ESSAYS IN FINANCIAL ECONOMICS

Sunjin Park

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Finance in the Kenan-Flagler Business School.

Chapel Hill
2017

Approved by:
Riccardo Colacito
Mariano M. Croce
Eric Ghysels
Anh Le
Christian Lundblad
ABSTRACT
Sunjin Park: Essays in Financial Economics
(Under the direction of Riccardo Colacito)

In the first chapter, titled "Global Macroeconomic Conditional Skewness and the Carry Risk Premium," I show that the time-variation in measures of global growth prospects constructed from the cross-section of individual macroeconomic forecasts can help explain currency markets. I show that conditional expectation and skewness of global economic growth have predictive ability in explaining the quarterly returns to carry trade and that the global skewness measure is particularly important in explaining a large cross-section of currencies. I provide the economic mechanism for the role of cross-sectional skewness in forecasts using a consumption-based asset pricing model with heterogeneous agents. In the second chapter, which is titled "Risk and Return Trade-off in the U.S. Treasury Market," we characterize the risk-return trade-off in the U.S. Treasury market through the lens of a discrete-time term structure model in which the conditional variances of bond yields feature a short-run component and a long-run component. Using Treasury yields data from January 1962 to August 2007, we find that the short-run volatility component of bond yields commands a positive risk premium whereas the long-run volatility component does not. In addition, for short-dated bonds, most of the variations in risk premiums are attributable to investors' changing attitudes toward risks. For longer-dated bonds, risk premiums reflect both the amount of risks bond investors face as well as their tolerance for risks over time.
ACKNOWLEDGMENTS

I would like to emphasize that my Ph.D. career at the Kenan-Flagler Business School would not have been possible without the support and encouragement of many wonderful people I met and have known.

I am indebted to my advisor Riccardo Colacito for his invaluable mentorship and dedication in directing my doctoral education. I am grateful to Eric Ghysels for his continued support and guidance since my Masters program and to Christian Lundblad for his valuable feedback and ongoing encouragement on various aspects. I am also indebted to Anh Le for his endless support and patience from the very beginning of my research career. Lastly, I thank Mariano Croce for serving as a committee member and providing insightful feedback.

I have benefited tremendously from the conversations with my Ph.D. colleagues as well as the faculty members in our department. I am appreciative of their comments on my research as well as suggestions as a researcher and a presenter, all of which I do not take for granted.

I am grateful to my parents, my brother and my friends for their encouragement and caring throughout my academic training in Chapel Hill over the past 12 years. Finally, I would like to emphasize my gratitude to my wife Youngeun for her endless love and for bringing to the world our soon-to-be-born baby girl.

I also acknowledge that faith was an integral part of every course of my life.
# TABLE OF CONTENTS

## LIST OF TABLES

vi

## LIST OF FIGURES

vii

## 1 GLOBAL MACROECONOMIC CONDITIONAL SKEWNESS AND THE CARRY RISK PREMIUM

1

### Introduction

1

### 1.1 Literature Review

3

### 1.2 Model

6

#### 1.2.1 Setup of the Economy

6

#### 1.2.2 Equilibrium and Solution of the Model

8

#### 1.2.3 Asset Pricing

10

#### 1.2.4 Skewness in Forecasts and Carry Risk Premium

11

#### 1.2.5 Dynamic model

16

### 1.3 Global Measures of Risk: Data Sources and Stylized Facts

19

### 1.4 Empirical Results

23

#### 1.4.1 Predictive Regressions of Currency Returns

24

#### 1.4.2 Robustness

33

#### 1.4.3 Trading Conditional on Global Skewness

34

### 1.5 Conclusion

36

## 2 RISK AND RETURN TRADE-OFF IN THE U.S. TREASURY MARKET

37

### Introduction

37

### 2.1 Motivating Exercises

41

### 2.2 Model

45

#### 2.2.1 The Risk-Neutral Dynamics and Bond Pricing

46

#### 2.2.2 The Time-Series Dynamics

46
LIST OF TABLES

1.1 Regression from simulated data ...................................................... 19
1.2 Start date of forecasts data and summary statistics of the number of forecasters ... 21
1.3 Predictive regression results of carry returns onto global measures .......................... 25
1.4 Predictive regressions of individual buckets of currencies onto global skewness ........... 30
1.5 Predictive regressions of individual buckets of currencies onto global expected growth .......................................................... 31
1.6 Regressions of carry trade returns onto global skewness and other explanatory variables .......................................................... 34
1.7 Summary statistics of the ordinary carry returns and the new strategy returns ............... 36

2.1 Predictability of weekly excess returns ............................................ 43
2.2 Validating model choices ................................................................. 56
2.3 Parameter estimates ...................................................................... 58
2.4 Two robustness checks of the model .................................................. 59
2.5 Risk premium decomposition ............................................................ 61

A.1 Calibration of the two-period model ................................................. 66
A.2 Calibration of the dynamic model ..................................................... 67
A.3 Details about foreign exchange rate data .......................................... 68
A.4 Predictive regression results with the exclusion of the recent crisis ....................... 69
A.5 Predictive regression results with global measure aggregated by taking the first principal component ......................................................... 70
A.6 Predictive regression results with global measure aggregated by taking the GDP-weighted average ......................................................... 71
A.7 Predictive regression results with alternative global measures .......................... 72
A.8 Predictive regression results controlling for other explanatory variables ................. 73
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Subjective probability density function for each agent</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>Global skewness, carry risk premia and interest rates</td>
<td>11</td>
</tr>
<tr>
<td>1.3</td>
<td>Individual agent’s consumption over the possible endowment shocks</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>Global measures of risks</td>
<td>22</td>
</tr>
<tr>
<td>1.5</td>
<td>Estimated beta loadings for different sets of currencies</td>
<td>27</td>
</tr>
<tr>
<td>1.6</td>
<td>Comparison of cumulative returns to the ordinary carry and the new strategy</td>
<td>35</td>
</tr>
<tr>
<td>2.1</td>
<td>Model-implied risk premium</td>
<td>62</td>
</tr>
<tr>
<td>2.2</td>
<td>Campbell-Shiller regression coefficients</td>
<td>64</td>
</tr>
<tr>
<td>2.3</td>
<td>Model-implied volatilities and realized volatilities</td>
<td>65</td>
</tr>
</tbody>
</table>
CHAPTER 1 GLOBAL MACROECONOMIC CONDITIONAL SKEWNESS AND THE CARRY RISK PREMIUM

Introduction

The carry trade is a well known investment strategy that exploits the profitability of borrowing in the low interest rate currencies to invest in the high interest rate currencies. In this paper I document that the time-variation in the distribution of global growth prospects has predictive power for carry trade returns. I study the macroeconomic risks that the carry trade investor faces. Interestingly, I find evidence that the time-variation in conditional skewness in global growth prospects has significant predictive power, namely that a one standard deviation decline in the skewness measure increases the next-quarter carry trade risk premium by 5.24% per annum. The novel contribution of the paper is that global macroeconomic conditional skewness plays an important role in the variation in the carry risk premium.

I empirically test if the variation in the cross-sectional measures of macroeconomic prospects explains the currency market. I collect analysts’ forecasts for the growth rates of real GDP for a list of major countries including nine of the G10 countries as well as China. The currencies of these countries constitute 85.75% of the total foreign exchange turnover\(^1\). The individual forecasts are contributed by analysts in different sectors of the economy and are collected primarily by Consensus Economics and Bloomberg. At each point in time and for each country, I construct measures of the cross-sectional mean, dispersion and skewness of the distribution of forecasts across analysts. Then for each quarter, I calculate the cross-sectional average of the means across countries and, similarly, the average of the dispersion and the average of the skewness across countries. This yields time-varying measures of the distribution of, what I shall refer to as, *global* growth prospects. The main empirical strategy proposed in this paper tests if my proposed global measures predict carry trade returns in the time-series.

\(^1\) Source: BIS (2016)
I find evidence that the time-variation in global conditional expected growth and global conditional skewness can help predict next-quarter carry trade returns. The estimated coefficients are negative, indicating that when global expected growth or global skewness is low or negative, subsequent carry trade returns tend to be high or positive, i.e., yielding a positive risk premium. Notably, global skewness appears to be the most robust predictor among the different moments, especially as I repeat the exercise with strategies based on a larger set of currencies of up to 33 developed and emerging markets. I conduct a series of robustness tests, such as forming dynamic and static portfolios, changing the number of currencies in the formation of portfolios, and aggregating country-specific measures, e.g., by taking the first principal component or by computing the GDP-weighted average across countries. I also try jointly regressing on the global skewness measure along with other known explanatory variables.

A key benefit of my approach is that it yields a time-varying proxy for conditional skewness of macroeconomic growth prospects. A skewness measure is related to, but has an interesting distinction from, the notion of disaster. Disasters are one-sided by nature and are often referred to as events that rarely happen. On the contrary, I observe frequent fluctuations of my skewness measure between positive and negative domains, even outside of times of heightened concerns about severe recessions. Moreover, my empirical strategy allows obtaining real-time measures based on a collection of professional forecasters’ views each time a survey is reported, thus revealing information about macroeconomic prospects that are otherwise not easy to detect. Furthermore, my results are robust to the exclusion of the Great Recession of 2008-09, confirming that they are not driven by extreme left-tail events.

I build a model in which agents have heterogeneous beliefs, so that it can be mapped directly to my empirical investigation. In this economy, there are two countries, each populated by three agents. In each country, one agent has the correct beliefs about the future growth rate of the economy, while the other two agents have expectations that are either larger (“the optimist”) or smaller (“the pessimist”) than the true growth rate. Depending on the specific degree of optimism and pessimism of those two agents, the cross-sectional distribution of beliefs within each country can take on any possible extent of skewness.

The presence of an agent with correct beliefs in each country is relevant because these agents will act as the marginal investors that pin down the equilibrium adjustment of the exchange rate.
Assuming that financial markets are complete, the exchange rate between the currencies of the two countries is equal to the ratio of marginal utilities of the two agents with correct beliefs by a simple no-arbitrage argument (as in Backus, Foresi, and Telmer (2001)).

Let us consider the situation in which the cross-sectional skewness is negative in one country and equal to zero in the other country. According to the definition that I adopt in my empirical approach, this situation corresponds to one in which the global skewness is negative. It is intuitive to conclude that the risk-free rate should be lower in the first country, in which the pessimist drives up the demand for the risk-free asset by a larger extent. Carry trade would thus involve borrowing in the currency of the first country with negative skewness and investing in the currency of the other country with zero skewness.

A key feature of the model with heterogeneous beliefs is that agents want to consume the most in states of the world that they think are the most likely. This means that the marginal investor consumes less than the pessimist in bad times. This helps explain why shorting the currency of the negatively skewed country is a risky strategy. In bad times, the marginal utility (consumption) of the marginal investor goes up (drops) more in the negatively skewed country. This, by no-arbitrage, results in an appreciation of the currency of this country. Equivalently, the carry trade is risky because it loses money in bad times. A similar argument can be used to show that it gains money in good times.

This example illustrates why the risk premium is higher in times in which the global skewness is more negative. Based on this idea, the model implies that carry trades are risky when the investor faces negatively skewed global prospects.

1.1 Literature Review

The cross-sectional measures of GDP forecasts have been previously considered in the literature primarily to explain domestic equity or bond risk premia. Campbell and Diebold (2009) document that expected business conditions, measured by taking the consensus forecasts, can predict next period stock returns. Bansal and Shaliastovich (2010) find that cross-sectional dispersion of forecasts informs us about confidence risk, which helps explain the equity risk premia. Buraschi and Whelan (2012) show that dispersion in forecasts can predict subsequent bond excess returns with the argument specifically about belief dispersion. Colacito, Ghysels, Meng, and
Siwasarit (2016) find that negative cross-sectional skewness precedes recessions and helps pre-
dict future stock returns. I base upon some papers that study the source of the cross-sectional
dispersion of GDP forecasts (Patton and Timmermann (2010) and Andrade, Crump, Eusepi,
and Moench (2016)) and interpret the cross-section of forecasts as macroeconomic disagreement
among forecasters.

I argue in this paper that my measure of macroeconomic conditional skewness is a global mea-
sure of risk. The implication for the currency market is that the global measure should affect the
stochastic discount factors of countries based on the exposure to the risk so that the movement of
the foreign exchange rate is also affected. This fits in with the literature following Lustig, Rous-
sanov, and Verdelhan (2011) that currency risk premia can be explained by the exposure to a
systematic risk. I focus on the risks about macroeconomic growth based on the evidence that
the currency risk premia can be explained by consumption growth risk as documented in Lustig
and Verdelhan (2007). Specifically given my model with disagreement, currency risk premia are
driven by the cross-sectional skewness in growth forecasts because the resultant risk-sharing
among agents will drive the stochastic discount factors and hence also drive the riskiness in the
foreign exchange rate.

The literature has been studying the dispersion of analysts’ forecasts and its asset pricing
implications. Anderson, Ghysels, and Juergens (2005) find that dispersion in analysts’ forecasts
about expected earnings is a priced factor in the equities market. Buraschi, Trojani, and Vedolin
(2014) provide evidence that belief disagreement, also constructed from earnings forecasts, can
explain the cross-section of corporate bond and stock returns.

More broadly, there is a large literature that shows the relevance of heterogeneity in agents in
asset pricing. Constantinides and Duffie (1996) incorporate incomplete consumption insurance in
multiple-agent economy and derive the pricing implication of the role of the cross-sectional vari-
ance of consumption across agents. Constantinides and Ghosh (2017) model uninsurable labor
income shocks as to generate countercyclical left skewness in the cross-section of consumption
growth which also affects aggregate prices. Although my model also similarly draws on the no-
tion of multiple-agent equilibrium, I focus on agents making decisions due to different beliefs.
Xiong and Yan (2010) present a model that illustrates how incorporating heterogeneous expec-
tations can explain empirical features such as the failure of the expectations hypothesis. In the
model of Dumas, Lewis, and Osambela (2016) home and foreign investors disagree because of their differing interpretations of home versus foreign news, and one implication of the model is the co-movement of returns and international capital flows.

We may relate the role of skewness to that of disaster risk. Farhi and Gabaix (2016) and Farhi, Fraiberger, Gabaix, Ranciere, and Verdelhan (2015) find that rare disaster risk can account for a large fraction of the carry trade risk premia. However, notice that a measure of skewness is not restricted to the notion of a rare, extreme event. In fact, my time-series of global macroeconomic conditional skewness tends to be low, well in advance of the onset of the recessions. Moreover, a skewed distribution in growth prospects has the further benefit that it measures both negative and positive directions of asymmetry, which cannot be captured by disaster risk.

The literature provides many competing explanations for currency risk premia, one of which emphasizes the role of commodities. In the model of Ready, Roussanov, and Ward (2014), the high interest rate countries, which correspond to the investment currencies, tend to be the commodity exporters, while the low interest rate countries, which correspond to the funding currencies, tend to be the exporters of the finished goods. The authors show empirical support that the strategy of sorting based on net exports in basic goods, which measure how much one specializes in producing and exporting basic commodities, yields high returns. Chen, Rogoff, and Rossi (2010) and Bakshi and Panayotov (2013) provide empirical evidence on the relationship between exchange rates and commodity prices.

Other explanations for the currency risk premia include the role of global foreign exchange volatility which rises precisely when the high interest rate currencies yield poor returns as shown in Menkhoff, Sarno, Schmeling, and Schrimpfl (2012). Londono and Zhou (2017) provide evidence that global currency variance risk premia can predict future currency movements, which they rationalize via global inflation risk. Yu (2013) proposes a sentiment-based, i.e. biased growth expectations, explanation for the currency movements. In the work of Gabaix and Maggiori (2015), since the net debtor country borrows from the financial market that has limited risk-bearing capacity, their currency requires a compensation. Della Corte and Krecetovs (2016) actually provide interesting evidence that the currency risk premia can be explained by current account uncertainty, which is measured by the cross-sectional dispersion of current account forecasts, dom-
inating other macro uncertainty variables like the dispersion in GDP forecasts. My paper looks specifically at the GDP forecasts and instead examines different moments of the distribution.

My paper also sheds some perspectives on the macro-finance literature that bridges international asset prices and consumption dynamics. Colacito and Croce (2013) provide evidence that the highly correlated long-run growth prospects can explain the Backus and Smith (1993) anomaly that the correlation between consumption differentials and exchange rate movements is low. Gourio, Siemer, and Verdelhan (2013) develop a standard real business cycle framework, in which the risk premia vary with the probability of a disaster that leads to a decline in investment. My measures of global risks are not directly from consumption or growth data, but they are derived from analysts’ views of future real economic growth prospects.

This paper is organized as follows. Section 2 introduces a model that yields testable predictions. Section 3 provides an explanation on the forecasts data and highlights stylized facts about the proposed global measures of risks. Section 4 presents the main currency predictability results. Lastly, section 5 provides concluding remarks.

1.2 Model

In the following section I focus on a static model with heterogeneous agents to highlight the economic mechanism of how the cross-sectional skewness in forecasts affects the riskiness of a currency trade. The static model can be understood as a snapshot of the dynamic version which follows subsequently.

1.2.1 Setup of the Economy

Consider a two-period, complete market economy with two countries, which I call home and foreign. The home country produces good $X$, and the foreign country produces good $Y$. The true data generating processes for the endowment goods $X$ and $Y$ are as follows

\[
\log X = \log X_0 + \varepsilon_X \quad \log Y = \log Y_0 + \varepsilon_Y \tag{1.1}
\]

where the endowment shocks $\varepsilon_X \sim \mathcal{N}(\mu, \sigma^2)$ and $\varepsilon_Y \sim \mathcal{N}(\mu, \sigma^2)$ have a correlation of $\rho$.

Each country is populated by three agents, denoted $AG_i$ for the home country and $AG_i^*$ for
the foreign country. Each agent forms a subjective probability density function about the endowment shocks. I assume that all agents correctly form the underlying distributional shape, the variance and the covariance of the shocks but that agents can have biased expectations about the mean of the shocks, which I will interchangeably refer to as the forecast or the prediction. For each country, there will be an optimist and a pessimist as well as an unbiased forecaster whose prediction coincides with the correct mean. Furthermore, for simplicity I assume that all agents correctly forecast the mean of the other country’s endowment shock. Mathematically, each home agent \( AG_i \) forms a joint probability distribution \( \pi_i(\varepsilon_X, \varepsilon_Y) \sim \mathcal{N}((\mu_i, \mu_j)', \Sigma) \), and each foreign agent \( AG_i^* \) forms \( \pi_i^*(\varepsilon_X, \varepsilon_Y) \sim \mathcal{N}((\mu_i^*, \mu_j^*)', \Sigma) \). In the coming discussion I will denote each agent by sorting agents within a country based on the means on a descending order: \( AG_1 \) forms the highest mean, while \( AG_3 \) forms the lowest mean.

Figure 1.1 presents an example in which the three home agents form much more negatively skewed set of forecasts about their endowment shocks relative to the foreign agents. I define \( \alpha = (\mu_1 + \mu_3 - 2 \times \mu_2)/(\mu_1 - \mu_3) \) (and similarly for \( \alpha^* \) for the foreign country) as a metric that summarizes the extent of skewness in forecasts. Later in the comparative statics exercise, I will examine the impact of a marginal change in \( \alpha \), but I do so by specifying the cross-section of forecasts that holds the cross-sectional variance fixed at a specified value. This is to suppress the
effect of variance, or dispersion, across forecasts. Notice that, because the cross-sectional variance across agents’ predictions are all fixed to be the same, negatively skewed set of forecasts must accompany a pessimist being extremely pessimistic as well as an optimist whose prediction is relatively close to that of the unbiased agent. Details of the calibration can be found in Table A.1.

I discipline the skewness of forecasts in each country \( \alpha \) and \( \alpha^* \) to be entirely determined by the global skewness \( \alpha^g \). Specifically,

\[
\alpha = \delta \alpha^g \\
\alpha^* = \delta^* \alpha^g
\]

In the language of the dynamic model, the time-series variation in the skewness in each country will be only driven by the global skewness and not by idiosyncratic shocks. Although this is a simplification, I impose it in the interest of obtaining a global measure of risk and abstracting away from idiosyncratic risks.

In terms of the preferences of the economy, I assume that agents have power reward functions with risk aversion parameter \( \gamma \) and subjective discount factor \( \beta \). I also assume for simplicity that agents have complete home bias in the consumption of goods. This means the home agents will only consume good \( X \), and the foreign agents will only consume good \( Y \). Finally, I assume that all six agents currently constitute the same, one-sixth share of the overall economy.

### 1.2.2 Equilibrium and Solution of the Model

The social planner optimizes the weighted average of the expected utility of each agent with Pareto weights \( \lambda_i \) and \( \lambda_i^* \). Since the model has only two periods, the planner forms the optimal allocation by choosing next period consumption for each agent \( C_i \) and \( C_i^* \)

\[
\Pi = \lambda_1 \mathbb{E}_1 \left[ \beta \frac{C_i}{1-\gamma} \right] + \lambda_2 \mathbb{E}_2 \left[ \beta \frac{C_i}{1-\gamma} \right] + \lambda_3 \mathbb{E}_3 \left[ \beta \frac{C_i}{1-\gamma} \right] \\
+ \lambda_i^* \mathbb{E}_1^* \left[ \beta \frac{C_i^*}{1-\gamma} \right] + \lambda_i^* \mathbb{E}_2^* \left[ \beta \frac{C_i^*}{1-\gamma} \right] + \lambda_i^* \mathbb{E}_3^* \left[ \beta \frac{C_i^*}{1-\gamma} \right]
\]

(1.2)

with each expectation \( \mathbb{E}_i \) or \( \mathbb{E}_i^* \) taken over the subjective distribution \( \pi_i \) or \( \pi_i^* \) formed by the home or foreign agent. For this two-period model in which each agent constitutes an equal share
of the economy, the Pareto weights attached to agents are all equal by assumption. This can be interpreted as a particular snapshot of the dynamic model in Section 1.2.5 at a point in which the Pareto weights happen to be all equal like in the steady state.

As a result of complete home bias, the social planner satisfies the following budget constraints

\[ X = X_1 + X_2 + X_3 \quad \text{and} \quad Y = Y_1 + Y_2 + Y_3 \]  

(1.4)

where \( X_i \) is the home agent \( AG_i \)'s optimal consumption of the home goods \( X \) next period, and \( Y_i \) is the foreign agent \( AG_i^* \)'s consumption of the foreign goods \( Y \) next period. I will use \( X_{i,0} \) and \( Y_{i,0} \) to denote the current period consumption, which is equal across all agents based on our assumption that all agents currently are of equal size.

Upon solving the above optimization problem we can write down the allocation next period as

\[ C_i = \frac{\pi_i^{1/\gamma}}{\sum_{i=1}^{3} \pi_i^{1/\gamma}} \times X \]  

(1.5)

\[ C_i^* = \frac{(\pi_i^*)^{1/\gamma}}{\sum_{i=1}^{3} (\pi_i^*)^{1/\gamma}} \times Y \]  

(1.6)

for \( i \in \{1, 2, 3\} \). For the above derivation I will let the interested reader to refer to the solution of the dynamic model in Section 1.2.5. The solution tells us how the allocation at each state depends on the subjective distributions of all agents in the same country. An agent will consume optimally based on how his perceived probability of a state differs from that of the other agents. At an intuitive level, each agent will consume more in the states that he thinks are more likely to occur. For example, the optimist of the country will consume a large fraction of the endowment in a good state of the world. All agents within the country agree to disagree on each other’s belief and write down the contract to allocate the endowment accordingly.
1.2.3 Asset Pricing

The home (or foreign) interest rate can be computed via the marginal utility of any home (foreign) agent. To see this, the subjective marginal utility of a home agent can be written as

\[ \tilde{M}_i = \beta \left( \frac{C_i}{C_{i,0}} \right)^{-\gamma} \]

\[ = \beta \left( \frac{1}{3} \right)^{\gamma} \exp \left\{ -\gamma \epsilon X \right\} \left( \pi_{1/\gamma} + \pi_{2/\gamma} + \pi_{3/\gamma} \right) \frac{1}{\pi_i} \]

which is useful for pricing the bond in the home country

\[ B = \mathbb{E}_i [\tilde{M}_i] = \mathbb{E}_i \left[ \tilde{M}_i \frac{\pi_i}{\pi} \right] \]

where the expectation \( \mathbb{E} \) is taken over the true underlying objective distribution \( \pi \). The second equality helps write the price of a bond in terms of the true, objective marginal utility \( M \) by appending an adjustment term \( \pi_i/\pi \) to the subjective marginal utility. Hence, upon adjusting for an agent’s misperception, asset pricing can be done by any agent in the economy. The price of the bond can be written as

\[ B = \mathbb{E}_i \left[ \beta \left( \frac{1}{3} \right)^{\gamma} \exp \left\{ -\gamma \epsilon X \right\} \left( \sum_{\pi_i} (\pi_i)^{1/\gamma} \right) \frac{1}{\pi} \right] \]

I denote the interest rate on the risk-free bond as \( i = -\log(B) \) and similarly for the foreign country \( i^* \).

With the assumption that financial markets are complete, the change in the real exchange rate between the two countries is

\[ \Delta s = \log M^* - \log M \]

where a rise in \( \Delta s \) refers to an appreciation of the foreign currency.
1.2.4 Skewness in Forecasts and Carry Risk Premium

From a currency investor’s perspective, the excess return on investing in the foreign currency and shorting the home currency can be written as $cxy = i^* - i + \Delta s$. Since the carry trade is taking a long position in the currency of the higher interest rate country while shorting the other currency, the carry risk premium in levels can be written as

$$
carry \text{ risk premium} = \begin{cases} 
\log \mathbb{E}[\exp \{i^* - i + \Delta s\}] & \text{if } i^* > i \\
\log \mathbb{E}[\exp \{-(i^* - i) - \Delta s\}] & \text{if } i^* < i 
\end{cases}
$$

Let us consider the comparative statics of varying global skewness $\alpha^g$, which will drive each country’s skewness in forecast. I specify $\delta > \delta^*$, so that the home country’s forecast skewness $\alpha$ is driven by global skewness $\alpha^g$ to a larger extent than the foreign country. This is to generate difference in riskiness between the two countries. In our exercise, any change in global skewness $\alpha^g$ will drive the home forecasts to be more skewed than the foreign forecasts whether it be negative or positive.

The panels in Figure 1.2 present the carry risk premium and the interest rates for different levels of global skewness. As skewness in forecasts becomes more negative, the carry risk

![Figure 1.2: Global skewness, carry risk premia and interest rates. Comparative statics of changing global skewness $\alpha^g$, which drives home forecasts to be more skewed than foreign ($\delta > \delta^*$). The left and right panels show the resulting carry risk premium and the interest rates for each country, respectively.](image-url)
premium rises. The time-series implication that is testable in the data is that if more negative global skewness tends to be followed by higher average excess returns on the carry.

The right panel indicates that the skewness in forecasts has implications for the sorting of interest rates. Notice that on the negative domain of skewness, the home interest rate is lower than the foreign rate. Recall that on the negative domain, the home skewness $\alpha$ is more negative than foreign skewness $\alpha^*$. In this region, the pessimist of the home country is so pessimistic that he will drive up the overall demand for the risk-free bond for precautionary motives and thus push down the equilibrium interest rate at home. The opposite happens in the positive domain in which $\alpha > \alpha^*$ because the optimist drives down the overall demand for bonds. Thus, the home interest rate is higher than the foreign interest rate on the positive domain.

From the investor’s point of view, the sorting of interest rates is important. When the skewness in forecasts is negative, the carry trade would involve investing in the foreign currency and shorting the home currency, while the long-short position would be swapped when the skewness in forecasts is positive instead. In addition to determining the investment strategy for the carry, the home and foreign interest rates also affect the level of the carry risk premium. The risk premium will be determined by the sign and magnitude of the foreign exchange rate component relative to the interest rate differential.

Let us consider the negative domain, in which case the home forecasts are more negatively
skewed than the foreign forecasts. Figure 1.3 plots individual log consumption growth rates over the possible realizations of the endowment shocks. Notice that the optimal consumption of the unbiased agent in the foreign country \((AG^*_2)\) is roughly in the middle of that of the optimist and the pessimist. On the contrary, the unbiased agent of the home country \((AG_2)\) consumes more "like" the optimist \(AG_1\). This is because the subjective belief of the unbiased agent \(AG_2\) is close to that of the optimist \(AG_1\) yet very far from that of the pessimist \(AG_3\) as observed in Figure 1.1.

The economic interpretation of the consumption path of \(AG_2\) is that in the bad state of the world \(AG_2\) ends up providing large insurance to the pessimist of the home country, resulting in a low consumption for himself. Recall the determinants of each agent’s optimal consumption shown in Equation (1.5). Because of the subjective probability density in the numerator, each agent will want to consume more in the states that he thinks are the most likely. Since the bad state of the world is the state in which the pessimist is more correct, the pessimist enjoys a larger share of the pie, leaving less for \(AG_2\) to consume.

What does the consumption of \(AG^*_2\) and \(AG_2\) imply for asset pricing? Recall that these agents are those with the unbiased predictions about the underlying distribution of endowment shocks. This makes their consumption directly applicable for computing the (objective) marginal utility of consumption in each country

\[
M = \beta \left( \frac{C_2}{C_{2,0}} \right)^{-\gamma} \quad \text{and} \quad M^* = \beta \left( \frac{C^*_2}{C^{*2}_{2,0}} \right)^{-\gamma}
\]

so I will often refer to these agents as the marginal agents within the respective countries. The above marginal utility of consumption will drive the growth rate in the foreign currency value as previously mentioned

\[
\Delta s = \log(M^*) - \log(M)
\]

Based on the previous discussion, let us consider what happens upon a bad endowment shock in both home and foreign countries. The home marginal agent \(AG_2\), relative to \(AG^*_2\) in the foreign country, consumes much less in this bad state of the world because he is entitled to provide large insurance to the pessimist. The marginal utility of the home agent, therefore, goes up much
more relative to the foreign agent’s. That makes the foreign exchange rate to depreciate, which would be a loss to the carry investor because his investment currency, the foreign currency, is valued less in terms of the home currency. Upon a good endowment shock in both countries, the home marginal agent now consumes relatively more than the foreign marginal agent, instead. This is because the home marginal agent’s consumption contract is relatively similar to that of the home optimist, both of which serves as the party highly disagreeing with the home pessimist. Since the home marginal agent happened to be quite correct in making prediction, he enjoys a large consumption along with the optimist in the same country. Consequently, $\Delta s$ would appreciate, delivering a positive return to the carry investor. Importantly, notice that the above carry trade is risky strategy to implement because the investor loses in the bad state of the world. The no-arbitrage argument suggests that prices would adjust so that there should be a high risk premium for implementing this risky strategy.

In order to fully characterize the riskiness of the strategy in this two-country environment, I will also need to consider what happens in the case of a bad endowment shock in one country and a good shock in another country. Consider the case of a bad shock at home but a good shock abroad. The foreign marginal utility of consumption will be low, while the home marginal utility of consumption will be high, thus pushing down $\Delta s$ with the same sign. In addition, the home marginal agent suffers even more so because he had to provide significant insurance to the pessimist based on the promised allocation. In other words, this combination of bad and good shock makes a bad news even worse. This causes $\Delta s$ to drop significantly, which contributes to an even higher risk for the carry strategy.

Now let us consider the case of global skewness $\alpha^g$ being positive instead. Here the home skewness in forecasts $\alpha$ is larger than the foreign $\alpha^*$, so keep in mind that the investment currency is the home currency. The important distinction is that here $AG_2$ serves more “like” the pessimist $AG_3$ in that he buys insurance from the optimist. That means the home marginal agent enjoys a large share of the pie upon a bad shock, which implies that a bad news is actually not too bad. As a result, the overall riskiness of the carry is not as high as the negative skewness case. Similarly by the argument of no-arbitrage, the less risky case of positive skewness would yield a lower carry risk premium compared to the negative skewness case.

A reader may note that in Figure 1.2 the downward slope of the risk premium on the positive
domain is not as steep. This is due to the construction of the carry, in that the interest rate differential $-(i^* - i)(> 0)$ increases with $\alpha^9$, which pushes up the slope of the curve on the positive domain. Although the kink on the negative line could be of interest in future work, I argue that for now the negative association by itself is the primary object of interest. In my empirical exercise, I find no conclusive evidence regarding the magnitude or the specific location of a kink. As long as each country’s skewness is at least partially exposed to global skewness, the negative association between the skewness in forecasts and the carry risk premium should exist, and I only focus on this in this paper.

In summary, the negative association between the carry risk premium and the skewness in forecasts can be rationalized by the appropriate compensation for the consumption risk that is being faced by the marginal agent. When global skewness becomes more negative, in which case the home forecasts become more negatively skewed, the home marginal agent has to provide significant insurance to the pessimist in the bad state of the world thus causing the carry trade to become a risky strategy. In the empirical section, I test for the model implication in the time-series to see if the expected excess returns on the carry tend to be higher following a more negative global skewness.

One comment to be made is that I assumed that the home forecast skewness is driven by global skewness to a larger extent than the foreign forecast skewness. I could alternatively consider the flip side $\delta < \delta^*$. The interesting observation is that the carry risk premium again has the same pattern across global skewness $\alpha^9$. The carry strategy will swap for the negative and positive skewness domains, but because of the switching of the strategy, the riskiness of the carry is identical. Hence, in terms of focusing on the carry risk premium, the more important assumption is $\delta \neq \delta^*$ regardless of the specific order. For the purpose of my study, I do not pin down which countries in the data correspond to the home or the foreign country in the model. As long as the exposure to global skewness is different across countries, the negative association between skewness in forecasts and the carry risk premium remains.

The final comment to be made regarding the exposure to global skewness in the time-series implication is that in my paper I am entirely excluding the discussion on idiosyncratic skewness in an individual country. I am supposing any skewness in a country’s forecasts is entirely driven by the global skewness factor. Although idiosyncratic skewness by itself is an interesting avenue
of research, my empirical work faces data limitation in that, for some countries in my sample, skewness is not tightly measured enough to tease out information about the idiosyncratic component. Hence, in my theory and empirical sections, I focus entirely on the systematic component based on the argument that for a large enough cross-section of currencies, only the systematic risk should be priced.

1.2.5 Dynamic model

The static model described above serves the purpose of highlighting the intuition of a model with heterogeneous beliefs. In this section, I discuss the dynamic model, which served as the basis of the static version in terms of finding the optimal allocation. I develop an economy in which skewness in forecasts varies over time and thus drives the variation in the carry risk premium.

I similarly model an economy with two countries, each of which is occupied by three agents. Importantly, I allow time-variation in the beliefs of the optimists and the pessimists \( \{\mu_{1,t}, \mu_{3,t}, \mu_{1,t}^*, \mu_{3,t}^*\} \), while the beliefs of the unbiased agents \( \mu_{2,t} \) and \( \mu_{2,t}^* \) are kept equal to the true average growth rate \( \mu \). Specifically I let the variation in the beliefs to be driven by the time-variation in skewness in forecasts

\[
\alpha_t = \frac{\mu_{1,t} + \mu_{3,t} - 2\mu_{2,t}}{\mu_{1,t} - \mu_{3,t}} \quad \quad \alpha_t^* = \frac{\mu_{1,t}^* + \mu_{3,t}^* - 2\mu_{2,t}^*}{\mu_{1,t}^* - \mu_{3,t}^*} \tag{1.14}
\]

In particular, I model the time-series such that the variable for global skewness follows an AR(1) process

\[
a_t^g = \rho a_{t-1}^g + \sigma a \varepsilon_{a,t} \quad \text{with} \quad \varepsilon_{a,t} \sim N(0,1) \tag{1.15}
\]

which drives each country’s skewness in forecasts determined by

\[
a_t = \delta a_t^g \quad a_t^* = \delta^* a_t^g \tag{1.16}
\]

For each individual country I employ the mapping \( \alpha_t = a_t/\sqrt{1+a_t^2} \) (and similarly for \( \alpha_t^* \)) to resolve the issue that \( \alpha_t \) that is described in Equation (1.14) must be bounded between -1 and 1. As in the static model, I assume \( \delta > \delta^* \), meaning that home forecast skewness \( \alpha_t \) is more exposed
to the global skewness $\alpha^g_t$ compared to foreign skewness $\alpha^*_t$. The calibration can be found in the Appendix in Table A.2. Although one can implement a more general time-series model of beliefs, I impose the above structure to keep it stylized so that I can highlight the role of time-varying skewness in forecasts.

I adopt a model with heterogeneous beliefs laid out in Anderson, Ghysels, and Juergens (2005). The social planner maximizes

$$
\Pi = \sum_{i=1}^{3} \lambda_{i,0} \left( \sum_{t=0}^{T} \mathbb{E}_{i,0} \left[ \beta \frac{C_{i,t+1}^{1-\gamma}}{1-\gamma} \right] \right) + \sum_{i=1}^{3} \lambda^*_i \left( \sum_{t=0}^{T} \mathbb{E}_{i*,0} \left[ \beta \frac{C^*_{i,t+1}^{1-\gamma}}{1-\gamma} \right] \right)
$$

(1.17)

with initial Pareto weights $\lambda_{i,0}$ and $\lambda^*_i$. Again with perfect home bias and the budget constraint

$$
X_{1,t} + X_{2,t} + X_{3,t} = X_t \quad \quad \quad \quad \quad \quad \quad Y_{i,t} + Y_{2,t} + Y_{3,t} = Y_t
$$

(1.18)

(1.19)

for every $t$, the planner chooses consumption $\{C_{i,t}, C^*_i\}$ to solve for the Pareto-optimal allocation.

The first order condition with respect to $X_{1,t}$ can be written as

$$
\lambda_{1,0} \pi_{1,t} (\omega | \omega^0) \beta^t C_{1,t}^{1-\gamma} - \lambda_{3,0} \pi_{3,t} (\omega | \omega^0) \beta^t C_{3,t}^{1-\gamma} = 0
$$

(1.20)

$$
\lambda_{1,0} \pi_{1,t} (\omega | \omega^0) X_{1,t}^{1-\gamma} = \lambda_{3,0} \pi_{3,t} (\omega | \omega^0) \left( X_t - X_{1,t} - X_{2,t} \right)^{-\gamma}
$$

(1.21)

where the second line simplifies the first line and applies the budget constraint. $\omega$ denotes a state, i.e. a realization of the home and foreign endowment shocks. $\omega^t$ denotes the history of the path of realizations up to time $t$. Let us recursively define Pareto weights as

$$
\lambda_{i,t} (\omega | \omega^{t-1}) = \frac{\lambda_{i,t-1} (\omega^{t-1}) \pi_{i,t} (\omega | \omega^{t-1})}{\sum_{i=1}^{3} \lambda_{i,t-1} (\omega^{t-1}) \pi_{i,t} (\omega | \omega^{t-1}) + \sum_{i=1}^{3} \lambda^*_i (\omega^{t-1}) \pi^*_i (\omega | \omega^{t-1})}
$$

(1.22)

$$
\lambda^*_i (\omega | \omega^{t-1}) = \frac{\lambda^*_i (\omega^{t-1}) \pi^*_i (\omega | \omega^{t-1})}{\sum_{i=1}^{3} \lambda_{i,t-1} (\omega^{t-1}) \pi^*_i (\omega | \omega^{t-1}) + \sum_{i=1}^{3} \lambda^*_i (\omega^{t-1}) \pi^*_i (\omega | \omega^{t-1})}
$$

(1.23)

for each agent $i$, where the denominator serves the role of normalization so that the sum of all Pareto weights equal to 1. Using this definition, we can simply write the above first order condi-
tion as

\[ \lambda_{1,t}(\omega|\omega^{t-1})X_{1,t}^{-\gamma} = \lambda_{3,t}(\omega|\omega^{t-1}) \left( X_t - X_{1,t} - X_{2,t} \right)^{-\gamma} \]  

(1.24)

A similar first order condition with respect to \( X_{2,t} \) yields

\[ \lambda_{2,t}(\omega|\omega^{t-1})X_{2,t}^{-\gamma} = \lambda_{3,t}(\omega|\omega^{t-1}) \left( X_t - X_{1,t} - X_{2,t} \right)^{-\gamma} \]  

(1.25)

Combining these two, we can write the solution as

\[
C_{i,t} = \frac{\lambda_{1,t}^{1/\gamma}(\omega|\omega^{t-1})}{\lambda_{1,t}^{1/\gamma}(\omega|\omega^{t-1}) + \lambda_{2,t}^{1/\gamma}(\omega|\omega^{t-1}) + \lambda_{3,t}^{1/\gamma}(\omega|\omega^{t-1})} \times X_t
\]  

(1.26)

Similarly repeat above for the foreign agents’ consumption to obtain

\[
C_{i,t}^* = \frac{\lambda_{1,t}^{*1/\gamma}(\omega|\omega^{t-1})}{\lambda_{1,t}^{*1/\gamma}(\omega|\omega^{t-1}) + \lambda_{2,t}^{*1/\gamma}(\omega|\omega^{t-1}) + \lambda_{3,t}^{*1/\gamma}(\omega|\omega^{t-1})} \times Y_t
\]  

(1.27)

Notice that in the static model, which would be a snapshot of the dynamic setup, I do not have to keep track of the Pareto weights. Instead only the subjective beliefs \( \pi_i \) matter as displayed in Equation (1.5). Here, however, consumption also depends on the previous Pareto weights \( \lambda_{i,t-1} \).

One can think of \( \lambda_{i,t-1} \) as a term that accumulates the history of the realization of past consumption. The optimal consumption is then determined by agents’ subjective beliefs, adjusted for the current standing of the Pareto weights. Nonetheless, one can imagine the point in time such that all Pareto weights happen to be all equal. In that specific case, Equation (1.26) simplifies to a form like Equation (1.5). Consequently, the intuition of the result from the static model can be similarly applied to the dynamic model in that I expect the carry risk premium to be high when global skewness is more negative.

I perform a brief simulation exercise. Instead of simulating the model for a very long number of periods, I fix the number of periods to \( T = 20 \) and start over and repeat 100 times. The purpose of this is that I do not want to consider periods in which one of the agents ends up being infinitesimally small (\( \lambda_{i,t} \approx 0 \)) which can happen after many periods.

Based on the simulated data, I regress the subsequent carry trade returns onto the current
Table 1.1: Regression of the subsequent carry trade returns on global skewness $\alpha^g_t$ based on simulation. The regressor is standardized. Statistical significance is calculated based on Newey-West standard errors.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.0076***</td>
<td>(0.0025)</td>
</tr>
<tr>
<td>global skewness</td>
<td>-0.0096***</td>
<td>(0.0022)</td>
</tr>
<tr>
<td>AdjR2</td>
<td>0.0087</td>
<td></td>
</tr>
</tbody>
</table>

level of global skewness $\alpha^g_t$. Table 1.1 shows that the loading on global skewness is negative and statistically different from zero with sizable t-statistics. The magnitude of the loading is comparable to the magnitude of the unconditional risk premium, which is also positive and different from zero.

The above predictive regression suggests a negative time-series relationship between the carry risk premium and global skewness. The result is in line with the earlier discussion on the static model, and moreover it provides testable time-series implications.

1.3 Global Measures of Risk: Data Sources and Stylized Facts

Constructing a measure of conditional skewness on macroeconomic growth prospects is challenging. The standard Pearson measure of skewness requires high frequency of data points for an appropriate time window given our interest in a conditional measure of skewness. For our purpose, measuring the time-varying skewness of macroeconomic variables then becomes far from obvious because most macroeconomic indicators are available only at quarterly frequency. Instead, I use the third moment from the cross-sectional distribution of individual forecaster’s macroeconomic forecasts. I construct global measures of risks by: (i) computing cross-sectional moments of each country and then (ii) aggregating across countries to obtain global expected growth, global uncertainty and global skewness. In terms of aggregation, I take the simple average across countries, i.e. I take the average of each country’s expected growth across countries, and so forth. I also supply alternative specifications of aggregation as robustness tests, using the 1st principal component or taking the weighted average based on GDP weights.

The two primary data sources that provide individual forecasts for various countries are Consensus Forecasts and Bloomberg. Consensus Forecasts is a monthly periodical that has been sur-
veying reputable institutions of their forecasts of future macroeconomic variables for the major countries of the world. The publication provides individual forecasts of real GDP growth rates broken down by each forecaster. Bloomberg is another popular data source that makes available individual forecasts, but since only post-2008 data were available to me I augment Bloomberg data to the data from Consensus Forecasts. In addition, I also include a few country-specific forecasts datasets, one of which is New Zealand and is kindly provided by the Reserve Bank of New Zealand. New Zealand’s forecasts data are available through Bloomberg but the data from Consensus Forecasts was not available to me. Since the dataset from the Reserve Bank of New Zealand does not make available the name of the institution, I do not augment Bloomberg forecasts data. For Sweden and Switzerland, whose forecasts data are available by Consensus Forecasts and Bloomberg, the number of analysts who cover these countries are not sufficient for the early part of the sample, so I also augment national sources. The respective forecasts data have been generously provided by the National Institute of Economic Research (NIER) in Sweden and the KOF Swiss Economic Institute. Lastly, I include China starting from the first quarter of 2008, for which I only have Bloomberg’s individual forecasts data. To alleviate the concern that China’s economic growth plays a large role in global markets, I augment the cross-sectional first moment of real GDP growth forecasts in China all the way back to the beginning of 2000 using an alternative forecasts survey called Blue Chip Economic Indicators. Note that other measures, such as uncertainty and skewness, are not available for China for the period from 2000 to 2007 because of the lack of data on forecasts broken down by individual forecaster. I end up with the following list of countries: United States, United Kingdom, Japan, New Zealand, Germany, France, Sweden, Canada, Italy, Spain, Switzerland, and China. I choose the sample period from the first quarter of 1995 up to the first quarter of 2015 at quarterly frequency with the exception of Switzerland which starts in the second quarter of 1998 and China as just described. Table 1.2 shows the sample period for each country and the descriptive statistics on the number of forecasters.

Individual analysts respond to the survey by providing their own forecasts of real GDP growth rates, for example, for the current and next calendar years. I linearly interpolate the 1-year horizon growth rate from the current quarter up to 1 year ahead, based on the number of quarters remaining until the end of the year. For each country at every point in time, I construct the
<table>
<thead>
<tr>
<th>Country</th>
<th>Start date</th>
<th>Num. of forecasters</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>1995.q1</td>
<td>29</td>
</tr>
<tr>
<td>UK</td>
<td>1995.q1</td>
<td>28</td>
</tr>
<tr>
<td>Japan</td>
<td>1995.q1</td>
<td>20</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1995.q1</td>
<td>44</td>
</tr>
<tr>
<td>Germany</td>
<td>1995.q1</td>
<td>30</td>
</tr>
<tr>
<td>France</td>
<td>1995.q1</td>
<td>19</td>
</tr>
<tr>
<td>Sweden</td>
<td>1995.q1</td>
<td>16</td>
</tr>
<tr>
<td>Canada</td>
<td>1995.q1</td>
<td>16</td>
</tr>
<tr>
<td>Italy</td>
<td>1995.q1</td>
<td>15</td>
</tr>
<tr>
<td>Spain</td>
<td>1995.q1</td>
<td>14</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2000.q1</td>
<td>19</td>
</tr>
<tr>
<td>China</td>
<td>2008.q1</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 1.2: Start date of forecasts data and summary statistics of the number of forecasters. Since the number of forecasters for each country is changing through time, I report the quantiles.

The reader may have noticed that the number of analysts making projections for a given country along with the number countries to aggregate are important empirical considerations. I provide justification at the end of Section 1.4.1 by presenting a Monte Carlo exercise to show that the given number of analyst coverage and country coverage is sufficient for my analysis on predicting future carry returns.

One may raise the concern that the proposed measure of skewness is not a representative agent’s belief of the distributional shape of the growth rate. Although I have demonstrated the economic mechanism of the role of cross-sectional skewness in forecasts, I provide the following intuition on why the measure can relate to macroeconomic skewness. My argument is that one may view the survey as a collective group view of the forecasters and is informative about the...
prospects of the variable being forecasted. Similar to how the median forecast can serve as the expectation of growth rate, the extent of how dispersed the predictions are can serve as the proxy for variance. Moreover, if one notes a pronounced asymmetry in the distribution of predictions in that a fraction of respondents are making very low (or very high) predictions, then one may infer that there are some beliefs that the growth rate can tank significantly (or boost significantly), while there still prevails non-extreme beliefs about growth. Analogously, a negatively (or positively) skewed distribution of macroeconomic prospects indicates there are some chances of left-tail (or right-tail) events while the remaining mass of the distribution is at the non-extreme part of the domain. Hence, our cross-sectional skewness of forecasts can arguably be interpreted as a measure of skewness risk about the macroeconomic prospects.

The first three plots in Figures 1.4 show the time series of my global measures. My global expected growth measure appears procyclical and drops significantly especially during the recent financial crisis. The global uncertainty measure on the other hand increases significantly during the crisis and notably remains elevated for a while after the end of that NBER-designated recession. Global skewness, which will serve as the measure of interest in our empirical analysis, dis-
plays an interesting pattern a number of quarters before the recessions. What we can observe is that global skewness tends to be low and negative a number of quarters before the onset of each recession. Intuitively, this means that a fraction of forecasters makes significantly pessimistic predictions relative to the non-pessimistic crowd at periods before a recession begins. When a bad event actually realizes, most of the survey respondents revise their predictions downward, so that the skewness of the distribution is no longer low. It is precisely this dynamic that I believe captures important time-series information about global macroeconomic risks.

Data on personal consumption and population are mostly from national sources and have been downloaded through Datastream. These data include: Federal Statistical Office of Germany, State Secretariat for Economic Affairs of Switzerland, Cabinet Office of Japan, Australian Bureau of Statistics, Statistics New Zealand and Statistics Norway. World Bank and IMF International Financial Statistics have been also used for population data.

Foreign exchange data are obtained primarily from Thomson Reuters through Datastream. I obtain foreign exchange spot rates and 3-month forward rates on 33 currencies from Thomson Reuters: United Kingdom, Japan, New Zealand, Australia, Sweden, Switzerland, Norway, Canada, South Africa, Singapore, Denmark, Euro, Austria, Belgium, Finland, France, Germany, Greece, Italy, Netherlands, Portugal, Spain, Ireland, South Korea, Czech Republic, Hungary, India, Malaysia, Mexico, Philippines, Poland, Taiwan, and Thailand. However, the data I have from Thomson Reuters only go back to the end of 1996, so for the period of 1995 through the third quarter of 1996 I use the 3-month interest rates and the spot rates for the major and Euro-joining currencies: United Kingdom, Japan, New Zealand, Australia, Sweden, Switzerland, Norway, Canada, South Africa, Singapore, Denmark, Austria, Belgium, Finland, France, Germany, Greece, Italy, Netherlands, Portugal, Spain, and Ireland. The sample period differs for different currencies either because of foreign exchange regimes, unreliable volatile periods, or data unavailability. The details about the foreign exchange data that I use can be found in Table A.3 in the Appendix.

1.4 Empirical Results

The empirical highlight of this paper is to show that my global measures of risks can predict carry trade returns. I provide robustness exercises in the subsequent section.
1.4.1 Predictive Regressions of Currency Returns

A common approach to understand the currency market is to study the returns to the carry trade. In practice this trading strategy involves taking long positions in currencies with high forward discount and taking short positions in those with low forward discount. This is roughly equivalent to forming long-short portfolios based on the aforementioned interest rate differentials, given that the covered interest rate parity approximately holds.

Based on the ordering of the forward discount, which is defined as the difference between the log forward rate $f^i_t$ and the log spot rate $s^i_t$, I separate currencies into usually five buckets from the highest to the lowest. For exercises that restrict the investable set of currencies to, a smaller number of currencies, say the G10, then I separate them into three buckets instead. The dynamic strategy means that I take long positions in the currencies in the high bucket and short positions in the currencies in the low bucket, while re-balancing the portfolio every quarter based on the sorting of currencies. For the static strategy, I instead form a static portfolio based on the time series average of the forward discount for each currency. As an example with the most actively traded currencies in the world, the static strategy would involve taking long positions on the Australian dollar, New Zealand dollar, and Norwegian Krone, while taking short positions on the Deutsche Mark (soon replaced by the Euro), Swiss Franc, and Japanese Yen. The predictive regression exercise is to regress the next-quarter carry trade returns onto one or more of our global measures of risks $X_t$:

$$x_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1} \quad (1.28)$$

where $x_{t+1}$ indicates the return on the carry trade strategy from time $t$ to the next quarter $t+1$, which consists of a long position in the high bucket and a short position in the low bucket. The currency excess return on a single foreign exchange rate $i$ is defined as $x_{t+1}^i = s_{t+1}^i - f_{t+1}^i$, and the portfolio return on a particular bucket is the average of the individual excess return $x_{t+1}^i$ for the currencies in the bucket. For the period before 1996.Q4, in which I use bonds data, the excess returns on the currency $i$ are defined as $x_{t+1}^i = i_t^i - i_t^{iUS} + \Delta s_{t+1}$ and the forward discount is defined as $f_{t+1}^i = i_t^i - i_t^{iUS}$.

Table 1.3 presents the main regression results of my exercise. For ease of interpretation I stan-
Table 1.3: Predictive regression results of next-quarter carry trade portfolio returns onto global measures. The top two panels are based on static carry trades, and the bottom two panels are based on dynamic carry trades. The regressors $x_t^g$, $v_t^g$, $sk_t^g$ correspond to global expected growth, global uncertainty, and global skewness, respectively, and all are standardized. Statistical significance is calculated based on Newey-West standard errors.
standardize all global measures so that each has a mean of 0 and a standard deviation of 1. Panels A and B show the results for the static carry, and panels C and D show for the dynamic carry. Panels A and C present the results for portfolios formed using the G10 currencies, one of which is the Euro which replaced the Deutsche Mark in 1999. Panels B and D correspond to portfolios formed using all 33 currencies, in which the set of currencies being considered is changing through time (See Table A.3 in Appendix). One can observe that the static carry trade returns based on the major currencies loads significantly on the global expected growth and global skewness with a negative sign and loads positively on global uncertainty, given the regression is done separately. If all three regressors are used altogether, then global expected growth and global skewness remain significant predictors of the carry trade returns. Moving onto the second panel, we can see that global uncertainty is no longer a significant predictor of the returns. Global skewness remains a reliably significant predictor of carry trade returns, while the loadings on global expected growth is not particularly significant. Likewise, the bottom panels C and D on the dynamic carry trade shows that the signs are consistent with those in the top panels A and B.

The economic magnitude of global conditional skewness is large. For the case of the dynamic carry trade based on a large set of currencies regressed on all three global measures, the loading on global skewness is -0.0131. Since my global measure is standardized, the coefficient suggests that a one standard deviation decline in global skewness indicates a rise in 5.24% risk premium per annum. Many of the other regressions suggest a per annum effect of at least 4%. Therefore, global conditional skewness risk appears to contribute to the time-variation in the carry risk premium with large economic significance.

One notable pattern is that global skewness seems to be a more robust predictor in explaining the carry trade based on a large list of currencies. Figure 1.5 presents a visualization of how the estimates change as we include more and more currencies in constructing portfolios. The two panels present the estimated loadings as well as the 90% confidence intervals plotted against the number of currencies that I use in forming portfolios. The left panel shows the beta estimates for global expected growth, and the right figure shows them for global skewness. We can see that as we utilize a larger number of currencies in constructing portfolios, we obtain more reliable negative estimates for global skewness. That means that the predictive ability of global skewness
Total number of currencies (unbalanced panel) $eta$ on global expected growth

Figure 1.5: Estimated beta loadings for different sets of currencies. Above are based on the dynamic carry portfolio. Estimated $\beta$s and their 90% confidence intervals are shown. We have an unbalanced panel of currencies, so the number of currencies changes over time.

becomes stronger in describing a larger set of currencies in the world, which includes not just the major currencies but also a number of emerging market currencies. On the contrary, global expected growth loadings become less negative as we consider a larger number of currencies. Hence we can argue that global expected growth has less predictive power in explaining the risk premia for a wider universe outside of the G10 currencies.

The negative loadings on global conditional expected growth and global conditional skewness inform us about the time-varying compensation for risk in the currency risk premium. When global conditional expected growth is low, the currency risk premium on the carry trade portfolio is high, meaning that there is a large risk premium arising from pessimistic prospects on global expected growth. Similarly, when global conditional skewness is low, or negative, the carry trade offers a high risk premium due to the perception of a negatively skewed distribution of global prospects, i.e., a high chance of a very significant downturn in the global economy. Conditional on such cases, the carry trade portfolio is considered risky, thus offering a high expected return.

What is notable is that the predictive power of global conditional skewness remains significant
when the recent crisis period is excluded. I define the recent crisis period consistent with the corresponding NBER recession, i.e., the fourth quarter of 2007 through the second quarter of 2009. Upon excluding 7 quarters of this period, I repeat the above exercise and find that carry trade returns load significantly on global conditional skewness (see Table A.4 in Appendix). I uncover that global conditional skewness has strong predictive ability for explaining the currency markets even during normal periods. This is a distinct feature from the disaster literature, as one might expect negative skewness to be only about the possibility of a very bad event. Instead, global skewness continues to explain next period currency returns in normal times.

The loading on global uncertainty merits some discussion as the literature has emphasized the role of volatility in explaining asset returns. From various regression exercises, I conclude that the direction of the predictive ability of global uncertainty appears consistent with the literature but not statistically strong in our context. The economic story of a positive loading is that when global uncertainty is perceived to be high, the carry trade portfolio tends to yield high expected excess returns, meaning that there is a large risk premium when global uncertainty is ex ante high. This is consistent with the argument that when agents expect high economic uncertainty, they require high compensation for investing those risky currencies. As noted before, however, the statistical significance of the loading on the second moment is not very pronounced. If all three regressors are included on the right hand side of the regression, the coefficient on global uncertainty is never statistically different from zero. Although global economic uncertainty does have explanatory power in currency returns, it is usually subsumed by the other moments. Note that global conditional skewness contains the information about the direction of the risks, in that a negative skewness is very different from a positive value. Hence, we may argue that skewness contains information about whether the impending uncertainty is good or bad. Since skewness effectively informs the sign of the uncertainty, the role that the second moment can play is relatively diminished in explaining returns. Given the relatively weaker explanatory power of global uncertainty, for subsequent analyses I exclude the results for it.

I have also repeated the work with alternative procedures of constructing global measures. Recall that I have taken the simple average across countries in aggregating country-specific values to single global measures. Instead I have tried taking the first principal component for countries, for which the entire history is available. This leaves 10 countries for analysis, ignoring Switzer-
land and China. The results are presented in Table A.5. I have alternatively tried an aggregation method of taking the average weighted by each country’s GDP share. The results are in Table A.6. These robustness exercises generally convey a consistent message that global expected growth and global skewness seem to have predictive ability in explaining carry trade returns.

I further show predictive regressions using alternative measures of skewness in Table A.7. I denote the component of global conditional skewness that is not explained by global expected growth as \( sk_{it}^{g,+} \). In other words, I regress global skewness onto global expected growth \( x_{it}^g \) and take the residuals that are not explained by \( x_{it}^g \). The first and third columns show that this measure still significantly predicts carry trade returns. The second and fourth columns use the alternative measure that is constructed as \( \sqrt[3]{sk_{it}^3} \times \sqrt{v_{it}} \). It measures the third moment raised to the power of one third and captures the direction of the uncertainty. I find evidence that this measure is also a significant predictor of the carry trade returns, which is consistent with the equity return results shown in Colacito, Ghysels, Meng, and Siwasarit (2016).

We may take a closer look at the carry trade regressions by examining the individual portfolios. Recall that carry trade is a high minus low strategy, i.e., it takes a long position in the high portfolio and takes a short position in the low portfolio. What we can instead study is to look at the returns on the high and low portfolios as well as the intermediate portfolios. We can take the time series of the returns on each portfolio, regress them onto our global measures of risks and compare the loadings across the portfolios.

Table 1.4 presents the results of regressing the individual portfolio returns onto global conditional skewness. The first column repeats the loadings on global skewness for reference, while the latter columns correspond to the high bucket portfolio, the middle, and the low, respectively. One can observe that the beta coefficients are negative and statistically significant for the high and medium portfolio, while the loading for the low portfolio is close to zero. A similar pattern holds in the case of the dynamic trading strategy.

Observing the results for the static portfolios based on a large list of currencies, we can observe an apparent pattern in the individual portfolio loadings. We can see that there is roughly a monotonic pattern in the loadings in that the high portfolio has the most negative loading, and the magnitude of the loading becomes smaller as we look at the subsequent portfolios. The pattern is similar with the dynamic portfolios, if we think of p4 through p2 as roughly similar port-
Table 1.4: Predictive regressions of individual buckets of currencies onto global skewness. Panels A and B are based on static carry trades, and panels B and C are based on dynamic carry trades. p1 indicates the portfolio of currencies with the lowest forward discount. p3 (or p5) indicates the portfolio with the largest forward discount, given the G10 currencies (or entire set of currencies). Statistical significance is calculated based on Newey-West standard errors.
### Table 1.5: Predictive regressions of individual buckets of currencies onto global expected growth.

Panels A and B are based on static carry trades, and panels B and C are based on dynamic carry trades. p1 indicates the portfolio of currencies with the lowest forward discount. p3 (or p5) indicates the portfolio with the largest forward discount, given the G10 currencies (or entire set of currencies). Statistical significance is calculated based on Newey-West standard errors.

<table>
<thead>
<tr>
<th>Panel A</th>
<th>Static carry: G10 currencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>carry</td>
</tr>
<tr>
<td>intercept</td>
<td>0.0042 (0.0049)</td>
</tr>
<tr>
<td>(x_t^g)</td>
<td>−0.0097*** (0.0026)</td>
</tr>
<tr>
<td>AdjR2</td>
<td>0.0356</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>Static carry: all currencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>carry</td>
</tr>
<tr>
<td>intercept</td>
<td>0.0052 (0.0043)</td>
</tr>
<tr>
<td>(x_t^g)</td>
<td>−0.0037 (0.0040)</td>
</tr>
<tr>
<td>AdjR2</td>
<td>−0.0056</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C</th>
<th>Dynamic carry: G10 currencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>carry</td>
</tr>
<tr>
<td>intercept</td>
<td>0.0085 (0.0054)</td>
</tr>
<tr>
<td>(x_t^g)</td>
<td>−0.0066*** (0.0025)</td>
</tr>
<tr>
<td>AdjR2</td>
<td>0.0099</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D</th>
<th>Dynamic carry: all currencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>carry</td>
</tr>
<tr>
<td>intercept</td>
<td>0.0138*** (0.0041)</td>
</tr>
<tr>
<td>(x_t^g)</td>
<td>0.0002 (0.0030)</td>
</tr>
<tr>
<td>AdjR2</td>
<td>−0.0126</td>
</tr>
</tbody>
</table>
folios. We can conclude that the the returns on the higher forward-discount portfolios load more negatively on global skewness, thus explaining the high minus low carry strategy that I showed earlier.

The results for regressing on global expected growth are presented in Table 1.5. Although we can find a similar monotonic pattern if the set of currencies was limited to the G10 currencies, this is not necessarily the case if we include many other currencies. The loadings on the 'high' bucket are not significantly negative and are not necessarily larger in magnitude than the others. Therefore, I argue that global conditional skewness seems to be the stronger predictor that produces a monotonic pattern when comparing across the portfolios formed on the forward discount.

In summary, the empirical results show that my measures of global expected growth and global skewness have predictive ability in explaining the carry trade returns. In particular, the explanatory power of global skewness becomes more robust as I include a larger number of currencies.

Before I conclude this section, I address the following concern about robustness with regards to whether I have sufficient analyst coverage in constructing a robust measure of global skewness. To be more precise, I need enough number of analysts projecting forecasts for a given country and also for a sufficient number of countries, over which I aggregate. To justify this I conduct the following Monte Carlo exercise. Taking the data of quantile-based skewness in forecasts for each country as given, for every point in time \( t \) and for every country \( c \) I generate \( I_{c,t} \) random analysts’ forecasts from the skewed distribution based on Matlab’s random number generation from a Pearson system that intakes the observed skewness \(^2\). I then calculate the quantile-based skewness of on the randomly generated \( I_{c,t} \) analysts’ forecasts. Repeat this for every country to construct the global measure of skewness and then repeat this for all \( t \). Run a time-series regression of the dynamic carry returns with the set of all available currencies onto the randomly generated measure of global skewness to obtain an estimate of the loading. Repeat this for a 1,000 simulations and then obtain the distribution of the estimated loading. The 90% confidence interval of the estimated loading is \((-0.0142, -0.0015884)\), indicating that the Monte Carlo simulation exercise indicates my given data of analyst coverage is sufficient in terms of justifying the predictive

\(^2\)I need a mapping between quantile-based skewness and the usual skewness which is the normalized third central moment. I use Matlab’s Pearson system random number generator to obtain a numerical mapping. As inputs, I vary the value of skewness and hold fix \( \mu = 0.02, \sigma = 0.0187, \text{kurt}=1.5 \).
1.4.2 Robustness

In order to provide evidence that global skewness is indeed a robust predictor, I consider a few variables known to have explanatory power for the foreign exchange market. The first variable I consider is the innovations to liquidity $\Delta\text{liquidity}_t$, where liquidity is proxied by the negative of the TED spread (LIBOR minus the 3-month Treasury Bill rate), retrieved from the FRED. The literature has documented that changes in liquidity can help predict subsequent carry trade returns as shown in Brunnermeier, Nagel, and Pedersen (2009) and Bakshi and Panayotov (2013).

Table 1.6 shows the results when next-quarter carry trade returns are regressed jointly on global skewness and the liquidity innovation. The predictive ability of the carry trade returns remains robust, when the liquidity channel is controlled for.

I also consider two other explanatory variables that are contemporaneous instead of lagged. I consider the innovations in foreign exchange volatility $\Delta\text{fxvol}_t$, in which foreign exchange volatility is defined as

$$\text{fxvol}_t = \frac{1}{9} \sum_{i=1}^{9} \sqrt{\left( \sum_{\tau \in \text{quarter } t} (\Delta s_{\tau}^{\text{daily}})^2 \right)}$$

for $i \in \{\text{G10 currencies}\}$. This variable is in line with Menkhoff, Sarno, Schmeling, and Schrimpf (2012) who find that the long-end of the carry trades tends to deliver low returns during periods of unexpected high global FX volatility. Following their argument, unexpected volatility proxy should be a contemporaneous variable instead of a predictor, namely that the timing of unexpected volatility $\Delta\text{fxvol}_{t+1}$ is consistent with the timing of the carry trade returns $\text{cxr}_{t+1}$.

Another explanatory variable of interest is the growth rate of a commodity index called the Commodity Research Bureau BLS Spot Index, retrieved from Datastream. Despite the proposed relationship between commodity prices and foreign exchange rates, there is mixed evidence of whether commodity prices can predict future foreign exchange rates (Chen, Rogoff, and Rossi (2010)). I instead consider the contemporaneous innovation $\Delta\text{commod}_{t+1}$.

Table 1.6 shows that the predictive ability of global skewness remains robust with the inclusion of either explanatory variable. Notice that these contemporaneous variables increase the $R^2$.
by a large extent. In addition, correlations among the covariates likely change the coefficient estimates. Nonetheless, global skewness appears to be a statistically significant predictor of future carry trade returns.

A similar exercise for the carry based on the G10 currencies is reported in the Appendix (Table A.8). The predictive power of global skewness given a control for the liquidity proxy remains significant. Given the control for the role of FX volatility or commodity, the role for global skewness is statistically weak, but the lack of power is due to the high $R^2$ arising from the contemporaneous variables.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: dynamic</th>
<th>Panel B: static</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.0136***</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td>$sk^2_t$</td>
<td>-0.0131***</td>
<td>-0.0173***</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(0.0046)</td>
</tr>
<tr>
<td>$\Delta\text{liquidity}_t$</td>
<td>2.8612*** (0.7725)</td>
<td>2.5495** (1.1047)</td>
</tr>
<tr>
<td>$\Delta\text{fxvol}_{t+1}$</td>
<td>-0.9988 (0.6031)</td>
<td>-1.4434*** (0.4860)</td>
</tr>
<tr>
<td>$\Delta\text{commod}_{t+1}$</td>
<td>0.1754 (0.1192)</td>
<td>0.2957*** (0.0973)</td>
</tr>
<tr>
<td>AdjR$^2$</td>
<td>0.1511</td>
<td>0.1773</td>
</tr>
</tbody>
</table>

Table 1.6: Regressions of carry trade returns onto global skewness and other explanatory variables. Above carry trades are formed based on all available set of currencies. The left three columns are based on the dynamic carry, and the right three columns are based on the static carry. $\Delta\text{liquidity}_t$ is a lagged innovations to liquidity, defined as the minus of TED spread. $\Delta\text{fxvol}_{t+1}$ is a contemporaneous innovations to the foreign exchange volatility constructed from the G10 currencies. $\Delta\text{commod}_{t+1}$ is a contemporaneous growth rate of the CRB BLS Spot Index.

### 1.4.3 Trading Conditional on Global Skewness

I have argued that the risk premium on the carry has a time-varying component that is dependent on the proposed global measures of risks. Given the observable predictive pattern, we can come up with a new trading strategy conditioning on the information at each point in time. Since global skewness negatively predicts future carry returns, I can swap the long-short position of the carry when global skewness is high. In particular, I follow the carry strategy if global skewness $sk^2_t$ is less than some threshold $\bar{s}$, but I swap the long-short positions of the carry if
Figure 1.6: Comparison of cumulative returns to the ordinary carry and the new strategy. The new strategy involves implementing the carry, except when $sk^d > \bar{s}$, in which case we swap the long-short positions. Above are based on the threshold $\bar{s} = \sqrt{\text{Var}(sk^d)}$. 
To avoid switching the strategy due to measurement error, I take the threshold \( \bar{s} \) to be a positive value instead of zero. For the following exercise I present the case of \( \bar{s} = \sqrt{\text{sk}_t^2} \).

<table>
<thead>
<tr>
<th></th>
<th>Panel A: G10</th>
<th>Panel B: All currencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Static</td>
<td>Dynamic</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0042</td>
<td>0.0055</td>
</tr>
<tr>
<td>Std.dev</td>
<td>0.0446</td>
<td>0.0477</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>0.0942</td>
<td>0.1152</td>
</tr>
</tbody>
</table>

Table 1.7: Summary statistics of the ordinary carry returns and the new strategy returns. The new strategy involves implementing the carry, \( \text{sk}_t^2 > \bar{s} \), except when \( \text{sk}_t^2 > \bar{s} \), in which case we swap the long-short positions. Above are based on the threshold \( \bar{s} = \sqrt{\text{sk}_t^2} \). All values are quarterly and shown in decimals.

Figure 1.6 shows the comparison between the cumulative returns on the carry trade and the cumulative returns on the new strategy for a particular specification. One can visually see that the new strategy tends to deliver a more stable time-series of cumulative returns. Table 1.7 provides summary statistics of the comparison between the ordinary carry and the new strategy. We can see that the new strategy that conditions on the information on global skewness yields more attractive Sharpe ratios.

1.5 Conclusion

In this paper I construct global measures of macroeconomic risks, constructed from the cross-section of GDP forecasts, and find that they have predictive ability in explaining carry trade returns. I motivate the discussion by building a consumption-based asset pricing model with heterogeneous agents to highlight the role of the cross-sectional skewness in forecasts, which is the novel contribution of this paper. I empirically find that the measures of global expected growth and global skewness negatively predict carry trade returns, and especially global skewness appears to be a robust predictor that can price a large set of currencies. Hence, I provide novel evidence that the carry risk premium is significantly driven by the variation in global macroeconomic skewness risk.
CHAPTER 2  RISK AND RETURN TRADE-OFF IN THE U.S. TREASURY MARKET
(with Eric Ghysels, Anh Le, and Haoxiang Zhu)

Introduction

Does higher risk lead to a higher expected excess return in the U.S. Treasuries market? If yes, is the short-run component or the long-run component of risk important in determining bond risk premiums. To answer these fundamental questions, we need an accurate measure of risk and an accurate account of returns predictability in the data. Most term structure models to date have difficulties providing both simultaneously. For example, Gaussian affine models entirely miss time-varying volatilities in the data, whereas affine models with stochastic volatility typically fail to generate the patterns of return predictability documented by Campbell and Shiller (1987).\footnote{See, for example, Dai and Singleton (2002) and Joslin and Le (2012). In the context of no-arbitrage affine term structure models, Joslin and Le (2012) discuss in depth why there must be trade-offs between fitting volatility and the predictability of bond returns.} Reasonable volatility processes, such as the GARCH model of Bollerslev (1986) or the EGARCH model of Nelson (1991), rarely make their way into the no-arbitrage term structure literature, partly because one loses the analytical solutions to bond prices if those stochastic volatilities are imposed on the risk-neutral dynamics of yields. Although the ARCH/GARCH literature and the term structure literature are both well established, synergies from bridging them have rarely been explored. It is unfortunate given the ample evidence that ARCH models can offer a good characterization of interest rate volatility (see, for example, K. G. Koedijk and Wolff, 1997; Brenner, Harjes, and Kroner, 1996; and Christiansen, 2005).

In this paper, we propose a discrete-time model that integrates the advantages of both the affine term structure models and the GARCH models of volatility. Not only do we retain the tractability of the affine models, we also inherit the ability of GARCH models to accurately capture time-varying volatilities of yields. The key to our approach is an asymmetric treatment of
conditional volatility under the physical (P) and risk-neutral (Q) measures. Since conditional variances under P and Q need not be the same in a discrete-time setup, we let the Q-conditional variances of yields be Gaussian (so analytical solutions to bond prices are retained), while letting the P-conditional variances follow a GARCH-type process.

In addition, following the ARCH-in-mean literature, pioneered by Engle, Lilien, and Robins (1987), we allow the P-conditional variances to affect the physical drifts of yields. More specifically, in the spirit of Engle and Lee (1999), we let the P conditional variances of yields to be driven by a long-run component and a short-run component, each of which follows its own GARCHI-like process with different degrees of persistence. From these two processes, we construct the short-run and long-run volatility components of two equal-weighted yield portfolios: a near-maturity portfolio and a far-maturity portfolio. These four volatility components allow us to differentiate the contributions to risk premiums of long-run and short-run volatility components of both the long end and short end of the yield curve. Through these four components, volatility, as a measure of risk, can potentially forecast future yields and bond excess returns, hence the risk-return trade-off. Importantly, because the feedback of volatility into the physical drifts happens entirely under P, it does not interfere with tractable bond pricing under Q.

Using weekly Treasury yields data from January 1962 to August 2007, we find a significantly positive relation between risk and return in the U.S. Treasury market. A higher conditional volatility this week predicts a lower yield level next week, thus a higher bond excess return. Notably, it is the short-run component of volatility, not the long-run component, that matters for return predictability. According to our estimates, the long-run component has a half-life of more than 60 years, whereas the short-run component has a half-life of about two years. Therefore, the risk-return relation is unlikely the results of transitory shocks or trading frictions, which we expect to move at much higher frequencies; nor is it driven by the extremely persistent time trend in volatility. The return-predicting short-run volatility component moves at roughly the business-cycle frequency, which we find reasonable and intuitive.

---

2 See, e.g., Le, Singleton, and Dai (2010) for a discrete-time stochastic volatility model in which P- and Q-conditional variances are distinct.

3 The long-run component may appear extremely persistent, but such persistence is not uncommon. For example, in a related context, Stock and Watson (2007) propose to include a random walk component to the (log) volatility of the inflation process. They find support for this model using U.S. inflation data over a similar sample period to ours. Given the intimate relation between nominal yields and inflation, it is not surprising that we detect a similar degree of persistence in the long-run volatility component of yields.
Moreover, the return-predicting power of the short-run volatility component predominantly comes from the short-end of the yield curve. Volatilities of far-maturity yields do not have additional predictive power for future yields once we control for volatility of the short end. Further, the slope and the curvature of the yield curve are not predicted by any volatility components that we study.

Putting all together, our main evidence is that a higher short-run volatility component of near-maturity yields predicts a lower yield level next week and thus a higher excess return. The economic magnitudes are large. For example, the volatility factor accounts for 14%, 42%, and 40% of the predictable components of weekly excess returns on one-year, five-year, and ten-year zero coupon bonds, respectively. (The other predictive factors are the principal components of yields.) Our findings are consistent with and complementary to results reported by Joslin (2013), who also finds a component of volatility risk that is important for explaining expected excess returns on bonds. Different from our study, Joslin (2013) does not distinguish the long-run and short-run components of volatility; nor does he differentiate the volatility of the short-maturity yields from that of the long-maturity yields.

Related Literature

Our paper contributes to the literature on term structure modeling. Most affine term structure models with stochastic volatilities imply that the conditional variances of yields are perfectly explained, or “spanned,” by yields themselves. By imposing this spanning condition, these affine stochastic volatility models are potentially restrictive in at least two aspects. First, there is considerable evidence that volatility is not fully spanned by yields (see Collin-Dufresne and Goldstein, 2002; Collin-Dufresne, Goldstein, and Jones, 2009; and Andersen and Benzoni, 2010). Second, imposing a spanning condition can induce a “cross-measures” tension, in the sense of Joslin and Le (2012), that could prevent a no-arbitrage model from fully capturing the predictability patterns of bond returns in the data. Our model addresses both issues by decoupling the risk-neutral conditional variances from their physical counterparts. As a result, not only do we

\footnote{Specifically, in an affine setup, volatility factors must be autonomous to remain strictly positive under both the \( P \) and \( Q \) measures. Applied to spanned volatilities, this autonomy requires that the \( P \) and \( Q \) feedback matrices share some common left-eigenvectors. Understandably, the resulting closeness between the \( P \) and \( Q \) feedback matrices limits the ability of the model to explain returns predictability.}
match the volatility dynamics of yields very well, our model also closely replicates the return predictability patterns documented by Campbell and Shiller (1987).

We emphasize that the departure from the spanned models is our key difference from previous attempts to build ARCH/GARCH volatility into no-arbitrage term structure models. For example, the volatility factors in Longstaff and Schwartz (1992) have been interpreted as following a GARCH process. Additionally, Heston and Nandi (1999) and Haubrich, Pennacchi, and Ritchken (2012) also use a GARCH process to explicitly model the volatility of interest rates. Nevertheless, in light of work by Dai and Singleton (2000), it is obvious that these models belong to the “completely affine class”. Because the GARCH volatility is one of the factors that determine bond yields in these models, it means that volatility is strictly “spanned” by yields. These models are thus subject to the critique by Andersen and Benzoni (2010) and others on spanned models.

Our modeling approach using a GARCH-like volatility process complements a few existing approaches to modeling interest rate volatility in a no-arbitrage setup. For example, compared with the unspanned stochastic volatility models by Collin-Dufresne and Goldstein (2002) and Collin-Dufresne, Goldstein, and Jones (2009), our model does not restrict model parameters to eliminate the spanning of conditional volatilities by yields. Our approach also differs from the “general stochastic volatility model” proposed by Trolle and Schwartz (2009) and the high-frequency approach proposed by Cieslak and Povala (2013). A GARCH-like model allows us to identify yield volatility dynamics with considerable precision over a long sample period, not restricted by the availability of high-frequency data.

There have been numerous attempts to go beyond the affine paradigm. Examples include regime-switching models (e.g. Bansal and Zhou, 2002; Bansal, Tauchen, and Zhou, 2003; Ang and Bekaert, 2002; and Dai, Singleton, and Yang, 2007), affine-quadratic models (e.g. Ahn, Dittmar, Gao, and Gallant, 2003; Leippold and Wu, 2002; and Ahn, Dittmar, and Gallant, 2002), and other nonlinear models (e.g. Ahn and Gao, 1999; and Feldhutter, Heyerdahl-Larsen, and Illeditsch, 2013). We complement these studies by offering a simple yet flexible model for interest rate volatility. For example, the fitted volatilities of regime-switching models will likely  

---

5 These restrictions require that the volatility factors of a model have certain mean reversion rates in order to result in an exact cancellation of the convexity effects. See Joslin (2013) for an in-depth discussion.
herit the Gaussian property of constant variances under each regime.\(^6\) Ahn, Dittmar, Gao, and Gallant (2003) conclude that the affine-quadratic models they consider are not “able to fully capture term structure volatility.” The approach proposed by Feldhutter, Heyerdahl-Larsen, and Illeditsch (2013) generates time-varying volatility through the convexity effect, although their approach in general does not allow for more flexible volatility structures nor the separation between long-run and short-run components of volatility. Additionally, since we retain the tractability of affine bond pricing, estimation of our models is much more convenient than most of the above models.

We add to the ARCH-in-mean literature in two important aspects. First, because both the yield volatility and the principal components of yields can forecast future yields, our model allows two determinants of the dynamics of risk premiums: the quantity of risk and the market price of risk. Evidence supporting the simultaneous presence of both channels can be found in Dai and Singleton (2000) and Dai and Singleton (2002) as well as our motivating exercises in the next section. By contrast, ARCH-in-mean models typically only allow the quantity of risk, but not the market price of risk, to explain risk premiums (see, for example, Engle, Lilien, and Robins, 1987; and Adrian and Rosenberg, 2008). Second, whereas the ARCH-in-mean models are typically applied to bond excess returns of individual maturities, our model is designed to match bond prices/yields as well as bond excess returns of all maturities, tied together by the no-arbitrage condition.\(^7\) Our model, therefore, offers a more coherent characterization of the risk-return trade-off in Treasury markets.

### 2.1 Motivating Exercises

In this section, we conduct several simple exercises that shed light on the potential relations between risks and returns in the bond markets. These will serve as guidance for us in designing our models for subsequent analysis.

\(^6\) For example, in the last figure of Dai, Singleton, and Yang (2007), the fitted volatilities effectively flip between two values – the constant volatilities under each regime.

\(^7\) To see how matching prices/yields is a much stronger requirement than matching excess returns, note that for a model with a stochastic discount factor \(M_{t+1}\) to match excess returns \(R^e_{t+1}\), the requirement is that \(E_t[M_{t+1}R^e_{t+1}] = 0\). However, any scaled version of such a discount factor will also match excess returns equally well. As a result, the requirement to match excess returns cannot pin down the conditional mean of the stochastic discount factor \(E_t[M_{t+1}]\), the one period bond price.
Our first exercise is to estimate an ARCH-in-mean model in the spirit of Engle, Lilien, and Robins (1987) (ELR):

\[ xr_{n,t+1} = \alpha + \delta h_t + e_{t+1} \]

(2.1)

where \( xr_{n,t+1} \) denotes the weekly excess return on an \( n \)-period zero coupon bond, starting at time \( t \) and realized one week later, at time \( t + 1 \). \( h_t \) denotes the conditional volatility of the shocks \( e_{t+1} \) and \( h_t \) is assumed to follow an ARCH process with thirteen weekly lags (one quarter):

\[ h_t^2 = \alpha_A + \sigma_A \sum_{i=1}^{L} w_i e_{t-i}^2, \]

(2.2)

where \( L = 13 \). In ELR’s implementation of the ARCH process, the loadings on the lag squared residuals (\( w_i \)'s) are set to fixed constants to avoid estimation uncertainty. ARCH models may indeed involve quite a few parameters, which led to the popularity of GARCH models which we will consider below. We adopt first an ARCH specification without parameter proliferation, opting for a flexible, but parsimonious lag structure. This is reminiscent of MIDAS polynomials and therefore follow a setup where the \( w_i \)'s are hyper-parameterized via a normalized beta probability density function, \( w_i = (L + 1 - i)\theta - 1 / \sum_{i=1}^{L}(L + 1 - i)^\theta - 1 \). This weighting scheme (through one parameter \( \theta \)) is parsimonious but nonetheless is known to reasonably capture volatility dynamics in the data.\(^8\)

The expected component of equation (2.1) says that the expected excess returns, \( E_t[xr_{n,t+1}] \), should be linearly related to bonds’ volatility as given by \( h_t \). If the estimate of \( \delta \) is positive and statistically significant, it suggests that there is a positive risk-return relationship in bonds markets.

We use zero yields data from Gurkanyak, Sack, and Wright (2007) (GSW).\(^9\) The data is sampled weekly, starting in January 1962 and ending in August 2007.\(^10\) The GSW dataset is useful

\(^8\) For more discussion on MIDAS polynomial specifications, see e.g. Ghysels, Sinko, and Valkanov (2007) and Ghysels (2013).


\(^10\) Our sample ends in August 2007 for at least two reasons. First, as is currently typical for the term structure literature, we avoid the era with near-zero interest rate in the wake of the global financial crisis. Second, recent work has shown that for equities the risk-return relation breaks down during financial crisis as flight-to-quality
for our purposes because it is considerably smoothed. Thus, volatility measures constructed from this data are less susceptible to the influence of outliers. Nonetheless, the short-maturity yields from the GSW data involve a high degree of extrapolation. As such, for maturities shorter than six months, we bootstrap zero yields from CRSP’s raw bond prices using the standard Fama-Bliss algorithm. We implement the estimations of equations (2.1) and (2.2) using QMLE.

Table 2.1 reports the results for twenty different maturities, from 6- to 120-month. Each row corresponds to a different maturity \( n \) in (2.1). The second column, labelled ARCH-M, reports the estimates of \( \delta \). Standard errors are calculated from the Newey-West covariance matrix, constructed using thirteen lags. Consistent with ELR, we find that the estimates of \( \delta \) are significantly positive across the entire maturity spectrum, suggesting a positive risk-return relationship.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>ARCH-M ( \delta )</th>
<th>GARCH-M ( \delta )</th>
<th>OLS: PC ( \delta_{PC1} \times 10^4 ), ( \delta_{PC2} \times 10^4 ), ( \delta_{PC3} \times 10^4 )</th>
<th>GLS: ARCH and PC ( \delta_{ARCH} ), ( \delta_{PC2} \times 10^4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6m</td>
<td>0.21***</td>
<td>0.19***</td>
<td>0.20, 0.10, -0.34</td>
<td>0.11*, -0.00</td>
</tr>
<tr>
<td>12m</td>
<td>0.23***</td>
<td>0.24***</td>
<td>0.27, 0.81*, 0.94</td>
<td>0.13*, 0.71***</td>
</tr>
<tr>
<td>18m</td>
<td>0.18***</td>
<td>0.19***</td>
<td>0.31, 1.23*, 1.47</td>
<td>0.17**, 1.11**</td>
</tr>
<tr>
<td>24m</td>
<td>0.16***</td>
<td>0.16***</td>
<td>0.35, 1.63*, 1.87</td>
<td>0.21***, 1.43*</td>
</tr>
<tr>
<td>30m</td>
<td>0.14***</td>
<td>0.14***</td>
<td>0.39, 2.02**, 2.14</td>
<td>0.22***, 1.69*</td>
</tr>
<tr>
<td>36m</td>
<td>0.12***</td>
<td>0.13***</td>
<td>0.41, 2.40**, 2.31</td>
<td>0.23***, 1.93*</td>
</tr>
<tr>
<td>42m</td>
<td>0.10**</td>
<td>0.11***</td>
<td>0.44, 2.79**, 2.39</td>
<td>0.22***, 2.17*</td>
</tr>
<tr>
<td>48m</td>
<td>0.09**</td>
<td>0.10***</td>
<td>0.46, 3.17***, 2.40</td>
<td>0.21***, 2.42*</td>
</tr>
<tr>
<td>54m</td>
<td>0.08**</td>
<td>0.09***</td>
<td>0.48, 3.55***, 2.34</td>
<td>0.19***, 2.69*</td>
</tr>
<tr>
<td>60m</td>
<td>0.08**</td>
<td>0.08***</td>
<td>0.50, 3.92**, 2.24</td>
<td>0.18***, 3.00*</td>
</tr>
<tr>
<td>66m</td>
<td>0.08**</td>
<td>0.08***</td>
<td>0.52, 4.30**, 2.08</td>
<td>0.16***, 3.34**</td>
</tr>
<tr>
<td>72m</td>
<td>0.08**</td>
<td>0.08***</td>
<td>0.54, 4.67**, 1.90</td>
<td>0.15**, 3.75**</td>
</tr>
<tr>
<td>78m</td>
<td>0.08**</td>
<td>0.08***</td>
<td>0.56, 5.03***, 1.67</td>
<td>0.14**, 4.21**</td>
</tr>
<tr>
<td>84m</td>
<td>0.09**</td>
<td>0.08**</td>
<td>0.58, 5.39***, 1.42</td>
<td>0.13**, 4.70**</td>
</tr>
<tr>
<td>90m</td>
<td>0.09**</td>
<td>0.08**</td>
<td>0.61, 5.75***, 1.14</td>
<td>0.11*, 5.21***</td>
</tr>
<tr>
<td>96m</td>
<td>0.10**</td>
<td>0.08***</td>
<td>0.63, 6.10***, 0.84</td>
<td>0.10*, 5.72***</td>
</tr>
<tr>
<td>102m</td>
<td>0.11***</td>
<td>0.08***</td>
<td>0.66, 6.45***, 0.50</td>
<td>0.10* 6.22***</td>
</tr>
<tr>
<td>108m</td>
<td>0.12***</td>
<td>0.09***</td>
<td>0.69, 6.79***, 0.15</td>
<td>0.10* 6.70***</td>
</tr>
<tr>
<td>114m</td>
<td>0.13***</td>
<td>0.09***</td>
<td>0.73, 7.12***, -0.23</td>
<td>0.09  7.13***</td>
</tr>
<tr>
<td>120m</td>
<td>0.13***</td>
<td>0.09**</td>
<td>0.76, 7.45***, -0.63</td>
<td>0.09  7.51***</td>
</tr>
</tbody>
</table>

Table 2.1: Predictability of weekly excess returns from ARCH volatilities, GARCH volatilities and yields PCs. *, **, *** denote the conventional significance levels of 1%, 5%, and 10%, respectively.
In the second exercise, we replace the ARCH process in (2.2) by a GARCH(1,1) process:

\[ h_t^2 = \alpha_G + \beta h_{t-1}^2 + \sigma_G e_t^2. \] (2.3)

All else is kept identical to the first exercise. The estimates of \( \delta \) are reported in the third column, labelled GARCH-M, of Table 2.1. Notably, the estimates of \( \delta \) are all positive and statistically significant. Note that the GARCH volatility can be written as an infinite sum of lagged squared residuals, \( h_t^2 = \text{constant} + \sigma_G \sum_{i=0}^{\infty} \beta^i e_{t-i}^2 \). The similar significant patterns between the ARCH-M and GARCH-M columns of Table 2.1 suggest that more recent volatility may be sufficient in capturing the risk premia in bond markets.

The \( \delta \) estimates for both the ARCH-M and GARCH-M columns have a similar pattern: they are highest (most positive) at the 12-month maturity, then decreasing as maturity increases to about five years, and finally flattening out for longer maturities. The weaker risk-return relation along the maturity spectrum suggests that (a) excess returns become less predictable as maturity lengthens, (b) volatility becomes less predictive for longer-dated bonds, or both (a) and (b). Either way, this evidence suggests that distinguishing short-maturity volatility and long-maturity volatility may prove fruitful for a coherent understanding of the risk-return relationship across all maturities.

In the two exercises implemented so far, we only allow the time-varying quantity of risk, \( h_t \), to predict excess returns. Absent from this setup is an independent role for time-varying market prices of risks in determining bond risk premia. This role is provided by the large literature on Gaussian term structure models, in which the quantity of risk (yields volatility) is assumed constant and thus all returns predictability is generated by the time variation in market prices of risks.

In the third exercise, we run a simple OLS regression that predicts weekly excess returns using three principal components (PCs) of yields:

\[ x_{r,n,t+1} = \alpha_{PC} + \delta_{PC1} PC1_t + \delta_{PC2} PC2_t + \delta_{PC3} PC3_t + e_{t+1}. \] (2.4)

In standard affine Gaussian term structure models, these three PCs (PC1-3) reasonably capture the time variation in the state variables, which also govern the market prices of risks implied by
these models. We construct the PCs from yields with maturities of 6 months, 1 year, 2 years, 3 years, 5 years, 7 years, and 10 years. As is standard, the first three PCs are the level, slope, and curvature of the yield curve, respectively.

Estimated coefficients for the OLS regression in (2.4) are reported in the columns under the heading “OLS:PC” of Table 2.1. Consistent with the established results in the literature (for example, Campbell and Shiller (1991) and Fama and Bliss (1987)), we find that the slope factor is strongly predictive of future excess returns, particularly so for the longer dated bonds.

In the last exercise, we combine the two sets of predictive variables, volatilities and yield PCs, in the following regression:

$$x_{r,n,t+1} = \alpha_{PC} + \delta_{ARCH} h_t + \delta_{PC1} PC1_t + \delta_{PC2} PC2_t + \delta_{PC3} PC3_t + e_{t+1}, \quad (2.5)$$

where we use the ARCH volatility $h_t$ implied from the first exercise. To save space, we only report estimates of $\delta_{ARCH}$ and $\delta_{PC2}$ in the last two columns of Table 2.1. We see that volatility is a significant predictor beyond the three PCs, and it is significantly if it is constructed from 1-year to 8-year yields.

Taken together, the exercises in this section suggest that (1) yields volatilities can forecast excess returns above and beyond the information embedded in the current yield curve, (2) information in more recent volatility seems sufficient in determining bond risk premia, and (3) volatilities constructed from short-maturity yields and long-maturity yields have different information content for excess returns. These three observations directly drive our modeling approach in the next section.

### 2.2 Model

In this section, we formally develop a model of the term structure of interest rates in discrete time.

---

12 We choose these maturities since they represent the most liquid segments of the yield curve. These maturities are also covered by the Federal Reserve Board’s H.15 releases.
2.2.1 The Risk-Neutral Dynamics and Bond Pricing

Our model specification for the risk-neutral dynamics is standard. The vector of state vari-
ables, $X_t$, follows a Gaussian VAR(1) under the risk-neutral ($Q$) measures, and the short rate, $r_t$, is a linear function of $X_t$:

\[ X_{t+1} = K_0^Q + K_1^Q X_t + \epsilon_{t+1}^Q \quad \text{with} \quad \epsilon_{t+1}^Q \sim N(0, \Sigma_X), \quad (2.6) \]

\[ r_t = \delta_0 + \delta_1^* X_t. \quad (2.7) \]

It immediately follows from the affine structure of the setup that zero-coupon yields at all matur-
ities are affine in $X_t$:

\[ y_{n,t} = A_{n,X} + B_{n,X} X_t \quad (2.8) \]

with the loadings $(A_{n,X}, B_{n,X})$ obtained from standard yield pricing recursions.

2.2.2 The Time-Series Dynamics

We assume that $X_t$ follows affine dynamics with conditionally Gaussian innovations under the
physical ($P$) measure:

\[ X_{t+1} = K_0 + K_1 X_t + K_V V_t + \epsilon_{t+1} \quad \text{with} \quad \epsilon_{t+1} \sim N(0, \Sigma_t), \quad (2.9) \]

where matrices $K_0$, $K_1$, and $K_V$ are $N \times 1$, $N \times N$, and $N \times M$, respectively. We now turn to the
specifications of the conditional variance, $\Sigma_t$, and the GARCH-in-mean term, $V_t$.

Dynamics of Conditional Variance $\Sigma_t$

Given the evidence provided in Section 2.1 that recent volatilities might contain sufficient in-
formation for excess returns, we explicitly model the long-run and short-run components of $\Sigma_t$ in
the spirit of Engle and Lee (1999) (EL):

\[
\Sigma_t = S_t + L_t, 
\]

\[
S_t = \rho_S S_{t-1} + \alpha (\epsilon_t \epsilon_t' - \Sigma_{t-1}),
\]

\[
L_t = \Sigma_X (1 - \rho_L) + \rho_L L_{t-1} + \phi (\epsilon_t \epsilon_t' - \Sigma_{t-1}),
\]

where \(\rho_S, \rho_L, \alpha,\) and \(\phi\) are all positive scalars. \(\Sigma_X\) is the same variance matrix in (2.6) and thus is positive semi-definite (psd). Clearly, the Gaussian models are obtained as a special case when \(\rho_S = \rho_L = \alpha = \phi = 0.\)

The interpretation of this model is straightforward. The total variance matrix \(\Sigma_t\) is decomposed into a short-run component, \(S_t,\) and a long-run component, \(L_t.\) This decomposition is a simple way to differentiate the impact of recent volatilities on returns dynamics from that of distant volatility information. Each component follows its own autoregressive process with different persistence, captured by \(\rho_S\) and \(\rho_L.\) Without loss of generality, we impose the restriction that \(\rho_S < \rho_L.\) In both equations (2.11) and (2.12), the last term, \((\epsilon_t \epsilon_t' - \Sigma_{t-1}),\) represents news about volatility. A piece of volatility news dissipates at a faster rate for \(S_t\) than for \(L_t.\)

In addition, the lack of the intercept term in the AR(1) process of \(S_t\) implies that the population mean of \(S_t\) is zero. In this sense, \(L_t\) is a low-frequency trend component of \(\Sigma_t,\) whereas \(S_t\) is a high-frequency, transitory component around zero.

Finally, to guarantee that \(\Sigma_t\) is strictly positive definite, we impose the restriction:

\[
1 > \rho_L > \rho_S > \alpha + \phi,
\]

in addition to the positivity requirement for \(\rho_S, \rho_L, \alpha,\) and \(\phi.\) Condition (2.13) is imposed by EL in their univariate setting. Along the lines of proofs in EL, we can show that condition (2.13) implies that \(\Sigma_t\) and \(L_t\) are positive definite in our multivariate model.\(^{13}\)

\(^{13}\)In particular, the long-term component, \(L_t,\) can be expressed as an infinite-order polynomial of the product \(\epsilon_t \epsilon_t'.\) Under condition (2.13), all coefficients of this polynomial are non-negative, which guarantees that \(L_t\) is positive definite. To see how the positive definiteness of \(\Sigma_t\) follows from this, we add equations (2.11) and (2.12) side by side to arrive at:

\[
\Sigma_t = \Sigma_X (1 - \rho_L) + (1 - \rho_L) L_{t-1} + (\rho_S - \alpha - \phi) \Sigma_{t-1} + (\alpha + \phi) \epsilon_t \epsilon_t'.
\]

Conditions (2.13) and the positivity of \(\alpha\) and \(\phi\) means that all the scalar loadings are positive. Hence, as long as \(L_{t-1}\) is positive definite, so is \(\Sigma_t,\) by induction.
We observe that our model naturally gives rise to unspanned stochastic volatility (USV) under the $\mathbb{P}$ measure. By construction, the conditional variance $\Sigma_t$, the sum of lagged “squared residuals,” is not spanned by $X_t$. In this regard, our model presents a significant departure from the traditional affine models with spanned volatilities. Collin-Dufresne and Goldstein (2002), Collin-Dufresne, Goldstein, and Jones (2009), Andersen and Benzoni (2010), among others, show the importance of allowing for USV in bond markets.

The GARCH-in-Mean Term $V_t$

As in the (G)ARCH-in-mean literature, the role of $V_t$ is to summarize volatility information relevant for forecasting excess returns. In our multivariate setting, the challenge is dimensionality. For an $N$-factor model ($N$ being the dimension of $X_t$), there are $N(N + 1)$ unique entries in the matrices $S_t$ and $L_t$. A typical three-factor model has 12 conditional variance and covariances. Clearly, including all of these elements in $V_t$ would make the model over-parameterized.

To keep the model parsimonious, we focus on the conditional volatilities of two particular portfolios of bond yields. The first portfolio is an equal-weighted portfolio of yields with maturities 1, 2, 3, 4, and 5 years. The second is an equal-weighted portfolio of yields with maturities 6, 7, 8, 9, and 10 years. This choice is motivated from the exercise in Section 2.1 that yield volatilities inferred from short to medium range of the yield curve contain information for excess returns beyond the yield curve factors; yield volatilities inferred from longer maturities of yields, less so. This separation is nondegenerate as long as $N \geq 2$. By separating the two yield portfolios by maturities, we will uncover the impacts of volatilities in the short and long ends of the yield curve. We emphasize that the choice of the cutoff point, 5 years, is not critical to our results. We also estimate model using a cutoff point of 3 years, and the results are similar.

With two yield portfolios and two horizons, $V_t$ contains four elements. These four entries can be explicitly derived from the conditional variance $\Sigma_t$ in the following way. We denote by $B_{1-5,X}$ and $B_{6-10,X}$ the weighting vectors of the two portfolios on the factor $X_t$. Thus, by (2.8), the conditional variance of the first yield portfolio is

$$
\text{Var}_t \left[ \frac{1}{5} \sum_{n=1}^{5} y_{n,t+1} \right] = B_{1-5,X} \Sigma_t B_{1-5,X}' .
$$

(2.14)
A similar calculation gives the conditional variance of the second yield portfolio, with maturities 6–10 years. The long-term variance of these two portfolios can be computed in a similar manner by replacing $\Sigma_t$ with $L_t$. Putting them together, we write

$$V_t = \begin{pmatrix}
\sqrt{B_{1-5,X} \Sigma_t B'_{1-5,X}} & - \sqrt{B_{1-5,X} L_t B'_{1-5,X}} \\
\sqrt{B_{6-10,X} \Sigma_t B'_{6-10,X}} & - \sqrt{B_{6-10,X} L_t B'_{6-10,X}}
\end{pmatrix}.$$  \hfill (2.15)

The first (second) element of $V_t$ is the short-run (long-run) volatility component of the yield portfolio with maturities 1–5 years. The third (fourth) element of $V_t$ is the short-run (long-run) volatility component of the yield portfolio with maturities 6–10 years. Note that the first element of $V_t$ is written this way, instead of “$\sqrt{B_{1-5,X} S_t B'_{1-5,X}}$,” because $S_t$ needs not be positive definite. But $L_t$ is.

**Implications on Bond Excess Returns**

Combining (2.8) and (2.9), we can write the one-period expected excess return on the $n$-period bond as:

$$E_t[x_{n,t+1}] = \text{constant} + (nB_{n,X} - (n - 1)B_{n-1,X}K_1 - B_{1,X})X_t$$

$$- (n - 1)B_{n-1,X}K_V V_t.$$  \hfill (2.16)

Clearly, we capture a volatility component as well as a pure yield curve component of bond risk premia. As $n$ varies, equation (2.16) gives us a term structure of risk-return relations across the maturity spectrum.

We note that the derivation of expected excess returns in (2.16) can also be obtained through the stochastic discount factor implied by our model. Specifically, given our specifications of the $P$ and $Q$ dynamics, the implied stochastic discount factor can be written as:

$$M_{t,t+1} = e^{-r_t \int_t^{t+1} f_t^Q(X_s) \, ds} \int_t^{t+1} f_t^P(X_s) \, ds.$$  \hfill (2.17)
where $f^Q_t(X_{t+1})$ and $f^P_t(X_{t+1})$ denote the conditional densities of $X_{t+1}$ given the time-$t$ information set under the $Q$ and $P$ measures, respectively. Since the state variables $X_t$ share a common support under both measures, $M_{t,t+1}$ defines a valid and strictly positive pricing kernel (which rules out arbitrage).

To summarize, our model is fully characterized by the risk neutral dynamics in (2.6) and (2.7), the time series dynamics in (2.9), and the associated volatility dynamics given by (2.10), (2.11), and (2.12). The full parameter set is given by: $\Theta_X = (\delta_0, \delta_1, K_0^Q, K_1^Q, \Sigma_X, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi)$.

### 2.2.3 Econometric Identification

#### Canonical Setup

To obtain econometric identification, we apply the standard rotations of the affine term structure literature (see, for example, Dai and Singleton, 2000). Since the state variables in our setup are not bounded, any rotation from $X$ to $Z = U_0 + U_1 X$ for any $(U_0, U_1)$ is admissible, as we show in Appendix A.2. In other words, for any affine transformation of $X$ to $Z$, we will obtain another observationally equivalent instance of our model, characterized by the same set of equations (2.6)-(2.13), with a different parameter set $\Theta_Z$. Explicit mappings between $\Theta_X$ and $\Theta_Z$ (as functions of $(U_0, U_1)$) are provided in Appendix A.2.

By rotating the state variables freely, we obtain econometric identification using the canonical setup of Joslin, Singleton, and Zhu (2011). Specifically, under this canonical setup, $\delta_1$ is a vector of ones, $K_0^Q$ is a vector of zeros, and $K_1^Q$ is of Jordan form. Thus, the risk-neutral means of the state variables are zeros, and the intercept term in the short rate equation, $\delta_0$, becomes the risk-neutral long-run mean of the short rate $r_t$. Adopting the notation of Joslin, Singleton, and Zhu (2011), we replace $\delta_0$ by $r^Q_\infty$ in the canonical setup. Thus, the econometrically identified parameter set becomes $\Theta_X = (r^Q_\infty, K_1^Q, \Sigma_X, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi)$.
Rotation to Observable Yield Portfolios

Consider a $J \times 1$ vector of yields $y_t$. The affine structure of bond yields,

$$y_t = A_X + B_X X_t,$$

(2.18)

means that any non-degenerate yields portfolios $\mathcal{P}_t$ characterized by a $N \times J$ loadings matrix $W$ must be affine in the states: $\mathcal{P}_t = WA + WB X_t$.

Our ability to freely rotate once again means that we can rotate $X$ to $\mathcal{P}$ and thus we can simply replace our canonical model by one in which $\mathcal{P}_t$ serves as state variables. As shown by Joslin, Singleton, and Zhu (2011) and Joslin, Le, and Singleton (2012), using yields portfolios, which are observable, as state variables greatly enhances the (numerical) identification of model parameters. With $\mathcal{P}_t$ as the state variables, we denote the parameter set by:

$$\Theta_{\mathcal{P}} = (r^Q, K_1^Q, \Sigma_{\mathcal{P}}, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi).$$

The parameters $(r^Q, K_1^Q, \rho_S, \alpha, \rho_L, \phi)$ are invariant to rotations and thus are identical across $\Theta_X$ and $\Theta_{\mathcal{P}}$. The remaining parameters are rotation-specific. For example, $\Sigma_X$ refers to the $Q$ conditional variance of the latent state variables under the canonical setup, whereas $\Sigma_{\mathcal{P}}$ refers to the $Q$ conditional variance of $\mathcal{P}_t$. As in Joslin, Singleton, and Zhu (2011), the first three parameters $(r^Q, K_1^Q, \Sigma_{\mathcal{P}})$ determine the loadings of bond yields with $\mathcal{P}_t$ as states. The next three parameters $(K_0, K_1, K_V)$ determine the $\mathcal{P}$ conditional mean of $\mathcal{P}_t$ (given by equation (2.9) replacing $X_t$ by $\mathcal{P}_t$). The last four parameters $(\rho_S, \alpha, \rho_L, \phi)$ together with $\Sigma_{\mathcal{P}}$ determine the dynamics of the $\mathcal{P}$ conditional variances of $\mathcal{P}_t$ (through equations (2.10), (2.11), (2.12) with $\Sigma_X$ replaced by $\Sigma_{\mathcal{P}}$). Because the volatility factors, $V_t$, are volatility components of observable yield portfolios, they remain invariant to factor rotations. (See Appendix A.2 for more details.)

Joslin, Le, and Singleton (2012) show that by letting $\mathcal{P}_t$ be the lower order principal components (PCs) of bond yields, estimation of the model is least sensitive to assumptions regarding the observational errors of bond yields. Using this observation, we will use a representation of our model with $\mathcal{P}_t$ being the first $N$ PCs of bond yields in our empirical implementation. As a result, subsequent mentions of state variables should be understood as references to the first $N$
2.2.4 Discussion of Modeling Choices

Our model can be viewed as a generalization of Gaussian no-arbitrage models, which can be recovered from our setup by setting $\Sigma_t$ to a constant matrix ($\Sigma_P$), and $K_V$ to zeros. As explained earlier, we choose an asymmetric approach in which only the $P$ conditional variances are stochastic, whereas the conditional variances under $Q$ remain constant. This asymmetric treatment is possible because diffusion invariance need not hold in a discrete-time model.

Keeping the $Q$-volatility constant has the benefit of parsimony. If generality were the objective, one would set the $Q$ conditional variances stochastic too, as long as analytical pricing remains feasible. One such model under $Q$ is provided by Le, Singleton, and Dai (2010) (a discrete-time counterpart to the stochastic volatility models in Dai and Singleton (2000)), in which the conditional variances are time-varying but affine in states in a way that affine pricing of yields is preserved.

We argue, however, that as long as yields are affine in states, the direct effects of alternative $Q$ dynamics on risk premiums are insensitive to the volatility structure under $Q$. That is, as far as risk premium is concerned, setting a constant variance under $Q$ is almost without loss of generality. The remaining of this subsection goes through the logic of this argument, based on analysis by Joslin and Le (2012).

Consider the expression for expected excess returns in (2.16). The direct channel that alternative $Q$ dynamics have on risk premiums is through the yields loadings $B$. The other parameters ($K_1, K_V$) come from the $P$ dynamics. Joslin and Le (2012) show that estimates of yields loadings $B$ are very similar across different affine models with distinct volatility structures. Intuitively, since the $Q$ dynamics is typically strongly identified in the data, minimizing cross-sectional pricing errors has priority in maximum-likelihood estimations. Thus, regardless of the volatility structure, estimates of yield loadings $B$ in affine models are typically very close to the unconstrained estimates obtained by regressing yields onto yields PCs.

Different $Q$ dynamics can also have indirect effects on the dynamics of risk premiums if they can somehow influence the estimates of the time series parameters $K_1$ and/or $K_V$ in (2.16). Joslin and Le (2012) provide one example in the context of affine stochastic volatility models,
in which the $Q$ conditional variances are at least partially priced in the cross-section of yields and thus are partially spanned by yields. For these models, Joslin and Le (2012) show that the positivity requirement under $Q$ for volatility can impose strong constraints on the time series parameters. Specifically, in an affine setup, volatility factors must be autonomous to remain strictly positive under both the $P$ and $Q$ measures. When applied to spanned volatilities, this autonomy requires that the $P$ and $Q$ feedback matrices share some common left-eigenvectors. The resulting closeness between the $P$ and $Q$ feedback matrices brings the model closer to the expectations hypothesis (in which the $P$ and $Q$ conditional means are the same). As a result, these constraints can prevent a model from fully explaining the risk-return relation in the data—the very focal point of our study.

There are at least two ways to address the tension documented by Joslin and Le (2012). One is our approach, in which the time series parameters ($K_1, K_V$) are unambiguously unconstrained since they are free parameters. The other is to adopt the risk-neutral setup of an affine stochastic volatility model but then impose constraints on the risk-neutral parameters so that our model exhibits (completely) unspanned stochastic volatility (USV, see Collin-Dufresne and Goldstein, 2002; and Collin-Dufresne, Goldstein, and Jones, 2009). But with volatility not priced in the cross-section, the yields pricing equation induced by USV and the yields pricing equation (2.8) of our setup are observationally equivalent. In both cases, what appear on the right hand side of (2.8) are pure yield curve factors and not stochastic volatility. It follows that either way the models’ implications for bond risk premia—the decomposition of expected excess returns into a volatility component and a pure yield curve component in (2.16)—will be similar.\footnote{The distinction between our model and a USV model is more pronounced for nonlinear securities such as calls or puts. This is not our focus, however.} Thus, for simplicity, we maintain the assumption that the conditional variance matrix is constant under $Q$.

\section{Results}

In this section we present the estimation results of the model of Section 2.2. As is standard, we use $N = 3$ factors. Again, the three factors are chosen to be the first three PCs of bond yields, denoted by $P_t$.\footnote{The distinction between our model and a USV model is more pronounced for nonlinear securities such as calls or puts. This is not our focus, however.}
2.3.1 Estimation

We use the same dataset over the same sample period as described in Section 2.1. We also adopt the same procedure in constructing the PCs of bond yields. Following Joslin, Le, and Singleton (2012), we assume that the first three PCs of bond yields are observed perfectly. As is standard, we assume that the remaining higher-order PCs, denoted by $\mathcal{P}_e$, are observed with i.i.d. uncorrelated Gaussian errors with one common variance. That is,

$$\mathcal{P}^o_{e,t} = \mathcal{P}_{e,t} + e_t \quad \text{and} \quad e_t \sim N(0, I\sigma_e^2),$$

(2.19)

where $\sigma_e$ is a scalar; and the superscript $^o$ indicates an observed quantity as opposed to a theoretical construct. Let $W_e$ denote the loading matrix corresponding to the higher-order yield PCs. Then, $\mathcal{P}^o_{e,t} = W_e g^o_t$, whereas the theoretical counterpart of $\mathcal{P}^o_{e,t}$ can be computed by $\mathcal{P}_{e,t} = W_e (A_P + B_P P_t)$. Recall that the yields loadings $A_P$ and $B_P$ can be obtained from the loadings $A_X$ and $B_X$ in (2.18) with necessary adjustments to account for the rotation from $X$ to $P$ (see Appendix A.2 for details).

The likelihood function of the observed data, $\mathcal{L}$, is given by:

$$\mathcal{L} = \sum_t f(\mathcal{P}_{t+1}|\mathcal{P}_t) + f(\mathcal{P}_{e,t+1}|\mathcal{P}_{t+1}),$$

(2.20)

where $f$ denotes log conditional density. The first term captures the density of the time-series dynamics and can be written as:

$$\sum_t f(\mathcal{P}_{t+1}|\mathcal{P}_t) = constant - \frac{T}{2} \log(|\Sigma_t|) - \frac{1}{2} \sum_t ||\Sigma_t^{-1/2}(\mathcal{P}_{t+1} - K_0 - K_1 P_t - K V V_t)||^2,$$

(2.21)

where $T$ denotes the sample length and $||.||$ denotes the $L^2$ norm. The second term of the likelihood function captures the density of the cross-sectional fit and can be expressed as:

$$\sum_t f(\mathcal{P}_{e,t+1}|\mathcal{P}_{t+1}) = constant - \frac{T}{2} \log(\sigma^2_e(J-N)) - \frac{1}{2\sigma_e^2} \sum_t ||\mathcal{P}_{e,t+1} - W_e (A_P + B_P P_{t+1})||^2,$$

(2.22)

Joslin, Le, and Singleton (2012) show that this assumption, as opposed to assuming individual yields are observed perfectly, guarantees that our estimates are close to those obtained from the Kalman filter with more general distributions of observational errors.
where \( J \) denotes the number of yields used in estimation.

Estimates of the model parameter set, \( \Theta_P \), and the standard deviation of pricing errors, \( \sigma_e \), are obtained by maximizing the likelihood of observed data \( L \). For inferences, we do not use classical MLE standard errors. Rather, we use standard procedures to convert our ML estimation problem into a set of GMM moment conditions. Standard errors that are robust to serial autocorrelation and heteroskedasticity in the residual errors are then obtained using the Newey and West (1987) matrix.

2.3.2 Model Diagnosis

Our first step in the estimation is to diagnose the individual effect of the four entries of \( V_t \) in our model. Recall that \( V_t \) affects the conditional means of the state variables through \( K_V \), a \( 3 \times 4 \) matrix. The three rows of \( K_V \) correspond to the three PCs used as state variables. The first two columns of \( K_V \) correspond to the first two elements of \( V_t \), which capture the effects of short-run and long-run components of the short end of the yield curve. The last two columns of \( K_V \) correspond to the last two elements of \( V_t \), which capture the effects of short-run and long-run components of the long end of the yield curve.

To understand the individual effect of each element of \( V_t \) for each element of the pricing state variables \( P_t \), we estimate 12 specifications of the model by allowing one entry of \( K_V \) to be nonzero at a time. For example, the first specification only allows the (1,1) entry of \( K_V \) to be a free parameter, and sets all other entries of \( K_V \) to be zero. This specification therefore solely examines the effect of the first entry of \( V_t \)—the short-run volatility component in near-maturity bond yields—on the level of the term structure of yields. Similarly, the second specification only allows the (1,2) entry of \( K_V \) to be a free parameter, and examines the effect of the first entry of \( V_t \) on the slope of yields.

Table 2.2 reports the results. We see that among specifications 1–12, only specifications 1 and 3 lead to a statistically significant estimate of \( K_V \). A higher first entry of \( V_t \), i.e. a higher short-run volatility component in near-maturity bond yields, predicts a lower PC1 next week. So does a higher third entry of \( V_t \), i.e., a higher short-run volatility component in far-maturity bond

---

\(^{16}\) Specifically, we take the first-order derivative of \( L \) with respect to each parameter being estimated. Let \( L_t \) be the time-\( t \) element of \( L \) (\( L = \sum_t L_t \)). This first-order condition gives rise to a moment condition of the form \( E \left[ \frac{\partial L_t}{\partial \theta} \right] = 0 \), where \( \theta \) denotes any given parameter being estimated.
Table 2.2: Validating model choices.

<table>
<thead>
<tr>
<th>Spec</th>
<th>$K_V(1,1)$</th>
<th>Estimate of non-zero entries</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spec 1:</td>
<td>$K_V(1,1)$</td>
<td>$-27.75^{***}$</td>
<td>0.003</td>
</tr>
<tr>
<td>Spec 2:</td>
<td>$K_V(1,2)$</td>
<td>7.42</td>
<td>0.319</td>
</tr>
<tr>
<td>Spec 3:</td>
<td>$K_V(1,3)$</td>
<td>$-20.99^{**}$</td>
<td>0.032</td>
</tr>
<tr>
<td>Spec 4:</td>
<td>$K_V(1,4)$</td>
<td>4.01</td>
<td>0.623</td>
</tr>
<tr>
<td>Spec 5:</td>
<td>$K_V(2,1)$</td>
<td>13.41</td>
<td>0.502</td>
</tr>
<tr>
<td>Spec 6:</td>
<td>$K_V(2,2)$</td>
<td>8.49</td>
<td>0.254</td>
</tr>
<tr>
<td>Spec 7:</td>
<td>$K_V(2,3)$</td>
<td>10.95</td>
<td>0.370</td>
</tr>
<tr>
<td>Spec 8:</td>
<td>$K_V(2,4)$</td>
<td>1.16</td>
<td>0.856</td>
</tr>
<tr>
<td>Spec 9:</td>
<td>$K_V(3,1)$</td>
<td>-8.62</td>
<td>0.612</td>
</tr>
<tr>
<td>Spec 10:</td>
<td>$K_V(3,2)$</td>
<td>-25.84</td>
<td>0.180</td>
</tr>
<tr>
<td>Spec 11:</td>
<td>$K_V(3,3)$</td>
<td>-31.98</td>
<td>0.159</td>
</tr>
<tr>
<td>Spec 12:</td>
<td>$K_V(3,4)$</td>
<td>-24.16</td>
<td>0.240</td>
</tr>
<tr>
<td>Spec 13:</td>
<td>$K_V(1,1)$</td>
<td>$-36.84^{**}$</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>$K_V(1,3)$</td>
<td>12.67</td>
<td>0.412</td>
</tr>
</tbody>
</table>

yield. Because a lower PC1 is associated with a higher bond excess returns, this evidence suggests that there is a positive risk-return relation in bond markets, but the risk must be measured as the short-run component.

More concretely, the effect of $V_t$ on bond risk premiums is spelled out exactly in equation (2.16). For each bond with $n$ periods to maturity, the effect of $V_t$ on its risk premium is the (negative of the) product of the exposure of the bond to each risk factor, as captured by the yield loading $B$, and the effect $V_t$ has on the forecast of future PCs, as captured by $K_V$. For specifications 1 and 3, for which the last two rows of $K_V$ are set to zeros, the last two elements of $B$ are inconsequential, i.e., $V_t$ only affects bond risk premium though exposure to level risk. Moreover, all yield loadings on the level factor are positive and precisely estimated (since the $Q$ dynamics is very strongly identified). As a result, for specifications 1 and 3, a negative and significantly estimated estimate for the non-zero entry of $K_V$ translates into a positive trade-off between risk and expected excess returns.

In specification 13, we allow both the (1,1) and (1,3) entries of $K_V$ to be free parameters, and find that only the (1,1) entry is significant. This further indicates that near-maturity bonds contain more volatility information about expected excess returns than far-maturity bonds do.

Moving beyond the above initial exercise, to find a statistically supported specification, we conduct a more comprehensive search by roaming over different variants of our model, each with
a different specification for $K_V$. Ideally, we would try all combinations of zero constraints on all 12 entries of $K_V$, giving us $2^{12} = 4096$ different specifications. However, given the computational burden, we consider combinations of rows and columns of $K_V$ and set the corresponding rows and columns to zero. With three rows and four columns, we end up with $2^3 \times 2^4 = 128$ different specifications of $K_V$. We use the BIC scores to compare across different specifications. Interestingly, specification 1, in which only the (1,1) entry of $K_V$ is set free, emerges as the most preferred candidate.

Based on the evidence provided in this subsection, our main focus for the remaining of the paper will be on specification 1.17

2.3.3 Parameter Estimates

Table 2.3 reports the full model estimation for specification 1, together with the p-values. To facilitate a comparison between feedback matrices under $P$ and $Q$, we report the $Q$ feedback matrix, $K_{1P}^Q$, with the PCs of yields $P_t$ used as state variables.18 As expected, the diagonal values of the $K_1$ and $K_{1P}^Q$ matrices are close to one, suggesting a high persistence of the PCs at the weekly frequency. Notably, all elements of $K_{1P}^Q$ are statistically significant. The differences between the two feedback matrices reveal the contributions of the PCs to bond risk premiums, which we will examine more closely in the next subsection. For now, we note that the differences between $K_1$ and $K_{1P}^Q$ and the statistical significance of the $K_V$ matrix means that both the PCs of bond yields and the short-run volatility component seem to have independent contributions to the dynamics of bond risk premiums.

For ease of interpretation, we report the annualized degrees of persistence, $\rho_{L}^{52}$ and $\rho_{S}^{52}$, instead of $\rho_L$ and $\rho_S$. The long-run volatility component $L_t$ is extremely persistent, with $\rho_L^{52}$ estimated to be 0.99, suggesting a half life of more than 60 years. The short-run volatility component is less persistent, and the estimated $\rho_S^{52}$ suggests that it has a half life of about two years. Given the highly persistent $L_t$, it is perhaps not surprising that the estimated $\Sigma_P$ is not statistically signif-

17 We also check whether our findings regarding the relative contributions of the short-maturity and far-maturity yield portfolios are sensitive to the five year cutoff. Specifically, we repeat all the exercises with a different construction of the yield portfolios. We let the short-maturity (long-maturity) portfolio be an equal-weighted portfolio of three (seven) yields with maturities ranging from one year to three (four to ten) years. All of our results in this subsection remain essentially the same.

18 That is, $K_{1P}^Q$ satisfies: $E_t^Q[P_{t+1}] = \text{constant} + K_{1P}^Q P_t$. $K_{1P}^Q$ is obtained from the $K_1^Q$ matrix and the yields loadings $B_X$ of the canonical model, using the rotations described in Appendix A.2.
<table>
<thead>
<tr>
<th>Estimates</th>
<th>P-vals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_0$</td>
<td>0.02*** 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.02*** 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.08*** 0.00</td>
</tr>
<tr>
<td>$K_1$</td>
<td>1.00*** 0.00 0.00* 0.00 0.75 0.05</td>
</tr>
<tr>
<td></td>
<td>0.00*** 0.99*** -0.01*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 -0.00 0.95*** 0.14 0.67 0.00</td>
</tr>
<tr>
<td>$K_V$</td>
<td>-27.75*** 0.00 0.00 0.00 0.01</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>$K_{1P}$</td>
<td>1.00*** 0.01*** 0.01*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.00*** 0.99*** -0.02*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.00*** -0.01*** 0.97*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td>$\Sigma_P$</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>-0.16 0.73 0.78 0.77</td>
</tr>
<tr>
<td></td>
<td>0.97 -0.06 1.74 0.77 0.83 0.77</td>
</tr>
<tr>
<td>$\rho_{L}^{52}$</td>
<td>0.99*** 0.00</td>
</tr>
<tr>
<td>$\rho_{S}^{52}$</td>
<td>0.73*** 0.00</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.04*** 0.00</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.04*** 0.00</td>
</tr>
<tr>
<td>$r_{Q}^{Q}$</td>
<td>0.10*** 0.00</td>
</tr>
<tr>
<td>$\sigma_e$ (bps)</td>
<td>7.22*** 0.00</td>
</tr>
</tbody>
</table>

Table 2.3: Parameter estimates of the model, only allowing the (1,1) entry of $K_V$ to be a free parameter.

significant: by (2.12), the point estimation of $\Sigma_P$ involves dividing by $(1 - \rho_L)$, a very small scaler, which leads to large standard errors. The estimated $\alpha$ and $\phi$ are both 0.04, so about 8% of the variance shock in each week, $\epsilon_t \epsilon'_t - \Sigma_t - 1$, enters next week’s conditional variance $\Sigma_t$.

Further, as is typically the case for three-factor models, our model prices bonds well, with standard deviation of pricing errors estimated to be about seven basis points.

As a robustness check we also implement specifications 3 and 13, with estimates reported in Table 2.4. We see that the estimates for those two specifications are essentially the same as those for specification 1 (Table 2.3). For specification 3 (top half of Table 2.4), a visible difference is that $\rho_S^{52} = 0.79$, suggesting slightly higher persistence of $S_t$ than in specification 1. This is probably because the higher persistence of volatility in the far-maturity yields makes the estimated $S_t$ more persistent through the GARCH-in-mean term $K_V V_t$. For specification 13 (bottom
<table>
<thead>
<tr>
<th>Estimates</th>
<th>P-vals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_0$</td>
<td>0.02** 0.01</td>
</tr>
<tr>
<td></td>
<td>-0.02*** 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.08*** 0.00</td>
</tr>
<tr>
<td>$K_1$</td>
<td>1.00*** 0.00 0.00* 0.00 0.44 0.10</td>
</tr>
<tr>
<td></td>
<td>0.00*** 0.99*** -0.01*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 -0.00 0.95*** 0.14 0.81 0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00 -20.99** 0.00 0.02</td>
</tr>
<tr>
<td>$K_V$</td>
<td>0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>$K_{1}^{p}$</td>
<td>1.00*** 0.01*** 0.01*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.00*** 0.99*** -0.02*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.00*** -0.01*** 0.97*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.49 0.66</td>
</tr>
<tr>
<td>$\Sigma_p$</td>
<td>-0.06 0.67 0.74 0.66</td>
</tr>
<tr>
<td></td>
<td>0.83 -0.12 1.50 0.67 0.72 0.67</td>
</tr>
<tr>
<td>$\rho_{L}^{S}$</td>
<td>0.99*** 0.00</td>
</tr>
<tr>
<td>$\rho_{S}^{L}$</td>
<td>0.79*** 0.00</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.04*** 0.00</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.04*** 0.01</td>
</tr>
<tr>
<td>$Q$</td>
<td>0.10*** 0.00</td>
</tr>
<tr>
<td>$r_{\infty}$</td>
<td>7.05*** 0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>P-vals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_0$</td>
<td>0.02*** 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.02*** 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.08*** 0.00</td>
</tr>
<tr>
<td>$K_1$</td>
<td>1.00*** 0.00 0.00* 0.00 0.85 0.04</td>
</tr>
<tr>
<td></td>
<td>0.00*** 0.99*** -0.01*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 -0.00 0.95*** 0.15 0.72 0.00</td>
</tr>
<tr>
<td></td>
<td>-36.84** 0.00 12.67 0.00 0.03 0.47</td>
</tr>
<tr>
<td>$K_V$</td>
<td>0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>$K_{1}^{p}$</td>
<td>1.00*** 0.01*** 0.01*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.00*** 0.99*** -0.02*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>-0.00*** -0.01*** 0.97*** 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>0.52 0.75</td>
</tr>
<tr>
<td>$\Sigma_p$</td>
<td>-0.18 0.69 0.77 0.75</td>
</tr>
<tr>
<td></td>
<td>0.95 -0.02 1.67 0.75 0.93 0.75</td>
</tr>
<tr>
<td>$\rho_{L}^{S}$</td>
<td>0.99*** 0.00</td>
</tr>
<tr>
<td>$\rho_{S}^{L}$</td>
<td>0.72*** 0.00</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.03*** 0.00</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.04*** 0.00</td>
</tr>
<tr>
<td>$r_{\infty}$</td>
<td>0.08*** 0.00</td>
</tr>
<tr>
<td>$\sigma_e$ (bps)</td>
<td>7.04*** 0.00</td>
</tr>
</tbody>
</table>

Table 2.4: Two robustness checks of the model. The top table shows the estimation for specification 3. The bottom table shows the estimation for specification 13.

half of Table 2.4), the estimates are essentially identical to those in Table 2.3, except for a larger (1,1) entry of $K_V$, suggesting that the short-run volatility component of far-maturity bond yields does not provide incremental information relative to the short-run volatility component of near-
maturity yields. Both specifications imply a similar magnitude of bond pricing errors.

The statistical significance of the short-run volatility component and the insignificance of long-run volatility component map well back to our motivating exercises in Section 2.1. The volatility process in the ARCH-in-mean equation (2.2) only collects recent histories of squared residuals; therefore, it is short-run by construction. The GARCH(1,1) volatility process in (2.3) collects a much longer history of squared residuals; therefore, it is a mixture of short-run and long-run volatility information. By mixing the return-relevant short-run volatility component and the return-irrelevant long-run component, a single-component GARCH volatility process suppresses the predictability of the former. Indeed, recall from the first two columns of Table 2.1 that ARCH volatilities show strong predictive power for excess returns across all maturities, but GARCH volatilities do so only for relatively short maturities. More generally, the same argument can apply to other term structure models with a single volatility factor: unless this single volatility factor is predominantly short-run, the risk-return trade-off can be contaminated by the long-run component and becomes hard to detect in the data.

2.3.4 Economic Significance

The evidence reveals that there is a significantly positive risk-return relation in bond markets. A natural next question is the economic magnitude of the effect of volatility for bond excess returns. From (2.16), we can decompose the predictive component of one-week excess return of an \( n \)-week bond, \( x_{n,t+1} \), into a \( \mathcal{P}_t \)-related component and a \( V_t \)-related component. For each maturity \( n \), we calculate the fraction

\[
\frac{\text{Var}[(n-1)B_{n-1,p}K_{V}V_t]}{\text{Var}[(nB_{n,p} - (n-1)B_{n-1,p}K_{1} - B_{1,p})\mathcal{P}_t - (n-1)B_{n-1,p}K_{V}V_t]}
\]

in sample as a proxy for the contribution of the volatility component \( V_t \) for bond excess returns. We calculate the equivalent fraction for the contribution of the PC component. In this calculation, we use the estimates from specification 1, that is, with only the (1,1) entry of \( K_V \) is set free and other entries are set to zero.

The contributions of the PC-related (\( V_t \)-related) components are reported in the first four rows (last row) of Table 2.5. Each column corresponds to a given bond maturity. The first three
Table 2.5: Risk premium decomposition.

<table>
<thead>
<tr>
<th></th>
<th>1-yr</th>
<th>2-yr</th>
<th>3-yr</th>
<th>4-yr</th>
<th>5-yr</th>
<th>6-yr</th>
<th>7-yr</th>
<th>8-yr</th>
<th>9-yr</th>
<th>10-yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>0.41</td>
<td>0.27</td>
<td>0.21</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.12</td>
<td>0.11</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>PC2</td>
<td>0.35</td>
<td>0.41</td>
<td>0.46</td>
<td>0.50</td>
<td>0.53</td>
<td>0.53</td>
<td>0.52</td>
<td>0.51</td>
<td>0.49</td>
<td>0.47</td>
</tr>
<tr>
<td>PC3</td>
<td>0.09</td>
<td>0.14</td>
<td>0.11</td>
<td>0.06</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>PC1-3</td>
<td>0.86</td>
<td>0.82</td>
<td>0.78</td>
<td>0.73</td>
<td>0.69</td>
<td>0.67</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>Vt-component</td>
<td>0.14</td>
<td>0.24</td>
<td>0.32</td>
<td>0.38</td>
<td>0.42</td>
<td>0.44</td>
<td>0.44</td>
<td>0.43</td>
<td>0.41</td>
<td>0.40</td>
</tr>
</tbody>
</table>

The table rows report the individual contributions of each of the three PCs. It is perhaps not surprising that the slope factor represents the most important contribution to risk premiums. This is consistent with established results in the literature (see, for example, Fama and Bliss, 1987; and Campbell and Shiller, 1991). The level factor seems most important for relatively short-dated bonds, whereas the curvature factor seems inconsequential across the entire maturity spectrum. Summing up the first three rows gives the overall contributions to risk premiums of all three PCs (because the PCs are uncorrelated).

Because of the correlation between $V_t$ and $P_t$, the sum of the last two rows can exceed one, but it is no higher than 1.11 for the 10 maturities, suggesting that the correlation is not a severe concern. In fact, the in-sample correlations between the $P_t$-related components and the $V_t$-related component is very close to zero for each bond maturity. This means that these components represent essentially independent channels through which risk premiums are determined.

Table 2.5 reveals that the volatility measure, $V_t$, is an important contributor to expected bond excess returns, with a magnitude comparable to that of yield PCs. About 14% of the predictive component of the excess returns of one-year zero-coupon bond can be attributed to $V_t$. This fraction increases with maturity, reaching its peak of 44% for the six-year and seven-year bond, and then slowly declines to 40% for the 10-year bond. The contribution by the PC components, by contrast, declines from 86% at one-year maturity to 65% at seven-year maturity, and then stabilizes at 65% for the remaining far end of the yield curve.

Figure 2.1 plots the demeaned time-series of the total expected excess returns and the volatility component for the 10-year zero-coupon bond. By construction, this volatility component is a linear function of the first entry of the $V_t$ vector. It has high time variation and captures a large fraction of expected bond excess returns. Figure 2.1 also shows a sizeable increase in risk premiums during the Fed experiment regime of the early 1980s. According to our model, much of this...
Figure 2.1: Model-implied (weekly) risk premium (demeaned) on 10-year zero coupon bond. In the model only the (1,1) entry of $K_V$ is estimated, and all other entries of $K_V$ are set to zero.

increase is attributable to an increase in the short-run volatility component—a result we find reasonable and intuitive.

Our results in this subsection imply that to fully characterize the dynamics of bond risk premiums, a model should allow for two channels: a time-varying market price of risk and a time-varying quantity of risk. Moreover, the market prices of risks must represent an independent source of time variation and cannot be subsumed by the quantity of risk. For example, the habit-based term structure model of Le, Singleton, and Dai (2010) allows for both channels but the time-varying market prices of risks depend exclusively on yield volatility. As a result, risk premiums implied by their model are solely determined by yield volatility. Based on our findings, their model is likely to miss a sizeable portion of time variation in bond risk premiums. Our results corroborate the findings of Le and Singleton (2013) in their analysis of structural term structure models.

Although it can be tempting to interpret the last row of Table 2.5 as the contribution by the quantity of risk, and the second last row as the contribution by the market price of risk, such an interpretation may not be entirely accurate. The reason is that, in principle, the market prices of risk can also depend on yield volatility.
2.3.5 Matching Return Predictability and Conditional Volatilities in the Data

Having examined the statistical and economic significance of the conditional volatility \( V_t \) for bond excess returns, we now take a step back and examine how well the model fits salient empirical patterns of return predictability and time-varying volatility in bond markets. We emphasize that a key contribution of our model is to well match both aspects of the data.

Figure 2.2 shows the Campbell and Shiller (1987) regression coefficients implied by the model, together with those implied by the data. As is well known, if the expectations hypothesis holds and thus risk premiums are not time-varying, the coefficients of this regression should be uniformly ones across all maturities. Instead, the coefficients obtained from the data are significantly negative and increasingly so as maturities increase. Our model does a relatively decent job in capturing this pattern of the data, arguably as well as the Gaussian affine term structure models do (see Dai and Singleton, 2002). A weakness of the Gaussian models is the constant-volatility assumption; thus, those models cannot match the conditional volatilities of yields.

Figure 2.3 plots the model-implied one-week ahead conditional volatilities of the first three PCs of the yield term structure, as well as those of the 1-year, 5-year, and 10-year yields. As a comparison, we also plot realized volatilities and conditional volatilities estimated from a univariate EGARCH model. At each point in time, the realized volatilities are computed using daily changes in yields over the preceding three months. The EGARCH model is implemented using weekly yields over the entire sample. Volatilities implied by our model are very close to those two commonly used volatility measures.

2.4 Conclusion

In this paper, we study the risk-return tradeoff in the U.S. Treasury markets through the lens of a new discrete-time no-arbitrage term structure model. The model combines the tractability of affine term structure models with the ability of GARCH models to deliver an accurate measure of yield volatility. Not only does this model fit yields and yield volatilities well across all maturities, it also closely replicates the returns predictability characterized by the Campbell and Shiller (1987) regressions. Moreover, this model also allows us to differentiate the contributions to risk premiums of long-run and short-run volatility components of both the short end and long end of
Using yields data from 1962 to 2007, we find a significantly positive relation between risk and return in U.S. Treasury markets. A higher conditional volatility this week predicts a higher expected excess return next week. Notably, it is the short-run component of volatility, not the long-run component, that matters for return predictability. Moreover, the return-predicting power of the short-run volatility component predominantly comes from the short-end of the yield curve. Volatilities of far-maturity yields do not have additional predictive power for future yields once we control for volatility of the short end. Volatilities have economically important effects on bond risk premiums. For example, the volatility factor accounts for 14%, 42%, and 40% of the predictable components of weekly excess returns on one-year, five-year, and ten-year zero coupon bonds, respectively. The other source of predictability comes from the principal components of yields.

Our results have important implications. We show that the volatility factor, a proxy for quantity of risk, and the principal components of yields, a proxy for the market price of risk, have comparable weights in determining bond risk premiums. Therefore, models that rule out either channel only provide an incomplete characterization of the risk premium dynamics. Furthermore, because the short-run volatility component, not the long-run component, is responsible for pre-
Figure 2.3: Model-implied volatilities and realized volatilities. In the model only the (1,1) entry of $K_V$ is estimated, and all other entries of $K_V$ are set to zero.

dicting returns, the risk-return trade-off may be missed by models with a single volatility factor.
APPENDIX

A.1 Appendix for: Global Macroeconomic Conditional Skewness and the Carry Risk Premium

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\delta$</th>
<th>$\delta^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>0.0187</td>
<td>0.3</td>
<td>0.98</td>
<td>3</td>
<td>1.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table A.1: Calibration of the two-period model. $\mu$ and $\sigma$ denote the mean and volatility of endowment shocks $\varepsilon_i$ for good $i \in \{X, Y\}$, and $\rho$ denotes correlation between them. $\beta$ denotes the subjective discount factor, and $\gamma$ denotes the risk aversion parameter. $\delta$ and $\delta^*$ determine how home forecast skewness $\alpha (= \delta \alpha^g)$ and foreign forecast skewness $\alpha^* (= \delta^* \alpha^g)$ are driven by global skewness $\alpha^g$.

Individual agents make subjective mean predictions $\mu_i$ or $\mu_i^*$ about the shock to the endowment good in their respective country. In the text I refer to $\alpha$ (or $\alpha^*$) as a metric that summarizes skewness in forecasts: $\alpha = (\mu_1 + \mu_3 - 2 \times \mu_2)/(\mu_1 - \mu_3)$. The baseline, no-skewness case of $\alpha = 0$ corresponds to $\mu_1, \mu_2, \mu_3 = (0.0215, 0.02, 0.0185)$. The comparative statics of changing $\alpha$ is done by adjusting $\mu_1$ and $\mu_3$ that hold the cross-sectional variance fixed to that of the baseline case.
Table A.2: Calibration of the dynamic model. $\mu$ and $\sigma$ denote the mean and volatility of endowment shocks $\varepsilon_i$ for good $i \in \{X,Y\}$, and $\rho$ denotes correlation between them. $\beta$ denotes the subjective discount factor, and $\gamma$ denotes the risk aversion parameter. The last two columns show the calibration of the time-series dynamics of skewness $a_t^g = \rho a_{t-1}^g + \sigma_a \varepsilon_{a,t}$. $\delta$ and $\delta^*$ determine how home forecast skewness $a_t = \delta a_t^g$ and foreign forecast skewness $a^* = \delta^* a_t^g$ are driven by global skewness $a_t^g$. I employ the mapping $\alpha_t = a_t / \sqrt{1 + a_t^2}$, and so on, to ensure that $\alpha_t$, $\alpha_t^*$, $a_t^g$ are all bounded between -1 and 1.

Individual agents make subjective mean predictions $\mu_{i,t}$ or $\mu_{i,t}^*$ about the shock to the endowment good in their respective country. I define $\alpha_t$ as a metric that summarizes skewness in forecasts: $\alpha_t = (\mu_{1,t} + \mu_{3,t} - 2\mu_{2,t})/(\mu_{1,t} - \mu_{3,t})$. When skewness in forecasts $\alpha_t$ is equal to zero, the beliefs are calibrated as $(\mu_{1,t}, \mu_{2,t}, \mu_{3,t}) = (0.03, 0.02, 0.01)$. Then changing $\alpha_t$ is done by adjusting $\mu_{1,t}$ and $\mu_{3,t}$ while holding the cross-sectional variance fixed to that of the zero-skewness case.
<table>
<thead>
<tr>
<th>Country</th>
<th>Source</th>
<th>Sample</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>Thomson Reuters</td>
<td>1995.Q1-1999.Q1</td>
<td>Danish krone almost pegged to Euro</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>WM/Reuters</td>
<td>1996.Q4-2015.Q2</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>WM/Reuters</td>
<td>1997.Q4-2015.Q2</td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td>WM/Reuters</td>
<td>1999.Q4-2015.Q2</td>
<td>Start 1999.Q4 because values are too volatile during the Asian crisis</td>
</tr>
<tr>
<td>Mexico</td>
<td>WM/Reuters</td>
<td>1996.Q4-2015.Q2</td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>WM/Reuters</td>
<td>1996.Q4-2015.Q2</td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>WM/Reuters</td>
<td>1996.Q4-2015.Q2</td>
<td></td>
</tr>
</tbody>
</table>

Table A.3: Details about foreign exchange rate data.
Table A.4: Predictive regression of next-quarter carry trade portfolio returns onto global measures with the exclusion of the recent crisis: 2007Q4-2009Q2. The top panel is based on static carry trades, and the bottom panel is based on dynamic carry trades. The regressors $x_t^g, v_t^g, sk_t^g$ correspond to global expected growth, global uncertainty, and global skewness, respectively. Statistical significance is calculated based on Newey-West standard errors.
Table A.5: Predictive regression of next-quarter carry trade portfolio returns onto global measures, each of which is aggregated by taking the first principal component. The top panel is based on static carry trades, and the bottom panel is based on dynamic carry trades. The regressors $x_t^g, v_t^g, sk_t^g$ correspond to global expected growth, global uncertainty, and global skewness, respectively. Statistical significance is calculated based on Newey-West standard errors.
### Table A.6: Predictive regression of next-quarter carry trade portfolio returns onto global measures, each which is aggregated by taking the GDP-weighted average. The top panel is based on static carry trades, and the bottom panel is based on dynamic carry trades. The regressors $x_t^g$, $v_t^g$, $sk_t^g$ correspond to global expected growth, global uncertainty, and global skewness, respectively. Statistical significance is calculated based on Newey-West standard errors.
Table A.7: Predictive regression of next-quarter carry trade portfolio returns onto alternative global measures. The top panel is based on static carry trades, and the bottom panel is based on dynamic carry trades. The regressor \( sk_t^{g,+} \) corresponds the component of global skewness that is orthogonal to (not explained by) global expected growth.
<table>
<thead>
<tr>
<th></th>
<th>Panel A: dynamic</th>
<th></th>
<th>Panel B: static</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.0081*</td>
<td></td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>(0.0045)</td>
<td></td>
<td>(0.0042)</td>
</tr>
<tr>
<td>$sk^{g}_{t}$</td>
<td>−0.0097***</td>
<td></td>
<td>−0.0093***</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td></td>
<td>(0.0034)</td>
</tr>
<tr>
<td>$\Delta\text{liquidity}_{t}$</td>
<td>4.0171***</td>
<td>3.5206**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.1048)</td>
<td>(1.5190)</td>
<td></td>
</tr>
<tr>
<td>$\Delta\text{fxvol}_{t+1}$</td>
<td>−1.3194*</td>
<td>−1.5548**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7062)</td>
<td>(0.6111)</td>
<td></td>
</tr>
<tr>
<td>$\Delta\text{commod}_{t+1}$</td>
<td>0.3081**</td>
<td>0.3028***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1208)</td>
<td>(0.1061)</td>
<td></td>
</tr>
<tr>
<td>AdjR2</td>
<td>0.1716</td>
<td>0.1291</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2242</td>
<td>0.2962</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1952</td>
<td>0.1818</td>
<td></td>
</tr>
</tbody>
</table>

Table A.8: Predictive regression results controlling for other explanatory variables. Above carry trades are formed based on G10 currencies. The left three columns are based on the dynamic carry, and the right three columns are based on the static carry. $\Delta\text{liquidity}_{t}$ is a lagged innovations to liquidity, defined as the minus of TED spread. $\Delta\text{fxvol}_{t+1}$ is a contemporaneous innovations to the foreign exchange volatility constructed from the G10 currencies. $\Delta\text{commod}_{t+1}$ is a contemporaneous growth rate of the CRB BLS Spot Index.
A.2 Appendix for: Risk and Return Trade-off in the U.S. Treasury Market

Rotation of Factors

This appendix describes the detailed steps of rotating one set of factors into another.

Summary of the model As explained in the text, our model is characterized by the following risk-neutral ($Q$) and time-series ($P$) dynamics of the latent factors $X_t$:

$$ r_t = \delta_0 + \delta_1^t X_t, \quad (A.1) $$
$$ X_{t+1} = K_0^Q + K_1^Q X_t + \epsilon^Q_{t+1}, \quad (A.2) $$
$$ X_{t+1} = K_0 + K_1 X_t + K_V V_t + \epsilon_{t+1}, \quad (A.3) $$

where $\epsilon^Q_{t+1} \sim N(0, \Sigma_X)$ and $\epsilon_{t+1} \sim N(0, \Sigma_t)$, with $\Sigma_t$ following the component model of Engle and Lee (1999):

$$ \Sigma_t = S_t + L_t, \quad (A.4) $$
$$ S_t = \rho_S S_{t-1} + \alpha (\epsilon_t \epsilon'_t - \Sigma_{t-1}), \quad (A.5) $$
$$ L_t = (1 - \rho_L) \Sigma_X + \rho_L L_{t-1} + \phi (\epsilon_t \epsilon'_t - \Sigma_{t-1}). \quad (A.6) $$

Combining (A.1) and (A.2), we see that yields at all maturities are affine in the state variables:

$$ y_{n,t} = A_{n,X} + B_{n,X} X_t, \quad (A.7) $$

with $A_{n,X}$ and $B_{n,X}$ computed from the standard recursive equations for bond pricing.

The volatility vector $V_t$ is given by:

$$ V_t = \begin{pmatrix} \sqrt{B_{1-5,X} \Sigma_t B_{1-5,X}'} - \sqrt{B_{1-5,X} L_t B_{1-5,X}'} \\ \sqrt{B_{1-5,X} L_t B_{1-5,X}'} \\ \sqrt{B_{6-10,X} \Sigma_t B_{6-10,X}'} - \sqrt{B_{6-10,X} L_t B_{6-10,X}'} \\ \sqrt{B_{6-10,X} L_t B_{6-10,X}'} \end{pmatrix}, \quad (A.8) $$
where $B_{1-5,X}$ and $B_{6-10,X}$ denote the weighting vectors for the equal-weighted yield portfolios with near maturities (1–5 years) and far maturities (6–10 years).

Our model is fully characterized by the following parameter set

$$
\Theta_X = (\delta_0, \delta_1, K_0^Q, K_1^Q, \Sigma_X, K_0, K_1, K_V, \rho_S, \alpha, \rho_L, \phi).
$$

**Rotation of factors**  Consider a rotation from factors $X_t$ to $Z_t = U_0 + U_1 X_t$ for any given pair $(U_0, U_1)$. For example, the new factors $Z_t$ could be the yield PCs. It is straightforward to see that both the risk-neutral and time-series dynamics will be of the same affine form:

$$
\begin{align*}
    r_t &= \tilde{\delta}_0 + \tilde{\delta}'_1 Z_t, \\
    Z_{t+1} &= \tilde{K}_0^Q + \tilde{K}_1^Q Z_t + \tilde{\epsilon}^Q_{t+1}, \\
    Z_{t+1} &= \tilde{K}_0 + \tilde{K}_1 Z_t + \tilde{K}_V \tilde{V}_t + \tilde{\epsilon}_{t+1},
\end{align*}
$$

where $\tilde{\epsilon}^Q_{t+1} \sim N(0, \tilde{\Sigma}_X)$ and $\tilde{\epsilon}_{t+1} \sim N(0, \tilde{\Sigma}_t)$. The mappings from the original model to the rotated model are as follows:

$$
\begin{align*}
    \tilde{\delta}_0 &= \delta_0 - \delta'_1 U_1^{-1} U_0, \\
    \tilde{\delta}'_1 &= \delta'_1 U_1^{-1}, \\
    \tilde{K}_0^Q &= U_0 + U_1 K_0^Q - U_1 K_1^Q U_1^{-1} U_0, \\
    \tilde{K}_1^Q &= U_1 K_1^Q U_1^{-1}, \\
    \tilde{\Sigma}_X &= U_1 \Sigma_X U_1', \\
    \tilde{K}_0 &= U_0 + U_1 K_0 - U_1 K_1 U_1^{-1} U_0, \\
    \tilde{K}_1 &= U_1 K_1 U_1^{-1}, \\
    \tilde{K}_V &= U_1 K_V.
\end{align*}
$$

Additionally, the conditional covariance matrix under $P$, as well as its long-run and short-run counterparts, are simply given by:

$$
\tilde{\Sigma}_t = U_1 \Sigma_t U_1', \quad \tilde{L}_t = U_1 L_t U_1', \quad \text{and} \quad \tilde{S}_t = U_1 S_t U_1'.
$$
Combining these with the dynamics in (A.4), (A.5), and (A.6), we see that the parameters \((\rho_S, \rho_L, \alpha, \phi)\) governing the volatility dynamics are invariant to rotations.

To calculate the yield loadings, we observe that
\[
y_{n,t} = A_{n,X} + B_{n,X} X_t = A_{n,X} + B_{n,X} U_1^{-1}(Z_t - U_0). \tag{A.21}
\]
Thus, the loadings with respect to the new state variable \(Z\) are given by:
\[
A_{n,Z} = A_{n,X} - B_{n,X} U_1^{-1} U_0 \quad \text{and} \quad B_{n,Z} = B_{n,X} U_1^{-1}.
\]

Finally, note that \(V_t\) is invariant to rotations in that \(\tilde{V}_t \equiv V_t\). Intuitively, this is due to the fact that \(V_t\) is measured by the volatilities of observable yields portfolios. More concretely, take one of the terms in the construction of \(\tilde{V}_t\), say \(B_{1-5,Z} \tilde{\Sigma}_t B'_{1-5,Z}\), we see that:
\[
B_{1-5,Z} \tilde{\Sigma}_t B'_{1-5,Z} = B_{1-5,X} U_1^{-1} U_1 \Sigma_t U_1 U_1^{-1} B'_{1-5,X} = B_{1-5,X} \Sigma_t B'_{1-5,X}.
\]
REFERENCES


