^{\mathbb{P}jk}$ and $\pi^{\mathbb{Q}jk}$ accounts for a large proportion of expected returns.

It can be shown that the market price of regime shift risk $\Gamma_{j,k} = \log\left(\frac{\pi^{\mathbb{P}jk}}{\pi^{\mathbb{Q}jk}}\right)$ (see equation (1.31) in Appendix 1.7.1 for derivation). Since $\pi^{\mathbb{P}jk}$ and $\pi^{\mathbb{Q}jk}$ are statistically significantly different, the market price of regime shift risk is nonzero, i.e. regime shift risk is priced. To motivate the labeling of $\Gamma_{s_t,s_{t+1}}$ as the market price of regime shift risk, consider a security which pays \$1 if the regime switches

from regime $s_t = j$ in month t to regime $s_{k+1} = k$ in month $t + 1$. This security has payoff $\mathbb{1}_{\{s_{t+1}=k\}}$ and has exposure only to the risk of shifting from regime j in month t to regime k in month $t + 1$. Conditional on the current regime $s_t = j$, its current price is

$$P_t^j = E_t^{\mathbb{Q}}[\mathbb{1}_{\{s_{t+1}=k\}}|s_t = j] = \pi^{\mathbb{Q}_{jk}}$$

Therefore, its log expected return is

$$\log \frac{E_t^{\mathbb{P}}[\mathbb{1}_{\{s_{t+1}=k\}}|s_t = j]}{P_t^j} = \log\left(\frac{\pi^{\mathbb{P}_{jk}}}{\pi^{\mathbb{Q}_{jk}}}\right) = \Gamma_{j,k} \qquad (1.29)$$

Thus, $\Gamma_{j,k}$ gives the log expected return per unit of regime shift risk exposure, and can therefore be interpreted as the market price of regime shift risk associated with switching from regime j to regime k.

For instance, a security which pays off \$1 if the regime switches from regime 2 (backwardation) today to regime 1 (contango) next month is priced at

$$P_t^{(2)} = E_t^{\mathbb{Q}}[\mathbb{1}_{\{s_{t+1}=1\}}|s_t = 2] = \pi^{\mathbb{Q}_{21}} = 1 - \pi^{\mathbb{Q}_{22}} = \$0.0187$$

Thus, investors are willing to pay only about 2 cents to hedge against the regime switching from backwardation to contango next month. The expected payoff of the security is

$$E_t^{\mathbb{P}}[\mathbb{1}_{\{s_{t+1}=1\}}|s_t = 2] = \pi^{\mathbb{P}_{21}} = 1 - \pi^{\mathbb{P}_{22}} = \$0.1944$$

So the security pays off about 19 cents on average. The price investors are willing to pay to hedge against regime shift risk is very low, reflecting the fact that they think the backwardation regime is considerably more persistent than it actually is.

Similarly, a security which pays off \$1 if the regime changes from regime 1

(contango) today to regime 2 (backwardation) next month is priced at

$$P_t^{(1)} = E_t^{\mathbb{Q}}[\mathbb{1}_{\{s_{t+1}=2\}}|s_t = 1] = \pi^{\mathbb{Q}_{12}} = 1 - \pi^{\mathbb{Q}_{11}} = \$0.0224$$

Thus, investors are willing to pay about 2 cents to hedge against the risk of the regime switching from contango today to backwardation next month. On average, the security pays off

$$E_t^{\mathbb{P}}[\mathbb{1}_{\{s_{t+1}=2\}}|s_t = 1] = \pi^{\mathbb{P}_{12}} = 1 - \pi^{\mathbb{P}_{11}} = \$0.0930$$

i.e. about 9 cents. Once again, the price investors are willing to pay to hedge against the risk of the regime changing is low, but it is closer to the actual expected payoff of the security. To summarize, I find that agents act as if both regimes are more persistent than they are, with the perceived overestimation of the regime persistence being even higher for the backwardation regime.

Why is the security which pays off $1 if the regime switches from regime 2 (backwardation) today to regime 1 (contango) next month so cheap? The potential seller of this security could be using it as a hedge against risk, if the state of the world when the seller would have to pay $1 is one in which the seller will profit from other sources, and the state of the world when the seller keeps the dollar is one when even a little more money would be helpful. For example, a commercial user who buys natural gas with 3-month contracts would benefit from selling this security. He would have to pay on the contract if the regime shifts from backwardation to contango. But if the regime shifts, he would profit by from then on being able to buy futures at a lower price. Hedging pressure could cause the price to fall below the 19-cent valuation.

The price of risk λ_1 is statistically significant and negative, and the loadings $B^{(n)}$ are positive. From equation (1.28), we can see that this implies that an increase in the level of futures prices (as measured by the factor X_t) decreases the expected log holding returns on the 4-9 month contracts in each regime. The expected log holding return loading for the n-month contract is $B^{(n-1)}\lambda_1$. According to my estimates, a positive one standard deviation shock to the level factor reduces the expected log holding return on the 4-9 month contracts in each regime by about 0.88%.

The estimates of the prices of risk $\lambda_{0,1}$ and $\lambda_{0,2}$ are not statistically significant. Moreover, using a Wald test I find that $\lambda_{0,1} - \lambda_{0,2}$ is not statistically significantly different from 0. Thus, I do not find considerable differences in the market pricing of factor risk in the two regimes.

1.6 Conclusion

In this paper I have provided a new way of characterizing risk in commodities futures markets, which tend to switch between periods of contango when the spread between the longer term futures and the shorter term futures is positive, and backwardation when the spread between the longer term futures and the shorter term futures is negative. I apply my framework to the natural gas futures market, where the risk agents face due to the possibility of switching between the two states of the market is particularly substantial. Motivated by the historically observed switches of natural gas futures prices between states of contango and backwardation, I propose and estimate a Markov regime-switching Gaussian affine term structure model with two regimes and estimate the states of the market from the data. In my model, the level and volatility of natural gas futures, as well as the risk premium, are regime dependent. I study the consequences of changes in regime on the risk

premium, and produce novel empirical estimates to characterize risk premia and the term structure of natural gas futures contracts. I find very strong evidence that there are regimes in the data. Moreover, I find that the regimes in my model correspond precisely to historically observed periods of contango and backwardation. I also find that regime switching risk is priced. I find that expected futures returns for most contracts change sign depending on which regime we are conditioning on. This is consistent with the claim that agents face significant risks from the possibility of a change in the regime. My results show that the market acts as if regimes are more persistent than they really are, which could be a result of hedging pressure. I find evidence that commercial users and commercial producers use natural gas futures contracts for purposes of hedging. In the contango regime, commercial producers may be trying to hedge their short positions in natural gas by selling 4-6 month contracts, while in the backwardation regime, commercial producers may be trying to hedge their short positions by selling 6-9 month contracts. I obtain analogous implications for commercial users. A separate contribution of the paper is to propose a new method for estimating Gaussian affine term structure models subject to regime-switching. My approach allows for computationally fast estimation and avoids the numerical difficulties that are common when using other maximum likelihood based methods in the literature.

1.7 Appendix

1.7.1 Relation between $\mathbb{P}$-dynamics and $\mathbb{Q}$-dynamics

By no arbitrage, an asset with payoff $g(X_{t+1})$ has a price in regime j equal to

$$P(X_t) = E_t^{\mathbb{P}}[M_{t,t+1}g(X_{t+1})|s_t = j] = E_t^{\mathbb{Q}}[g(X_{t+1})|s_t = j] \qquad (1.30)$$

$$
\begin{aligned}
\pi^{\mathbb{Q}_{jk}} =& E_t^{\mathbb{Q}}\left[\mathbb{1}_{\{s_{t+1}=k\}}|s_t = j\right] = E_t^{\mathbb{P}}\left[\mathbb{1}_{\{s_{t+1}=k\}}M_{t,t+1}|s_t = j\right] \\
=& E_t^{\mathbb{P}}\left[\mathbb{1}_{\{s_{t+1}=k\}}\exp\left(-\Gamma_{s_t,s_{t+1}} - \frac{1}{2}\lambda_{t,s_t}^2 - \lambda_{t,s_t}\sigma_{s_t}^{-1}v_{t+1}\right)|s_t = j\right] \\
=& E_t^{\mathbb{P}}\left[\exp\left(-\frac{1}{2}\lambda_{t,s_t}^2 - \lambda_{t,s_t}\sigma_{s_t}^{-1}v_{t+1}\right)|s_t = j\right] \times \\
& E_t^{\mathbb{P}}\left[\mathbb{1}_{\{s_{t+1}=k\}}\exp\left(-\Gamma_{s_t,s_{t+1}}\right)|s_t = j\right] \\
=& \exp\left(-\frac{1}{2}\lambda_{t,j}^2\right)E_t^{\mathbb{P}}\left[\exp\left(-\lambda_{t,s_t}\sigma_{s_t}^{-1}v_{t+1}\right)|s_t = j\right]\pi^{\mathbb{P}_{jk}}\exp(-\Gamma_{j,k}) \\
=& \exp\left(-\frac{1}{2}\lambda_{t,j}^2\right)\exp\left(\frac{1}{2}Var_t(-\lambda_{t,s_t}\sigma_{s_t}^{-1}v_{t+1}|s_t = j)\right)\pi^{\mathbb{P}_{jk}}\exp\left(-\Gamma_{j,k}\right) \\
=& \exp\left(-\frac{1}{2}\lambda_{t,j}^2\right)\exp\left(\frac{1}{2}\lambda_{t,j}^2\sigma_j^{-2}\sigma_j^2\right)\pi^{\mathbb{P}_{jk}}\exp\left(-\Gamma_{j,k}\right) = \pi^{\mathbb{P}_{jk}}\exp\left(-\Gamma_{j,k}\right)
\end{aligned}
$$

Therefore,

$$\Gamma_{j,k} = \log\left(\frac{\pi^{\mathbb{P}_{jk}}}{\pi^{\mathbb{Q}_{jk}}}\right) \qquad (1.31)$$

By no arbitrage, an asset with payoff $g(X_{t+1})$ has a price in regime j equal

to

$$P(X_t) = E_t \left[M_{t,t+1} g(X_{t+1}) | s_t = j \right]$$

$$= E_t \left[\exp \left(-\Gamma_{s_t,s_{t+1}} - \frac{1}{2} \lambda_{t,s_t}^2 - \lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left(-\frac{1}{2} \lambda_{t,j}^2 \right) E_t^{\mathbb{P}} \left[\exp \left(-\Gamma_{s_t,s_{t+1}} - \lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left(-\frac{1}{2} \lambda_{t,j}^2 \right) \left[\sum_{k=1}^{2} \pi^{\mathbb{P}_{jk}} \exp(-\Gamma_{j,k}) \right] \times$$

$$E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left(-\frac{1}{2} \lambda_{t,j}^2 \right) \left[\sum_{k=1}^{2} \pi^{\mathbb{P}_{jk}} \frac{\pi^{\mathbb{Q}_{jk}}}{\pi^{\mathbb{P}_{jk}}} \right] E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left(-\frac{1}{2} \lambda_{t,j}^2 \right) \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}_{jk}} \right] E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left(-\frac{1}{2} \lambda_{t,j}^2 \right) E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left(-\frac{1}{2} \lambda_{t,j}^2 \right) E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t} \sigma_{s_t}^{-1} \left(X_{t+1} - \mu_{s_t} - \Phi X_t \right) \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left(-\frac{1}{2} \lambda_{t,j}^2 \right) \int g(X_{t+1}) \exp \left(-\lambda_{t,j} \sigma_j^{-1} (X_{t+1} - \mu_j - \Phi X_t) \right) (2\pi)^{-1/2} \sigma_j^{-1}$$

$$\exp \left(-\frac{1}{2\sigma_j^2} (X_{t+1} - \mu_j - \Phi X_t)^2 \right) dX_{t+1}$$

$$= (2\pi)^{-1/2} \sigma_j^{-1} \int g(X_{t+1}) \exp \left(-\frac{1}{2} \left[\frac{1}{\sigma_j^2} (X_{t+1} - \mu_j - \Phi X_t)^2 + \right. \right.$$

$$\left. \left. + 2\lambda_{t,j} \sigma_j^{-1} (X_{t+1} - \mu_j - \Phi X_t) + \lambda_{t,j}^2 \right] \right) dX_{t+1}$$

$$= (2\pi)^{-1/2}\sigma_j^{-1}\int g(X_{t+1})\exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_j}(X_{t+1}-\mu_j-\Phi X_t)+\lambda_{t,j}\right]^2\right)dX_{t+1}$$

$$= (2\pi)^{-1/2}\sigma_j^{-1}\int g(X_{t+1})\exp\left(-\frac{1}{2}\left[\frac{X_{t+1}-\mu_j-\Phi X_t+\sigma_j\lambda_{t,j}}{\sigma_j}\right]^2\right)dX_{t+1}$$

$$= (2\pi)^{-1/2}\sigma_j^{-1}\int g(X_{t+1})\exp\left(-\frac{1}{2\sigma_j^2}[X_{t+1}-\mu_j-\Phi X_t+\sigma_j\lambda_{t,j}]^2\right)dX_{t+1}$$

$$= (2\pi)^{-1/2}\sigma_j^{-1}\int g(X_{t+1})\exp\left(-\frac{1}{2\sigma_j^2}[X_{t+1}-\mu_j-\Phi X_t+\lambda_{0,j}+\lambda_1 X_t]^2\right)dX_{t+1}$$

$$= (2\pi)^{-1/2}\sigma_j^{-1}\int g(X_{t+1})\exp\left(-\frac{1}{2\sigma_j^2}[X_{t+1}-(\mu_j-\lambda_{0,j})-(\Phi-\lambda_1)X_t]^2\right)dX_{t+1}$$

$$= E_t^{\mathbb{Q}}(g(X_{t+1})|s_t=j)$$

where in the last line I used equation (1.30).

Therefore, under the $\mathbb{Q}$-measure,

$$X_{t+1}|s_t=j \sim^{\mathbb{Q}} N((\mu_j-\lambda_{0,j})+(\Phi-\lambda_1)X_t,\sigma_j^2) \tag{1.32}$$

or, equivalently,

$$X_{t+1}|s_t=j \sim^{\mathbb{Q}} N(\mu_j^{\mathbb{Q}}+\Phi^{\mathbb{Q}}X_t,\sigma_j^2) \tag{1.33}$$

where

$$\mu_j^{\mathbb{Q}} \equiv \mu_j-\lambda_{0,j} \tag{1.34}$$

and

$$\Phi^{\mathbb{Q}} \equiv \Phi-\lambda_1 \tag{1.35}$$

Hence, under the $\mathbb{Q}$-measure, X_{t+1} follows the dynamics

$$X_{t+1}=\mu_j^{\mathbb{Q}}+\Phi^{\mathbb{Q}}X_t+v_{t+1}^{\mathbb{Q}} \tag{1.36}$$

where $v_{t+1}^{\mathbb{Q}}|s_t = j \sim^{\mathbb{Q}} N(0, \sigma_j^2)$ under the $\mathbb{Q}$-measure.

1.7.2 Calculating expected returns

$$
\begin{aligned}
E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t = j\right] &= E_t^{\mathbb{P}}\left[f_{t+1}^{(n-1)} - f_t^{(n)}|s_t = j\right] \\
&= E_t^{\mathbb{P}}\left[A_{s_{t+1}}^{(n-1)} + B^{(n-1)}X_{t+1} - A_{s_t}^{(n)} - B^{(n)}X_t|s_t = j\right] \\
&= E_t^{\mathbb{P}}\left[A_{s_{t+1}}^{(n-1)} + B^{(n-1)}(\mu_{s_t} + \Phi X_t + v_{t+1}) - A_{s_t} - \right. \\
&\left. B^{(n)}X_t|s_t = j\right] = \pi^{\mathbb{P}j1}A_1^{(n-1)} + \pi^{\mathbb{P}j2}A_2^{(n-1)} + B^{(n-1)}\mu_j + \\
&(B^{(n-1)}\Phi - B^{(n)})X_t - A_j^{(n)}
\end{aligned}
\tag{1.37}
$$

The futures price is

$$
F_t^{(n)j} = e^{A_j^{(n)} + B^{(n)}X_t}
$$

$$
\begin{aligned}
f_t^{(n)j} &= \log F_t^{(n)j} = \log E_t^{\mathbb{Q}}\left[F_{t+1}^{(n-1)}|s_t = j\right] = \\
&= \log\left(\sum_{k=1}^2 \pi^{\mathbb{Q}jk} E_t^{\mathbb{Q}}\left[F_{t+1}^{(n-1)k}|s_t = j\right]\right) \\
&= \log\left(\left[\sum_{k=1}^2 \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right] E_t^{\mathbb{Q}}\left[e^{B^{(n-1)}X_{t+1}}|s_t = j\right]\right) \\
&= \log\left(\sum_{k=1}^2 \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right) + \log E_t^{\mathbb{Q}}\left[e^{B^{(n-1)}X_{t+1}}|s_t = j\right] \\
&= \log\left(\sum_{k=1}^2 \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right) + \log E_t^{\mathbb{Q}}\left[e^{B^{(n-1)}(\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}}X_t + v_{t+1})}|s_t = j\right] \\
&= \log\left(\sum_{k=1}^2 \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right) + \log\left[e^{B^{(n-1)}(\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}}X_t) + \frac{1}{2}B^{(n-1)2}\sigma_j^2}\right] \\
&= \log\left(\sum_{k=1}^2 \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right) + B^{(n-1)}(\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}}X_t) + \frac{1}{2}B^{(n-1)^2}\sigma_j^2
\end{aligned}
$$

Therefore,

$$f_t^{(n)j} = A_j^{(n)} + B^{(n)}X_t = \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk}e^{A_k^{(n-1)}}\right) + B^{(n-1)}\mu_j^{\mathbb{Q}} + \frac{1}{2}B^{(n-1)^2}\sigma_j^2 +$$

$$B^{(n-1)}\Phi^{\mathbb{Q}}X_t$$

The above equation implies the following recursions:

$$A_j^{(n)} = \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk}e^{A_k^{(n-1)}}\right) + B^{(n-1)}\mu_j^{\mathbb{Q}} + \frac{1}{2}B^{(n-1)^2}\sigma_j^2$$

or equivalently

$$A_j^{(n)} = \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk}e^{A_k^{(n-1)}}\right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2}B^{(n-1)^2}\sigma_j^2 \qquad (1.38)$$

and

$$B^{(n)} = B^{(n-1)}\Phi^{\mathbb{Q}} \qquad (1.39)$$

or equivalently

$$B^{(n)} = B^{(n-1)}(\Phi - \lambda_1) \qquad (1.40)$$

$$E_t^{\mathbb{P}}\left[f_{t+1}^{(n-1)}|s_t=j\right]=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}E_t^{\mathbb{P}}\left[f_{t+1}^{(n-1)k}|s_t=j\right]$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}E_t^{\mathbb{P}}\left[A_k^{(n-1)}+B^{(n-1)}X_{t+1}|s_t=j\right]$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}\left(A_k^{(n-1)}+B^{(n-1)}E_t^{\mathbb{P}}\left[X_{t+1}|s_t=j\right]\right)$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}\left(A_k^{(n-1)}+B^{(n-1)}E_t^{\mathbb{P}}\left[\mu_{s_t}+\Phi X_t|s_t=j\right]\right)$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}\left(A_k^{(n-1)}+B^{(n-1)}(\mu_j+\Phi X_t)\right)$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}A_k^{(n-1)}+\left(\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}\right)B^{(n-1)}(\mu_j+\Phi X_t)$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}A_k^{(n-1)}+B^{(n-1)}(\mu_j+\Phi X_t)$$

$$E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t=j\right]=E_t^{\mathbb{P}}\left[f_{t+1}^{(n-1)}-f_t^{(n)}|s_t=j\right]=E_t^{\mathbb{P}}\left[f_{t+1}^{(n-1)}|s_t=j\right]-f_t^{(n)j}$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}A_k^{(n-1)}+B^{(n-1)}(\mu_j+\Phi X_t)-$$

$$\log\left(\sum_{k=1}^{2}\pi^{\mathbb{Q}_{jk}}e^{A_k^{(n-1)}}\right)-B^{(n-1)}(\mu_j^{\mathbb{Q}}+\Phi^{\mathbb{Q}}X_t)-\frac{1}{2}B^{(n-1)^2}\sigma_j^2$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}A_k^{(n-1)}+B^{(n-1)}(\mu_j-\mu_j^{\mathbb{Q}})+B^{(n-1)}(\Phi-\Phi^{\mathbb{Q}})X_t-$$

$$\log\left(\sum_{k=1}^{2}\pi^{\mathbb{Q}_{jk}}e^{A_k^{(n-1)}}\right)-\frac{1}{2}B^{(n-1)^2}\sigma_j^2$$

$$=\sum_{k=1}^{2}\pi^{\mathbb{P}_{jk}}A_k^{(n-1)}-\log\left(\sum_{k=1}^{2}\pi^{\mathbb{Q}_{jk}}e^{A_k^{(n-1)}}\right)+B^{(n-1)}\lambda_{0,j}+$$

$$B^{(n-1)}\lambda_1 X_t-\frac{1}{2}B^{(n-1)^2}\sigma_j^2 \tag{1.41}$$

$$E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|\mathcal{F}_t\right] = \sum_{j=1}^{2} E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t = j\right] P(s_t = j|\mathcal{F}_t)$$

$$= \sum_{j=1}^{2}\left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right) + B^{(n-1)}\lambda_{0,j} + \right.$$

$$\left. B^{(n-1)}\lambda_1 X_t - \frac{1}{2}B^{(n-1)2}\sigma_j^2\right] \times P(s_t = j|\mathbb{F}_t)$$

We can also derive an expression for $\frac{E_t^{\mathbb{P}}[F_{t+1}^{(n-1)}|s_t=j]}{F_t^{(n)j}}$.

$$F_t^{(n)j} = E_t^{\mathbb{Q}}[F_{t+1}^{(n-1)}|s_t = j] = \sum_{k=1}^{2} \pi^{\mathbb{Q}jk} E_t^{\mathbb{Q}}[F_{t+1}^{(n-1)k}|s_t = j]$$

$$= \sum_{k=1}^{2} \pi^{\mathbb{Q}jk} E_t^{\mathbb{Q}}[e^{A_k^{(n-1)}+B^{(n-1)}X_{t+1}}|s_t = j]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right] E_t^{\mathbb{Q}}\left[e^{B^{(n-1)}X_{t+1}}|s_t = j\right]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right] E_t^{\mathbb{Q}}\left[e^{B^{(n-1)}(\mu_{s_t}^{\mathbb{Q}}+\Phi^{\mathbb{Q}}X_t+v_{t+1})}|s_t = j\right]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right] e^{B^{(n-1)}(\mu_j^{\mathbb{Q}}+\Phi^{\mathbb{Q}}X_t)+\frac{1}{2}B^{(n-1)2}\sigma_j^2} \qquad (1.42)$$

$$E_t^{\mathbb{P}}[F_{t+1}^{(n-1)}|s_t = j] = \sum_{k=1}^{2} \pi^{\mathbb{P}jk} E_t^{\mathbb{P}}[F_{t+1}^{(n-1)k}|s_t = j]$$

$$= \sum_{k=1}^{2} \pi^{\mathbb{P}jk} E_t^{\mathbb{P}}[e^{A_k^{(n-1)}+B^{(n-1)}X_{t+1}}|s_t = j]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{A_k^{(n-1)}}\right] E_t^{\mathbb{P}}\left[e^{B^{(n-1)}X_{t+1}}|s_t = j\right]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{A_k^{(n-1)}}\right] E_t^{\mathbb{P}}\left[e^{B^{(n-1)}(\mu_{s_t}^{\mathbb{P}}+\Phi^{\mathbb{P}}X_t+v_{t+1})}|s_t = j\right]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{A_k^{(n-1)}}\right] e^{B^{(n-1)}(\mu_j^{\mathbb{P}}+\Phi^{\mathbb{P}}X_t)+\frac{1}{2}B^{(n-1)^2}\sigma_j^2} \qquad (1.43)$$

Then

$$\frac{E_t^{\mathbb{P}}[F_{t+1}^{(n-1)}|s_t = j]}{F_t^{(n)j}} = \frac{\left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{A_k^{(n-1)}}\right] e^{B^{(n-1)}(\mu_j^{\mathbb{P}}+\Phi^{\mathbb{P}}X_t)+\frac{1}{2}B^{(n-1)^2}\sigma_j^2}}{\left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right] e^{B^{(n-1)}(\mu_j^{\mathbb{Q}}+\Phi^{\mathbb{Q}}X_t)+\frac{1}{2}B^{(n-1)^2}\sigma_j^2}}$$

$$= \frac{\left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{A_k^{(n-1)}}\right]}{\left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right]} e^{B^{(n-1)}(\mu_j^{\mathbb{P}}-\mu_j^{\mathbb{Q}})+B^{(n-1)}(\Phi^{\mathbb{P}}-\Phi^{\mathbb{Q}})X_t}$$

$$= \frac{\left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{A_k^{(n-1)}}\right]}{\left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right]} e^{B^{(n-1)}(\lambda_{0,j}+\lambda_1 X_t)}$$

$$= \frac{\left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{A_k^{(n-1)}}\right]}{\left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{A_k^{(n-1)}}\right]} e^{B^{(n-1)}\sigma_j \lambda_{t,j}} \qquad (1.44)$$

Therefore,

$$\frac{E_t^{\mathbb{P}}[F_{t+1}^{(n-1)} - F_t^{(n)}|s_t = j]}{F_t^{(n)j}} = \frac{E_t^{\mathbb{P}}[F_{t+1}^{(n-1)}|s_t = j] - F_t^{(n)j}}{F_t^{(n)j}}$$

$$= \frac{E_t^{\mathbb{P}}[F_{t+1}^{(n-1)}|s_t = j]}{F_t^{(n)j}} - 1$$

$$= \frac{\left[\sum_{k=1}^{2} \pi^{\mathbb{P}}_{jk} e^{A_k^{(n-1)}}\right]}{\left[\sum_{k=1}^{2} \pi^{\mathbb{Q}}_{jk} e^{A_k^{(n-1)}}\right]} e^{B^{(n-1)}\sigma_j \lambda_{t,j}} - 1 \qquad (1.45)$$

1.7.3 EM algorithm for first stage estimation

In the first stage I estimate the system of equations (1.14) and (1.17). It is known that in the absence of regime-switching, maximum likelihood estimation of this system is equivalent to OLS estimation equation by equation. Conditional on parameters, the inference about the regime $\Pr(s_t = j|\mathcal{F}_T)$[10] can be obtained using the Hamilton filtering and smoothing algorithm. This suggests estimation via the EM algorithm. Let θ denote the vector of parameters to be estimated, $\theta \equiv \{vec(A'), vec(B), \Phi, \Omega, \{\mu_j, \sigma_j\}_{j=1}^2\}$ where $vec(A')$ and $vec(B)$ are as defined in Section 1.4.2. First, I initialize the algorithm with an initial guess for the vector of parameters, and compute the corresponding smoothed probabilities. Then each iteration l of the algorithm proceeds as follows. First, I update inference for the regression parameters equation by equation. An updated estimate $\hat{\theta}^{(l)}$ is derived as a solution to the first-order conditions for maximization of the likelihood function, where the conditional regime probabilities $\Pr(s_t|Y, \theta)$ are replaced with the smoothed probabilities $\Pr(s_t = j|Y, \theta^{(l-1)})$ computed in the previous iteration, for $Y = \{Y_{t1}, Y_{t2}, X_t\}_{t=1}^T$ defined below. Conditional on knowing the smoothed probabilities, a closed form solution for the regression parameters of each regime-

---

[10] $\mathcal{F}_T$ represents information available up to time T

switching equation can be obtained by linear regression in which the observations are weighted by the smoothed probability that they came from the corresponding regime. Details are shown below. Next, I update inference about the smoothed probabilities $\Pr(s_t = j|Y, \theta^{(l)})$, where I am conditioning on the parameter vector estimate obtained in the current iteration instead of the unknown parameter vector θ.

The regime-switching system I estimate is of the form

$$\underset{(1\times1)}{Y_{t1}} = \underset{(1\times1)}{\mu_{s_t}} + \underset{(1\times1)}{\Phi} \underset{(1\times1)}{X_t} + \underset{(1\times1)}{\varepsilon_t} \qquad \varepsilon_t|s_t \sim N(0, \sigma^2_{s_t}) \tag{1.46}$$

$$\underset{(7\times1)}{Y_{t2}} = \underset{(7\times1)}{A_{s_t}} + \underset{(7\times1)}{B} \underset{(1\times1)}{X_t} + \underset{(7\times1)}{u_t} \qquad u_t|s_t \sim N(0, \Omega) \tag{1.47}$$

Equation (1.14) when stacked across all maturities $n = 3, \ldots, 9$ is of the form of the above equation (1.47) with $Y_{t2} = (f_t^{(3)}, f_t^{(4)}, f_t^{(5)}, f_t^{(6)}, f_t^{(7)}, f_t^{(8)}, f_t^{(9)})'$, while equation (1.17) is of the form of equation (1.46) with $Y_{t1} = X_{t+1}$.

I estimate the vector system using a partially restricted algorithm equation by equation. The algorithm for a single equation is described in Appendix 1.7.4. Suppose at the previous iteration of the algorithm I have estimates $\theta^{(\ell)}$ and $\Pr(s_t = j|\theta^{(\ell)}, Y)$ for $Y = \{Y_{t1}, Y_{t2}, X_t\}_{t=1}^T$. Iteration $(\ell + 1)$ of the algorithm works as follows.

Step 1. Update inference for Y_{t2} regression parameters.

1a) Taking each $n = 1, \ldots, 7$ one at a time starting with $n = 1$, construct

$$\omega_n^{2(\ell)} = \text{ row } n, \text{ col. } n \text{ element of } \Omega^{(\ell)}$$

$$\lambda_{nt1}^{(\ell)} = \frac{\sqrt{Pr(s_t = 1|Y, \theta^{(\ell)})}}{\omega_n^{(l)}}$$

$$\lambda_{nt2}^{(\ell)} = \frac{\sqrt{Pr(s_t = 2|Y, \theta^{(\ell)})}}{\omega_n^{(l)}}$$

$$Y_{t2}^{(n)} = n^{\text{th}} \text{ element of } Y_{t2}$$

$$A_{in}^{(\ell)} = n^{\text{th}} \text{ element of } A_i^{(\ell)}$$

$$B_n^{(\ell)} = n^{\text{th}} \text{ element of } B^{(\ell)}$$

For $t = 1, \ldots, T$, define

$$\tilde{y}_{nt}^{(\ell)} = \lambda_{nt1}^{(\ell)} Y_{t2}^{(n)}$$

$$\tilde{x}_{nt}^{(\ell)} = \lambda_{nt1}^{(\ell)} X_t$$

$$\tilde{z}_{n1t}^{(\ell)} = \lambda_{nt1}^{(\ell)}$$

$$\tilde{z}_{n2t}^{(\ell)} = 0$$

and

$$\tilde{y}_{n,T+t}^{(\ell)} = \lambda_{nt2}^{(\ell)} Y_{t2}^{(n)}$$

$$\tilde{x}_{n,T+t}^{(\ell)} = \lambda_{nt2}^{(\ell)} X_t$$

$$\tilde{z}_{n1,T+t}^{(\ell)} = 0$$

$$\tilde{z}_{n2,T+t}^{(\ell)} = \lambda_{nt2}^{(\ell)}$$

Conditional on knowing $\lambda_{nt1}^{(\ell)}$ and $\lambda_{nt2}^{(\ell)}$, a closed-form solution for $(\hat{A}_1^{(n)}, \hat{A}_2^{(n)}, \hat{B}^{(n)})'$ can be found by performing an OLS regression on an artificial sample of size $2T$,

$$\tilde{y}_{nt}^{(\ell)} = A_1^{(n)} \tilde{z}_{n1t}^{(\ell)} + A_2^{(n)} \tilde{z}_{n2t}^{(\ell)} + B^{(n)} \tilde{x}_{nt}^{(\ell)} + \tilde{u}_{nt}, \ t=1, \ 2, \ \ldots, \ 2T$$

Construct

$$\hat{u}_{nt1}^{(\ell+1)} = Y_{t2}^{(n)} - \hat{A}_{1n}^{(\ell+1)} - \hat{B}_n^{(\ell+1)} X_t$$

$$\hat{u}_{nt2}^{(\ell+1)} = Y_{t2}^{(n)} - \hat{A}_{2n}^{(\ell+1)} - \hat{B}_n^{(\ell+1)} X_t$$

1b) For each $n = 1, ..., 7$ calculate

$$\omega_n^{(\ell+1)} = \left\{ \frac{1}{T} \left(\sum_{t=1}^T \hat{u}_{nt1}^{2(\ell+1)} Pr(s_t = 1 | Y, \theta^{(\ell)}) + \sum_{t=1}^T \hat{u}_{nt2}^{2(\ell+1)} Pr(s_t = 2 | Y, \theta^{(\ell)}) \right) \right\}^{1/2}$$

Step 2. Update the inference for the Y_{t1} parameters. This involves the analogous steps to those above using the partially restricted algorithm for a single equation as described in Appendix 1.7.4. The factor variance is updated as

$$\sigma_1^{2(\ell+1)} = \frac{\sum_{t=1}^T Pr(s_t = 1 | Y, \theta^{(\ell)})(Y_{t1} - \mu_1^{(\ell+1)} - \Phi^{(\ell+1)} X_t)^2}{\sum_{t=1}^T Pr(s_t = 1 | Y, \theta^{(\ell)})}$$

$$\sigma_2^{2(\ell+1)} = \frac{\sum_{t=1}^T Pr(s_t = 2 | Y, \theta^{(\ell)})(Y_{t1} - \mu_2^{(\ell+1)} - \Phi^{(\ell+1)} X_t)^2}{\sum_{t=1}^T Pr(s_t = 2 | Y, \theta^{(\ell)})}$$

Step 3. Update the inference about the transition probabilities. The transition probabilities are updated as

$$\hat{\pi}^{\mathbb{P}_{ij}(\ell+1)} = \frac{\sum_{t=2}^T Pr(s_t = j, s_{t-1} = i | Y_T, \theta^{(\ell)})}{\sum_{t=2}^T Pr(s_{t-1} = i | Y_T, \theta^{(\ell)})}$$

Specifically,

$$\hat{\pi}^{\mathbb{P}_{11}(\ell+1)} = \frac{\sum_{t=2}^T \pi^{\mathbb{P}_{11}(\ell)} \frac{Pr(s_t=1|Y_T,\theta^{(\ell)})Pr(s_{t-1}=1|Y_{t-1},\theta^{(\ell)})}{Pr(s_t=1|Y_{t-1},\theta^{(\ell)})}}{\sum_{t=2}^T Pr(s_{t-1} = 1 | Y_T, \theta^{(\ell)})}$$

$$\hat{\pi}^{\mathbb{P}_{21}(\ell+1)} = \frac{\sum_{t=2}^T \pi^{\mathbb{P}_{21}(\ell)} \frac{Pr(s_t=1|Y_T,\theta^{(\ell)})Pr(s_{t-1}=2|Y_{t-1},\theta^{(\ell)})}{Pr(s_t=1|Y_{t-1},\theta^{(\ell)})}}{\sum_{t=2}^T Pr(s_{t-1} = 2 | Y_T, \theta^{(\ell)})}$$

Step 4. Update the inference about smoothed probabilities. This step calculates the smoothed probabilities $Pr(s_t = j|Y, \theta^{(l+1)})$ using the Hamilton filtering and smoothing algorithms, which are described in Appendix 1.7.5. The initial probability vector ρ is updated as

$$\rho_j^{(l+1)} = Pr(s_1 = j|Y_T, \theta^{(\ell)})$$

1.7.4   EM algorithm for scalar regression

Here I present the general form of the EM algorithm I use for estimation of each equation from my regime-switching vector system. Suppose the variances and some but not all of the parameters change with the regime, that is

$$y_t = x_t'\beta + z_t'c_{s_t} + \sigma_{s_t}v_t$$

for y_t a scalar, x_t an $(m \times 1)$ vector, z_t an $(r \times 1)$ vector, and $v_t \sim N(0,1)$. Thus

$$\eta_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{ \frac{-(y_t - x_t'\beta - z_t'c_1)^2}{2\sigma_1^2} \right\} \\ \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{ \frac{-(y_t - x_t'\beta - z_t'c_2)^2}{2\sigma_2^2} \right\} \end{bmatrix}$$

$$\frac{\partial \log \eta_t}{\partial \beta'} = \begin{bmatrix} \frac{(y_t - x_t'\beta - z_t'c_1)x_t'}{\sigma_1^2} \\ \frac{(y_t - x_t'\beta - z_t'c_2)x_t'}{\sigma_2^2} \end{bmatrix} \tag{1.48}$$

$$\frac{\partial \log \eta_t}{\partial c_1'} = \begin{bmatrix} \frac{(y_t - x_t'\beta - z_t'c_1)z_t'}{\sigma_1^2} \\ 0 \end{bmatrix}$$

$$\frac{\partial \log \eta_t}{\partial c_2'} = \begin{bmatrix} 0 \\ \frac{(y_t - x_t'\beta - z_t'c_2)z_t'}{\sigma_2^2} \end{bmatrix}$$

$$\frac{\partial \log \eta_t}{\partial \sigma_1^2} = \begin{bmatrix} -\frac{1}{2\sigma_1^2} + \frac{(y_t - x_t'\beta - z_t'c_1)^2}{2\sigma_1^4} \\ 0 \end{bmatrix}$$

$$\frac{\partial \log \eta_t}{\partial \sigma_2^2} = \begin{bmatrix} 0 \\ -\frac{1}{2\sigma_2^2} + \frac{(y_t - x_t'\beta - z_t'c_2)^2}{2\sigma_2^4} \end{bmatrix}.$$

The MLE for $\theta = (\beta', c_1', c_2', \sigma_1^2, \sigma_2^2)'$ satisfies

$$\sum_{t=1}^{T} \left(\frac{\partial \log \eta_t}{\partial \theta'} \right)' \hat{\xi}_{t|T} = 0.$$

Define

$$\begin{bmatrix} \lambda_{t1} \\ \lambda_{t2} \end{bmatrix} = \begin{bmatrix} \sigma_1^{-1}\sqrt{\Pr(s_t = 1|Y)} \\ \sigma_2^{-1}\sqrt{\Pr(s_t = 2|Y)} \end{bmatrix}$$

for $Y = \{y_t, x_t, z_t\}_{t=1}^{T}$ the full set of observed data. Then the FOC associated with choice of β (using equation (1.48)) can be written

$$\left(\sum_{t=1}^{T} x_t y_t \lambda_{t1}^2 + \sum_{t=1}^{T} x_t y_t \lambda_{t2}^2 \right) = \left(\sum_{t=1}^{T} x_t x_t' \lambda_{t1}^2 + \sum_{t=1}^{T} x_t x_t' \lambda_{t2}^2 \right) \beta$$
$$+ \left(\sum_{t=1}^{T} x_t z_t' \lambda_{t1}^2 \right) c_1 + \left(\sum_{t=1}^{T} x_t z_t' \lambda_{t2}^2 \right) c_2.$$

Take the analogous FOC for choice of c_1 and c_2 and stack the three equations

together:

$$
\left[
\begin{array}{c}
\left(\sum_{t=1}^{T} x_t y_t \lambda_{t1}^2 + \sum_{t=1}^{T} x_t y_t \lambda_{t2}^2 \right) \\
\left(\sum_{t=1}^{T} z_t y_t \lambda_{t1}^2 \right) \\
\left(\sum_{t=1}^{T} z_t y_t \lambda_{t2}^2 \right)
\end{array}
\right]
=
\tag{1.49}
$$

$$
\left[
\begin{array}{ccc}
\left(\sum_{t=1}^{T} x_t x_t' \lambda_{t1}^2 + \sum_{t=1}^{T} x_t x_t' \lambda_{t1}^2 \right) & \left(\sum_{t=1}^{T} x_t z_t' \lambda_{t1}^2 \right) & \left(\sum_{t=1}^{T} x_t z_t' \lambda_{t2}^2 \right) \\
\left(\sum_{t=1}^{T} z_t x_t' \lambda_{t1}^2 \right) & \left(\sum_{t=1}^{T} z_t z_t' \lambda_{t1}^2 \right) & 0 \\
\left(\sum_{t=1}^{T} z_t x_t' \lambda_{t2}^2 \right) & 0 & \left(\sum_{t=1}^{T} z_t z_t' \lambda_{t2}^2 \right)
\end{array}
\right]
\left[
\begin{array}{c}
\beta \\
c_1 \\
c_2
\end{array}
\right].
$$

Conditional on knowing λ_{t1} and λ_{t2}, a closed-form solution for $(\hat{\beta}', \hat{c}_1', \hat{c}_2')'$ can be found by performing a single OLS regression on an artificial sample of size $2T$,

$$
\tilde{y}_t = \tilde{x}_t' \beta + \tilde{z}_{t1}' c_1 + \tilde{z}_{t2}' c_2 + \tilde{v}_t \quad t = 1, 2, ..., 2T,
$$

where for $t = 1, 2, ..., T$ I have defined

$$
\tilde{y}_t = y_t \lambda_{t1}
$$

$$
\tilde{x}_t = x_t \lambda_{t1}
$$

$$
\tilde{z}_{t1} = z_t \lambda_{t1}
$$

$$
\tilde{z}_{t2} = 0
$$

whereas the next T observations (denoted $T + t$ for $t = 1, ..., T$) are from

$$
\tilde{y}_{T+t} = y_t \lambda_{t2}
$$

$$
\tilde{x}_{T+t} = x_t \lambda_{t2}
$$

$$\tilde{z}_{T+t,1} = 0$$

$$\tilde{z}_{T+t,2} = z_t \lambda_{t2}.$$

The OLS coefficients from this artificial system are given by

$$
\begin{bmatrix} \hat{\beta} \\ \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{2T} \tilde{x}_t \tilde{x}_t' & \sum_{t=1}^{2T} \tilde{x}_t \tilde{z}_{t1}' & \sum_{t=1}^{2T} \tilde{x}_t \tilde{z}_{t2}' \\ \sum_{t=1}^{2T} \tilde{z}_{t1} \tilde{x}_t' & \sum_{t=1}^{2T} \tilde{z}_{t1} \tilde{z}_{t1}' & \sum_{t=1}^{2T} \tilde{z}_{t1} \tilde{z}_{t2}' \\ \sum_{t=1}^{2T} \tilde{z}_{t2} \tilde{x}_t' & \sum_{t=1}^{2T} \tilde{z}_{t2} \tilde{z}_{t1}' & \sum_{t=1}^{2T} \tilde{z}_{t2} \tilde{z}_{t2}' \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{2T} \tilde{x}_t \tilde{y}_t \\ \sum_{t=1}^{2T} \tilde{z}_{t1} \tilde{y}_t \\ \sum_{t=1}^{2T} \tilde{z}_{t2} \tilde{y}_t \end{bmatrix}
$$

$$
= \begin{bmatrix} \left(\sum_{t=1}^{T} x_t x_t' \lambda_{t1}^2 + \sum_{t=1}^{T} x_t x_t' \lambda_{t2}^2\right) & \left(\sum_{t=1}^{T} x_t z_t' \lambda_{t1}^2\right) & \left(\sum_{t=1}^{T} x_t z_t' \lambda_{t2}^2\right) \\ \left(\sum_{t=1}^{T} z_t x_t' \lambda_{t1}^2\right) & \left(\sum_{t=1}^{T} z_t z_t' \lambda_{t1}^2\right) & 0 \\ \left(\sum_{t=1}^{T} z_t x_t' \lambda_{t2}^2\right) & 0 & \left(\sum_{t=1}^{T} z_t z_t' \lambda_{t2}^2\right) \end{bmatrix}^{-1} \times
$$

$$
\begin{bmatrix} \left(\sum_{t=1}^{T} x_t y_t \lambda_{t1}^2 + \sum_{t=1}^{T} x_t y_t \lambda_{t2}^2\right) \\ \left(\sum_{t=1}^{T} z_t y_t \lambda_{t1}^2\right) \\ \left(\sum_{t=1}^{T} z_t y_t \lambda_{t2}^2\right) \end{bmatrix}
$$

which will be recognized as a closed-form solution to the FOC for the MLE as given in equation (1.49).

Thus an EM algorithm would work as follows. At the previous step I have calculated estimates $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\xi}_{t|T}$, from which I can construct λ_{t1} and λ_{t2}. I then use these to construct $\{\tilde{y}_t, \tilde{x}_t, \tilde{z}_{t1}, \tilde{z}_{t2}\}_{t=1}^{2T}$ and do an OLS regression of $\tilde{y}_t$ on $\tilde{x}_t, \tilde{z}_{t1}, \tilde{z}_{t2}$ to get new estimates of β, c_1, c_2.

Taking first order conditions for σ_1^2 and σ_2^2 results in the following expressions for the next step estimates:

$$
\hat{\sigma}_1^2 = \frac{\sum_{t=1}^{T} (y_t - x_t'\hat{\beta} - z_t'\hat{c}_1)^2 \Pr(s_t = 1|Y)}{\sum_{t=1}^{T} \Pr(s_t = 1|Y)}
$$

$$\hat{\sigma}_2^2 = \frac{\sum_{t=1}^{T}(y_t - x_t'\hat{\beta} - z_t'\hat{c}_2)^2 \Pr(s_t = 2|Y)}{\sum_{t=1}^{T}\Pr(s_t = 2|Y)}.$$

1.7.5   Filtering and smoothing algorithm

Let

$$\xi_t = \begin{bmatrix} \mathbb{1}\{s_t = 1\} \\ \mathbb{1}\{s_t = 2\} \end{bmatrix} \tag{1.50}$$

Let $\hat{\xi}_{t|\tau} = E(\xi_t|Y_\tau)$. Then

$$\hat{\xi}_{t|\tau} = \begin{bmatrix} Pr(\xi_t = e_1|Y_\tau) \\ Pr(\xi_t = e_2|Y_\tau) \end{bmatrix} \tag{1.51}$$

where Y_τ consists of information available up to time τ, $e_1 = (1,0)'$, $e_2 = (0,1)'$.
Let y_t be the vector of dependent variables of all the equations. Let η_t be the vector of densities of y_t conditional on ξ_t and Y_{t-1}:

$$\eta_t = \begin{bmatrix} p(y_t|\theta_1, Y_{t-1}) \\ p(y_t|\theta_2, Y_{t-1}) \end{bmatrix} = \begin{bmatrix} p(y_t|\xi_t = e_1, Y_{t-1}) \\ p(y_t|\xi_t = e_2, Y_{t-1}) \end{bmatrix} \tag{1.52}$$

where θ has been dropped on the right hand side for brevity.

In my model,

$$\eta_{t1} = (2\pi)^{(-(K+N)/2)}|\Psi_1|^{-1/2} \times$$

$$\exp\left[-\frac{1}{2}\left[\begin{pmatrix} Y_{t1} \\ Y_{t2} \end{pmatrix} - \begin{pmatrix} \mu_1 + \Phi X_t \\ A_1 + B X_t \end{pmatrix}\right]' \Psi_1^{-1}\left[\begin{pmatrix} Y_{t1} \\ Y_{t2} \end{pmatrix} - \begin{pmatrix} \mu_1 + \Phi X_t \\ A_1 + B X_t \end{pmatrix}\right]\right]$$

$$\eta_{t2} = (2\pi)^{(-(K+N)/2)} |\Psi_2|^{-1/2} \times$$

$$\exp\left[-\frac{1}{2}\left[\begin{pmatrix} Y_{t1} \\ Y_{t2} \end{pmatrix} - \begin{pmatrix} \mu_2 + \Phi X_t \\ A_2 + B X_t \end{pmatrix}\right]' \Psi_2^{-1}\left[\begin{pmatrix} Y_{t1} \\ Y_{t2} \end{pmatrix} - \begin{pmatrix} \mu_2 + \Phi X_t \\ A_2 + B X_t \end{pmatrix}\right]\right]$$

where

$$\Psi_j = \begin{bmatrix} \sigma_j^2 & 0_{K \times N} \\ 0_{N \times K} & \Omega_j \end{bmatrix}$$

and $K = 1$ and $N = 7$.

The density of y_t conditional on Y_{t-1} is given by $p(y_t | Y_{t-1}) = \eta_t' \hat{\xi}_{t|t-1} = \mathbb{1}_2'(\eta_t \odot \hat{\xi}_{t|t-1})$ where $\odot$ signifies element-wise matrix multiplication. The contemporaneous inference $\hat{\xi}_{t|t}$ about the unobserved state vector ξ_t is given in matrix notation by the filtering recursions

$$\hat{\xi}_{t|t} = \frac{\eta_t \odot \hat{\xi}_{t|t-1}}{\mathbb{1}_2'(\eta_t \odot \hat{\xi}_{t|t-1})} \tag{1.53}$$

$$\hat{\xi}_{t+1|t} = P \cdot \hat{\xi}_{t|t} \tag{1.54}$$

where P is the matrix of transition probabilities. The recursion is initialized with

$$\hat{\xi}_{1|0} = \rho$$

The smoothed inference about the unobserved state vector ξ_t is given by

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \left(P'(\hat{\xi}_{t+1|T}(\div)\hat{\xi}_{t+1|t})\right) \tag{1.55}$$

where the sign $(\div)$ denotes element-by-element division. The smoothed probabilities

$\hat{\xi}_{t|T}$ are found by iteration on equation (1.55) backward for $t = T-1, T-2, \cdots, 1$. This iteration is started with $\hat{\xi}_{T|T}$, which is obtained from equation (1.53) for $t = T$.

1.7.6 Estimation of number of factors

I use the Eigenvalue Ratio and Growth Ratio tests proposed in Ahn and Horenstein (2013) in order to estimate the number of factors in my model. Let Y be the $T \times N$ matrix containing the demeaned futures price data, with T=234 and $N = 7$, and let $\hat{\lambda}_k$ denote the kth largest eigenvalue of the covariance matrix $(Y'Y)/NT$. The Eigenvalue Ratio criterion function $ER(k)$ is the ratio of two adjacent eigenvalues of $(Y'Y)/NT$:

$$ER(k) \equiv \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}}, k = 1, 2, \ldots, k_{max} \tag{1.56}$$

where k is the number of factors used, and k_{max} is a specified maximum number of factors.

The Growth Ratio criterion function $GR(k)$ is given by

$$GR(k) \equiv \frac{\log(1 + \hat{\lambda}_k^*)}{\log(1 + \hat{\lambda}_{k+1}^*)} \tag{1.57}$$

where $V(k) = \sum_{j=k+1}^{m} \hat{\lambda}_j$ and $\hat{\lambda}_k^* = \hat{\lambda}_k / V(k)$.

The estimators of the true number of factors r are the maximizers of $ER(k)$ and $GR(k)$:

$$\hat{r}_{ER} = max_{1 \le k \le k_{max}} ER(k) \tag{1.58}$$

$$\hat{r}_{GR} = max_{1 \le k \le k_{max}} GR(k) \tag{1.59}$$

These estimators are called the ER and GR estimators, respectively.

1.8 Figures
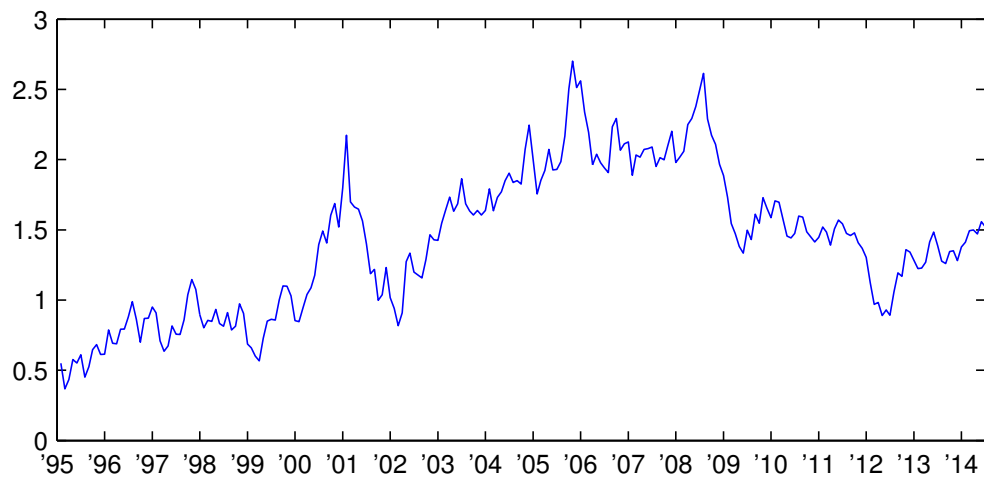
Figure 1.1. Log of the observed 3-month natural gas futures price
The natural gas futures price is measured in dollars per million Btu.
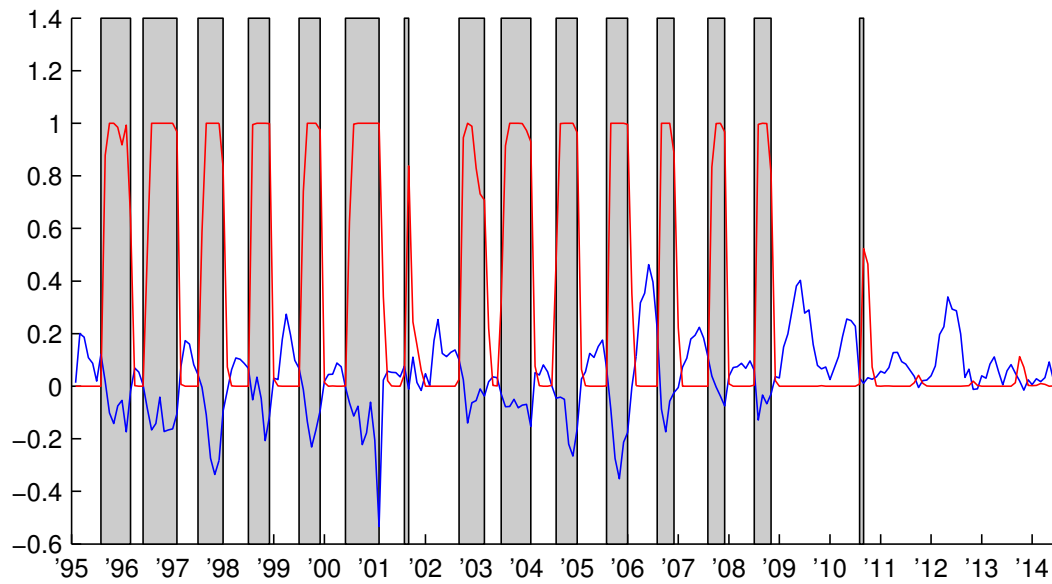
Figure 1.2. Spread between the log of the 9-month futures price and the log of the 3-month futures price vs. smoothed probability of regime 2 (the backwardation regime)
The spread is shown in blue, and the smoothed probability of regime 2 (the backwardation regime) is shown in red. Shaded areas represent the backwardation regime.
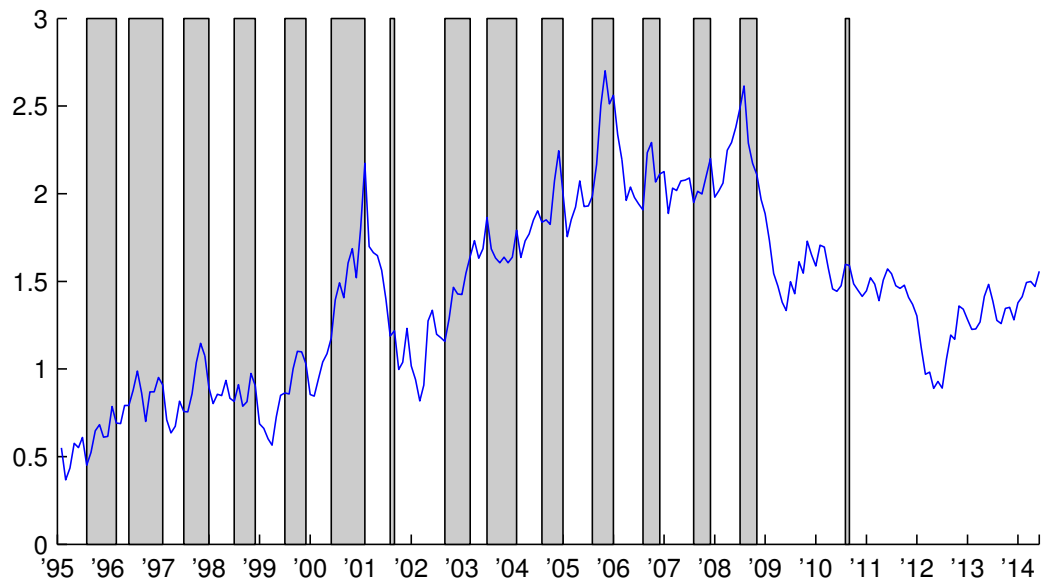
Figure 1.3. Log of the observed 3-month natural gas futures price
Shaded areas represent the backwardation regime.
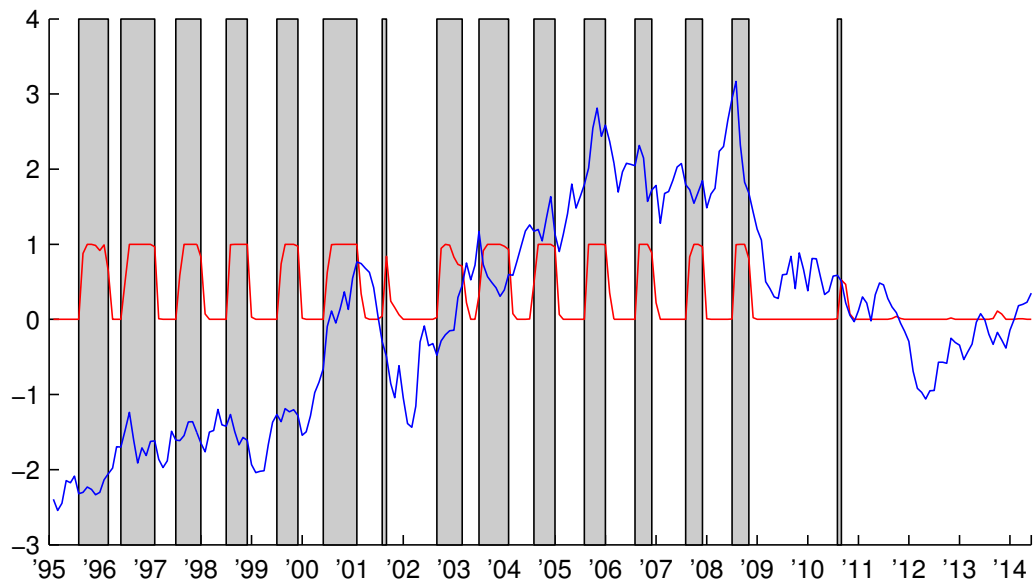
Figure 1.4. First principal component vs. smoothed probability of the backwardation regime
Shaded areas represent the backwardation regime. The first principal component is shown in blue, and the smoothed probability of the backwardation regime is shown in red.
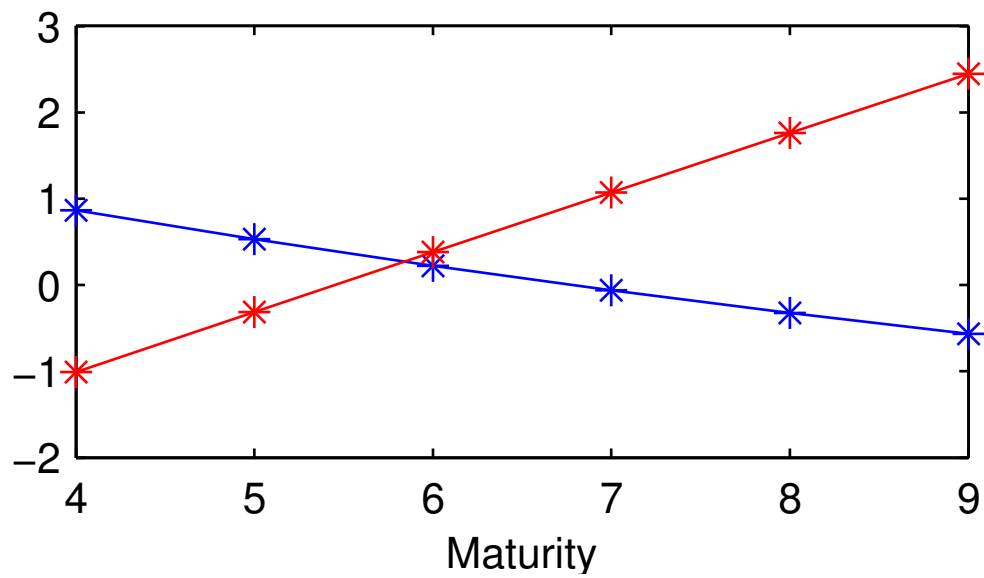
Figure 1.5. Expected monthly holding returns (%) conditional on each regime Expected monthly holding returns (%) conditional on the contango regime (regime 1, in blue) and conditional on the backwardation regime (regime 2, in red), averaged over each regime.

1.9 Tables

Table 1.1. First stage reduced form parameter estimates

	Regime 1		Regime 2
μ	0.0365		-0.0412
	(0.0189)		(0.0292)
Φ		0.9785	
		(0.0113)	
$A_{(3)}$	1.3803		1.4891
	(0.0058)		(0.0097)
$A_{(4)}$	1.3926		1.4743
	(0.0037)		(0.0058)
$A_{(5)}$	1.4120		1.4555
	(0.0031)		(0.0044)
$A_{(6)}$	1.4328		1.4309
	(0.0030)		(0.0045)
$A_{(7)}$	1.4524		1.4064
	(0.0031)		(0.0050)
$A_{(8)}$	1.4692		1.3863
	(0.0036)		(0.0055)
$A_{(9)}$	1.4786		1.3778
	(0.0049)		(0.0070)
$B_{(3)}$		0.3724	
		(0.0035)	
$B_{(4)}$		0.3760	
		(0.0022)	
$B_{(5)}$		0.3789	
		(0.0018)	
$B_{(6)}$		0.3808	
		(0.0017)	
$B_{(7)}$		0.3809	
		(0.0019)	

Table 1.1. First stage reduced form parameter estimates (continued)

	Regime 1		Regime 2
$B_{(8)}$		0.3797	
		(0.0021)	
$B_{(9)}$		0.3770	
		(0.0028)	
σ^2	0.0528		0.0584
	(0.0061)		(0.0098)
$\Omega_{(3)}$		0.0048	
		(0.0005)	
$\Omega_{(4)}$		0.0018	
		(0.0002)	
$\Omega_{(5)}$		0.0013	
		(0.0001)	
$\Omega_{(6)}$		0.0013	
		(0.0001)	
$\Omega_{(7)}$		0.0014	
		(0.0001)	
$\Omega_{(8)}$		0.0017	
		(0.0002)	
$\Omega_{(9)}$		0.0029	
		(0.0003)	
$\pi^{\mathbb{P}_{11}}$		0.9053	
		(0.0252)	
$\pi^{\mathbb{P}_{22}}$		0.8023	
		(0.0480)	
ρ_1		1.0000	
logL		2721.6	

Asymptotic standard errors are in parentheses

Table 1.2. Second stage parameter estimates

	Regime 1		Regime 2
μ	0.0357		-0.0396
	(0.0189)		(0.0292)
Φ		0.9785	
		(0.0113)	
$A_{(3)}$	1.3743		1.4959
	(0.0046)		(0.0083)
$B_{(3)}$		0.3767	
		(0.0020)	
σ^2	0.0528		0.0586
	(0.0061)		(0.0098)
$\Omega_{(3)}$		0.0048	
		(0.0005)	
$\Omega_{(4)}$		0.0018	
		(0.0002)	
$\Omega_{(5)}$		0.0013	
		(0.0001)	
$\Omega_{(6)}$		0.0013	
		(0.0001)	
$\Omega_{(7)}$		0.0014	
		(0.0001)	
$\Omega_{(8)}$		0.0017	
		(0.0002)	
$\Omega_{(9)}$		0.0029	
		(0.0003)	
$\pi^{\mathbb{P}_{11}}$		0.9070	
		(0.0252)	
$\pi^{\mathbb{P}_{22}}$		0.8060	
		(0.0479)	
$\pi^{\mathbb{Q}_{11}}$		0.9776	
		(0.0231)	
$\pi^{\mathbb{Q}_{22}}$		0.9813	
		(0.0383)	
λ_0	-0.0008		0.0257
	(0.0189)		(0.0295)
λ_1		-0.0233	
		(0.0114)	

Table 1.3. Wald t-statistics

$\mathbb{H}_0$	Estimate	Std error	Wald t-stat
Reduced-form parameters (1st stage)			
$A_{(4),1} - A_{(4),2} = 0$	-0.0817	0.0063	-13.0459
$A_{(5),1} - A_{(5),2} = 0$	-0.0435	0.0053	-8.1938
$A_{(6),1} - A_{(6),2} = 0$	0.0019	0.0056	0.3393
$A_{(7),1} - A_{(7),2} = 0$	0.0460	0.0057	8.0412
$A_{(8),1} - A_{(8),2} = 0$	0.0829	0.0059	14.0888
$A_{(9),1} - A_{(9),2} = 0$	0.1008	0.0079	12.7908
Reduced-form parameters (2nd stage)			
$\mu_1 - \mu_2 = 0$	0.0753	0.0357	2.1076
$\sigma_1^2 - \sigma_2^2 = 0$	-0.0057	0.0117	-0.4903
$A_{(3),1} - A_{(3),2} = 0$	-0.1216	0.0087	-13.9022
Structural parameters (2nd stage)			
$\lambda_{0,1} - \lambda_{0,2} = 0$	0.0489	0.0360	1.3575
$\pi^{\mathbb{P}_{11}} - \pi^{\mathbb{Q}_{11}} = 0$	-0.0706	0.0351	-2.0078
$\pi^{\mathbb{P}_{22}} - \pi^{\mathbb{Q}_{22}} = 0$	-0.1753	0.0605	-2.8996

Table 1.4. Reduced-form versus model implied values for $A_{(n),j}$ and $B^{(n)}$

	First stage estimates	Model implied values
$A_{(3),1}$	1.3803	1.3743
$A_{(3),2}$	1.4891	1.4959
$A_{(4),1}$	1.3926	1.3948
$A_{(4),2}$	1.4743	1.4734
$A_{(5),1}$	1.4120	1.4142
$A_{(5),2}$	1.4555	1.4515
$A_{(6),1}$	1.4328	1.4326
$A_{(6),2}$	1.4309	1.4304
$A_{(7),1}$	1.4524	1.4502
$A_{(7),2}$	1.4064	1.4099
$A_{(8),1}$	1.4692	1.4670
$A_{(8),2}$	1.3863	1.3901
$A_{(9),1}$	1.4786	1.4831
$A_{(9),2}$	1.3778	1.3711
$B_{(3)}$	0.3724	0.3767
$B_{(4)}$	0.3760	0.3774
$B_{(5)}$	0.3789	0.3781
$B_{(6)}$	0.3808	0.3788
$B_{(7)}$	0.3809	0.3794
$B_{(8)}$	0.3797	0.3801
$B_{(9)}$	0.3770	0.3808

1.10 Acknowledgement

Chapter 1, in full, is currently being prepared for submission for publication of the material. Zhecheva, Irina Y. The dissertation author was the primary investigator and author of this material.

1.11   Bibliography

Adrian, T., Crump, R., and Moench, E. (2013) "Pricing the Term Structure with Linear Regressions." *Journal of Financial Economics* 110(1), 110-138.

Ahn, S. and Hornstein, A. (2013) "Eigenvalue Ratio Test for the Number of Factors." *Econometrica* 81, 1203-1227.

Almansour, A. (2016) "Convenience yield in commodity price modeling: A regime switching approach." *Energy Economics* 53 (2016) 238-247.

Ang, A. and Bekaert, G. (2002) "Regime Switches in Interest Rates." *Journal of Business & Economic Statistics*, American Statistical Association, vol. 20(2), 163-82.

Ang, A. and Piazzesi, M. (2003) "A No-arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics* 50, 745-787.

Ang, A., Bekaert, G., and Wei, M. (2008) "The Term Structure of Real Rates and Expected Inflation." *Journal of Finance*, 63: 797-849.

Bansal, R. and Zhou, H. (2002) "Term Structure of Interest Rates with Regime Shifts." *Journal of Finance*, 57, 1997-2043.

Casassus, J. and Collin-Dufresne, P. (2006) "Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates." *Journal of Finance* 60, 2283-2331.

Chen, R.-R. and Scott, L. (1993) "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates." *Journal of Fixed Income*, Vol. 3, No. 3: 14-31.

Dai, Q. and Singleton, K.J. (2000) "Specification Analysis of Affine Term Structure Models." *Journal of Finance*, 55, 1943-1978.

Dai, Q., Singleton, K., and Yang, W. (2007) "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields." *Review of Financial Studies*, Vol. 20, 1669-1706.

Diez de los Rios, A. (2015) "A New Linear Estimator for Gaussian Dynamic Term Structure Models." *Journal of Business and Economic Statistics*, Vol. 33, Iss. 2.

Duffee, G. R. (2002) "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57, 405-443.

Duffee, G. R. (2009) "Forecasting with the Term Structure: the Role of No-Arbitrage," Working Paper June, Haas School of Business, University of California, Berkeley.

Duffee, D. and Kan, R. (1996) "A Yield-factor Model of Interest Rates." *Mathematical Finance*, 6, 379-406.

Duffee, G. R. and Stanton, R. H. (2008) "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation." *Journal of Financial Econometrics*, 6 (1), 108-142.

Gargano, A. and Timmermann, A. (2014) "Predictive Dynamics in Commodity Prices." *International Journal of Forecasting* 30(3), p. 825-843.

Gibson, R. and Schwartz, E.S. (1990) "Stochastic convenience yield and the pricing of oil contingent claims." *Journal of Finance*, 45 (3), 959-976.

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

Hamilton, J. D. (1994) "Time Series Analysis." Princeton, NJ: Princeton University Press.

Hamilton, J. D. and Wu, J. C. (2012) "Identification and Estimation of Gaussian Affine Term Structure Models." *Journal of Econometrics*, 168, no. 2: 315-331.

Hamilton, J. D. and Wu, J. C. (2014) "Risk Premia in Crude Oil Futures Prices." *Journal of International Money and Finance*, 42, 9-37.

Hamilton, J. D. (2016) "Macroeconomic Regimes and Regime Shifts." *Handbook of Macroeconomics*, Vol. 2A., 163-201, edited by Harald Uhlig and John Taylor, Amsterdam: Elsevier.

Joslin, S., Singleton, K.J., and Zhu, H. (2011) "A New Perspective on Gaussian DTSMs." *Review of Financial studies*, 24, 926-970.

Kim, D. H. (2008) "Challenges in macro-finance modeling." BIS Working Paper No. 240, FEDS Working Paper No. 2008-06.

Kim, D. H. and Orphanides, A. (2005) "Term structure estimation with survey data on interest rate forecasts." *Federal Reserve Board, Finance and Economics Discussion Series 2005-48*.

Krolzig, H. (1997) "Markov-Switching Vector Autoregressions." Springer-Verlag Berlin Heidelberg.

Linn, S. and Zhu, Z. (2004) "Natural Gas Prices and the Gas Storage Report: Public News and Volatility in Energy Futures Markets." *Journal of Futures Markets*, Vol. 24, Issue 3, 283-313.

Mankiw, N.G. and Miron, J. A. (1986) "The Changing Behavior of the Term Structure of Interest Rates." *Quarterly Journal of Economics*, 101, 211-228.

Piazzesi, M. (2010) "Affine term structure models." In *Handbook of Financial Econometrics*, edited by Ait-Sahalia, Y. and Hansen, L.P., pp. 691-766. New York: Elsevier.

Schwartz, Eduardo S. (1997) "The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging." *Journal of Finance*, 52, 923-973.

Schwartz, Eduardo S. and Smith, J. E. (2000) "Short-Term Variations and Long-Term Dynamics in Commodity Prices." *Management Science*, 46, 893-911.

Suenaga, H., Smith, A., and Williams, J. (2008) "Volatility Dynamics of NYMEX Natural Gas Futures Prices." *Journal of Futures Markets*, Vol. 28, Issue 5, 438-463.

Vasicek, O. (1977) "An equilibrium characterization of the term structure." *Journal of Financial Economics*, 5(2):177-188.

Chapter 2

Bond Risk Premia in the Presence of Regime Switching

Abstract

Bond yields are known to behave quite differently during different periods
and to change their behavior abruptly. Regime-switching models are very well suited
to match these properties of bond markets. I develop and estimate a dynamic
term structure model with regime switching, and investigate the consequences
of changes in regime on expected excess returns, using zero-coupon bond yield
data from 1990 to 2007. Statistical tests provide very strong evidence for this
regime-switching model. I find that one regime corresponds to episodes when
rates are falling and the volatility of yields is lower, whereas the second regime
corresponds to non-decreasing rates and higher volatility. I find that, on average, it
is always profitable for investors to be long in either short term or long term bonds.
According to my model, expected excess returns on all bonds are higher conditional
on the decreasing rates regime. Even though the yield curve is flatter in that regime,
investors would want to hold more long term bonds because rates are decreasing. I
also find that a higher yield spread is associated with higher expected excess returns.
A separate contribution of this paper is to provide an estimation procedure that is

considerably simpler and less prone to numerical difficulties than other methods in the literature. To that end, I adapt the new econometric framework I proposed in the first chapter of my dissertation to the setting of bond yields.

2.1 Introduction

Bond yields are known to behave quite differently during different periods and to change their behavior abruptly. While some changes may be very short term, others may persist for many periods. Regime-switching models are very well suited to match these properties of bond markets. Previous research has provided evidence that two-state regime-switching models do much better at describing interest rate dynamics than single regime models, e.g. Ang and Bekaert (2002), Bansal and Zhou (2002), and Dai, Singleton, and Yang (2007), among others.

In this paper, I model bond yields using a regime-switching Gaussian affine term structure model. I use a three factor model with two regimes, and assume the factors are observed and equal to the first three principal components of bond yields with maturities 3 months, 6 months, 1 year, 2 years, 4 years, 5 years, 9 years, and 10 years. As is typical in the bond yield literature, the factors correspond to the level, slope, and curvature of the yield curve. I allow the level and volatility of the vector of factors, the level of bond yields, the level of the short rate, the market price of factor risk, and the market price of regime shift risk to depend on the regime. I use the zero-coupon yield dataset constructed by Gurkaynak, Sack, and Wright (2007) for the period from 1990 to 2007. My sample period ends before the Zero Lower Bound (ZLB) period, during which it is not appropriate to apply Gaussian affine term structure models, as pointed out by Wu and Xia (2016).

I find very strong evidence for a regime-switching model in which one regime is characterized by falling rates and lower volatility of yields, while the other regime

corresponds to non-decreasing rates and higher volatility. This result differs from other papers studying regime switching in bond yields in that regimes are not found to be associated with recessions and expansions. The main empirical finding of this paper is that on average it is always profitable for investors to be long in bonds of any maturity over my sample period. Investors would want to hold more long term bonds in the regime in which rates are decreasing, even though on average the yield curve in that regime is flatter.

Gaussian affine term structure models are extensively employed to describe the cross-section of yields. Some seminal papers in the literature on Gaussian affine term structure models are Vasicek (1977), Duffie and Kan (1996), Dai and Singleton (2000), Duffee (2002), and Piazzesi (2010). Dai et al. (2007), Bansal and Zhou (2002), and Ang and Bekaert (2002), among others, propose and estimate regime-switching models for the term structure of interest rates.

Most methods in the literature rely on numerical maximization of the likelihood function with respect to all the parameters. As a result, these methods often run into numerical issues due to highly non-linear and badly behaved likelihood surfaces, and estimation can be very difficult or even computationally infeasible. Multiple researchers have reported encountering such difficulties, e.g. Kim and Orphanides (2005), Duffee (2002), Ang and Piazzesi (2003), Kim (2008), Duffee and Stanton (2008), Duffee (2009), and Ang and Bekaert (2002). The estimation in Dai et al. (2007) and Ang and Bekaert (2002) is performed using maximum-likelihood based methods based on an iterative procedure developed by Hamilton (1989). Bansal and Zhou (2002) employ a two-step efficient method of moments estimator. Significant numerical issues are often encountered when using these methods.

One of the contributions of this paper is to provide an estimation procedure that is considerably simpler and less prone to numerical difficulties. To that end, I

adapt the new econometric framework for estimation of regime-switching Gaussian affine term structure models that I developed in Chapter 1 of my dissertation to the setting of bond yields. I use a two stage estimation method. In the first stage, I use a regression based iterative approach to estimate the reduced-form parameters. The no-arbitrage restrictions are not imposed in this stage. In the second stage, I exploit the no-arbitrage restrictions and estimate the prices of risk and risk-neutral transition probabilities via minimum-chi-square estimation. The minimum-chi-square procedure chooses the values of the prices of risk and risk-neutral transition probabilities so that the values for the recursive bond pricing parameters implied by the no-arbitrage restrictions most closely fit the unrestricted first-stage estimates. The numerical component of my method is significantly simpler computationally than maximizing the likelihood function numerically with respect to all the parameters, as is typically done by other methods. Thus, with my approach I avoid many of the numerical difficulties encountered with other methods in the literature.

The rest of the paper is organized as follows. Section 2.2 presents the model framework, Section 2.3 describes the estimation approach, Section 2.4 gives empirical results, and Section 2.5 concludes.

2.2 Model

Let $P_t^{(n)}$ denote the price of an n-period zero-coupon bond at time t, and let $p_t^{(n)}$ denote the log price: $p_t^{(n)} \equiv \log P_t^{(n)}$. The corresponding yield is

$$y_t^{(n)} = -n^{-1} p_t^{(n)} \tag{2.1}$$

Let $rx_{t+1}^{(n-1)}$ denote the one-month log excess holding return on an n-period zero-coupon bond. It is defined as

$$rx_{t+1}^{(n-1)} = p_{t+1}^{(n-1)} - p_t^{(n)} - r_t \tag{2.2}$$

This is the return on buying an n-period zero-coupon bond in period t and then selling it as an $(n-1)$-period zero-coupon bond in period $t+1$.

The Gaussian affine term structure framework assumes that there are K factors, denoted by X_t, relevant for bond pricing. The log of the bond price, and hence the yield, are assumed to be a function of these factors. The factors X_t follow a first-order Gaussian vector autoregression under the $\mathbb{P}$-measure:

$$X_{t+1} = \mu_{s_t} + \Phi X_t + v_{t+1} \ , \ v_{t+1}|s_t \sim N(0, \Sigma_{s_t}) \tag{2.3}$$

Here s_t denotes the regime at time t, $s_t \in \{1, 2\}$. I allow the intercept parameter μ to change with the regime. I assume that the slope parameter Φ is independent of the regime. [1] Under no-arbitrage, there exists a pricing kernel $M_{t,t+1}$ such that

$$P_t^{(n)} = E_t \left[M_{t,t+1} P_{t+1}^{(n-1)} | s_t = j \right] \tag{2.4}$$

if the regime at time t is j. Following Dai et al. (2007), I assume that the pricing kernel is exponentially affine:

$$M_{t,t+1} = \exp \left[-r_t - \Gamma_{s_t, s_{t+1}} - \frac{1}{2} \lambda_{t,s_t}' \lambda_{t,s_t} - \lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} v_{t+1} \right] \tag{2.5}$$

---

[1] In my empirical application, I find that allowing Φ to change with the regime leads to a non-stationary process for the factor under one of the regimes. To avoid this issue, I restrict Φ to be regime independent.

where r_t is the short rate. The terms $\Gamma_{s_t,s_{t+1}}$ and λ_{t,s_t} allow for the possibility of non-zero expected excess returns in equilibrium. The term $\Gamma_{s_t,s_{t+1}}$ is referred to as the market price of regime shift risk, and can be interpreted as the excess log expected return per unit of regime shift risk exposure. The term λ_{t,s_t} is referred to as the market price of factor risk. The market price of factor risk can be thought of as the excess log expected return required per unit of factor risk exposure (Dai et al. (2007)). I also assume that the market price of factor risk λ_{t,s_t} is an affine function of the vector of factors X_t:

$$\lambda_{t,s_t} = \Sigma_{s_t}^{-1/2}\left(\lambda_{0,s_t} + \lambda_1 X_t\right) \tag{2.6}$$

I assume that the slope component λ_1 of the market price of risk is regime independent.[2] Under assumptions (2.3), (2.5), and (2.6), it can be shown that the price $P(X_t)$ in regime j of any asset whose payoff is a function of the factors $g(X_{t+1})$ can be computed as

$$P(X_t) = E_t^{\mathbb{P}}[M_{t,t+1}g(X_{t+1})|s_t = j] = \exp(-r_t^j)E_t^{\mathbb{Q}}[g(X_{t+1})|s_t = j] \tag{2.7}$$

Under the risk-neutral $\mathbb{Q}$-measure, the factor X_t follows a Gaussian vector autoregression:

$$X_{t+1} = \mu_{s_t}^{\mathbb{Q}} + \Phi^{\mathbb{Q}}X_t + v_{t+1}^{\mathbb{Q}} \tag{2.8}$$

In Appendix 2.6.1 it is shown that

$$\mu_j^{\mathbb{Q}} = \mu_j - \lambda_{0,j} \tag{2.9}$$

---

[2]The slope component of the market price of risk λ_1 is related to the slope parameter Φ from the factor VAR through the relation $\Phi^{\mathbb{Q}} = \Phi - \lambda_1$. I assume $\Phi^{\mathbb{Q}}$ to be regime independent in order to obtain closed form solutions for bond prices. Since I have also assumed Φ to be regime independent, it follows that λ_1 is regime independent.

$$\Phi^{\mathbb{Q}} = \Phi - \lambda_1 \tag{2.10}$$

and $v_{t+1}^{\mathbb{Q}}|s_t = j \sim^{\mathbb{Q}} N(0, \Sigma_j)$. I also assume that the short rate is an affine function of the vector of factors:

$$r_t = y_t^{(1)} = \delta_{0,s_t} + \delta_1' X_t \tag{2.11}$$

I constrain the loadings δ_1 of the short rate on the vector of factors X_t to be the same across regimes in order to be able to obtain closed-form solutions for bond prices. Equations (2.3), (2.4), (2.5), and (2.11) imply that zero-coupon bond yields and the log of bond prices are affine in the vector of factors as well:

$$y_t^{(n)} = \frac{1}{n} A_{s_t}^{(n)} + \frac{1}{n} B^{(n)'} X_t \tag{2.12}$$

Similarly for bond prices,

$$p_t^{(n)} = -A_{s_t}^{(n)} - B^{(n)'} X_t \tag{2.13}$$

I allow the intercept term $A_{s_t}^{(n)}$ to be different across regimes. Since $\Phi^{\mathbb{Q}}$ and δ_1 are assumed to be regime independent, from equation (2.38) it follows that the factor loadings $B^{(n)}$ are regime independent. This makes it possible to obtain exact closed-form solutions for bond prices.

I assume that there are two regimes that govern the dynamic properties of the vector of factors X_t. The unobserved regime variable s_t is presumed to follow a two-state Markov chain, with the risk-neutral probability of switching from regime $s_t = j$ to regime $s_{t+1} = k$ given by $\pi^{\mathbb{Q}jk}, 1 \le j, k \le 2$, with $\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} = 1$, for $j = 1, 2$. I assume that the risk-neutral transition probabilities $\pi^{\mathbb{Q}jk}$ and the historic transition probabilities $\pi^{\mathbb{P}jk}$ are regime independent, and allow $\pi^{\mathbb{P}jk} \ne \pi^{\mathbb{Q}jk}$. Agents

are presumed to know the history of the vector of factors X_t and of the regime. For tractability, I assume that the Markov process governing regime changes is conditionally independent of the process for X_t. The econometrician is presumed to observe X_t but not the state s_t.

The model implies the following recursive relations for the bond pricing parameters $A_j^{(n)}$ and $B^{(n)}$ (derived in Appendix 2.6.2 in equations (2.37) and (2.39)):

$$A_j^{(n)} = \delta_0^j - \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}_{jk}} e^{-A_k^{(n-1)}}\right) + B^{(n-1)'}(\mu_j - \lambda_{0,j}) - \frac{1}{2}B^{(n-1)'}\Sigma_j B^{(n-1)} \quad (2.14)$$

for $j = 1, 2$ and

$$B^{(n)} = \delta_1 + (\Phi - \lambda_1)' B^{(n-1)} \tag{2.15}$$

with initial conditions $A_j^{(1)} = \delta_0^j$ and $B^{(1)} = \delta_1$. These recursions represent non-linear cross-equation no-arbitrage restrictions on $A_j^{(n)}$ and $B^{(n)}$. The restrictions are not used or imposed in the initial reduced-form estimation, but are exploited in the second stage of the estimation described below.

2.3   Estimation procedure

I use the zero-coupon yield data set constructed by Gurkaynak, Sack, and Wright (2007) for the period from 1990 to 2007. Gurkaynak et al. (2007) provide parameters of fitted Nelson-Siegel-Svensson curves based on which zero-coupon yields can be calculated. I use these parameters to compute yields for maturities 1 month, 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 9 years, and 10 years. I assume that the vector of factors X_t is observed, and it consists of the first three principal components extracted from standardized end-of-month bond yields with maturities 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 9 years, and 10

years. I take the 1-month yield as the risk-free rate. My empirical analysis uses annualized yield data because as pointed out by Lemke (2006), for monthly yields the measurement error variance would be very small, which might lead to numerical difficulties. I propose a two-step procedure for estimating the model parameters. First, I estimate the reduced-form parameters based on equations (2.12) and (2.3).

2.3.1 Estimation of reduced-form parameters via regime-switching VAR's

Annualized yields [3] are given by

$$y_{t,A}^{(n)} = 1200 y_t^{(n)} = \frac{1200}{n} A_{s_t}^{(n)} + \frac{1200}{n} B^{(n)'} X_t \tag{2.16}$$

where $y_t^{(n)}$ represent monthly yields. Based on equations (2.16) and (2.3), I propose to estimate the following regime-switching regressions:

$$y_{t,A}^{(n)} = a_{s_t}^{(n)} + b^{(n)'} X_t + u_t^{(n)}, n = 1m, 3m, 6m, 12m, 24m, 36m, 60m, 108m, 120m \tag{2.17}$$

where

$$\left(u_t^{(1)}, u_t^{(3)}, u_t^{(6)}, u_t^{(12)}, u_t^{(24)}, u_t^{(36)}, u_t^{(60)}, u_t^{(108)}, u_t^{(120)} \right) | s_t \sim N(0, \Omega) \tag{2.18}$$

---

[3]We have to multiply by 1200 since yields in the dataset are expressed in percentages.

and

$$\Omega \equiv \begin{pmatrix} [\omega^{(1)}]^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \omega^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \omega^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \omega^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \omega^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \omega^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \omega^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \omega^2 \end{pmatrix} \tag{2.19}$$

jointly with the regime-switching regression for X_t:

$$X_{t+1} = \mu_{s_t} + \Phi X_t + v_{t+1} \;,\; v_{t+1}|s_t \sim N(0, \Sigma_{s_t}) \tag{2.20}$$

as a vector system of regime-switching equations.

In the first stage, I estimate the vector system of regime-switching equations in (2.17)-(2.20) using the EM algorithm. The details of the algorithm I use can be found in Appendix 1.7.3-1.7.5 of Chapter 1 of this dissertation. The general vector version of the EM algorithm is given in Hamilton (2016).

2.3.2 Minimum-chi-square estimation of structural parameters

In the second stage, I back out the prices of risk and risk-neutral transition probabilities by minimum-chi-square estimation.

Comparing equations (2.16) and (2.17), we can see that the first stage estimates of the recursive bond pricing parameters $A_j^{(n)}$ and $B^{(n)}$ can be backed

out from

$$\hat{A}_j^{(n)} = \frac{n}{1200} \hat{a}_j^{(n)} \tag{2.21}$$

and

$$\hat{B}^{(n)} = \frac{n}{1200} \hat{b}^{(n)} \tag{2.22}$$

The minimum-chi-square procedure chooses the values of $\lambda_{0,1}$, $\lambda_{0,2}$, λ_1, $\pi_{11}^{\mathbb{Q}}$, and $\pi_{22}^{\mathbb{Q}}$ so that the values for $A_j^{(n)}$ and $B^{(n)}$ implied by equations (2.14) and (2.15) most closely fit the unrestricted estimates (2.21)-(2.22). Minimum-chi-square estimation in the context of single regime affine term structure models is discussed in Hamilton and Wu (2012).

Let π denote the vector of reduced-form parameters (VAR coefficients, variance-covariance matrix of the vector of factors, measurement error variances, and $\mathbb{P}$-measure regime-switching probabilities). Let $\mathcal{L}(\pi; Y)$ denote the log-likelihood for the entire sample, and let $\hat{\pi} =$arg max $\mathcal{L}(\pi; Y)$ denote the full-information maximum likelihood estimate. If $\hat{R}$ is a consistent estimate of the information matrix,

$$R = -T^{-1} E \left[\frac{\partial^2 \mathcal{L}(\pi; Y)}{\partial \pi \partial \pi'} \right] \tag{2.23}$$

then θ can be estimated by minimizing the chi-square statistic

$$T \left[\hat{\pi} - g(\theta) \right]' \hat{R} \left[\hat{\pi} - g(\theta) \right] \tag{2.24}$$

As noted by Hamilton and Wu (2012), the variance of $\hat{\theta}$ can be approximated with

$$\hat{Var}(\hat{\theta}) \equiv T^{-1} (\hat{\Gamma}' \hat{R} \hat{\Gamma})^{-1} \tag{2.25}$$

for $\hat{\Gamma} = \frac{\partial g(\theta)}{\partial \theta'} |_{\theta = \hat{\theta}}$.

The second stage of the estimation in this case consists of minimizing the distance between the unrestricted maximum likelihood estimates of the coefficients $A_j^{(n)}$ and $B^{(n)}$ (from the regime-switching regressions) and the values of $A_j^{(n)}$ and $B^{(n)}$ implied by the no-arbitrage restrictions. According to equations (2.14) and (2.15), the latter are predicted to be functions of θ, a vector of structural parameters defined as follows:

$$\theta = \left(\mu_1, \mu_2, \Phi, \delta_{0,1}, \delta_{0,2}, \delta_1, \Sigma_1, \Sigma_2, \omega^{2(1)}, \omega^2, \pi^{\mathbb{P}_{11}}, \pi^{\mathbb{P}_{22}}, \lambda_{0,1}, \lambda_{0,2}, \lambda_1, \pi^{\mathbb{Q}_{11}}, \pi^{\mathbb{Q}_{22}}\right)$$

(2.26)

Let $\hat{\pi}$ be the vector of the unrestricted maximum likelihood estimates from the regime-switching VAR:

$$\hat{\pi} = \left(\mu_1, \mu_2, \Phi, [a_j^{(3m)}; a_j^{(6m)}; a_j^{(12m)}; a_j^{(24m)}; a_j^{(36m)}; a_j^{(60m)}; a_j^{(108m)}; a_j^{(120m)}; a_j^{(1m)}],\right.$$
$$\left.[b^{(3m)}; b^{(6m)}; b^{(12m)}; b^{(24m)}; b^{(36m)}; b^{(60m)}; b^{(108m)}; b^{(120m)}; b^{(1m)}], \Sigma_1, \Sigma_2, \omega^{2(1)}, \omega^2,\right.$$
$$\left.\pi^{\mathbb{P}_{11}}, \pi^{\mathbb{P}_{22}}\right)$$

(2.27)

and

$$g(\theta) = \left(\mu_1, \mu_2, \Phi, \left[\frac{1200 A_j^{(3m)}(\theta)}{3}; \frac{1200 A_j^{(6m)}(\theta)}{6}; \frac{1200 A_j^{(12m)}(\theta)}{12}; \frac{1200 A_j^{(24m)}(\theta)}{24};\right.\right.$$
$$\frac{1200 A_j^{(36m)}(\theta)}{36}; \frac{1200 A_j^{(60m)}(\theta)}{60}; \frac{1200 A_j^{(108m)}(\theta)}{108}; \frac{1200 A_j^{(120m)}(\theta)}{120};$$
$$\left.\frac{1200 A_j^{(1m)}(\theta)}{1}\right], \left[\frac{1200 B^{(3m)}(\theta)}{3}; \frac{1200 B^{(6m)}(\theta)}{6}; \frac{1200 B^{(12m)}(\theta)}{12}; \frac{1200 B^{(24m)}(\theta)}{24};\right.$$
$$\frac{1200 B^{(36m)}(\theta)}{36}; \frac{1200 B^{(60m)}(\theta)}{60}; \frac{1200 B^{(108m)}(\theta)}{108}; \frac{1200 B^{(120m)}(\theta)}{120};$$
$$\left.\left.\frac{1200 B^{(1m)}(\theta)}{1}\right], \Sigma_1, \Sigma_2, \omega^{2(1)}, \omega^2, \pi^{\mathbb{P}_{11}}, \pi^{\mathbb{P}_{22}}\right)$$

(2.28)

$A_j^{(n)}(\theta)$ and $B^{(n)}(\theta)$ are defined by the no-arbitrage restrictions in equations (2.14) and (2.15).

Then I obtain $\hat{\theta}$ as

$$\hat{\theta} \equiv \operatorname{argmin}_\theta \{T \left[\hat{\pi} - g(\theta)\right]' \hat{R} \left[\hat{\pi} - g(\theta)\right]\} \tag{2.29}$$

In this way I obtain estimates of the prices of risk $\lambda_{0,1}$, $\lambda_{0,2}$, and λ_1 and of the risk-neutral transition probabilities $\pi^{\mathbb{Q}_{11}}$ and $\pi^{\mathbb{Q}_{22}}$ as part of the vector $\hat{\theta}$. I also obtain second-stage estimates of the reduced-form parameters.

Rothenberg (1973, pp. 24-25) shows that when the reduced-form estimate is the unrestricted MLE and the weighting matrix is the associated information matrix (which is the case in this paper), the resulting minimum-chi-square estimate is asymptotically equivalent to full-information MLE. Hamilton and Wu (2012) also show that the variance $\hat{Var}(\hat{\theta})$ given in equation (2.25) above is identical to the usual asymptotic variance for the MLE obtained from second derivatives of the log-likelihood function directly with respect to θ.

2.4 Empirical results

Table 2.1 shows the first stage reduced-form parameter estimates, Table 2.2 shows the second stage reduced-form estimates, Table 2.3 shows the second stage structural parameter estimates, and Table 2.4 shows Wald t-statistics for the hypothesis tests that there is no regime switching in the regime-dependent parameters. I find very strong evidence that there is regime switching in the U.S. Treasury bond yields data. The hypotheses of no regime switching in the intercept parameters in the yield equations are strongly rejected. I also find that the level and variance of the first factor change with the regime. As is typical in the bond yield

literature, the factors are found to correspond to the level, slope, and curvature of the yield curve. Figure 2.1 plots the first factor with the periods corresponding to regime 2 shaded. The level $\mu_{1,1}$ of the first factor in regime 1 is negative and statistically significant, indicating that this regime corresponds to episodes when rates are falling. On the other hand, regime 2 corresponds to periods when rates are generally either rising or relatively constant. Based on these results, I refer to regime 1 as the "decreasing rates" regime and regime 2 as the "non-decreasing rates" regime. Contrary to previous studies, I do not find that regimes correspond to recessions and expansions. I also find that the variance of the level factor is statistically significantly lower in regime 1. Figure 2.2 plots the spread between the 10-year yield and the 1-month yield with the periods corresponding to the "non-decreasing rates" regime shaded. We can see that the spread is generally higher in that regime.

Figure 2.3 shows the observed and model-implied term structure of average yields conditional on each regime. To construct the observed term structure of average yields, I first assign each month to a regime based on the smoothed probability of the regime for that month. In other words, if $\text{Prob}(s_t = j | \mathcal{F}_T) > 0.5$, I assign month t to regime j. Then for each maturity from $n = 1, \ldots, 120$, I compute the average yield over all months identified as regime j, for each $j = 1, 2$. To construct the model-implied term structure of average yields, I use the model-implied values of the recursion parameters $A_j^{(n)}$ and $B^{(n)}$ and simulate 1000 time series of bond yields of each maturity, with the same length as that of my dataset (215 months). Then I calculate the mean yield for each simulated series of each maturity, conditional on either regime 1 or regime 2, and plot the mean yields averaged over all simulations. We can see that the average observed and model-implied term structures match really well. The average yield curve is flatter in the

"decreasing rates" regime. Moreover, yields are higher on average in that regime. The average yield curve is considerably steeper in the "non-decreasing rates" regime. This is consistent with my finding that the spread between the 10-year and the 1-month yield is generally higher in the "non-decreasing rates" regime.

Figure 2.4 shows expected excess returns conditional on each regime. I find that the expected excess returns on bonds of all maturities are positive in each regime. Moreover, expected excess returns are higher in the "decreasing rates" regime. These results imply that over my sample period, on average it is always profitable for investors to be long in bonds of any maturity, and that they would want to hold larger positions in longer term bonds in the "decreasing rates" regime. Thus, even though the yield curve is flatter in the "decreasing rates" regime, on average investors would be better off holding more long term bonds in that regime because rates are decreasing.

I find significant differences in the market pricing of risk between the two regimes. In particular, the price of risk corresponding to the third (curvature) factor is statistically significantly different between the two regimes. Since there is a statistically significant element in each row of either λ_0 or λ_1, I conclude that level, slope, and curvature risks are all priced. This evidence supports the view that investors require risk compensation for shocks to the level, slope, and curvature of the yield curve.

The loading of level risk on the slope factor is negative and statistically significant. Thus, changes in the slope of the yield curve drive variation in the price of level risk. Figure 2.10 shows the loadings of expected one-month excess holding log returns on each factor. These loadings represent the effect of a one-standard-deviation change in each factor on the expected one-month excess holding log returns. As can be seen in Figure 2.10, expected excess log returns load positively

on the slope factor with coefficients increasing approximately linearly in maturity. Therefore, the model implies that a higher slope (yield spread) is associated with higher expected excess log returns. This is consistent with the finding in Adrian et al. (2013) for a single regime model, as well as with prior evidence on the predictive power of yield spreads for bond returns as, e.g., in Campbell and Shiller (1991). From Figure 2.10, we can see that a one standard deviation increase in the slope factor increases the annualized expected excess log return on the 10-year bond by about 5%. I also find that the loading of slope risk on the curvature factor is statistically significant, which implies that the current curvature factor forecasts future expected excess returns. This is consistent with results in Cochrane and Piazzesi (2008). From Figure 2.10 we can also see that a one standard deviation decline in the curvature factor increases the annualized expected excess log return on the 10-year Treasury bond by about 6%.

Figures 2.5 and 2.6 show observed and model-implied one-month excess holding log returns for maturities 24 months and 120 months, while Figures 2.7 and 2.8 plot the observed and model-implied yields for these two maturities. We see that the model does a very good job of matching the actual yields and returns. Figures 2.7 and 2.8 also plot the term premium.

2.5   Conclusion

In this paper, I study the consequences of changes in regime on bond risk premia, using Treasury bond yield data form 1990 to 2007. The main empirical finding of the paper is that over my sample period, on average it is always profitable for investors to be long in either short term or long term bonds. Moreover, they would want to hold more long term bonds in periods when rates are decreasing, even though the yield curve is flatter on average during those periods. I also find

that a higher yield spread is associated with higher expected excess returns. A one standard deviation increase in the slope factor, which is related to the slope of the term structure, increases the annualized expected excess log return on the 10-year Treasury bond by about 5%. A separate contribution of this paper is to provide an estimation procedure that is considerably simpler than existing methods and avoids most of the numerical difficulties encountered with other methods in the literature.

2.6 Appendix

2.6.1 Relation between $\mathbb{P}$-dynamics and $\mathbb{Q}$-dynamics

By no arbitrage, an asset with payoff $g(X_{t+1})$ has a price in regime j equal to

$$P(X_t) = E_t^{\mathbb{P}}[M_{t,t+1}g(X_{t+1})|s_t = j] = \exp(-r_t^j)E_t^{\mathbb{Q}}[g(X_{t+1})|s_t = j] \qquad (2.30)$$

$$
\begin{aligned}
\pi^{\mathbb{Q}_{jk}} &= E_t^{\mathbb{Q}}\left[\mathbb{1}_{\{s_{t+1}=k\}}|s_t = j\right] = \exp(r_t^j)E_t^{\mathbb{P}}\left[\mathbb{1}_{\{s_{t+1}=k\}}M_{t,t+1}|s_t = j\right] \\
&= \exp(r_t^j)E_t^{\mathbb{P}}\left[\mathbb{1}_{\{s_{t+1}=k\}}\exp\left(-r_t - \Gamma_{s_t,s_{t+1}} - \frac{1}{2}\lambda'_{t,s_t}\lambda_{t,s_t} - \lambda'_{t,s_t}\Sigma_{s_t}^{-1/2}v_{t+1}\right)|s_t = j\right] \\
&= \exp(r_t^j)\exp(-r_t^j)E_t^{\mathbb{P}}\left[\mathbb{1}_{\{s_{t+1}=k\}}\exp\left(-\Gamma_{s_t,s_{t+1}} - \frac{1}{2}\lambda'_{t,s_t}\lambda_{t,s_t} - \lambda'_{t,s_t}\Sigma_{s_t}^{-1/2}v_{t+1}\right)|s_t = j\right] \\
&= E_t^{\mathbb{P}}\left[\exp\left(-\frac{1}{2}\lambda'_{t,s_t}\lambda_{t,s_t} - \lambda'_{t,s_t}\Sigma_{s_t}^{-1/2}v_{t+1}\right)|s_t = j\right]E_t^{\mathbb{P}}\left[\mathbb{1}_{\{s_{t+1}=k\}}\exp\left(-\Gamma_{s_t,s_{t+1}}\right)|s_t = j\right] \\
&= \exp\left(-\frac{1}{2}\lambda'_{t,j}\lambda_{t,j}\right)E_t^{\mathbb{P}}\left[\exp\left(-\lambda'_{t,s_t}\Sigma_{s_t}^{-1/2}v_{t+1}\right)|s_t = j\right]\pi^{\mathbb{P}_{jk}}\exp(-\Gamma_{j,k}) \\
&= \exp\left(-\frac{1}{2}\lambda'_{t,j}\lambda_{t,j}\right)\exp\left(\frac{1}{2}Var_t(-\lambda'_{t,s_t}\Sigma_{s_t}^{-1/2}v_{t+1}|s_t = j)\right)\pi^{\mathbb{P}_{jk}}\exp\left(-\Gamma_{j,k}\right) \\
&= \exp\left(-\frac{1}{2}\lambda'_{t,j}\lambda_{t,j}\right)\exp\left(\frac{1}{2}\lambda'_{t,j}\lambda_{t,j}\right)\pi^{\mathbb{P}_{jk}}\exp\left(-\Gamma_{j,k}\right) = \pi^{\mathbb{P}_{jk}}\exp\left(-\Gamma_{j,k}\right)
\end{aligned}
$$

Therefore,

$$\Gamma_{j,k} = \log\left(\frac{\pi^{\mathbb{P}_{jk}}}{\pi^{\mathbb{Q}_{jk}}}\right) \qquad (2.31)$$

$$P(X_t) = E_t^{\mathbb{P}} \left[M_{t,t+1} g(X_{t+1}) | s_t = j \right]$$

$$= E_t^{\mathbb{P}} \left[\exp \left(-r_t - \Gamma_{s_t,s_{t+1}} - \frac{1}{2} \lambda_{t,s_t}' \lambda_{t,s_t} - \lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp(-r_t^j) \exp \left(-\frac{1}{2} \lambda_{t,j}' \lambda_{t,j} \right) E_t^{\mathbb{P}} \left[\exp \left(-\Gamma_{s_t,s_{t+1}} - \lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp(-r_t^j) \exp \left(-\frac{1}{2} \lambda_{t,j}' \lambda_{t,j} \right) \left[\sum_{k=1}^{2} \pi^{\mathbb{P}_{jk}} \exp(-\Gamma_{jk}) \right] \times$$

$$E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} v_{t+1} \right) g(X_{t+1}) | s_t = j \right] =$$

$$= \exp(-r_t^j) \exp \left(-\frac{1}{2} \lambda_{t,j}' \lambda_{t,j} \right) \left[\sum_{k=1}^{2} \pi^{\mathbb{P}_{jk}} \frac{\pi^{\mathbb{Q}_{jk}}}{\pi^{\mathbb{P}_{jk}}} \right] \times$$

$$E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} v_{t+1} \right) g(X_{t+1}) | s_t = j \right] =$$

$$= \exp(-r_t^j) \exp \left(-\frac{1}{2} \lambda_{t,j}' \lambda_{t,j} \right) \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}_{jk}} \right] E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp(-r_t^j) \exp \left(-\frac{1}{2} \lambda_{t,j}' \lambda_{t,j} \right) E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} v_{t+1} \right) g(X_{t+1}) | s_t = j \right])$$

$$= \exp(-r_t^j) \exp \left(-\frac{1}{2} \lambda_{t,j}' \lambda_{t,j} \right) E_t^{\mathbb{P}} \left[\exp \left(-\lambda_{t,s_t}' \Sigma_{s_t}^{-1/2} \left(X_{t+1} - \mu_{s_t} - \Phi X_t \right) \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp(-r_t^j) \exp \left(-\frac{1}{2} \lambda_{t,j}' \lambda_{t,j} \right) \int g(X_{t+1}) \exp \left(-\lambda_{t,j}' \Sigma_j^{-1/2} (X_{t+1} - \mu_j - \Phi X_t) \right)$$

$$(2\pi)^{-K/2} |\Sigma_j|^{-1/2} \exp \left(-\frac{1}{2} (X_{t+1} - \mu_j - \Phi X_t)' \Sigma_j^{-1} (X_{t+1} - \mu_j - \Phi X_t) \right) dX_{t+1}$$

$$= \exp(-r_t^j)(2\pi)^{-K/2} |\Sigma_j|^{-1/2} \int g(X_{t+1}) \exp \left(-\frac{1}{2} \left[(X_{t+1} - \mu_j - \Phi X_t)' \Sigma_j^{-1} \right. \right.$$

$$\left. \left. (X_{t+1} - \mu_j - \Phi X_t) + 2\lambda_{t,j}' \Sigma_j^{-1/2} (X_{t+1} - \mu_j - \Phi X_t) + \lambda_{t,j}' \lambda_{t,j} \right] \right) dX_{t+1}$$

$$= \exp(-r_t^j)(2\pi)^{-K/2} |\Sigma_j|^{-1/2} \int g(X_{t+1}) \exp \left(-\frac{1}{2} \left[\Sigma_j^{-1/2} (X_{t+1} - \mu_j - \Phi X_t) + \lambda_{t,j} \right]' \right.$$

$$\left. \left[\Sigma_j^{-1/2} (X_{t+1} - \mu_j - \Phi X_t) + \lambda_{t,j} \right] \right) dX_{t+1} =$$

$$= \exp(-r_t^j)(2\pi)^{-K/2} |\Sigma_j|^{-1/2} \int g(X_{t+1}) \exp \left(-\frac{1}{2} \left[\Sigma_j^{-1/2} (X_{t+1} - \mu_j - \Phi X_t + \Sigma_j^{1/2} \lambda_{t,j}) \right]' \right.$$

$$\left. \left[\Sigma_j^{-1/2} (X_{t+1} - \mu_j - \Phi X_t + \Sigma_j^{1/2} \lambda_{t,j}) \right] \right) dX_{t+1}$$

$$= \exp(-r_t^j)(2\pi)^{-K/2}|\Sigma_j|^{-1/2} \int g(X_{t+1}) \exp\left(-\frac{1}{2}\left(X_{t+1} - \mu_j - \Phi X_t + \lambda_{0,j} + \lambda_1 X_t\right)' \Sigma_j^{-1}\right.$$

$$= \left. (X_{t+1} - \mu_j - \Phi X_t + \lambda_{0,j} + \lambda_1 X_t)\right) dX_{t+1}$$

$$= \exp(-r_t^j)(2\pi)^{-K/2}|\Sigma_j|^{-1/2} \int g(X_{t+1}) \exp\left(-\frac{1}{2}\left[X_{t+1} - (\mu_j - \lambda_{0,j}) - (\Phi - \lambda_1)X_t\right]' \Sigma_j^{-1}\right.$$

$$\left. [X_{t+1} - (\mu_j - \lambda_{0,j}) - (\Phi - \lambda_1)X_t]\right) dX_{t+1} = \exp(-r_t^j)E_t^{\mathbb{Q}}(g(X_{t+1})|s_t = j)$$

Therefore, under the $\mathbb{Q}$-measure,

$$X_{t+1}|s_t = j \sim^{\mathbb{Q}} N((\mu_j - \lambda_{0,j}) + (\Phi - \lambda_1)X_t, \Sigma_j) \tag{2.32}$$

or, equivalently,

$$X_{t+1}|s_t = j \sim^{\mathbb{Q}} N(\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}}X_t, \Sigma_j) \tag{2.33}$$

where

$$\mu_j^{\mathbb{Q}} \equiv \mu_j - \lambda_{0,j} \tag{2.34}$$

and

$$\Phi^{\mathbb{Q}} \equiv \Phi - \lambda_1 \tag{2.35}$$

Hence, under the $\mathbb{Q}$-measure, X_{t+1} follows the dynamics

$$X_{t+1} = \mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}}X_t + v_{t+1}^{\mathbb{Q}} \tag{2.36}$$

where $v_{t+1}^{\mathbb{Q}}|s_t = j \sim^{\mathbb{Q}} N(0, \Sigma_j)$ under the $\mathbb{Q}$-measure.

2.6.2   Calculating expected excess returns

The bond price is

$$P_t^{(n)j} = e^{-A_j^{(n)} - B^{(n)'}X_t}$$

$$p_t^{(n)j} = \log P_t^{(n)j} = \log E_t^{\mathbb{Q}} \left[e^{-r_t^j} P_{t+1}^{(n-1)} | s_t = j \right] =$$

$$= \log \left(e^{-r_t^j} \sum_{k=1}^{2} \pi^{\mathbb{Q}jk} E_t^{\mathbb{Q}} \left[P_{t+1}^{(n-1)k} | s_t = j \right] \right)$$

$$= -r_t^j + \log \left(\left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right] E_t^{\mathbb{Q}} \left[e^{-B^{(n-1)'} X_{t+1}} | s_t = j \right] \right)$$

$$= -r_t^j + \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right) + \log E_t^{\mathbb{Q}} \left[e^{-B^{(n-1)'} X_{t+1}} | s_t = j \right]$$

$$= -r_t^j + \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right) + \log E_t^{\mathbb{Q}} \left[e^{-B^{(n-1)'} (\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t + v_{t+1})} | s_t = j \right]$$

$$= -r_t^j + \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right) + \log \left[e^{-B^{(n-1)'} (\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t) + \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}} \right]$$

$$= -r_t^j + \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right) - B^{(n-1)'} (\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t) + \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}$$

$$= -\delta_0^j - \delta_1' X_t + \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right) - B^{(n-1)'} (\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t) +$$

$$\frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}$$

Therefore,

$$p_t^{(n)j} = -A_j^{(n)} - B^{(n)'} X_t = -\delta_0^j - \delta_1' X_t + \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right) - B^{(n-1)'} \mu_j^{\mathbb{Q}} +$$

$$\frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)} - B^{(n-1)'} \Phi^{\mathbb{Q}} X_t$$

The above equation implies the recursions:

$$A_j^{(n)} = \delta_0^j - \log \left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}} \right) + B^{(n-1)'} \mu_j^{\mathbb{Q}} - \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}$$

or equivalently

$$A_j^{(n)} = \delta_0^j - \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right) + B^{(n-1)'}(\mu_j - \lambda_{0,j}) - \frac{1}{2}B^{(n-1)'}\Sigma_j B^{(n-1)} \quad (2.37)$$

and

$$B^{(n)'} = \delta_1' + B^{(n-1)'}\Phi^{\mathbb{Q}} \quad (2.38)$$

or equivalently

$$B^{(n)} = \delta_1 + (\Phi - \lambda_1)'B^{(n-1)} \quad (2.39)$$

The initial conditions are $A_j^{(1)} = \delta_0^j$, $B^{(1)} = \delta_1$.

$$\begin{aligned}
E_t^{\mathbb{P}}\left[p_{t+1}^{(n-1)}|s_t = j\right] &= \sum_{k=1}^{2} \pi^{\mathbb{P}jk} E_t^{\mathbb{P}}\left[p_{t+1}^{(n-1)k}|s_t = j\right] \\
&= \sum_{k=1}^{2} \pi^{\mathbb{P}jk} E_t^{\mathbb{P}}\left[-A_k^{(n-1)} - B^{(n-1)'}X_{t+1}|s_t = j\right] \\
&= -\sum_{k=1}^{2} \pi^{\mathbb{P}jk}\left(A_k^{(n-1)} + B^{(n-1)'}E_t\left[X_{t+1}|s_t = j\right]\right) \\
&= -\sum_{k=1}^{2} \pi^{\mathbb{P}jk}\left(A_k^{(n-1)} + B^{(n-1)'}E_t\left[\mu_{s_t} + \Phi X_t|s_t = j\right]\right) \\
&= -\sum_{k=1}^{2} \pi^{\mathbb{P}jk}\left(A_k^{(n-1)} + B^{(n-1)'}(\mu_j + \Phi X_t)\right) \\
&= -\sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - \left(\sum_{k=1}^{2} \pi^{\mathbb{P}jk}\right)B^{(n-1)'}(\mu_j + \Phi X_t) \\
&= -\sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - B^{(n-1)'}(\mu_j + \Phi X_t)
\end{aligned}$$

$$E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t = j\right] = E_t^{\mathbb{P}}\left[p_{t+1}^{(n-1)} - p_t^{(n)} - r_t|s_t = j\right] = E_t^{\mathbb{P}}\left[p_{t+1}^{(n-1)}|s_t = j\right] - p_t^{(n)j} - r_t^j$$

$$= -\sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - B^{(n-1)'}(\mu_j + \Phi X_t) + \delta_{0,j} + \delta_1' X_t$$

$$- \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right) + B^{(n-1)'}(\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t) - \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}$$

$$- \delta_{0,j} - \delta_1' X_t = -\sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - B^{(n-1)'}(\mu_j - \mu_j^{\mathbb{Q}})$$

$$- B^{(n-1)'}(\Phi - \Phi^{\mathbb{Q}}) X_t - \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right) - \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}$$

$$= -\sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right) - B^{(n-1)'} \lambda_{0,j}$$

$$- B^{(n-1)'} \lambda_1 X_t - \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}$$

$$E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|\mathcal{F}_t\right] = \sum_{j=1}^{2} E_t^{\mathbb{P}}\left[rx_{t+1}^{(n-1)}|s_t = j\right] P(s_t = j|\mathcal{F}_t)$$

$$= \sum_{j=1}^{2}\left[-\sum_{k=1}^{2} \pi^{\mathbb{P}jk} A_k^{(n-1)} - \log\left(\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right) - B^{(n-1)'} \lambda_{0,j}\right.$$

$$\left. - B^{(n-1)'} \lambda_1 X_t - \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}\right] \times P(s_t = j|\mathbb{F}_t)$$

We can also derive an expression for $e^{-r_t^j} \frac{E_t^{\mathbb{P}}[P_{t+1}^{(n-1)}|s_t=j]}{P_t^{(n)j}}$.

$$P_t^{(n)j} = E_t^{\mathbb{Q}}[e^{-r_t^j} P_{t+1}^{(n-1)}|s_t = j] = e^{-r_t^j} \sum_{k=1}^{2} \pi^{\mathbb{Q}jk} E_t^{\mathbb{Q}}[P_{t+1}^{(n-1)k}|s_t = j]$$

$$= e^{-r_t^j} \sum_{k=1}^{2} \pi^{\mathbb{Q}jk} E_t^{\mathbb{Q}}[e^{-A_k^{(n-1)} - B^{(n-1)'} X_{t+1}}|s_t = j]$$

$$= e^{-r_t^j} \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right] E_t^{\mathbb{Q}} \left[e^{-B^{(n-1)'} X_{t+1}}|s_t = j\right]$$

$$= e^{-r_t^j} \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right] E_t^{\mathbb{Q}} \left[e^{-B^{(n-1)'}(\mu_{s_t}^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t + v_{t+1})}|s_t = j\right]$$

$$= e^{-r_t^j} \left[\sum_{k=1}^{2} \pi^{\mathbb{Q}jk} e^{-A_k^{(n-1)}}\right] e^{-B^{(n-1)'}(\mu_j^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t) + \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}} \qquad (2.40)$$

$$E_t^{\mathbb{P}}[P_{t+1}^{(n-1)}|s_t = j] = \sum_{k=1}^{2} \pi^{\mathbb{P}jk} E_t^{\mathbb{P}}[P_{t+1}^{(n-1)k}|s_t = j]$$

$$= \sum_{k=1}^{2} \pi^{\mathbb{P}jk} E_t^{\mathbb{P}}[e^{-A_k^{(n-1)} - B^{(n-1)'} X_{t+1}}|s_t = j]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{-A_k^{(n-1)}}\right] E_t^{\mathbb{P}} \left[e^{-B^{(n-1)'} X_{t+1}}|s_t = j\right]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{-A_k^{(n-1)}}\right] E_t^{\mathbb{P}} \left[e^{-B^{(n-1)'}(\mu_{s_t}^{\mathbb{P}} + \Phi^{\mathbb{P}} X_t + v_{t+1})}|s_t = j\right]$$

$$= \left[\sum_{k=1}^{2} \pi^{\mathbb{P}jk} e^{-A_k^{(n-1)}}\right] e^{-B^{(n-1)'}(\mu_j^{\mathbb{P}} + \Phi^{\mathbb{P}} X_t) + \frac{1}{2} B^{(n-1)'} \Sigma_j B^{(n-1)}} \qquad (2.41)$$

Then

$$e^{-r_t^j}\frac{E_t^{\mathbb{P}}[P_{t+1}^{(n-1)}|s_t=j]}{P_t^{(n)j}} =$$

$$\frac{e^{-r_t^j}\left[\sum_{k=1}^2 \pi^{\mathbb{P}_{jk}}e^{-A_k^{(n-1)}}\right]e^{-B^{(n-1)'}(\mu_j^{\mathbb{P}}+\Phi^{\mathbb{P}}X_t)+\frac{1}{2}B^{(n-1)'}\Sigma_j B^{(n-1)}}}{e^{-r_t^j}\left[\sum_{k=1}^2 \pi^{\mathbb{Q}_{jk}}e^{-A_k^{(n-1)}}\right]e^{-B^{(n-1)'}(\mu_j^{\mathbb{Q}}+\Phi^{\mathbb{Q}}X_t)+\frac{1}{2}B^{(n-1)'}\Sigma_j B^{(n-1)}}}$$

$$=\frac{\left[\sum_{k=1}^2 \pi^{\mathbb{P}_{jk}}e^{-A_k^{(n-1)}}\right]}{\left[\sum_{k=1}^2 \pi^{\mathbb{Q}_{jk}}e^{-A_k^{(n-1)}}\right]}e^{-B^{(n-1)'}(\mu_j^{\mathbb{P}}-\mu_j^{\mathbb{Q}})-B^{(n-1)'}(\Phi^{\mathbb{P}}-\Phi^{\mathbb{Q}})X_t}$$

$$=\frac{\left[\sum_{k=1}^2 \pi^{\mathbb{P}_{jk}}e^{-A_k^{(n-1)}}\right]}{\left[\sum_{k=1}^2 \pi^{\mathbb{Q}_{jk}}e^{-A_k^{(n-1)}}\right]}e^{-B^{(n-1)'}(\lambda_{0,j}+\lambda_1 X_t)}$$

$$=\frac{\left[\sum_{k=1}^2 \pi^{\mathbb{P}_{jk}}e^{-A_k^{(n-1)}}\right]}{\left[\sum_{k=1}^2 \pi^{\mathbb{Q}_{jk}}e^{-A_k^{(n-1)}}\right]}e^{-B^{(n-1)'}\Sigma_j^{1/2}\lambda_{t,j}} \tag{2.42}$$

2.7 Figures
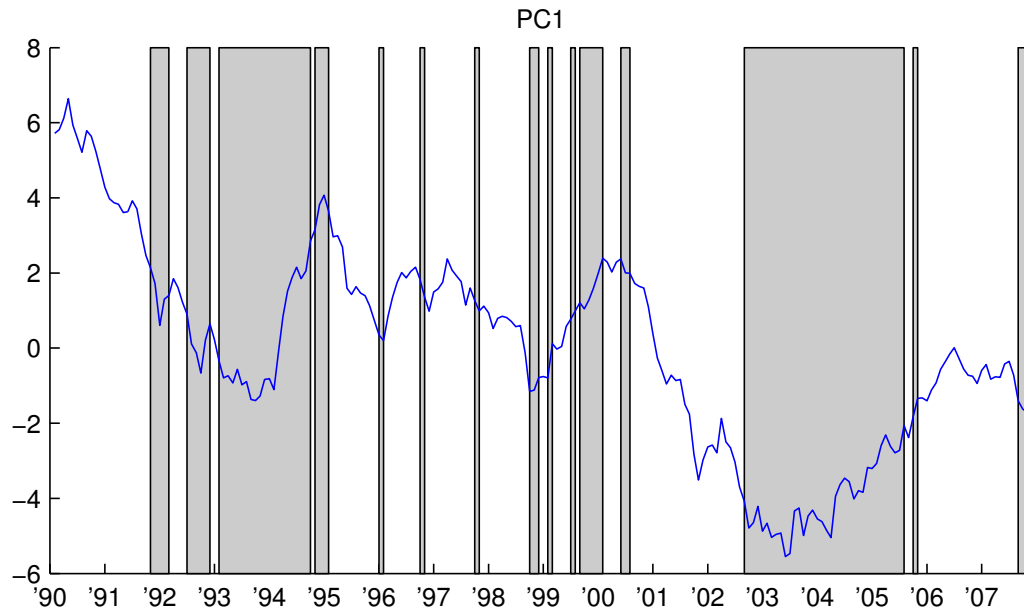
Figure 2.1. First principal component of bond yields
Shown in blue. Shaded areas represent regime 2 (non-decreasing rates).
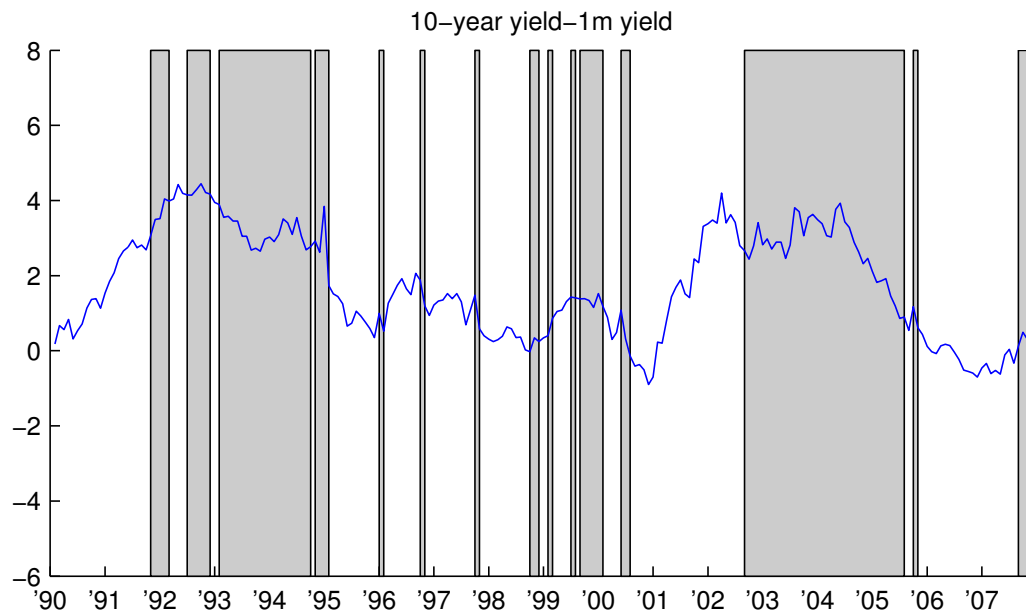
Figure 2.2. Spread between the 10-year yield and the 1-month yield
Shown in blue. Units are annualized percentage. Shaded areas represent regime 2
(non-decreasing rates).
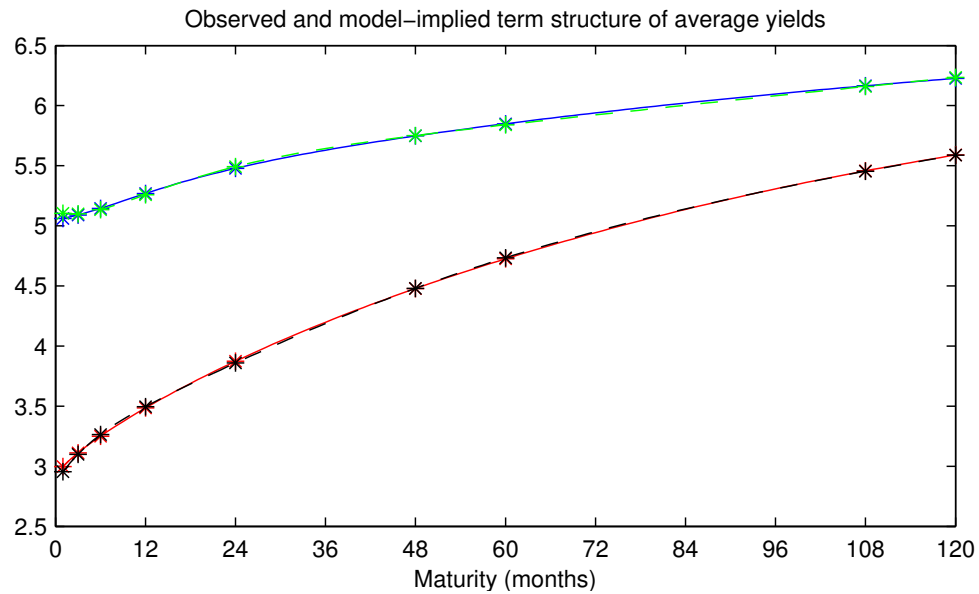
Figure 2.3. Observed and model-implied term structure of average yields conditional on each regime

The term structures of observed mean yields conditional on each regime are shown in solid lines in blue for regime 1 (decreasing rates) and red for regime 2 (non-decreasing rates). They are obtained by computing sample means after allocating dates to regimes based on the smoothed probabilities. The term structures of average simulated yields conditional on each regime are shown in dotted lines in green for regime 1 and black for regime 2. They are computed by taking the average of the means of time series of yields simulated from the model.
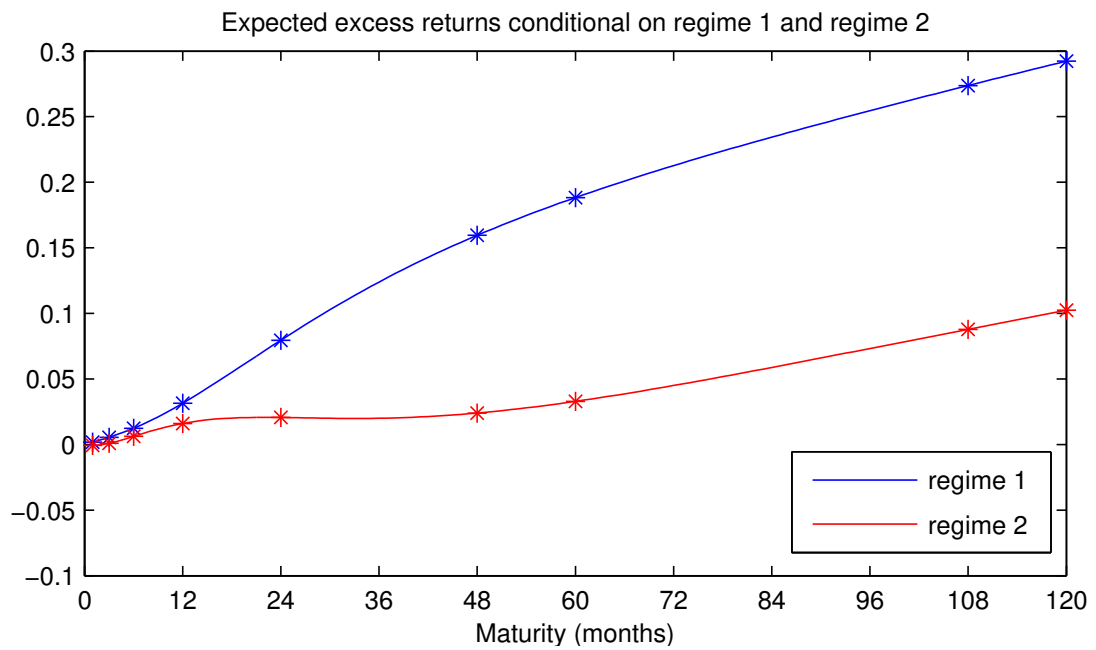
Figure 2.4. Monthly expected excess one-month holding returns (%) conditional on regime 1 (decreasing rates) and regime 2 (non-decreasing rates)
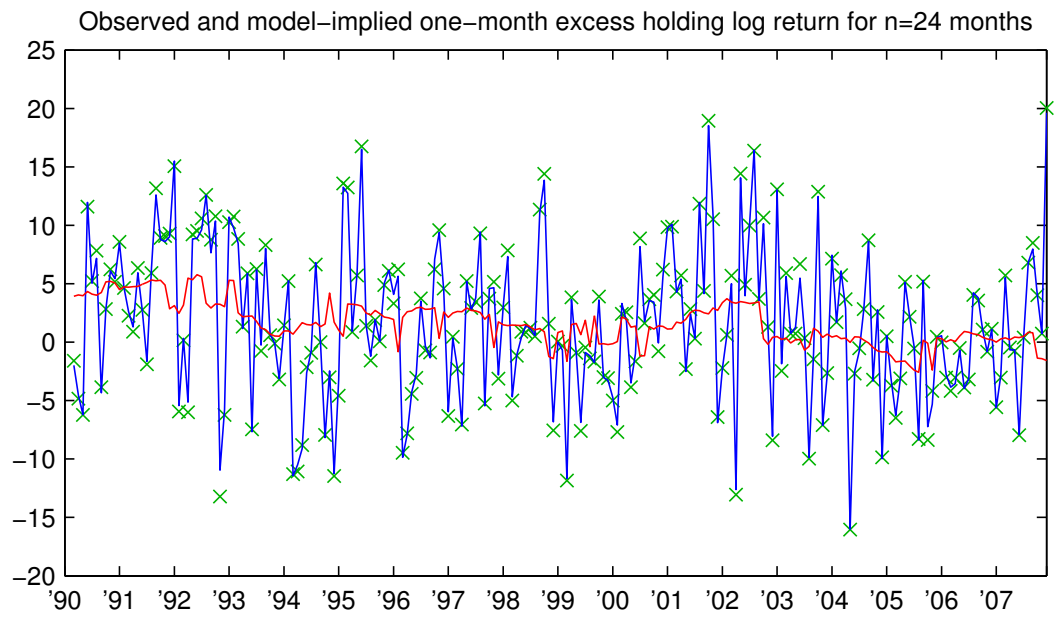
Figure 2.5. Observed and model-implied one-month excess holding log return for the 24-month bond
Model-implied return is in green, observed return is in blue, expected return is in red. Returns are in annualized percentages.
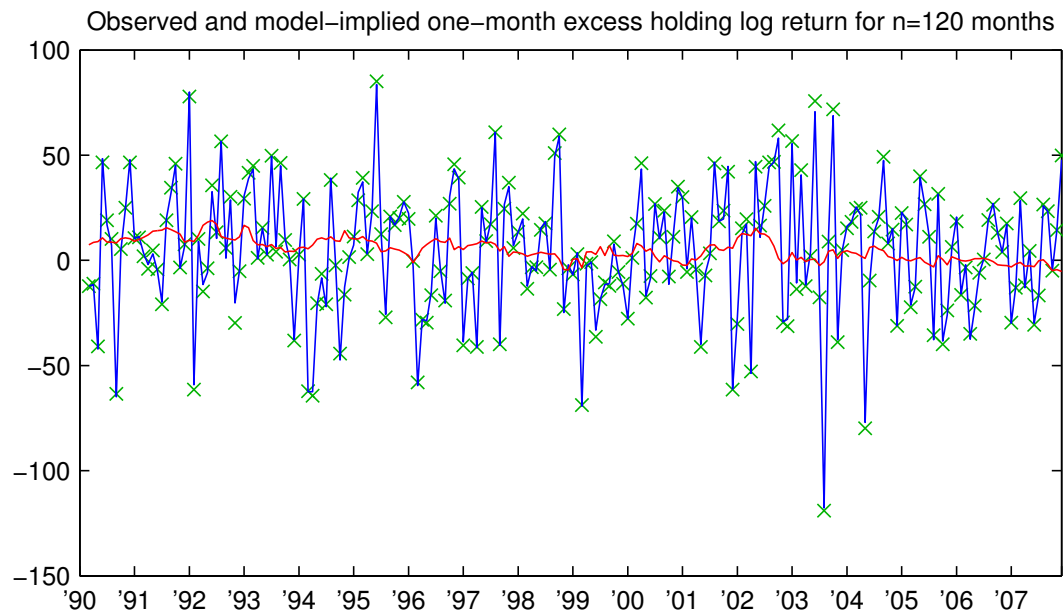
Figure 2.6. Observed and model-implied one-month excess holding log return for the 120-month bond

Model-implied return is in green, observed return is in blue, expected return is in red. Returns are in annualized percentages.
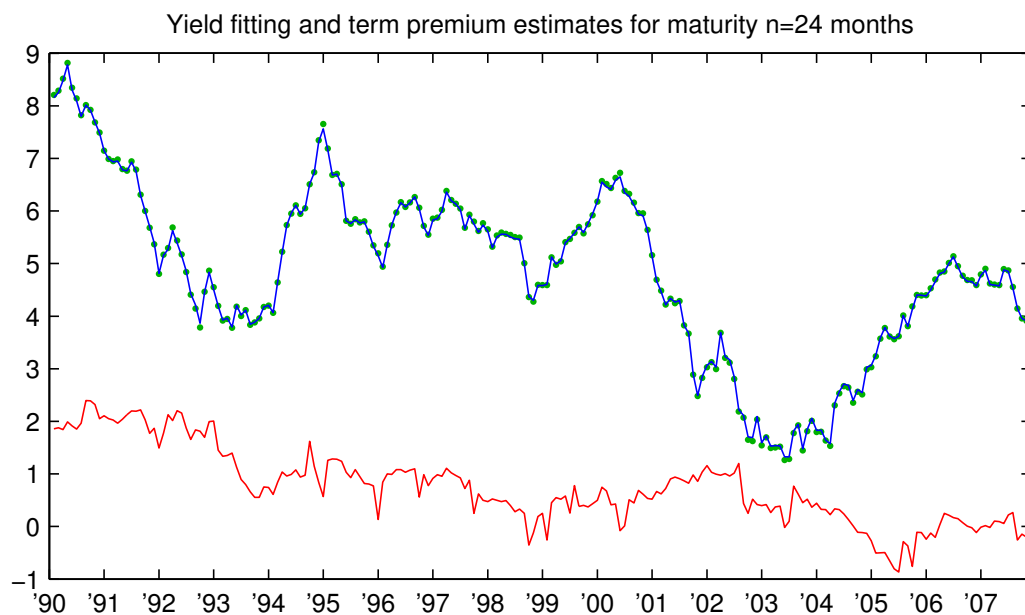
Figure 2.7. Yield fitting and term premium estimates for the 24-month bond Model-implied (fitted) yield is in green, observed yield is in blue, and term premium is in red, in annualized percentage terms.
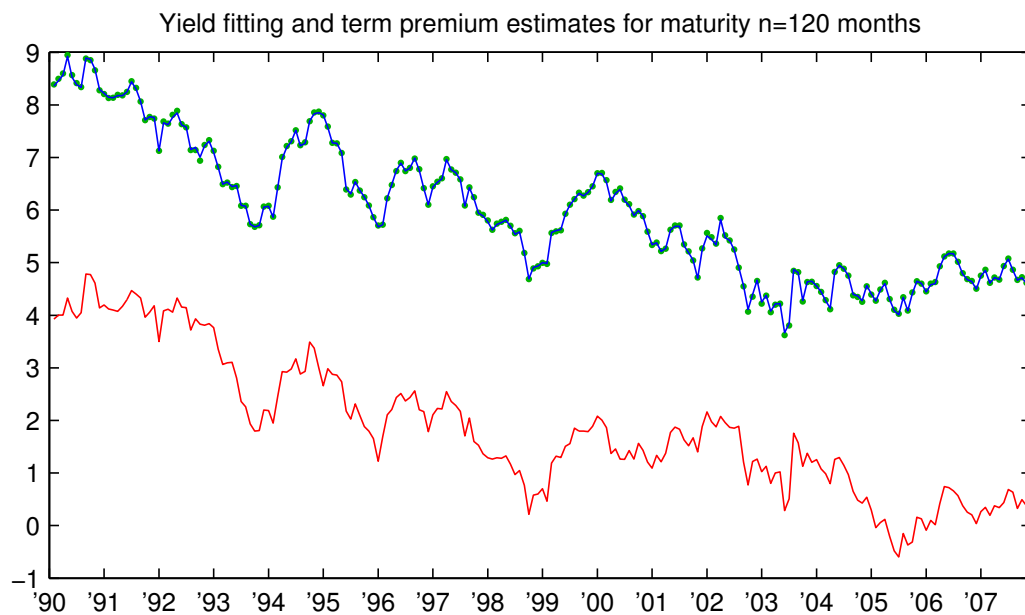
Figure 2.8. Yield fitting and term premium estimates for the 120-month bond Model-implied (fitted) yield is in green, observed yield is in blue, and term premium is in red, in annualized percentage terms.
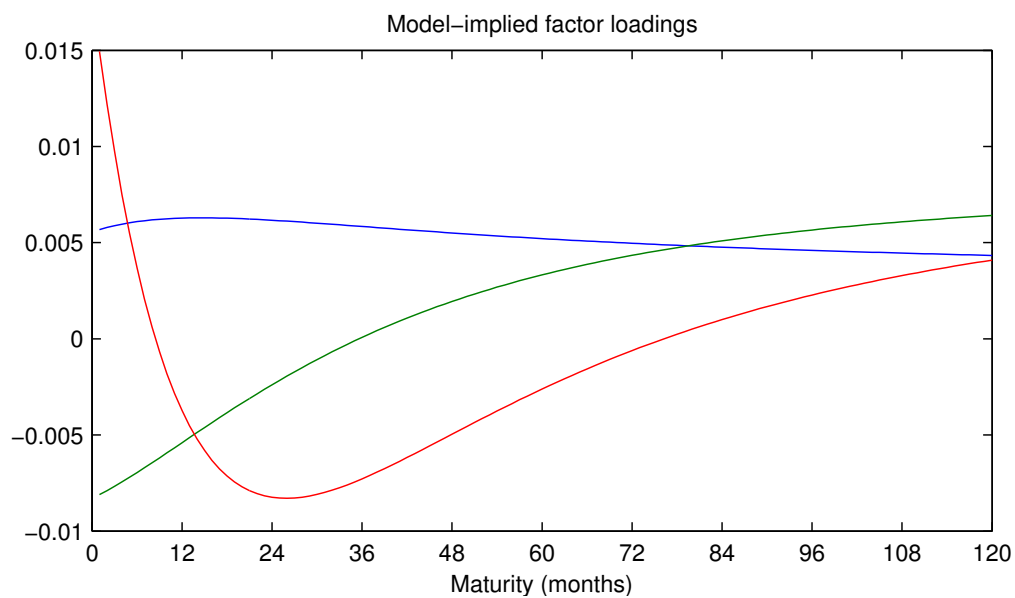
Figure 2.9. Model-implied factor loadings

The figure plots the model-implied loadings of annualized yields on the factors. Each line represents the response of the term structure to a one standard deviation shock in the given factor. The graph shows loadings for the level factor (first principal component, in blue), the slope factor (second principal component, in green), and the curvature factor (third principal component, in red).
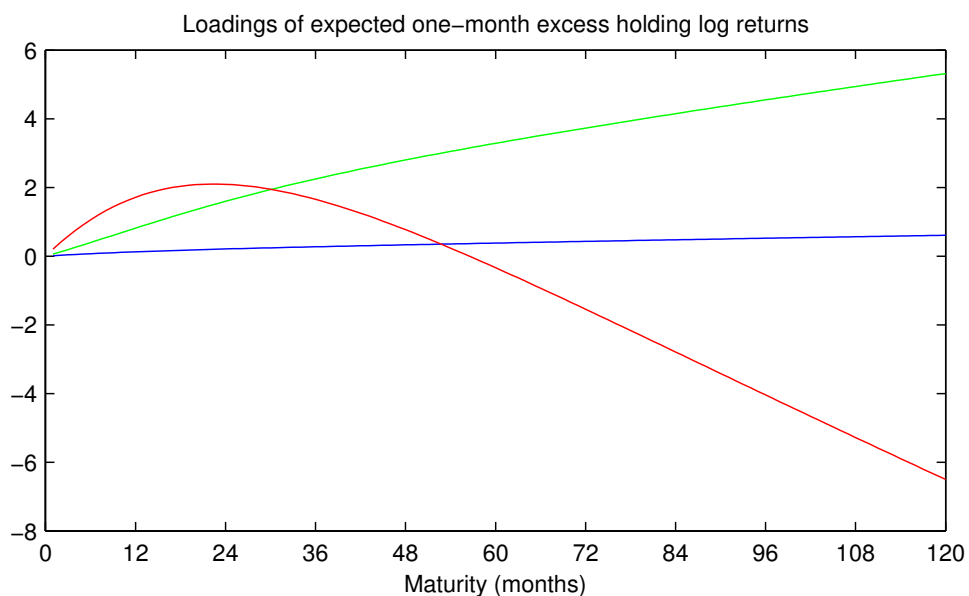
Figure 2.10. Loadings of expected one-month excess holding log returns on each factor

Sensitivity of expected one-month excess holding log returns to one-standard-deviation changes in the factors. The figure shows loadings for the level factor (first principal component, in blue), the slope factor (second principal component, in green), and the curvature factor (third principal component, in red). The loadings are for returns in annualized percentage terms.

2.8 Tables

Table 2.1. First stage reduced-form parameter estimates

	Regime 1				Regime 2
μ	-0.1081				0.0616
	(0.0369)				(0.0561)
	0.0143				-0.0277
	(0.0174)				(0.0255)
	0.0018				-0.0023
	(0.0076)				(0.0113)
Φ		0.9943	-0.0460	-0.3479	
		(0.0122)	(0.0328)	(0.1768)	
		-0.0064	0.9872	0.4102	
		(0.0056)	(0.0154)	(0.0815)	
		0.0028	-0.0018	0.8492	
		(0.0025)	(0.0067)	(0.0361)	
$A^{(1)}$	0.0036				0.0034
	(9.10E-06)				(1.12E-05)
$A^{(3)}$	0.0108				0.0106
	(4.82E-06)				(5.96E-06)
$A^{(6)}$	0.0218				0.0220
	(9.56E-06)				(1.18E-05)
$A^{(12)}$	0.0452				0.0460
	(1.93E-05)				(2.38E-05)
$A^{(24)}$	0.0965				0.0967
	(3.83E-05)				(4.69E-05)
$A^{(48)}$	0.2101				0.2079
	(7.69E-05)				(9.41E-05)
$A^{(60)}$	0.2706				0.2680
	(9.64E-05)				(0.0001)
$A^{(108)}$	0.5281				0.5296
	(0.0002)				(0.0002)
$A^{(120)}$	0.5951				0.5983
	(0.0002)				(0.0002)

Table 2.1. First stage reduced-form parameter estimates (continued)

	Regime 1			Regime 2
$B^{(1)}$	0.0005	-0.0006	0.0013	
	(2.65E-06)	(7.75E-06)	(4.42E-05)	
$B^{(3)}$	0.0015	-0.0019	0.0024	
	(1.40E-06)	(4.13E-06)	(2.37E-05)	
$B^{(6)}$	0.0031	-0.0035	0.0018	
	(2.79E-06)	(8.15E-06)	(4.64E-05)	
$B^{(12)}$	0.0063	-0.0055	-0.0034	
	(5.66E-06)	(1.66E-05)	(9.54E-05)	
$B^{(24)}$	0.0122	-0.0046	-0.0158	
	(1.11E-05)	(3.24E-05)	(0.0002)	
$B^{(48)}$	0.0220	0.0079	-0.0221	
	(2.24E-05)	(6.54E-05)	(0.0004)	
$B^{(60)}$	0.0262	0.0165	-0.0164	
	(2.80E-05)	(8.17E-05)	(0.0005)	
$B^{(108)}$	0.0402	0.0546	0.0315	
	(5.00E-05)	(0.0001)	(0.0008)	
$B^{(120)}$	0.0432	0.0640	0.0457	
	(5.57E-05)	(0.0002)	(0.0009)	

Table 2.1. First stage reduced-form parameter estimates (continued)

	Regime 1			Regime 2		
Σ	0.1398	0.0390	-0.0182	0.2191	0.0661	-0.0285
	(0.0177)	(0.0069)	(0.0030)	(0.0336)	(0.0129)	(0.0057)
	0.0390	0.0319	-0.0070	0.0661	0.0451	-0.0064
		(0.0040)	(0.0014)		(0.0070)	(0.0023)
	-0.0182	-0.0070	0.0060	-0.0285	-0.0064	0.0090
			(0.0008)			(0.0014)
ω^2	0.0005					
	(1.61E-05)					
$[\omega^{(1)}]^2$	0.0153					
	(0.0015)					
$\pi^{\mathbb{P}}$	0.8831			0.1657		
	(0.0286)					
	0.1169			0.8343		
					(0.0406)	
ρ	1.0000			0		

logL=4487.02

Table 2.2. Second stage reduced-form parameter estimates

	Regime 1			Regime 2		
μ	-0.1089			0.0603		
	(0.0369)			(0.0561)		
	0.0150			-0.0268		
	(0.0174)			(0.0255)		
	-0.0006			-0.0029		
	(0.0076)			(0.0113)		
Φ	0.9896	-0.0425	-0.3474			
	(0.0122)	(0.0328)	(0.1768)			
	-0.0076	0.9852	0.4107			
	(0.0056)	(0.0154)	(0.0815)			
	0.0054	-0.0047	0.8494			
	(0.0025)	(0.0067)	(0.0361)			
δ_0	0.0036			0.0034		
	(3.86E-06)			(4.74E-06)		
δ_1	0.0005	-0.0007	0.0012			
	(6.44E-07)	(1.72E-06)	(1.25E-05)			
Σ	0.1398	0.0393	-0.0178	0.2191	0.0663	-0.0285
	(0.0177)			(0.0335)		
	0.0393	0.0320	-0.0067	0.0663	0.0450	-0.0065
	(0.0068)	(0.0040)		(0.0129)	(0.0069)	
	-0.0178	-0.0067	0.0062	-0.0285	-0.0065	0.0090
	(0.0030)	(0.0014)	(0.0008)	(0.0057)	(0.0023)	(0.0014)

Table 2.2. Second stage reduced-form parameter estimates (continued)

	Regime 1		Regime 2
ω^2	0.0005		
	(1.61E-05)		
$[\omega^{(1)}]^2$	0.0153		
	(0.0015)		
$\pi^{\mathbb{P}}$	0.8831	0.1657	
	(0.0286)		
	0.1169	0.8343	
		(0.0406)	

Table 2.3. Second stage structural parameter estimates

	Regime 1			Regime 2
λ_0	-0.1632			-0.0452
	(0.0369)			(0.0561)
	0.0081			0.0162
	(0.0174)			(0.0256)
	0.0148			-0.0338
	(0.0076)			(0.0114)
λ_1		-0.0148	-0.1060	-0.0982
		(0.0122)	(0.0328)	(0.1767)
		0.0031	0.0021	0.1599
		(0.0056)	(0.0154)	(0.0815)
		-0.0015	0.0008	-0.0141
		(0.0025)	(0.0067)	(0.0362)
$\pi^{\mathbb{Q}}$		0.9109		0.0984
		(0.0056)		
		0.0891		0.9016
				(0.0070)

Table 2.4. Wald t-statistics

$\mathbb{H}_0$	Wald t-statistic
Reduced-form parameters (first stage)	
$A_1^{(1)} - A_2^{(1)} = 0$	14.6751
$A_1^{(3)} - A_2^{(3)} = 0$	25.3251
$A_1^{(6)} - A_2^{(6)} = 0$	-14.0303
$A_1^{(12)} - A_2^{(12)} = 0$	-27.6478
$A_1^{(24)} - A_2^{(24)} = 0$	-4.5481
$A_1^{(48)} - A_2^{(48)} = 0$	17.9280
$A_1^{(60)} - A_2^{(60)} = 0$	16.8128
$A_1^{(108)} - A_2^{(108)} = 0$	-5.4761
$A_1^{(120)} - A_2^{(120)} = 0$	-10.7335
Reduced-form parameters (second stage)	
$\mu_{1,1} - \mu_{1,2} = 0$	-2.3216
	1.2466
	0.1509
$\Sigma_{11}^1 - \Sigma_{11}^2 = 0$	-2.0819
$\Sigma_{12}^1 - \Sigma_{12}^2 = 0$	-1.8370
$\Sigma_{22}^1 - \Sigma_{22}^2 = 0$	-1.6235
$\Sigma_{13}^1 - \Sigma_{13}^2 = 0$	1.6466
$\Sigma_{23}^1 - \Sigma_{23}^2 = 0$	-0.0441
$\Sigma_{33}^1 - \Sigma_{33}^2 = 0$	-1.7614
$\delta_{0,1} - \delta_{0,2} = 0$	30.0960
Structural parameters (second stage)	
$\lambda_{0,1} - \lambda_{0,2} = 0$	-1.6182
	-0.2417
	3.2822
$\pi^{\mathbb{P}_{11}} - \pi^{\mathbb{Q}_{11}} = 0$	-0.9554
$\pi^{\mathbb{P}_{22}} - \pi^{\mathbb{Q}_{22}} = 0$	-1.6321

Table 2.5. Reduced-form versus model implied values for the recursion parameters $A_j^{(n)}$

	Regime 1		Regime 2	
	First stage	Model	First stage	Model
$A^{(1)}$	0.0036	0.0036	0.0034	0.0034
$A^{(3)}$	0.0108	0.0108	0.0106	0.0106
$A^{(6)}$	0.0218	0.0218	0.0220	0.0220
$A^{(12)}$	0.0452	0.0451	0.0460	0.0460
$A^{(24)}$	0.0965	0.0965	0.0967	0.0967
$A^{(48)}$	0.2101	0.2102	0.2079	0.2081
$A^{(60)}$	0.2706	0.2703	0.2680	0.2682
$A^{(108)}$	0.5281	0.5276	0.5296	0.5291
$A^{(120)}$	0.5951	0.5958	0.5983	0.5983

Table 2.6. Reduced-form versus model implied values for the recursion parameters $B^{(n)}$

	First stage			Model implied		
$B^{(1)'}$	0.0005	-0.0006	0.0013	0.0005	-0.0007	0.0012
$B^{(3)'}$	0.0015	-0.0019	0.0024	0.0015	-0.0019	0.0024
$B^{(6)'}$	0.0031	-0.0035	0.0018	0.0030	-0.0035	0.0018
$B^{(12)'}$	0.0063	-0.0055	-0.0034	0.0063	-0.0054	-0.0037
$B^{(24)'}$	0.0122	-0.0046	-0.0158	0.0123	-0.0048	-0.0165
$B^{(48)'}$	0.0220	0.0079	-0.0221	0.0220	0.0078	-0.0198
$B^{(60)'}$	0.0262	0.0165	-0.0164	0.0261	0.0166	-0.0131
$B^{(108)'}$	0.0402	0.0546	0.0315	0.0401	0.0547	0.0296
$B^{(120)'}$	0.0432	0.0640	0.0457	0.0433	0.0641	0.0409

2.9   Acknowledgement

Chapter 2, in part, is currently being prepared for submission for publication of the material. Zhecheva, Irina Y. The dissertation author was the primary investigator and author of this material.

2.10   Bibliography

Adrian, T., Crump, R., and Moench, E. (2013) "Pricing the Term Structure with Linear Regressions." *Journal of Financial Economics* 110(1), 110-138.

Ang, A. and Bekaert, G. (2002) "Regime Switches in Interest Rates." *Journal of Business & Economic Statistics*, American Statistical Association, vol. 20(2), 163-82.

Ang, A. and Piazzesi, M. (2003) "A No-arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics* 50, 745-787.

Bansal, R. and Zhou, H. (2002) "Term Structure of Interest Rates with Regime Shifts." *Journal of Finance*, 57, 1997-2043.

Campbell, J. Y. and Shiller, R. J. (1991) "Yield Spreads and Interest Rate Movements: A Bird's Eye View." *Review of Economic Studies* 58, 495-514.

Cochrane, J. and Piazzesi, M. (2008) "Decomposing the yield curve." Working paper, University of Chicago, Chicago, IL.

Dai, Q., Singleton, K., and Yang, W. (2007) "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields." *Review of Financial Studies*, Vol. 20, 1669-1706.

Duffee, G. R. (2002) "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57, 405-443.

Duffee, G. R. (2009) "Forecasting with the Term Structure: the Role of No-Arbitrage," Working Paper June, Haas School of Business, University of California, Berkeley.

Duffee, G. R. and Stanton, R. H. (2008) "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation." *Journal of Financial Econometrics*, 6 (1), 108-142.

Gray, S. (1996) "Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process." *Journal of Financial Economics* 42:27-62.

Gurkaynak, R. S., Sack, B., and Wright, J. H. (2007) "The US treasury yield curve: 1961 to the present." *Journal of Monetary Economics* 54, 2291-2304.

Hamilton, J. D. (1988) ""Rational Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates." *Journal of Economic Dynamics and Control*.

Hamilton, J. D. and Wu, J. C. (2012) "Identification and Estimation of Gaussian Affine Term Structure Models." *Journal of Econometrics*, 168, no. 2: 315-331.

Hamilton, J. D. (2016) "Macroeconomic Regimes and Regime Shifts." *Handbook of Macroeconomics*, Vol. 2A., 163-201, edited by Harald Uhlig and John Taylor, Amsterdam: Elsevier.

Kim, D. H. (2008) "Challenges in macro-finance modeling." BIS Working Paper No. 240, FEDS Working Paper No. 2008-06.

Kim, D. H. and Orphanides, A. (2005) "Term structure estimation with survey data on interest rate forecasts." *Federal Reserve Board, Finance and Economics Discussion Series 2005-48*.

Lemke, W. (2006) "Term Structure Modeling and Estimation in a State Space Framework," no. 565 in *Lecture Notes in Economics and Mathematical Systems*, Heidelberg: Springer.

Hamilton, J. D. (1988) "Rational Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates." *Journal of Economic Dynamics and Control*.

Wu, J. C. and Xia, F. D. (2016) "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound." *Journal of Money, Credit and Banking* 48, 2-3, 253-291.

Chapter 3

Forecasting the Commodity Price Index: Does Aggregating Forecasts of Sub-components Improve Forecast Accuracy?

Abstract

Forecasting commodity prices is important for policymakers, investors, and the general public. Commodity prices have exhibited big fluctuations over the past few years, and it is of wide interest to find out which method produces the best forecasts. In this paper I address the question of what is the best way to forecast the IMF aggregate commodity price index. I investigate whether it is better to use the individual components (indirect approach), or just forecast the aggregate directly (direct approach). I evaluate the forecasting performance of exponential smoothing, autoregressive, and vector autoregressive models. I find that aggregating component forecasts increases forecasting accuracy when using double exponential smoothing, triple exponential smoothing, or an AR(p) model. For the AR(1) model and the simple exponential smoothing method, aggregating sub-index forecasts does not help forecast the price index. However, whether the direct or the indirect forecasting approach is used, none of the models considered generate more accurate

out-of-sample forecasts than those based on an AR(1) model specification.

3.1 Introduction

Predicting the movements of commodity prices is important for policymakers, financial investors, and the general public. Commodity prices have exhibited big fluctuations over the past decade, and it is of wide interest to find out which method produces the best forecasts. Another reason why predicting commodity prices is of interest is that commodity prices may affect current and future inflation. There is literature examining the usefulness of commodity prices in predicting inflation, e.g. Chen et al. (2014) and Ishimwe (2015). In this paper I evaluate the forecasting performance of exponential smoothing, autoregressive, and vector autoregressive models. I study whether the forecasting accuracy of forecasting the aggregate commodity price index can be improved by aggregating forecasts of sub-indices as opposed to forecasting the aggregate index directly. To my knowledge, this is the first paper looking at this aggregation issue for a major aggregate commodity price index.

Many macroeconomic variables are aggregates which can be expressed as a sum of components. Some examples are inflation, GDP, industrial production, and unemployment. When forecasting such variables, one can either forecast the aggregate directly or forecast the disaggregate components and then combine the disaggregate forecasts. The forecast aggregation issue is of considerable practical importance but there are conflicting results in the empirical literature regarding the benefits of forecast aggregation. Theoretical results suggest that generally forecast aggregation should tend to improve forecast accuracy. However, depending on the dataset and model considered, some researchers have found that aggregating component forecasts improves forecast accuracy whereas others find that forecast

aggregation does not help to improve the forecast of the aggregate. Thus, it is not clear whether in a particular application forecast aggregation would be helpful.

The aggregation issue is often encountered when forecasting inflation since prices of the component series from which the aggregate price index is constructed are available. The performance of aggregate versus disaggregate forecasting strategies has been studied in the context of Euro area inflation. There are multiple studies investigating whether it is better to forecast Euro area aggregates directly or to aggregate forecasts of the member states. Hubrich (2005) studies whether aggregating forecasts of sub-indices of the Harmonized Index of Consumer Prices (HICP) improves forecast accuracy relative to forecasting the aggregate HICP directly. She uses a range of univariate and multivariate linear time series models such as AR, VAR, and random walk with drift. Hubrich finds that aggregating forecasts by component does not necessarily help forecast year-on-year inflation twelve months ahead. There are similar studies about U.S. inflation. Zellner and Tobias (2000) study whether disaggregation is useful in forecasting median GDP growth in industrialized countries. Fair and Shiller (1990) use disaggregated components for forecasting US real GNP growth.

In this paper, I compare the forecasting accuracy of simple exponential smoothing, double exponential smoothing, triple exponential smoothing, AR(p), AR(1), and VAR(1) models when forecasting the log of the aggregate commodity index by using sub-index forecasts versus forecasting the log of the aggregate commodity index directly.

3.2   Data

I use data for the IMF aggregate global primary commodity price index and its 52 sub-indices corresponding to 52 primary commodities. Among the primary

commodities, non-fuel, crude oil petroleum, natural gas, energy, food, metals, beverages, agricultural, and industrial inputs are included. The IMF commodity price index is a world export-earnings-weighted price index. The IMF commodity price series are benchmark prices which are representative of the global market. They are determined by the largest exporter of a given commodity. The IMF commodity prices are end-of-month prices measured at close and denominated in U.S. dollars. The data for the aggregate index as well as its component sub-indices is available on the IMF website.[1] The weights used to calculate the aggregate index from individual sub-index prices can be found in Table 2 "Indices of Market Prices for Non-Fuel and Fuel Commodities" and Table 6 "Specifications for Commodity prices" on the IMF website. The weights are based on 2002-2004 average world export earnings. The IMF provides a single set of weights which can be used to calculate the aggregate index from its components for all the historical data. These are the weights I use to aggregate the individual sub-index forecasts into a combined forecast for the aggregate index. In other words, I use constant weights throughout the estimation samples.

I use monthly data from January 1992 to February 2017. The data I use starts from January 1992 since that is the earliest date for which data for all 52 sub-indices is available. The 52-item breakdown is the most detailed breakdown available for the aggregate commodity index.

Figure 3.1 shows the monthly aggregate IMF commodity index, measured in nominal US dollar terms. Figure 3.2 presents a plot of the commodity price sub-indices as well as the aggregate index. There is noticeable comovement of the individual sub-indices with the aggregate index. Prices were relatively stable until 2006. Then they rose sharply until mid-2008 and fell sharply (except textiles)

---

[1]http://www.imf.org/external/np/res/commod/index.aspx

during the financial crisis. Commodity prices started rising again after March 2009 until mid-2011. After that they remained relatively stable until 2015. Then they declined throughout 2015 and started rising again in 2016.

3.3 Empirical framework and methodology

Let P_t^{aggr} represent the level of the aggregate commodity index. The aggregate commodity index can be considered as a contemporaneously aggregated variable that can be represented as the weighted sum of the prices P_t^j of its N disaggregated sub-indices at each time t, where $N = 52$. The aggregate can be expressed as

$$P_t^{aggr} = \sum_{j=1}^{N} w_j P_t^j \tag{3.1}$$

for $t = 1, \ldots, T$, where w_j, $j = 1, \ldots, N$ are the aggregation weights. I assume that the aggregation weights are constant over time, which is indeed the case for the IMF commodity index data. Each weight $w_j > 0$ and $\sum_{j=1}^{N} w_j = 1$. The variable I am interested in forecasting is the log of the monthly aggregate commodity index, p_t^{aggr}:

$$p_t^{aggr} \equiv \log P_t^{aggr} \tag{3.2}$$

The direct forecast of p_t^{aggr} is obtained by forecasting p_t^{aggr} directly and is denoted $\hat{p}_t^{aggr}$.

An indirect forecast $\hat{P}_{sub,t}^{aggr}$ of the level of the aggregate index can be obtained by aggregating the N sub-index forecasts $\hat{P}_t^j$ $(j = 1, \ldots, N)$:

$$\hat{P}_{sub,t}^{aggr} = \sum_{j=1}^{N} w_j \hat{P}_t^j \tag{3.3}$$

I am interested in an indirect forecast $\hat{p}_{sub,t}^{aggr}$ of the log of the aggregate index, which

can be obtained as

$$\hat{p}_{sub,t}^{aggr} = \log\left(\hat{P}_{sub,t}^{aggr}\right) = \log\left(\sum_{j=1}^{N} w_j \hat{P}_t^j\right) \qquad (3.4)$$

For the approach using sub-index data (the indirect approach), for a given model, I fit the model to the log of each sub-index.[2] Then I calculate forecasts $\hat{p}_t^j$ for the logs of the sub-indices. Next, I construct forecasts $\hat{P}_t^j$ for the levels of the sub-indices by exponentiating the forecasts of the logs:

$$\hat{P}_t^j = \exp(\hat{p}_t^j) \qquad (3.5)$$

I use the weights w_j to generate a combined (indirect) level forecast $\hat{P}_{sub,t}^{aggr}$ for the aggregate index as the weighted average of the level forecasts $\hat{P}_t^j$ of the sub-indices, as in equation (3.3). Then I compute an indirect forecast $\hat{p}_{sub,t}^{aggr}$ for the log of the aggregate index by taking the log of the combined level forecast $\hat{P}_{sub,t}^{aggr}$ for the aggregate index, as in equation (3.4). Finally, I compute RMSFE of the indirect forecast $\hat{p}_{sub,t}^{aggr}$ for the log of the aggregate index over all the recursive estimation windows.

For the direct approach, I estimate the given model on the log of the aggregate index, p_t^{aggr},[3] and calculate RMSFE for the forecast $\hat{p}_t^{aggr}$ of the log of the aggregate index.

For each model, I construct one set of forecasts by using sub-index data and another set of forecasts by using the direct approach. Then I compare the forecast accuracy of the two approaches.

I use data from January 1992 to December 1997 as the initial estimation

---

[2]With the exception of the AR(1) model, which is fit to the log difference.
[3]With the exception of the AR(1) model, which is fit to the log difference.

sample. For each model, I generate recursive pseudo out-of-sample forecasts with an expanding estimation window. Using data up to time T (where initially T corresponds to December 1997), I estimate each model and compute h-step ahead forecasts (for time $T+h$) of the log of the aggregate commodity index using the direct approach and using the indirect approach, for each horizon $h = 1, \ldots, 12$ months. Next, I expand the estimation sample by adding the next monthly observation. Then given data up to time $T + 1$, I re-estimate the model and calculate h-step ahead forecasts for $h = 1, \ldots, 12$ (i.e. for time $T+1+h$). I repeat this process up to the last estimation sample, which ends in February 2016. I evaluate out-of-sample forecast accuracy through Root Mean Square Forecast Error (RMSFE).

I denote by $\hat{p}^{aggr}_{T+h|T}$ the direct h-step ahead forecast of the log of the aggregate index given information up to time T, and by $\hat{p}^{aggr}_{sub,T+h|T}$ the indirect h-step ahead forecast of the log of the aggregate index.

I evaluate the forecasting performance of simple exponential smoothing, double exponential smoothing, triple exponential smoothing, an AR(1) model, an AR(p) model, and a VAR(1) model.

3.3.1 Simple exponential smoothing

Simple exponential smoothing models the level component l_t of a series.[4] The simple exponential smoothing method can be expressed in terms of a forecast equation and a smoothing equation for the level called level equation. For a commodity sub-index log price p^j_t, simple exponential smoothing can be written as

$$\text{Forecast equation: } \hat{p}^j_{t+h|t} = l_t, h = 1, 2, 3, \ldots \tag{3.6}$$

$$\text{Level equation: } l_t = \alpha p^j_t + (1 - \alpha)l_{t-1} = \alpha p^j_t + (1 - \alpha)\hat{p}^j_{t|t-1} \tag{3.7}$$

---

[4]The following description of the method is based on the textbook by Hyndman (2013).

where l_t is the estimated level of the series at time t. The h-step ahead forecast for the sub-index log price $\hat{p}^j_{t+h|t}$ at any time $t + h$, $h = 1, 2, \ldots$ is the estimated level at time t. The level equation shows that the estimated level l_t of the series at time t is a weighted average of the current value p^j_t and the previous level l_{t-1}. In other words, the estimated level at time t is a weighted average of the current value of the sub-index log price and the within-sample 1-step ahead forecast for time t. Simple exponential smoothing is useful for forecasting time series with no trend or seasonal component.

3.3.2   Holt's double exponential smoothing

Holt's double exponential smoothing method (1957) extends simple exponential smoothing to accommodate smoothing and forecasting of a series with a trend.[5] In addition to the forecast equation and the smoothing equation for the level, there is now also a second smoothing equation for the trend. For a commodity sub-index log price p^j_t, double exponential smoothing can be expressed as

Forecast equation: $\hat{p}^j_{t+h|t} = l_t + h b_t$ $\qquad\qquad$ (3.8)

Level equation: $l_t = \alpha p^j_t + (1 - \alpha)(l_{t-1} + b_{t-1}) = \alpha p^j_t + (1 - \alpha)\hat{p}^j_{t|t-1}$ $\qquad$ (3.9)

Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$ $\qquad\qquad$ (3.10)

where l_t is the level of the series at time t, b_t is an estimate of the trend (slope) of the series at time t, α ($0 \leq \alpha \leq 1$) is the smoothing parameter for the level, and β^* ($0 \leq \beta^* \leq 1$) is the smoothing parameter for the trend. Again, the level equation means that the level l_t is a weighted average of the current value of the sub-index log price p^j_t and the within-sample 1-step ahead forecast for time t. The

---

[5]The following description of the method is based on the textbook by Hyndman (2013).

trend equation can be interpreted as the trend b_t being a weighted average of the change in the level $(l_t - l_{t-1})$ and the previous estimate of the trend b_{t-1}.

The forecast equation (3.8) shows that the forecast function is trending, unlike the flat forecast function associated with simple exponential smoothing. The h-step ahead forecasts are a linear function of h.

3.3.3   Holt-Winters triple exponential smoothing

Holt-Winters triple exponential smoothing (Holt (1957) and Winters (1960)) extends Holt's double exponential smoothing method to time series exhibiting linear trend and additive or multiplicative seasonality.[6] In addition to a forecast equation and smoothing equations for the level l_t and trend b_t, triple exponential smoothing adds a smoothing equation for the seasonal component, denoted by s_t. The corresponding smoothing parameters are α, β^*, and γ. Let m denote the period of seasonality, i.e. the number of seasons in a year. For monthly data, $m = 12$. There is an additive version and a multiplicative version of the method. The version is chosen depending on the nature of the seasonality (additive or multiplicative). For the IMF data, the seasonal variations are relatively constant throughout the series (i.e. additive seasonality), so I use the additive method. With this method, in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component adds up to approximately zero.

The additive triple exponential smoothing method can be expressed in terms

---

[6]The following description of the method is based on the textbook by Hyndman (2013).

of the following equations:

$$\text{Forecast equation: } \hat{p}^{j}_{t+h|t} = l_t + h b_t + s_{t-m+h^+_m} \tag{3.11}$$

$$\text{Level equation: } l_t = \alpha(p^j_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{3.12}$$

$$\text{Trend equation: } b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \tag{3.13}$$

$$\text{Seasonal factor: } s_t = \gamma(p^j_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \tag{3.14}$$

where $h^+_m \equiv \lfloor (h-1) \bmod m \rfloor + 1$ so that the estimates of the seasonal indices used for forecasting are from the last year of the sample (Hyndman (2013)). From the level equation, we can see that the estimated level is a weighted average of the seasonally adjusted current value of the sub-index log price $(p^j_t - s_{t-m})$ and the forecast $(l_{t-1} + b_{t-1})$ for time t in the absence of seasonality. The trend equation is the same as for double exponential smoothing. The seasonal factor equation expresses the seasonal component as a weighted average of the current seasonal index $(p^j_t - l_{t-1} - b_{t-1})$ and s_{t-m}, which is the seasonal index of the same season in the previous year.

3.3.4 Autoregressive models AR(p)

For a commodity sub-index log price p^j_t, the autoregressive AR(p) model can be written as:

$$p^j_t = \alpha^j + \sum_{i=1}^{K} \beta^j_i p^j_{t-i} + \epsilon^j_t \tag{3.15}$$

For each series p^j_t and for each estimation window, I estimate the parameters by OLS. The maximum lag length for the AR model is specified ex ante and the number of lags K is chosen according to the Akaike Information Criterion (AIC).

The 1-step ahead forecast is computed as:

$$\hat{p}^j_{t+1|t} = \alpha^j + \sum_{i=1}^{K} \beta^j_i p^j_{t-i+1} \tag{3.16}$$

The h-step ahead forecast is computed as

$$\hat{p}^j_{t+h|t} = \alpha^j + \sum_{i=1}^{K} \beta^j_i \hat{p}^j_{t+h-i|t} \tag{3.17}$$

3.3.5 Vector autoregression VAR(1)

Let $\bar{p}_t$ be an $(N \times 1)$ vector of all commodity sub-index log prices (N=52): $\bar{p}_t = (p^1_t, p^2_t, \ldots, p^N_t)$, c an $(N \times 1)$ vector of constants, A_1 an $(N \times N)$ matrix of coefficients, and ϵ_t an $(N \times 1)$ vector of disturbances. I estimate a VAR(1) model on the log-level of the series $\bar{p}_t$.

$$\bar{p}_t = c + A_1 \bar{p}_{t-1} + \epsilon_t \tag{3.18}$$

Then the 1-step ahead forecast is

$$\bar{p}_{t+1|t} = c + A_1 \bar{p}_t \tag{3.19}$$

The h-step ahead forecast is

$$\bar{p}_{t+h|t} = c + A_1 p_{t+h-1|t} \tag{3.20}$$

3.4   Empirical results

Table 3.1 shows the Root Mean Square Forecast Error (RMSFE) for the direct and indirect forecasts when using an AR(1) model. I find that using an AR(1) model, the forecasting accuracy from the disaggregated approach and the direct approach is almost the same, i.e. using disaggregate information does not help improve the accuracy of the forecast of the log of the aggregate. Figure 3.3 plots the RMSFE for each horizon $h = 1, \ldots, 12$ months for each of the two approaches. The blue line represents the RMSFE obtained when using the indirect approach with an AR(1) model. The red line represents the RMSFE obtained when using the AR(1) model and the direct approach. The blue line and red line overlap very closely, indicating that the two forecasting approaches have almost identical forecasting accuracy. The upward slope of the lines indicates that the forecasting performance of each approach deteriorates as the horizon increases.

Table 3.2 shows the forecasting performance of the simple exponential smoothing (SES) model for each of the two approaches. Again, I find that the forecasting accuracy is almost identical no matter whether sub-index information is used or the log of the aggregate is forecasted directly. Hence, aggregating sub-index forecasts does not help forecast the log of the aggregate commodity price index at any of the horizons under consideration. This can also be seen in Figure 3.4, which shows the RMSFE for each horizon from $h = 1, \ldots, 12$ months for each of the two approaches. The lines representing the RMSFE for each of the two approaches overlap, indicating that the two approaches have almost identical forecasting accuracy as measured by RMSFE. Again, the forecasting performance of each of the two methods gets worse as the horizon increases.

Table 3.3 shows the RMSFE for the two forecasting approaches when

using Holt's double exponential smoothing method. In this case, I find that using disaggregate forecasts to forecast the log of the aggregate commodity index improves forecasting accuracy at each horizon relative to forecasting the log of the aggregate directly. This can also be seen from Figure 3.5. As the forecast horizon increases, the difference between the forecasting accuracy of the two methods increases. In other words, for forecasts further out into the future, the disaggregated approach provides greater improvement in RMSFE relative to the direct approach.

Table 3.4 shows the forecasting performance of Holt-Winters triple exponential smoothing for both approaches. As with double exponential smoothing, triple exponential smoothing has greater forecasting accuracy at each horizon when combining sub-index forecasts as opposed to forecasting the log of the aggregate index directly. This can be seen more clearly in Figure 3.6. Again, the forecasting accuracy gain of the disaggregated approach relative to the direct approach increases for horizons further out into the future.

Table 3.5 shows the RMSFE for the AR(p) model for the disaggregated approach and for the direct approach. Figure 3.7 depicts the RMSFE for each approach and for each horizon. For horizons from 1 to 4 months, the forecasting accuracy of the two approaches is approximately the same. For horizons 5 to 12 months, combining sub-index forecasts to forecast the log of the aggregate index improves forecasting accuracy relative to forecasting the log of the aggregate directly. The accuracy gain increases as the forecast horizon increases.

Finally, Table 3.6 shows the RMSFE for a VAR(1) model on logs of the 52 sub-index series. The RMSFE for the VAR(1) model is higher at each forecast horizon than the RMSFE for the disaggregated approach when using any of the other methods.

When using double exponential smoothing and triple exponential smoothing,

aggregating forecasts of the individual sub-indices increases forecasting accuracy at any horizon $h = 1, \ldots, 12$ months. For the AR(p) model, combining sub-index forecasts improves forecasting accuracy relative to forecasting the log of the aggregate directly for horizons $h \geq 5$ months. For the AR(1) model and the simple exponential smoothing model, using the indirect approach does not improve the forecasting performance and results in almost identical RMSFE as forecasting the aggregate index directly. Overall, no matter whether I use disaggregate forecasts to forecast the log of the aggregate commodity index or I forecast the log of the aggregate index directly, none of the models considered generate more accurate out-of-sample forecasts than those based on an AR(1) model specification. When forecasting the log of the aggregate commodity index directly, for horizons $h \geq 8$ months, the simple exponential smoothing model has almost identical forecast accuracy as the AR(1) model. The VAR(1) model has the worst forecast accuracy among the models considered.

3.5 Conclusion

In this paper, I compare the forecasting performance of simple exponential smoothing, Holt's double exponential smoothing, Holt-Winters triple exponential smoothing, AR(p), AR(1), and VAR(1) models when forecasting the log of the aggregate commodity index by using sub-index forecasts (indirect approach) versus forecasting the log of the aggregate commodity index directly (direct approach). I find that when using double exponential smoothing and additive triple exponential smoothing, aggregating forecasts from the individual sub-indices increases forecasting accuracy at any horizon $h = 1, \ldots, 12$ months. For the AR(p) model, using sub-index forecasts to forecast the log of the aggregate improves forecasting accuracy relative to forecasting the log of the aggregate directly for horizons $h \geq 5$

months. On the other hand, for the AR(1) model and the simple exponential smoothing model, the indirect approach does not improve the forecasting performance and results in almost identical RMSFE as the direct approach. When using either forecasting approach, none of the models considered generate more accurate out-of-sample forecasts than the AR(1) model. For the direct approach, for horizons $h \geq 8$ months, the simple exponential smoothing model has almost identical forecast accuracy as the AR(1) model.
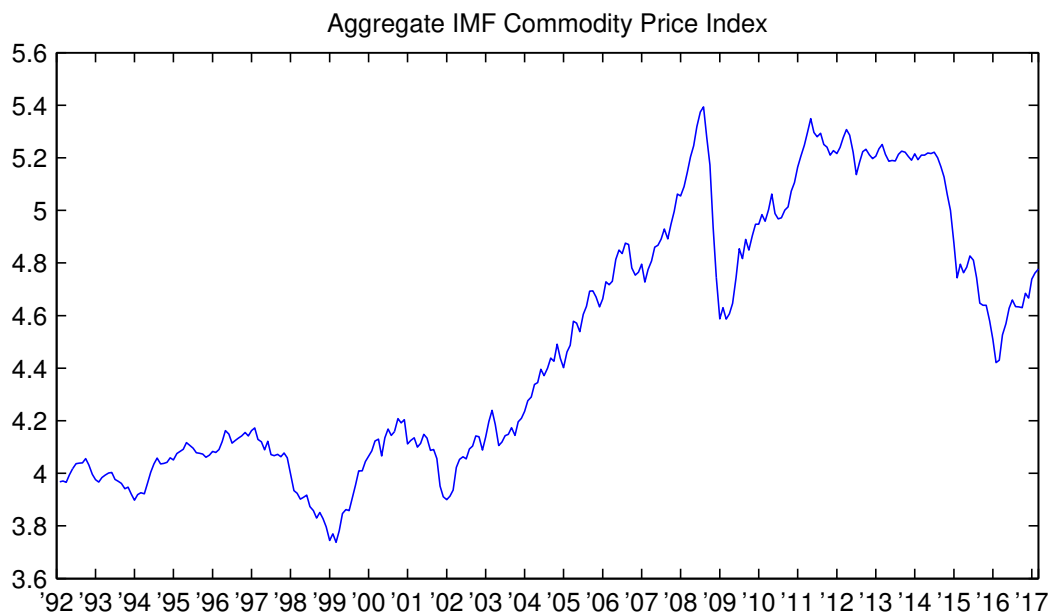
3.6 Figures

Figure 3.1. Log of the monthly aggregate IMF commodity index
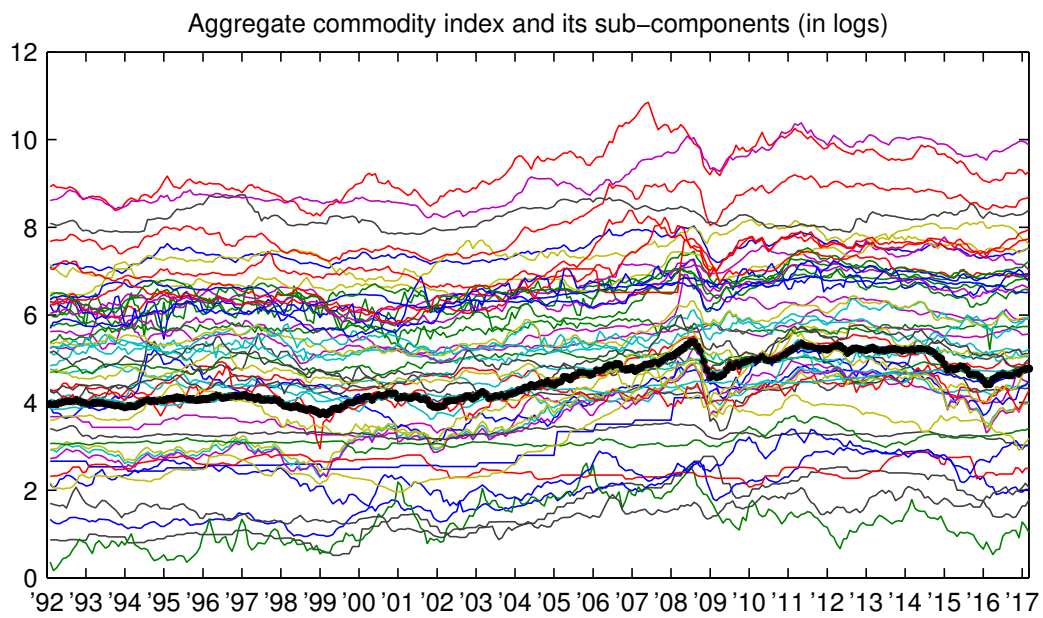The index is measured in nominal US dollar terms.

Figure 3.2. Log of the monthly aggregate IMF commodity index and its 52 sub-indices

The thick black line represents the aggregate index. Prices are measured in nominal US dollar terms.
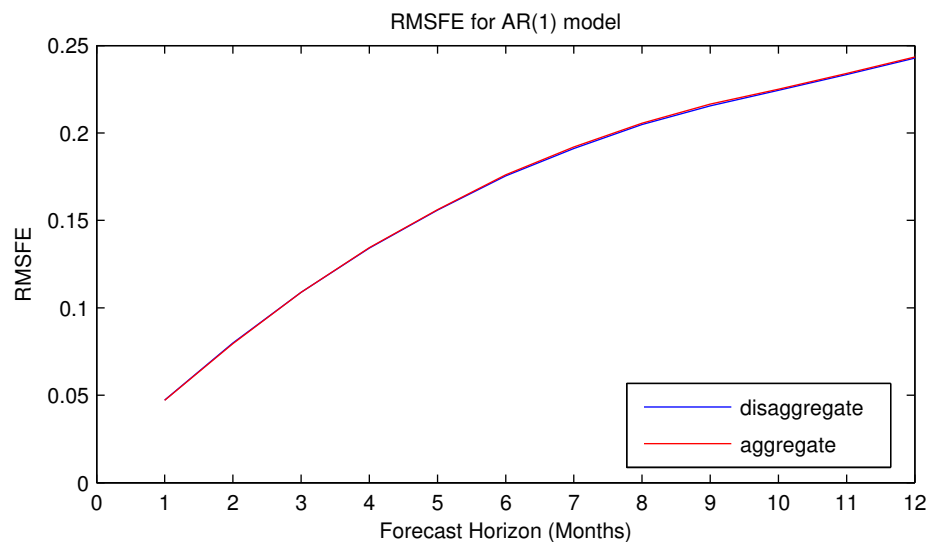
Figure 3.3. RMSFE for the indirect and the direct approach when using an AR(1) model

Root mean square forecast error (RMSFE) obtained when the AR(1) forecasts for the commodity sub-indices are aggregated to obtain a forecast for the aggregate index (in blue) versus RMSFE obtained when the aggregate commodity price index is forecasted directly using an AR(1) model (in red).
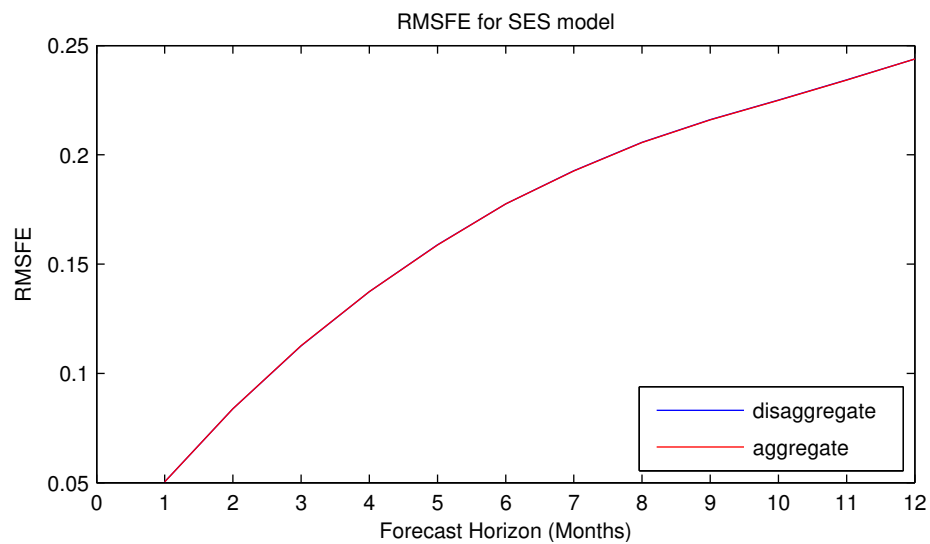
Figure 3.4. RMSFE for the indirect and the direct approach when using a simple exponential smoothing model

Root mean square forecast error (RMSFE) obtained when the simple exponential smoothing (SES) forecasts for the commodity sub-indices are aggregated to obtain a forecast for the aggregate index (in blue) versus RMSFE obtained when the aggregate commodity price index is forecasted directly using a SES model (in red).
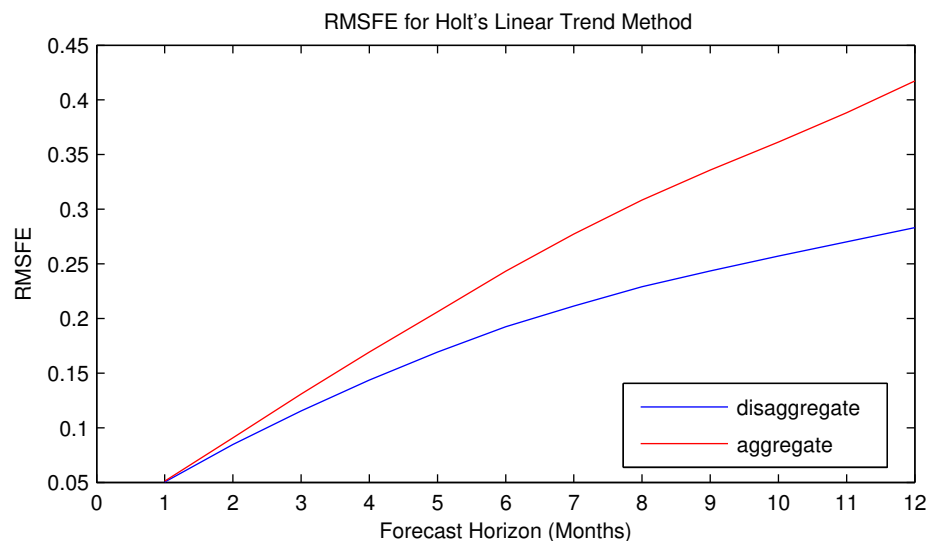
Figure 3.5. RMSFE for the indirect and the direct approach when using Holt's double exponential smoothing
Root mean square forecast error (RMSFE) obtained when the individual forecasts for the commodity sub-indices obtained by using Holt's double exponential smoothing method are aggregated to obtain a forecast for the aggregate index (in blue) versus RMSFE obtained when the aggregate commodity price index is forecasted directly using Holt's double exponential smoothing (in red).
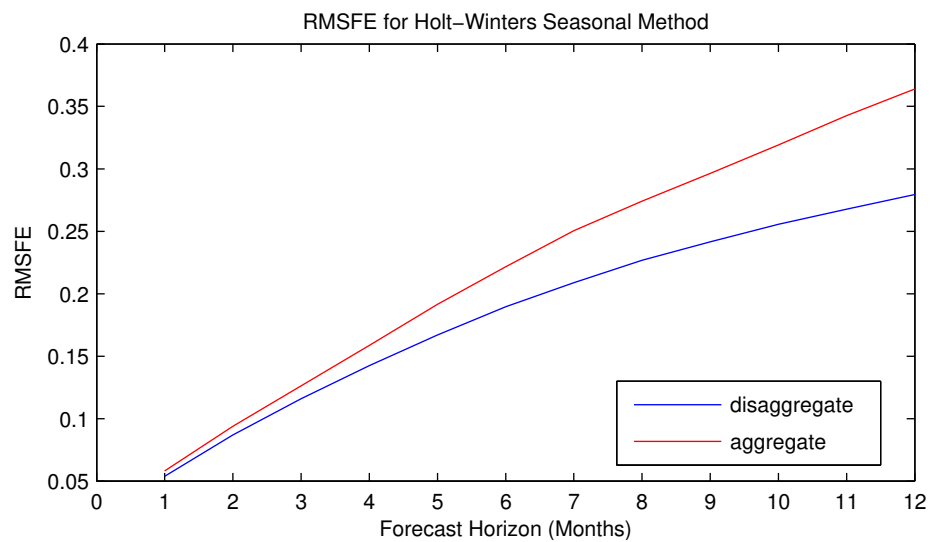
Figure 3.6. RMSFE for the indirect and the direct approach when using additive triple exponential smoothing

Root mean square forecast error (RMSFE) obtained when the individual forecasts for the commodity sub-indices obtained by using additive triple exponential smoothing are aggregated to obtain a forecast for the aggregate index (in blue) versus RMSFE obtained when the aggregate commodity price index is forecasted directly using additive triple exponential smoothing (in red).
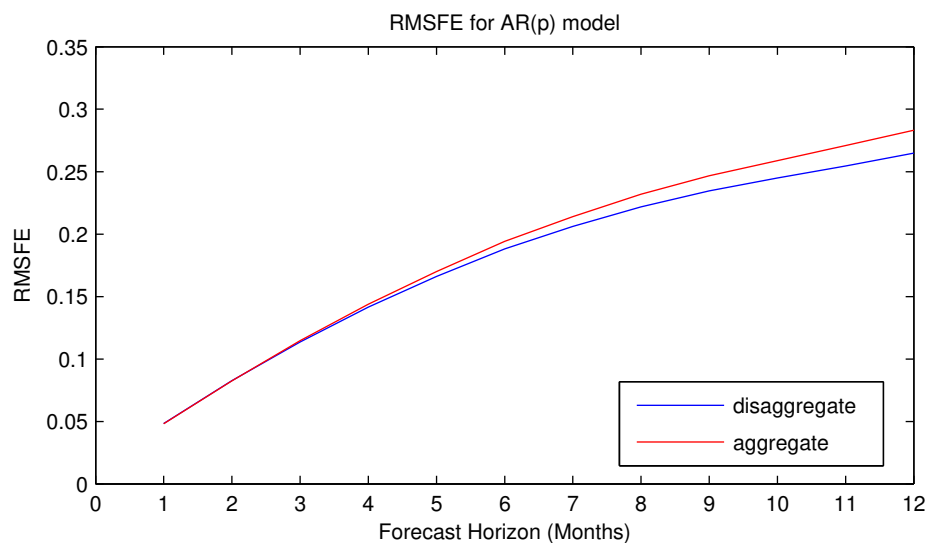
Figure 3.7. RMSFE for the indirect and the direct approach when using an AR(p) model

Root mean square forecast error (RMSFE) obtained when the AR(p) forecasts for the commodity sub-indices are aggregated to obtain a forecast for the aggregate index (in blue) versus RMSFE obtained when the aggregate commodity price index is forecasted directly using an AR(p) model (in red).
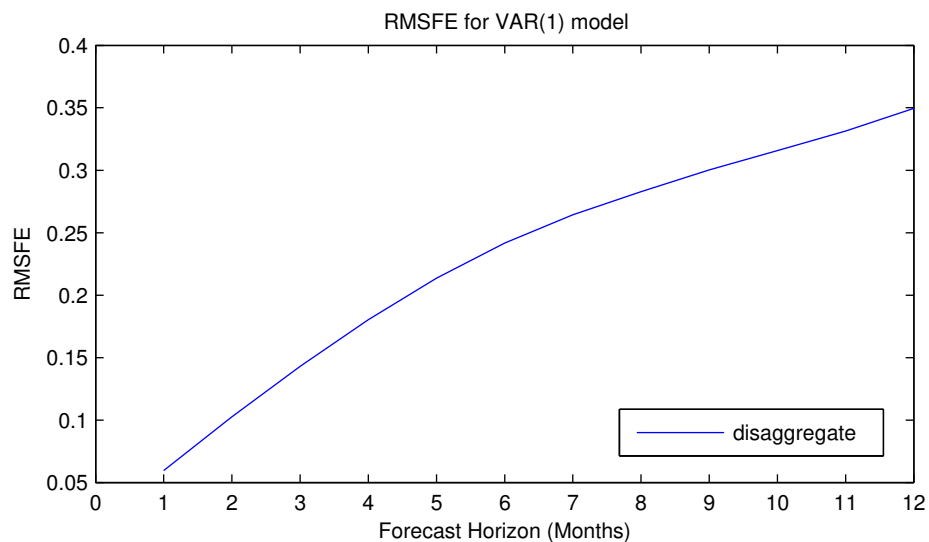

Figure 3.8. RMSFE for the indirect method when using a VAR(1) model

Root mean square forecast error (RMSFE) obtained when the VAR(1) forecasts for the individual commodity sub-indices are aggregated to obtain a forecast for the aggregate index.

3.7   Tables

Table 3.1. RMSFE for the AR(1) model

Horizon (month)	Disaggregate (52 items)	Aggregate
1	0.0473	0.0470
2	0.0799	0.0795
3	0.1089	0.1088
4	0.1342	0.1343
5	0.1559	0.1561
6	0.1755	0.1761
7	0.1911	0.1919
8	0.2048	0.2056
9	0.2155	0.2165
10	0.2244	0.2251
11	0.2334	0.2341
12	0.2429	0.2434

Table 3.2. RMSFE for the simple exponential smoothing model

Horizon (month)	Disaggregate (52 items)	Aggregate
1	0.0503	0.0504
2	0.0838	0.0838
3	0.1126	0.1126
4	0.1374	0.1374
5	0.1588	0.1588
6	0.1776	0.1775
7	0.1927	0.1926
8	0.2057	0.2056
9	0.2161	0.2160
10	0.2250	0.2249
11	0.2343	0.2342
12	0.2439	0.2438

Table 3.3. RMSFE for Holt's double exponential smoothing

Horizon (month)	Disaggregate (52 items)	Aggregate
1	0.0503	0.0511
2	0.0847	0.0909
3	0.1154	0.1309
4	0.1437	0.1692
5	0.1693	0.2061
6	0.1925	0.2433
7	0.2114	0.2772
8	0.2290	0.3083
9	0.2436	0.3358
10	0.2571	0.3615
11	0.2701	0.3883
12	0.2831	0.4173

Table 3.4. RMSFE for the Holt-Winters additive triple exponential smoothing

Horizon (month)	Disaggregate (52 items)	Aggregate
1	0.0539	0.0581
2	0.0870	0.0939
3	0.1159	0.1263
4	0.1425	0.1586
5	0.1670	0.1916
6	0.1897	0.2217
7	0.2088	0.2505
8	0.2268	0.2742
9	0.2416	0.2964
10	0.2555	0.3191
11	0.2676	0.3426
12	0.2795	0.3640

Table 3.5. RMSFE for the AR(p) model

Horizon (month)	Disaggregate (52 items)	Aggregate
1	0.0483	0.0481
2	0.0827	0.0825
3	0.1136	0.1147
4	0.1416	0.1438
5	0.1661	0.1699
6	0.1882	0.1942
7	0.2062	0.2140
8	0.2218	0.2319
9	0.2346	0.2467
10	0.2450	0.2588
11	0.2546	0.2708
12	0.2649	0.2832

Table 3.6. RMSFE for the VAR(1) model

Horizon (month)	Disaggregate (52 items)
1	0.0595
2	0.1027
3	0.1431
4	0.1803
5	0.2136
6	0.2418
7	0.2643
8	0.2828
9	0.3003
10	0.3157
11	0.3314
12	0.3496

3.8 Bibliography

Benalal, N., Del Hoyo, J. L. D., Landau, B., Roma, M. and Skudelny, F. (2004)."To aggregate or not to aggregate? euro area inflation forecasting," Working Paper Series 374, European Central Bank.

Bermingham, C. and D'Agostino, A. (2014) "Understanding and forecasting aggregate and disaggregate price dynamics" *Empirical Economics*, vol. 46, issue 2, 765-788.

Chen, Y., Turnovsky, S. J, and Zivot, E. (2014) "Forecasting inflation using commodity price aggregates" *Journal of Econometrics* Volume 183, Issue 1, 117-134.

Espasa, A., Senra, E., and Albacete, R. (2002), Forecasting inflation in the European Monetary Union: A disaggregated approach by countries and by sectors, *European Journal of Finance*, 8 (4), 402-421.

Fair, R. C. and Shiller, J. (1990) "Comparing information in forecasts from econometric models," *The American Economic Review* 80(3): 375-389.

Gargano, A. and Timmermann, A. (2014) "Predictive Dynamics in Commodity Prices." *International Journal of Forecasting* 30(3), 825-843.

Granger, C. W. J. (1988) "Aggregation of time series variables - a survey," Discussion Paper 1, Federal Reserve Bank of Minneapolis.

Groen, J. J. J. and Pesenti, P. A. (2011) "How easy is it to forecast commodity prices?" *Liberty Street Economics blog*, Federal Reserve Bank of New York.

Groen, J. J. J. and Pesenti, P. A. (2011) "Commodity prices, commodity currencies, and global economic developments." In T. Ito and A. Rose (Eds.), *Commodity prices and markets, East Asia Seminar on Economics*, Vol. 20, 15-42. Chicago: University of Chicago Press.

Hendry, D. F. and Hubrich, K. (2006) "Forecasting economic aggregates by disaggregates," European Central Bank Working Paper Series No. 589.

Hendry, D. F. and Hubrich, K. (2011) "Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate," *Journal of Business and Economic Statistics*, 29:2, 216-227.

Holt, C. C. (1957) "Forecasting trends and seasonals by exponentially weighted averages," Carnegie Institute of Technology, Pittsburgh ONR memorandum no. 52.

Hubrich, K. (2005) "Forecasting euro area inflation: Does aggregating

forecasts by HICP component improve forecast accuracy?" *International Journal of Forecasting* 21, 119-136.

Hyndman, R. J. and Athanasopoulos, G. (2013) "Forecasting: principles and practice," OTexts.

Ishimwe, B. (2015) "Can the IMF commodity price index forecast U.S. inflation?," Universite Catholique de Louvain, Louvain School of Management, Masters Thesis.

Kohn, R. (1982) "When is an aggregate of a time series efficiently forecast by its past?," *Journal of Econometrics* (18): 337-349.

Lütkepohl, H. (1984a) "Forecasting contemporaneously aggregated vector ARMA processes," *Journal of Business & Economic Statistics* 2(3): 201-214.

Lütkepohl, H. (1984b) "Linear transformations of vector ARMA processes," *Journal of Econometrics* (26): 283-293.

Lütkepohl, H. (1987) "Forecasting aggregated vector ARMA processes," Vol. 284: Springer Science & Business Media.

Lütkepohl, H. (2010) "Forecasting aggregated time series variables: A survey," *OECD Journal: Journal of Business Cycle Measurement and Analysis*, Vol. 2010/2.

Marcellino, M., Stock., J. H., and Watson, M. W. (2003) "Macroeconomic forecasting in the Euro area: Country specific versus area-wide information," *European Economic Review* 47, 1-18.

Marcellino, M. (2004) "Forecast pooling for European macroeconomic variables," *Oxford Bulletin of Economics and Statistics*, 66(1), 91-112.

Rose, D. E. (1977) "Forecasting aggregates of independent ARIMA processes," *Journal of Econometrics* (5): 323-345.

Theil, H. (1954) "Linear Aggregation of Economic Relations," North Holland, Amsterdam.

Tiao, G. C. and Guttman, I. (1980) "Forecasting contemporal aggregates of multiple time series," *Journal of Econometrics* (12): 219-230.

Timmermann, A. (2006) "Forecast combinations." In: Elliott, C.G.G., Timmermann, A. (Eds.), "Handbook of Economic Forecasting," "Handbook of Economic Forecasting," vol. 1. Elsevier, 135-196.

Winters, P. R. (1960) "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, 6 (3): 324-342.

Zellner, A. and Tobias, J. (2000) "A note on aggregation, disaggregation and forecasting performance," *Journal of Forecasting* 19: 457-469.