Essays on Non-linearities in Stock and Bond Returns: A Density-Based Approach

By

Jiening Pan

Doctor of Philosophy

Economics

Esfandiar Maasoumi, Ph.D.
Advisor

Francisco Barillas, Ph.D.
Committee Member

Elena Pesavento, Ph.D.
Committee Member

Zhanwei (Vivian) Yue, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Essays on Non-linearities in Stock and Bond Returns: A Density-Based Approach

By

Jiening Pan

M.A., Emory University, 2014
M.S., Emory University, 2010
B.S., Zhejiang University, 2008

Advisor: Esfandiar Maasoumi, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics
2015

Abstract

Essays on Non-linearities in Stock and Bond Returns: A Density-Based Approach
By Jiening Pan

The dissertation consists of three essays that revolve around non-linearities embedded in asset returns.

In the first essay "The Role of Slope Heterogeneity in Bond Excess Returns Predictability", I investigates bond excess return forecastability using current forward rates. The dynamics of excess return are modeled non-parametrically. Estimation shows heterogeneous slopes for independent variables, indicating the existence of non-linearity. Empirically, I find this non-linearity plays an important role in excess return prediction both in- and out-of-sample. By including non-linearity in the model, the in-sample R^2 jumps to as high as 91%. Meanwhile, lagged forward rates are no longer statistically significant, in contrast to the results documented in previous research. The out-of-sample forecasts also favor the non-parametric model. Findings in this paper suggest a potential important information source embedded in the current forward rates cross-section. Information associated with non-linearity is largely ignored in the existing literature as it is averaged out by linear model settings.

The second essay "Do Non-Linearities Matter in the Yield Curve?" tries to answer the question that do non-yield variables contain information beyond what is contained in the yield curve? Using a non-linear factor extracted from the yield curve, I find non-yield factors, which are constructed from a large panel of macro-finance data, are no longer significant in predicting future bond excess returns both in- and out-of-sample. Moreover, my non-linear factor generates countercyclical and business cycle frequency bond risk premia. The findings underscore the importance of non-linearities embedded in the term structure, suggesting a fully spanned term structure model with non-linear state factors may be capable of matching features observed in the data.

In the third essay "A Test on Asymmetric Dependence" (joint work with Prof. Maasoumi, Lei Jiang and Ke Wu), we provide a model-free test for asymmetric dependence between stock and market returns, based on the Kullback-Leibler mutual information measure. Our test has greater power in small samples than previous tests of asymmetric correlation proposed by Hong, Tu and Zhou (2007). Empirically, we find that asymmetric dependence is a prevailing phenomenon in most commonly used portfolios.

Essays on Non-linearities in Stock and Bond Returns: A Density-Based Approach

By

Jiening Pan

M.A., Emory University, 2014
M.S., Emory University, 2010
B.S., Zhejiang University, 2008

Advisor: Esfandiar Maasoumi, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics
2015

Acknowledgements

I owe my deepest gratitude to my advisor, Prof. Esfandiar Maasoumi, for his guidance and support during my Ph.D. studies. These pages would be impossible without his kindly mentorship over my stay at Emory.

I am heartily grateful to my committee members: Dr. Francisco Barillas, Dr. Elena Pesavento and Dr. Vivian Yue for their encouraging words, thoughtful criticism, and help on preparing my job market interviews.

I appreciate Dr. Kaiji Chen, Dr. David Jacho-Chavez, Dr. Tao Zha and other economics faculty members for their discussions and comments on my research.

I would also like to thank Brenda Baxter, Marie Browne, Alexa DeVetter, Stephanie Gray and other economics staff members for their assistant. I wish to thank Grace Song for her help on revising this dissertation.

I would like to acknowledge my parents, my father Xiaotian Pan and my mother Suqin Zhou for their constant support, encouragement and understanding.

I received numerous guidance and assistance from many people over the years of my study. It is impossible to list and thank all. The dissertation is dedicated to all of them.

Contents

List of Figures

List of Tables

Preface

The dissertation contains three essays that center on non-linearities associated with financial assets. More specifically, the first two chapters focus on non-linearities embedded in the yield curve of the U.S. Treasury bonds. I investigate how non-linearities affect future bond excess return predictability and their implications in dynamic term structure modeling. The third chapter proposes a non-linear measure of exceedance dependence and provides a model free test for asymmetric dependence between stock portfolio and market returns.

Parametric models have dominated financial economics research over past decades, mainly because they help us decompose and understand the complicated dynamics of financial variables, as well as more tractable and elegant solutions they provide. However, the theoretical attractiveness of parametric models does not guarantee the usefulness in practice. For example in the class of fixed income securities, Aït-Sahalia (1996) tests almost all existing parametric models of interest rate dynamics and claims none of them describes the true data generating process. For the equity securities, various empirical studies show that the classical Capital Asset Pricing Model (CAPM) developed by Sharpe (1964) and Lintner (1965) fail to fully capture the risk premium of stock returns. In other words, the CAPM market beta, which only considers the linear correlation between individual stock and market returns, is not a perfect proxy for the market risk factor.

These findings suggest that commonly used parametric models in finance are misspecified, which implies potential inconsistent estimation and unreliable testing results. On the other hand, the density-based non-parametric approach as in Nadaraya (1964) and Watson (1964), which let the data speak for themselves, provides an ideal alternative.

Non-linearities in the yield curve

Starting from Vasicek (1977) and Cox, Ingersoll, and Ross (1985) and many others, the dynamics of interest rates are usually modeled using parametric diffusion processes. Extending the previous work of Cox, Ingersoll, and Ross (1985) and Heath, Jarrow, and Morton (1992), Duffie and Kan (1996) develop a class of arbitrage-free multi-factor models of the term structure of interest rates, which is known as the affine term structure models (ATSMs). Although ATSMs have closed form solution and they are easy to estimate from an empirical perspective, later studies show that this class of models fails to match certain patterns observed in data. For example, the forecast errors in Dai and Singleton (2000) are negatively correlated with the slope of the yield curve, which produces smaller model-implied bond excess returns when the slope of yield curve is steeper.[1] Duffee (2002) further modifies Duffie and Kan (1996) and Dai and Singleton (2000) model by separating the compensation of interest risk and interest rate volatility and proposes the "essentially affine" term structure models, which later becomes the benchmark model in dynamic term structure literature. ATSMs share one commonality: the model dynamics are fully captured by a set of state variables and all yields are **linear** functions of the same state variables. Thus an important question in macro-finance term structure literature is that what should be used as the state variable in a dynamic term structure model?

Empirical research in financial economics has uncovered significant forecastable variation in bond excess returns. Fama and Bliss (1987) report that the one-year excess return of the n-year bond can be explained by the spread between the n-year forward rate and one-year yield. Stambaugh (1988) finds that two- to six-month bond excess returns can be predicted by one- to six-month forward rates, and recently, Cochrane and Piazzesi (2005) show that a linear combination of five forward rates can be used to forecast one-year excess bond return across different maturities. These results suggest that the current yield curve contains substantial information about future interest rates. Therefore an assumption usually

[1] Empirical results documented in Campbell and Shiller (1991) support the opposite, i.e., steeper current yield slope suggests larger future bond excess returns.

adopted in the bond literature is that bond yields or their linear combinations serve as the state variables. In other words, the model is fully spanned by bond yields. However, more recent research show that various non-yield variables can provide additional predictability, which questions the fully spanned assumption. Ang and Piazzesi (2003) modify the original continuous-time Duffie and Kan model into discrete-time setting. They combine bond yields with inflation and real productivity factors in a no-arbitrage vector autoregression (VAR) framework and use this model to study the joint dynamics of macroeconomic activity and inflation. One of the major conclusions in their paper is that macroeconomic factors improve forecast accuracy of yields. Cooper and Priestley (2009) report that the output gap has strong predictive power of bond excess returns both in- and out-of-sample. Ludvigson and Ng (2009) apply dynamic factor analysis (DFA) to a large panel of macro-finance data and find that factors closely related to real output and inflation can improve forecasting power for future excess returns. They also show that non-yield factors are the key ingredient in generating countercyclical risk premia. Fontaine and Garcia (2012) construct a liquidity factor to measure funding conditions confronted by financial intermediaries. They find that bond excess returns can be predicted by this factor, which suggests that liquidity risks might be an important source of bond excess return predictability.

As pointed out in Ludvigson and Ng (2009) and Joslin, Priebsch, and Singleton (2014), the failure of the fully spanned ATSMs is due to the invertibility between the state variables and yields. To resolve this conflict, Duffee (2011) and Joslin, Priebsch, and Singleton (2014) propose an ATSM in which part of the state vector is not spanned by the yield curve. Duffee (2011) assumes the existence of "hidden" factors that are undetectable from the cross section of yields but that have significant predictive power with respect to future bond excess returns. Joslin, Priebsch, and Singleton (2014) allow for macroeconomic risks that cannot be spanned by the yield curve. Two linear combinations of smoothed industrial production and inflation are chosen as the unspanned factors. They show that the unspanned factors can explain a substantial portion of the variation in forward terms premia.

Although the results of the unspanned ATSM look promising, the potential problem associated with this approach is that economic theory does not tell us which variables should be used as unspanned factors. For example, Cieslak and Povala (2013) argue the source of the unspanned risks is inflation. They construct a factor which is equal to the weighted average of past inflation, where the weights are estimated from survey data. However, it appears that the weights are chosen such that the dynamics of the factor follow the same pattern as the dynamics of the yields after smoothing.

In Chapter one and two, I show that non-linearities in the yield curve can provide an alternative modeling approach to resolve this conflict. Non-linearities in bond yield dynamics are well documented in the finance literature. For example, Gray (1996) concludes that a generalized regime switching (RS) model that nests generalized autoregressive conditional heteroskedasticity (GARCH) structures is able to generate both mean-reversion and conditional heteroskedasticity pattern for the short rate. Ang and Bekaert (2002) claim that an ATSM with RS features can replicate some of patterns found by the NP studies. Dai, Singleton, and Yang (2007) build an affine RS model with a closed form solution. Ang, Bekaert, and Wei (2008) consider different inflation regimes and construct a RS affine model for the nominal yield curve. Moreover, by allowing for non-linearities, invertibility between state variables and yields generally will not exist.

The first essay, "The Role of Slope Heterogeneity in Bond Excess Return Prediction", investigates the impact of non-linearities on bond excess return predictability. Similar to Cochrane and Piazzesi (2005), I also use one-year forward rates as the predictor, but I allow the conditional expectation of bond excess returns to take an unknown functional form and estimate it using the non-parametric local constant kernel estimator. The results show a strong heterogeneous pattern for slopes evaluated at different values of forward rates. I also find significant improvement in excess return predictability when non-linearities are retained. Moreover, the lagged forward rates no longer help predict future excess returns. The out-of-sample results also support the in-sample findings. I conclude a potential important information source is embedded in the current forward rates cross section that is largely ignored in the existing literature, as a linear model setting averages out the non-linearities.

The second essay,"Do non-linearities matter in the yield curve?" further addresses the research question whether non-yield variables can provide additional predictability on bond excess returns when non-linearities in the yield curve have been taken into account. By using a non-linear factor extracted from the bond cross section, I show empirically that non-yield factors in Ludvigson and Ng (2009) are no longer significant in predicting future bond excess returns both in- and out-of-sample. In addition, I link the risk associated with the non-linear factor to the fundamental macroeconomic shocks by showing that the non-linear factor is able to generate countercyclical and business cycle frequency bond risk premia. Results indicate that excess returns associated with the non-linear risk are compensations for a risk-averse investor to bear unexpected macroeconomic shocks, which is consistent with economic theory. My findings also signal the possibility of building a dynamic term structure model with fully spanned non-linear state factors to jointly model the dynamics of bond yields and macroeconomic variables.

Non-linearities in stock returns

Since Sharpe (1964) and Lintner (1965) proposed the CAPM in 1960s, one prevailing assumption adopted in empirical asset pricing literature is that the individual stock and market returns follow joint normal distribution. Joint normality is attractive in at least two aspects. First, the dependence between two variables can be fully described by their linear correlation (CAPM Betas), independence is equivalent to 0 in linear correlation. Second, the return distribution is symmetric in market up and downturn. Therefore under normality assumption, same factor that is equal to the linear correlation between the individual stock return and the market return, capture the market risk in both market up and downturn.

However, empirical evidence in various studies suggests that individual stocks exhibit stronger co-movements with the market during market downturn than in market upturn (Ball and Kothari, 1989; Ang, Chen, and Xing, 2006; Longin and Solnik, 2001; Kroner and Ng, 1998). Moreover, similar characteristics are also documented for stock portfolios (Kroner and Ng, 1998; Conrad, Gultekin, and Kaul, 1991). These results indicate that a very important aspect in stock return distribution is missing if normality is assumed, i.e., asymmetric dependence. Ignoring such asymmetry will lead to underestimation of portfolio's market risk during market downturn, which causes suboptimal asset allocation and utility loss for a risk averse investor.

There are two primary questions associated with the asymmetric dependence in stock returns. The first is how should "dependence" be defined. As we are deviating from the normality, linear correlation is no longer equivalent to dependence. In fact, it is quite easy to find examples such that a correlation coefficient of 0 does not imply two random variables are independent. And the second, is it possible to test whether the difference of dependence on different domains of a distribution is statistically significant. Many attempts have been made to answer these two questions. Ang and Chen (2002) first come up with a test on asymmetric correlation and conclude that asymmetries are significantly higher for stocks with various characteristics such as small size, high book-to-market ratio and past losers. However, their test is that the test is model dependent. It is only able to judge whether the asymmetries observed in the data can be explained by a given model. Observing this shortcoming, Hong, Tu, and Zhou (2007) propose a model free test on asymmetric correlation. Based on this test, they conclude that it is not common to observe statistically significant asymmetries in portfolio returns. However, linear correlation is not an ideal measure for dependence once the underlying distribution is not normal. In fact, we need an alternative dependence measure which is able to capture non-linear dependence related with all higher order moments.

In the third essay, "A Test on Asymmetric Dependence", I first propose a new exceedance dependence measure, which is constructed by a modified Kullback-Leibler mutual information measure. This dependence measure is defined under the distribution perspective, which summarizes information of the whole distribution, including all existing moments. Then I develop a model-free test on asymmetric dependence of stock and market returns. The sampling distribution of the test statistic is obtained via consistent bootstrap method. Comparing with existing asymmetry test based on linear correlation, this new test possesses greater power in small samples. Moreover in contrast to findings documented in the existing literature, I find asymmetries in dependence is a prevailing phenomenon in commonly used portfolios sorted by size, book-to-market ratio and momentum.

References

Aït-Sahalia, Y. 1996. Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9:385–426.

Ang, A., and G. Bekaert. 2002. Short rate nonlinearities and regime switches. *Journal of Economic Dynamics & Control* 26:1243–74.

Ang, A., G. Bekaert, and M. Wei. 2008. The term structure of real rates and expected inflation. *Journal of Finance* 63:797–849.

Ang, A., and J. Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63:443–94.

Ang, A., J. Chen, and Y. Xing. 2006. Downside risk. *Review of Financial Studies* 19:1191–239.

Ang, A., and M. Piazzesi. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50:745–87.

Ball, R., and S. P. Kothari. 1989. Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics* 25:51–74.

Campbell, J. Y., and R. J. Shiller. 1991. Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies* 58:495–514.

Cieslak, A., and P. Povala. 2013. Expected returns in treasury bonds. *Working Paper* Northwestern University.

Cochrane, J. H., and M. Piazzesi. 2005. Bond risk premia. *American Economic Review* 95:138–60.

Conrad, J., M. N. Gultekin, and G. Kaul. 1991. Asymmetric predictability of conditional variances. *Review of Financial Studies* 4:597–622.

Cooper, I., and R. Priestley. 2009. Time-varying risk premiums and the output gap. *Review of Financial Studies* 22:2801–33.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53:385–407.

Dai, Q., and K. J. Singleton. 2000. Specification analysis of affine term structure models. *Journal of Finance* 55:1943–78.

Dai, Q., K. J. Singleton, and W. Yang. 2007. Regime shifts in a dynamic term structure model of u.s. treasury bond yields. *Review of Financial Studies* 20:1669–706.

Duffee, G. R. 2002. Term premia and interest rate forecasts in affine models. *Journal of Finance* 57:405–43.

———. 2011. Information in (and not in) the term structure. *Review of Financial Studies* 24:2895–934.

Duffie, D., and R. Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6:379–406.

Fama, E. F., and R. R. Bliss. 1987. The information in long-maturity forward rates. *American Economic Review* 77:680–92.

Fontaine, J.-S., and R. Garcia. 2012. Bond liquidity premia. *Review of Financial Studies* 25:1207–54.

Gray, S. F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42:27–62.

Heath, D., R. Jarrow, and A. Morton. 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* 60:77–105.

Hong, Y., J. Tu, and G. Zhou. 2007. Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20:1547–81.

Joslin, S., M. Priebsch, and K. J. Singleton. 2014. Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance* 69:1197–233.

Kroner, K. F., and V. K. Ng. 1998. Modeling asymmetric comovements of asset returns. *Review of Financial Studies* 11:817–44.

Lintner, J. 1965. Security prices, risk and maximal gains from diversification. *Journal of Finance* 20:587–616.

Longin, F., and B. Solnik. 2001. Extreme correlation of international equity markets. *Journal of Finance* 56:649–76.

Ludvigson, S., and S. Ng. 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22:5027–67.

Nadaraya, E. A. 1964. On estimating regression. *Theory of Probability and its Applications* 9:141–2.

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:425–42.

Stambaugh, R. F. 1988. The information in forward rates: Implications for models of the term structure. *Journal of Financial Economics* 21:41–70.

Vasicek, O. 1977. An equilibrium characterisation of the term structure. *Journal of Financial Economics* 5:177–88.

Watson, G. S. 1964. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* 26:359–72.

Chapter 1

The role of slope heterogeneity in bond excess return predictability

Abstract

This paper investigates bond excess return forecastability using current forward rates. The dynamics of excess return are modeled non-parametrically. Estimation shows heterogeneous slopes for independent variables, indicating the existence of non-linearity. Empirically, I find this non-linearity plays an important role in excess return prediction both in- and out-of-sample. By including non-linearity in the model, the in-sample R^2 jumps to as high as 91%. Meanwhile, lagged forward rates are no longer statistically significant, in contrast to the results documented in previous research. The out-of-sample forecasts also favor the non-parametric model. Findings in this paper suggest a potential important information source embedded in the current forward rates cross-section. Information associated with non-linearity is largely ignored in the existing literature as it is averaged out by linear model settings.

JEL Classification: E0, E4, G0, G1.

Key words: Bond term structure, in- and out-of-sample forecasting, non-linearity, non-parametric methods, bootstrap.

I Introduction

The forecastability of variation in excess returns of U.S. Treasury bonds has been widely documented in the literature. Fama and Bliss (1987) find that the one-year excess return of the n-year bond can be explained by the spread between the n-year forward rate and one-year yield. Stambaugh (1988) reports excess returns forecastability on a shorter time horizon by regressing two- to six-month bond excess return on one- to six-month forward rates. Campbell and Shiller (1991) show that excess returns of Treasury bonds are forecastable by yield spreads. More recently, Cochrane and Piazzesi (2005) use a linear combination of five forward rates to forecast one-year excess bond return across different maturities simultaneously and report substantial forecastibility. Researchers generally accept the failure of the expectations hypothesis that future bond excess returns are not predictable; however, disagreement arises with respect to the choice of predictors. There is no doubt that current bond prices (and their linear functions) have predictive power in future excess returns. In addition, empirical results show many other variables also contain substantial excess return forecastability. Cochrane and Piazzesi (2005) notice that the lagged forward rates help improve excess return prediction. Ludvigson and Ng (2009) and Cooper and Priestley (2009) report that macroeconomic variables contain substantial information about future excess returns that is orthogonal to the current bond term structure. Fontaine and Garcia (2012) and Cieslak and Povala (2013) argue that bond risk premia are compensation for investors to bear liquidity and inflation risks, respectively.

In standard affine term structure models (ATSMs),[1] bond yields or their linear combinations are chosen as the state vector. To ensure solution tractability and computation feasibility, the state vector is assumed to evolve linearly. Therefore, standard ATSMs implicitly assume that a linear regression model with linear factors extracted from the bond cross-section is **correctly specified**. To accommodate ATSMs with the findings discussed above, researchers suggest the state vector should also incorporate lagged yields and non-

[1] For reference, see e.g., Duffie and Kan (1996), Dai and Singleton (2000) and Duffee (2002), among others.

yield variables (Ang and Piazzesi, 2003; Duffee, 2011b; Joslin, Priebsch, and Singleton, 2014; Feunou and Fontaine, 2014). However, this approach receives much criticism as these newly-introduced variables are lack of theoretical foundation (Duffee, 2013).

I argue that the linear model setting, i.e., the effect of each individual predictor on excess return remains constant over the domain of that predictor, may be misspecified. In fact, the literature has reported rich non-linearities in bond yield dynamics. Gray (1996) concludes that a generalized regime switching (RS) model that nests generalized autoregressive conditional heteroskedasticity (GARCH) structures is able to generate both mean-reversion and conditional heteroskedasticity pattern for the short rate. Aït-Sahalia (1996a,b) and Stanton (1997) model the short rate dynamics with a non-parametric (NP) kernel density estimator and reject almost "every parametric model of spot rate [previously] proposed in the literature" such as Vasicek (1977) and Cox, Ingersoll, and Ross (1985). Ang and Bekaert (2002) claim that an ATSM with RS features can replicate some of patterns found by the NP studies. Dai, Singleton, and Yang (2007) build an affine RS model with a closed form solution. Ang, Bekaert, and Wei (2008) consider different inflation regimes and construct a RS affine model for the nominal yield curve.

In this paper, I re-examine the excess bond return predictability with commonly used predictors after giving up the usual assumption of linearity. Same as in Fama and Bliss (1987) and Cochrane and Piazzesi (2005), current one-year forward rates are chosen as predictors. I use NP regression techniques to take into account non-linearities in the yield curve for the following reasons. First, the RS approach requires a presumed number of regimes. To ensure computation feasibility, the existing literature usually assumes the total number of regimes equals two, in which underestimating the number of regimes may cause a misspecification issue. Second, as Aït-Sahalia (1996b) argues "a single continuous but nonlinear function that preserves time-homogeneity" would be equivalent to multi-regime, piece-wise linear functions. Therefore, if one does not care about policy implications in different regimes, the NP approach would be at least as good as the RS setting.

By allowing for non-linearities, on average 91% of variation in bond excess returns can be explained by current forward rates. Compared to the linear regression results for individual bond in Cochrane and Piazzesi (2005), the NP setting captures a greater portion of variation, beyond what is documented in the existing literature.

Following Racine (1997), I then construct a studentized test statistic and test the joint significance of independent variables using the bootstrap method. Two competing models (one uses only the current forward rates and the other uses both current and one-month lagged forward rates) are examined. Results suggest that lagged forward rates fail to provide additional predictive power once non-linearities are preserved.

To investigate whether in-sample results are due to overfitting, I extend the sample period and conduct pseudo out-of-sample forecasting for different models. I find that prediction generated by the NP model outperform commonly used linear models, especially when the excess return volatilities are high. The findings in this paper reveal that the non-linearity could be a potentially important information source which is largely ignored in the existing literature, as the non-linear effect is averaged out in a misspecified linear regression model.

The rest of this article is organized as follows. In the next section, I introduce the notation in this paper. Section 3 introduces the econometric tools used in this paper and discusses the testing significance of explanatory variables under the NP setting. Section 4 lays out the models' in-sample performance. Section 5 compares the performance of different models by examining their out-of-sample forecasting. Section 6 concludes and discusses potential future research.

II Econometric framework

II.1 Notation

In this paper, each period represents one year. Consider a zero-coupon bond that will mature in n years with a final payoff of \$1. I use $p_t^{(n)}$ to denote its log price:

$$p_t^{(n)} = \text{log price of } n\text{-year zero coupon bond at time } t.$$

Its one-period continuous compound yield is simply written as

$$y_t^{(n)} = -\frac{1}{n} p_t^{(n)}.$$

The one-year log forward rate for this bond is defined as the arbitrage-free one-year rate between time $t+n-1$ and $t+n$,

$$f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)}.$$

I use $r_{t+1}^{(n)}$ to denote the log one-year holding period return from buying an n-year bond at time t and selling it as an $n-1$ bond after one year at $t+1$

$$r_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}.$$

The log one-year excess return for n-year bond $rx_{t+1}^{(n)}$ is defined as borrowing at the current short rate at period t, buying n-year bond and then selling after one year to repay the debt:

$$rx_{t+1}^{(n)} \equiv r_{t+1}^{(n)} - y_t^{(1)}. \tag{1.1}$$

Variables in bold font represent vectors across maturity, e.g.,

$$\begin{aligned} \mathbf{rx}_{t+1} &\equiv \left[rx_{t+1}^{(2)} \quad rx_{t+1}^{(3)} \quad rx_{t+1}^{(4)} \quad rx_{t+1}^{(5)} \right]^\top, \\ \mathbf{f}_t &\equiv \left[y_t^{(1)} \quad f_t^{(2)} \quad f_t^{(3)} \quad f_t^{(4)} \quad f_t^{(5)} \right]^\top. \end{aligned}$$

II.2 The model

I consider two models in this paper. One only uses current forward rates to predict future bond excess returns. The conditional mean of excess return for the n-year bond only depends on $\mathbf{f}_t$, following some unknown function $g_n(\cdot)$,

$$E_t\left(rx_{t+1}^{(n)}\right) = g_n(\mathbf{f}_t).$$

The other model assumes that the conditional mean function of the n-year excess return should also include p lagged forward rates, i.e.,

$$E_t\left(rx_{t+1}^{(n)}\right) = h_n(\mathbf{f}_t, \mathbf{f}_{t-\frac{1}{12}}, \ldots, \mathbf{f}_{t-\frac{p}{12}}).$$

For simplicity, I only consider the case for $p = 1$ in this paper. In order to avoid model misspecification and retain non-linearities, I do not assume any specific format for $g_n(\cdot)$ and $h_n(\cdot)$ except stationarity.

I run regression of excess return $rx_{t+1}^{(n)}$ on all forward rates,

$$rx_{t+1}^{(n)} = g_n(\mathbf{f}_t) + \varepsilon_{t+1}^{(n)}, \tag{1.2}$$

$$rx_{t+1}^{(n)} = h_n(\mathbf{f}_t, \mathbf{f}_{t-\frac{1}{12}}) + \epsilon_{t+1}^{(n)}. \tag{1.3}$$

(1.2) and (1.3) (denoted as NP model) are estimated non-parametrically by the local constant kernel estimator.[2] I adopt the commonly used Gaussian kernel in this paper, whose bandwidth is selected via least square cross validation (Stone, 1984).

To directly compare the impact of non-linearity, the unrestricted excess return regression model[3] in Cochrane and Piazzesi (2005) (CP hereafter)

$$rx_{t+1}^{(n)} = \widetilde{\mathbf{f}}_t' \boldsymbol{\beta_n} + u_{t+1}^{(n)}, \quad \text{where} \quad \widetilde{\mathbf{f}}_t \equiv \begin{bmatrix} 1 & \mathbf{f}_t' \end{bmatrix}^\top . \tag{1.4}$$

serves as the benchmark. The vector $\boldsymbol{\beta_n}$ in (1.4) is estimated using OLS.

II.3 Test joint significance of predictors

Unlike linear regression models, a non-parametric model does not have a specific functional form for its conditional mean. Moreover, the slope coefficient for an independent variable under the non-parametric setting is defined point-wisely over the domain of that variable. For a typical empirical exercise, the estimates will no longer yield a constant.

[2]Briefly described in Appendix 1.A. Readers can refer to Li and Racine (2006) for details.
[3]Denoted as linear model, or LM.

Following Racine (1997), I construct the test statistic with the estimated conditional mean and associated estimated partial derivatives. I define the non-parametric slope coefficient $\beta^n_{j,t}$ at the observation $\mathbf{f}_t$ as the partial derivative of the conditional mean function g_n with respect to j-year forward rate $f^{(j)}_t$, i.e.,

$$\beta^n_{j,t} = \frac{\partial g_n(\mathbf{f}_t)}{\partial f^{(j)}_t}. \tag{1.5}$$

If the variable $f^{(j)}_t$ does not have predictive power, equivalently we will have

$$\beta^n_j = 0, \quad a.s.$$

The subscript t is dropped under the assumption of stationarity. Let λ_n be the expectation of the sum of the squared slopes over all forward rates $f^{(j)}$ for the n-year excess return,

$$\lambda_n = E\left[\sum_{j=1}^{5} \left(\beta^n_j\right)^2\right]. \tag{1.6}$$

If all β^j_n's are jointly insignificant, λ_n will be indistinguishable from 0. Therefore, the following hypothesis is tested:

$$\begin{aligned} H_0 &: \quad \lambda_n = 0 \\ H_1 &: \quad \lambda_n > 0 \end{aligned}.$$

λ_n is estimated by its unconditional mean

$$\hat{\lambda}_n = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=1}^{5}\left(\hat{\beta}^n_{j,t}\right)^2\right], \tag{1.7}$$

where $\hat{\beta}^n_{j,t}$ is directly estimated from (1.2).[4]

As noted in Rilstone (1991) and Robinson (1991, 1994), the sampling distribution of $\hat{\lambda}_n$ under the asymptotic theory produces misleading finite sample inferences. Racine (1997) suggests using the sampling distribution constructed from the bootstrap method. The

[4]Readers can refer to the references in Racine (1997) for technical details.

algorithm in Racine (1997) is designed for the *i.i.d* data. Using the block bootstrap method introduced in Künsch (1989), I modify Racine's original algorithm to take into account the serial dependence features in the data. Details of the algorithm are presented in the Appendix 1.B.

III Excess return forecasts

One-year holding period excess returns are constructed from the monthly Fama-Bliss bond prices data set available at the Center for Research in Securities Prices (CRSP). All prices refer to zero-coupon U.S. Treasury bond prices. For bonds with maturities greater than one year, the Fama-Bliss prices are interpolated from yields on coupon bonds. One- through five-year prices are used in this paper. 1964/01-2003/12 is chosen as the in-sample period, the same as in CP.

III.1 In-sample analysis: current forward rates only

Fitted bond excess returns are plotted along with the realized excess returns (solid blue lines) in Figure 1.1. Results for bonds with different maturities are displayed in four subplots. The estimated $\hat{g}_n$ of the EMH model (1.2) and $\widetilde{\mathbf{f}}_t'\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ of the model (1.4) are used as the NP[5] (red dashed line) and linear fitting (black dotted-dashed line) respectively in each subplot.

[Figure 1.1 about here.]

Consistent with results documented in CP and many other researches, both models show that a portion of variation in bond excess return can be explained by the current forward rates. However, results also show that the linear model performs poorly when the volatility in excess return is high. For example, when the excess returns fluctuate dramatically, as they did during the periods of the energy crisis in the 1970s, high inflation in the 1980s, and after the burst of the dot-com bubble around 2000, the linear model fails to generate

[5] I do not report the estimation of (1.3) because it produces almost identical fitting results as $\widehat{g}_n$.

excess return dynamics with correct direction and magnitude. Figure (1.2) explains why the NP model outperforms the linear model. The slope coefficient $\beta_{j,t}^{(n)}$ is estimated pointwisely over the domain of each independent variable $f_t^{(j)}$ and plotted in different panels in Figure (1.2) for all bonds. The estimation shows a clear heterogeneous pattern for almost all forward rates over their domain. The shape of the slope curves are similar across different maturities, differing only with respect to some maturity dependent constants. For example, the slope of the one-year yield ($y_t^{(1)}$, the first element in $\mathbf{f}_t$) exhibits a clear mean-reversion pattern for all excess returns, which is consistent with findings in Aït-Sahalia (1996a,b) and Stanton (1997). The slope estimates also show an interesting RS-like pattern for the two- to five-year forward rates. For example the estimation suggests there are at least three regimes in the three-year forward rates: 2-5%, 8.5-11% and 12-15%, which exceeds what is usually used in the RS literature. The step-like patterns suggest a constant slope estimation within each regime and different levels across regimes.

[Figure 1.2 about here.]

I plot the time series of forward rates in Figure 1.3. Interestingly, regimes where the linear model creates a poor fit correspond to extreme forward rate values: the forward rates reached the maximum and stayed in the range between 10-15% during the high inflation era in the 1980s; almost all forward rates plummeted and stayed in the regime between 2-4% in the post-NASDAQ collapse period.

[Figure 1.3 about here.]

The goodness-of-fit is measured in R^2 for both models and reported in Table 1.1. By allowing slope heterogeneity, the range of R^2's of the NP model falls between 0.89 and 0.91, which is well above the results reported in CP.[6] In column three to five of Table 1.1, I report the p-values of the test on the joint significance of all current forward rates as proposed

[6] In CP, the R^2's of the unrestricted model (as in (1.4)), which only contains the current forward rates, range between 32% to 37%.

in Section II.3. Under the null hypothesis that excess returns are not predictable, the one year excess return will have an MA(12) structure. Correspondingly I keep 12 lags in the bootstrap exercise. I also set the block length to be 6 and 18 in the bootstrap to examine the robustness of the finding. All p-values are calculated based on 199 bootstrap replications. Results indicate the null hypothesis of no predictability is overwhelmingly rejected at the 5% level for all maturities and block lengths.

[Table 1.1 about here.]

III.2 In-sample analysis: with lagged variables

CP also examine a model with the lagged forward rates,

$$rx_{t+1}^{(n)} = b_n \boldsymbol{\gamma}^\top \left[\alpha_0 \mathbf{f}_t + \alpha_1 \mathbf{f}_{t-\frac{1}{12}} + \cdots + \alpha_{p-1} \mathbf{f}_{t-\frac{p}{12}} \right] + \varepsilon_{t+1}^{(n)} \quad . \tag{1.8}$$

Once three additional lags are included as regressors, the adjusted R^2 increases from 35% to 44%, suggesting the improvement in the marginal predictive power is economically significant. Estimations also show both $\boldsymbol{\gamma}^\top$and $\boldsymbol{\alpha}^\top$ are statistically significant. Their finding shows current prices do not contain all information that is useful in predicting future excess returns under the linear setting. As pointed out in the previous section, the improvement in excess return predictability hints at a rich information structure that is largely omitted in the current forward rates cross section under the linear setting. In other words, information is not fully revealed if a linear model is assumed. Once we allow for slope heterogeneity, we may no longer be able to conclude that lagged forward rates help predict future excess returns.

I complete the in-sample exercise by inspecting estimations from (1.3). The in-sample R^2's of the NP estimation $\hat{h}_n$ are reported in column 6 of Table 1.1. The average R^2 increases from 89% to 93% once the lagged forward rates are included, indicating an economically small improvement in excess return predictability. The joint significance of lagged forward rates are tested by examining the hypothesis $H_0 : \lambda_{n,L} = 0$, where $\lambda_{n,L}$ is defined as the expectation of the sum of squared slope coefficients on lagged forward rates: $\lambda_{n,L} = E\left[\sum_{j=1}^{5}\left(\beta^{n}_{j,t-\frac{1}{12}}\right)^2\right]$. P-values are calculated based on the empirical distribution constructed under the algorithm similar to that which is proposed in Appendix 1.B. I set the block length to be 12 in the bootstrap, and the results are shown in the column 7. Interestingly, all p-values are greater than 20%, which makes the null hypothesis of no predictive power far from being rejected at the 5% level for all n. Based on this finding, I focus on the model with only the current forward rates for the remainder of this article.

III.3 Summary and implications

I emphasize two aspects of the in-sample findings. First, by taking into account the non-linear effect, a large predictable portion of variation in bond excess returns is discovered in-sample. By comparing the results from the NP model with the linear model, I show this predictability is directly linked to the slope heterogeneity in the NP model. Second, theory tells us that lagged prices $\mathbf{P}_{t-1}$ are time-(t-1) conditional expectation of future stochastic discount factors (SDF) M,

$$P^{(n)}_{t-1} = E_{t-1}(M_{t+n-1}),$$

so it is awkward to include them in the time-t state vector to form expectations on future excess return. In contrast to the findings in CP (in which the authors show that lagged forward rates have additional predictive power in future excess returns), I show empirically that the information contained in lagged forward rates has already embedded non-linearly in the current bond prices. The results of lagged regression in this paper are consistent with requirements imposed by a reasonable SDF.

IV Out-of-sample prediction

Readers may worry about the possibility of in-sample over-fitting as the NP model produces suspiciously large R^2's. Clark and McCracken (2013) underscore the importance of pseudo out-of-sample forecasts in over-fitting detection and model evaluation. If the high R^2 in the NP model does relate to information that provides additional predictability, as one can imagine, the NP model will generate more accurate out-of-sample forecasts. In this section, I evaluate and compare the out-of-sample forecasts generated by different models.

The out-of-sample performance of three models are evaluated and compared. In addition to the NP (1.2) and linear models (1.4) discussed in the previous sections, I also examine the model that assumes yields follow a simple random walk (denoted as RW hereafter), in which

$$y_{t+1}^{(n)} = y_t^{(n)} + v_{t+1}, \quad v_{t+1} \sim i.i.d.\ N(0, \sigma^2).$$

Under the RW model, the time-t expectation of $t{+}1$ yield is the time-t observation itself, i.e.,

$$\widehat{y}_{t+1}^{(n)} = y_t^{(n)}. \tag{1.9}$$

Substituting (1.9) into the conditional expectation of (1.1), the expected excess return implied by the RW model is written

$$\widehat{rx}_{t+1}^{(n)} = n(y_t^{(n)} - y_t^{(n-1)}) + y_t^{(n-1)} - y_t^{(1)}, \quad n = 2, 3, 4, 5. \tag{1.10}$$

The RW model for the yield dynamics is a simple benchmark widely used in the literature. Duffee (2002) finds that the yields generated by the "complete affine" term structure model proposed in Dai and Singleton (2000) could not beat the random walk benchmark. Moreover, Duffee (2011a) extends the out-of-sample periods in Duffee (2002) and Diebold and Li (2006) and concludes both models fail to beat the random walk forecasts. The RW model is interesting also because it only uses information from the last period. The NP, linear and RW model essentially correspond to whole sample with different weights, whole sample with the same weight and latest observation only cases, respectively.

All models are estimated recursively, with the initial estimation using the in-sample data between 1964/01 and 2003/12 (observations from 1965/01 to 2003/12 for the dependent variable and 1964/01 to 2002/12 for independent variables). The forecast excess returns for 2004/01 are calculated from the estimated model and forward rates observed in

2003/01. Models are then re-estimated using data from 1964/01 to 2004/01 (observations from 1965/01 to 2004/01 for the dependent variable and 1964/01 to 2003/01 for independent variables) and forecasts are computed for 2004/02 and so on. The out-of-sample period covers the time frame between 2004/01 and 2010/12.

[Figure 1.4 about here.]

Figure 1.4 plots the forecasts from all three models along with the realized excess returns over the out-of-sample period. In this 7-year out-of-sample period, 2007/12/01-2009/06/01 is designated as a recession by the National Bureau of Economic Research (NBER). Realized excess returns follow a downward trend from 2004 to the beginning of 2007 but suddenly jumps to a peak in less than six months and generally remains at a relatively high level during the recession.

The forecasts from the RW model fail to match either the magnitude or dynamics of excess returns in this period. Although showing a gradual decline, the RW model consistently produces exaggerated forecasts in the pre-recession period; on the other hand, it consistently underestimates bond excess returns during the great recession in 2008. It is because the RW forecasts from (1.10) only use the time-t yield spreads to forecast excess returns one year later, which will cause severe problem if a policy shock hits the economy and structure breaks occur afterwards. In November 2008, the US Federal Reserve launched the first round of Quantitative Easing (QE). They accelerated the pace of asset purchasing in the second half of 2009. As a result, excess returns plunged with yields of long term bond. For example, the estimated 5-year excess return in October 2009, which is computed from the yield spreads observed in October 2008, before the QE started, blows out in the opposite direction. The forecasts from the linear model do not look promising either, especially in the post 2007 period. By assuming a constant slope for each independent variable, the linear model implicitly assumes an identical marginal effect for changes in forward rates in boom and recession periods. Essentially only the average effect (across potential structure changes) of an independent variable is retained under the linear setting, which conflicts with the empirical findings documented in Gray (1996) and Dai, Singleton, and Yang (2007), among others. As indirect evidence, linear forecasts in Figure 1.4 do indicate that the volatility in excess returns is underestimated. The NP model produces the most accurate forecasts among the three models, particularly in the post QE period, where the other two models fail. The discussion above indicates it is crucial to assume heterogeneous slopes, especially under the case of potential structure breaks.

To quantitatively evaluate models' performance, I adopt Diebold and Mariano (1995)(DM) test[7] to compare different models' mean squared predictive errors (MSPE). Table 1.2 presents results from the out-of-sample test. Under the columns labeled "MSPE", I report the MSPE's of the three models. The linear and RW model produce MSPE's of about the same magnitude, which on average is more than twice as large as that of the NP model. I use the DM test statistic to evaluate whether the difference in MSPE's from different models is statistically significant. P-values of the test are reported in the last two columns of Table 1.2. For both LM vs. NP and RW vs. NP cases, the null hypothesis of equal MSPE's is immediately rejected for all excess returns, which supports the qualitatively finding that the NP model produces more accurate forecasts.

[Table 1.2 about here.]

V Concluding remarks

This paper contributes to the literature by showing that the non-linearity in the yield curve, which is introduced by allowing heterogeneous slopes defined on different parts of the forward rates domain, plays an important role in bond excess return forecastability. Both the in- and out-of-sample results show that the constant-slope setting essentially takes into consideration only the average marginal effect of an independent variable, which will result in the failure of matching excess returns in the high volatility period.

This paper offers two major contributions to the literature. First, a greater portion of

[7]Readers can refer to Appendix 1.C for details about the test.

predictable variation in excess returns are discovered when the non-linearity in the current bond term structure is incorporated. In contrast to the existing literature, I also show that lagged prices do not contribute to return predictability under the NP setting. Second, the out-of-sample results mitigate the concern of in-sample overfitting. The NP model does quite well in the out-of-sample period. It produces a much smaller MSPE than the competing linear model and the simple RW benchmark. The fitted excess returns in the post recession period also provide indirect evidence that supports non-linearities.

The analysis in this paper is still considered as incomplete in many aspects. Bond yields are the only predictor used in this paper. Should this configuration be viewed as complete? The EMH says yes. However, it has been widely documented that many macroeconomic variables could provide additional predictive power. Ludvigson and Ng (2009) point out "specifications using pure financial variables omit pertinent information about future bond returns associated with macroeconomic fundamentals". Similar results are found in the hidden factor model under the affine framework. Barillas (2010) estimates two versions of hidden factor models with the Kalman filter. One only uses the current yields as observables while the other one uses yields and macro indicators. He finds the hidden factor contributes little in the former while the predictive power increases dramatically after including macro variables. An interesting extension in the future is to investigate the relation between the information embedded in heterogeneous slopes and that in macro economic variables.

1.A Appendix: Local constant estimator

In this paper, the non-parametric regression is implemented with the local constant kernel estimator, which was originally proposed in Nadaraya (1964) and Watson (1964). For the time series data, the general model being considered is

$$y_t = g(\mathbf{x}_t) + \varepsilon_{t+1}, \quad t = 1, 2, \ldots, T,$$

where $\mathbf{x}_t$ denotes $n \times 1$ independent variables $(x_{1t}, x_{2t}, \cdots, x_{nt})$. The conditional mean function $g(\cdot)$ is assumed to be smooth. Let $(y_t, \mathbf{x}_t)$ form a stationary joint distribution $f(\mathbf{x}, y)$. By definition, the conditional mean

$$\begin{aligned} g(\mathbf{x}_t) \equiv E(y_t|\mathbf{x}_t) &= \int y f(y|\mathbf{x}) dy \\ &= \int y \frac{f(\mathbf{x}, y)}{f(\mathbf{x})} dy \\ &= \frac{1}{f(\mathbf{x})} \int y f(\mathbf{x}, y) dy \end{aligned}$$

The unknown densities $f(\mathbf{x})$ and $f(\mathbf{x}, y)$ are then replaced by the estimates $\hat{f}(\mathbf{x})$ and $\hat{f}(\mathbf{x}, y)$ respectively, which gives the non-parametric estimates of

$$\hat{g}(\mathbf{x}_t) = \frac{1}{\hat{f}(\mathbf{x})} \int y \hat{f}(\mathbf{x}, y) dy \tag{1.11}$$

The local constant estimator calculates $\hat{f}(\mathbf{x})$ and $\hat{f}(\mathbf{x}, y)$ in the following manner:

$$\hat{f}(\mathbf{x}) = \frac{1}{T h_1 h_2 \cdots h_n} \sum_{t=1}^{T} K\left(\frac{\mathbf{x} - \mathbf{x}_t}{h}\right) \tag{1.12}$$

$$\hat{f}(\mathbf{x}, y) = \frac{1}{T h_0 h_1 h_2 \cdots h_n} \sum_{t=1}^{T} K\left(\frac{\mathbf{x} - \mathbf{x}_t}{h}\right) K\left(\frac{y - y_t}{h_0}\right) \tag{1.13}$$

Substituting (1.12) and (1.13) into (1.11), the unknown conditional mean function $\hat{g}$ is simply given by

$$\hat{g}(\mathbf{x}) = \frac{\sum_{t=1}^{T} Y_t K\left(\frac{\mathbf{x}-\mathbf{x}_t}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{\mathbf{x}-\mathbf{x}_t}{h}\right)},$$

where $K\left(\frac{\mathbf{x}-\mathbf{x}_t}{h}\right) \equiv \prod_{i=1}^{n} k\left(\frac{x_i - x_{it}}{h_i}\right)$. $k(\cdot)$ is a kernel function which gives weight to each point on the domain of the probability density based on how far it is from actual data points. The literature suggests that different kernel functions $K(\cdot)$ have very little impact on estimation results Epanechnikov (1969). The bandwidth parameter h_i is also known as the smoothing parameter. When implementing the nonparametric estimation, one bandwidth is chosen for each explanatory variable to minimize some loss functions. Following the literature, the bandwidth vector $\mathbf{h} = (h_1, h_2, \cdots, h_5)^\top$ is selected via least square cross validation. Details of least square cross validation and its properties can be found in references such as Li and Racine (2006).

1.B Appendix: Bootstrap algorithm and decision rule

In this section, I present the detailed bootstrap procedures for jointly testing the significance of current forward rates. The algorithm for testing the joint significance of lagged forward rates is essentially the same, so I skip the details here. As pointed out in much of the literature[8], statistical inferences based on the pivotal method are more reliable than the results obtained directly from the bootstrap without pivotaling. In this paper, instead of bootstrapping $\hat{\lambda}_n$ directly, a studentized statistic

$$\hat{t}_n = \frac{\hat{\lambda}_n}{s.e.(\hat{\lambda}_n)}$$

is bootstrapped and calculated. A block bootstrap algorithm with fixed block length l is designed for highly cross- and auto-correlated right-hand variables.

The resampling algorithm is developed using the restricted wild bootstrap method, which generally follows Racine (1997). The detailed procedure is listed as follows:

[8] See, for example, Beran (1988), Hall (1986) and Horowitz (2001), among others.

1. All forward rates do not have explanatory power jointly under the null hypothesis, so a null data sample $\left\{rx_{t+1}^{(n)}, \mathbf{f}_t^*\right\}_{t=1}^{T}$ can be created by randomly shuffling right-hand variables in blocks (size l) with replacement. Estimate the restricted conditional mean $\widehat{rx}_{t+1}^{(n)*} = \hat{g}_n\left(\mathbf{f}_t^*\right)$.

2. Generate residuals $\hat{\varepsilon}_{t+1}^{(n)} = rx_{t+1}^{(n)} - \widehat{rx}_{t+1}^{(n)*}$, then recenter around them 0.

3. Divide $\left\{\hat{\varepsilon}_{t+1}^{(n)}\right\}_{t=1}^{T}$ into $T-l+1$ overlapping blocks, in which the j^{th} block covers the residuals from period j to $j+l-1$. Construct the empirical distribution function (ECDF) F_n for these blocks by assigning an equal probability of $\frac{1}{T-l+1}$ to each block.

4. Randomly create a bootstrap residual sample with replacement under F_n, denoted as $\left\{\hat{\varepsilon}_{t+1}^{(n)*}\right\}_{t=1}^{T}$.

5. For each $\hat{\varepsilon}_{t+1}^{(n)*}$, define $\hat{\epsilon}_{t+1}^{(n)*} = \hat{\varepsilon}_{t+1}^{(n)*} \cdot v$, where

$$v = \begin{cases} -(\sqrt{5}-1)/2 & \text{w/ prob.}(\sqrt{5}+1)/2\sqrt{5} \\ (\sqrt{5}+1)/2 & \text{w/ prob.}(\sqrt{5}-1)/2\sqrt{5} \end{cases}$$

to take into account possible heteroskedasticity (Wu, 1986). Generate a null bootstrap data sample in which the dependent variable is obtained from $\widetilde{rx}_t^{(n)} = \widehat{rx}_t^{(n)*} + \hat{\epsilon}_t^{(n)*}$ and the independent variables $\mathbf{f}_t$ are inherited from the original sample.

6. Estimate (1.2) non-parametrically for the data set $\left\{\widetilde{rx}_{t+1}^{(n)}, \mathbf{f}_t\right\}_{t=1}^{T}$, use (1.7) to calculate $\hat{\lambda}_{n,i}$ for the i^{th} bootstrap data sample.

7. For each bootstrap sample, resample $\left(\widetilde{rx}_{t+1}^{(n)}, \mathbf{f}_t\right)$ by pair with the same block length l, compute $\hat{\lambda}^*$ based on this resample. Repeat this procedure for B_1 times, to obtain a sequence $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \cdots, \hat{\lambda}_{B_1}^*$. The $s.e.\left(\hat{\lambda}_{n,i}\right)$ in step 6 is the standard error of $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \cdots, \hat{\lambda}_{B_1}^*$.

8. Compute the studentized statistic $\hat{t}_{n,i}$ for the bootstrap data sample in step 6.

9. Repeat steps 4-8 independently B_2 times and create a sequence $\{\hat{t}_{n,i}\}_{i=1}^{B_2}$, construct the ECDF $\hat{F}_n^*$ of $\{\hat{t}_{n,i}\}_{i=1}^{B_2}$.

10. Compute $\hat{t}$ for the original sample and report its percentile under $\hat{F}_n^*$, the null will be rejected if $\hat{t}$ is located within the top α-percentile for a given level of significance α.

1.C Appendix: Diebold and Mariano (D-M) test

Let $\varepsilon^i_{t+1|t}$ be the $t+1$ prediction error associated with the model i at t, i.e.,

$$\varepsilon^i_{t+1|t} = rx_{t+1|t} - \widehat{rx}^i_{t+1|t}, \quad i = \text{NP, LM, RW}.$$

The accuracy of model forecasts is measured by the square loss function

$$L(\varepsilon^i_{t+1|t}) = \left(\varepsilon^i_{t+1|t}\right)^2.$$

The D-M test proposed in Diebold and Mariano (1995) compares two models' expected loss, which is the **mean squared predictive error** (MSPE) for the models. A model with smaller expected loss would be considered outperform. To determine if model i predicts better than model j, we may test the following null hypothesis

$$H_0 : E\left[\left(\varepsilon^i_{t+1|t}\right)^2\right] = E\left[\left(\varepsilon^j_{t+1|t}\right)^2\right]$$

against the one-sided alternative

$$H_1 : E\left[\left(\varepsilon^i_{t+1|t}\right)^2\right] < E\left[\left(\varepsilon^j_{t+1|t}\right)^2\right]$$

The D-M test is based on the loss difference $d_t = \left(\varepsilon^j_{t+1|t}\right)^2 - \left(\varepsilon^i_{t+1|t}\right)^2$, noting the original hypothesis could be rewritten as

$$\begin{aligned} H_0 :& \quad E[d] = 0 \\ H_1 :& \quad E[d] > 0 \end{aligned} \tag{1.14}$$

Diebold and Mariano (1995) use the following test statistic

$$\hat{S} = \frac{\bar{d}}{\left(\widehat{var}\left(\bar{d}\right)\right)^{1/2}}$$

where

$$\begin{aligned}\bar{d} &= \tfrac{1}{T}\textstyle\sum_{t=1}^{T} d_t \\ \widehat{var}\left(\bar{d}\right) &= \tfrac{1}{T}\left(\hat{\gamma}_0 + 2\textstyle\sum_{k=1}^{T-k}\hat{\gamma}_k\right), \quad \gamma_j = cov(d_t, d_{t-k})\end{aligned}.$$

They proved that under the null hypothesis, $\hat{S}$ will converge asymptotically in distribution to a normal random variable for a stationary process.

Table 1.1: Summary of model's in-sample performance

	Current forward rates only				With lags	
n	R^2	6	12	18	R^2	12
2	0.90	0.04	0.02	0.01	0.90	0.32
3	0.91	0.04	0.02	0.04	0.91	0.23
4	0.89	0.04	0.00	0.00	0.93	0.35
5	0.89	0.03	0.02	0.00	0.97	0.32

Note: This table provides the in-sample performance of the non-parametric model under two specifications. Both R^2 and p-values obtained from block bootstraps are reported. For the NP model only with the current forward rates (1.2), bootstrap with block lengths of 6, 12 and 18 are considered. For the NP model includes one-lagged forward rates (1.3), block length of 12 is used in the bootstrap. The in-sample period spans from January 1964 to December 2003.

Table 1.2: Summary of models' out-of-sample performance

	Mean square predictive error			p-value of DM test	
n	RW	LM	NP	LM vs. NP	RW vs. NP
2	3.78	3.09	1.36	0.00	0.00
3	9.33	11.15	4.66	0.00	0.00
4	14.26	21.62	9.18	0.00	0.01
5	28.79	32.56	13.51	0.00	0.00

Note: This table compares the out-of-sample performance of the non-parametric (NP) model (1.2), linear model (LM) and random walk model (RW) using the Diebold and Mariano (DM) test. Column 2-4 report the out-of-sample mean square predictive errors (MSPE) of all three models are reported for bonds with different maturities. In column 5 and 6, we report the p-values of the hypothesis that the MSPE of NP is equal to that of the competing model against one side alternative. The out-of-sample period spans from January 2004 to December 2010.

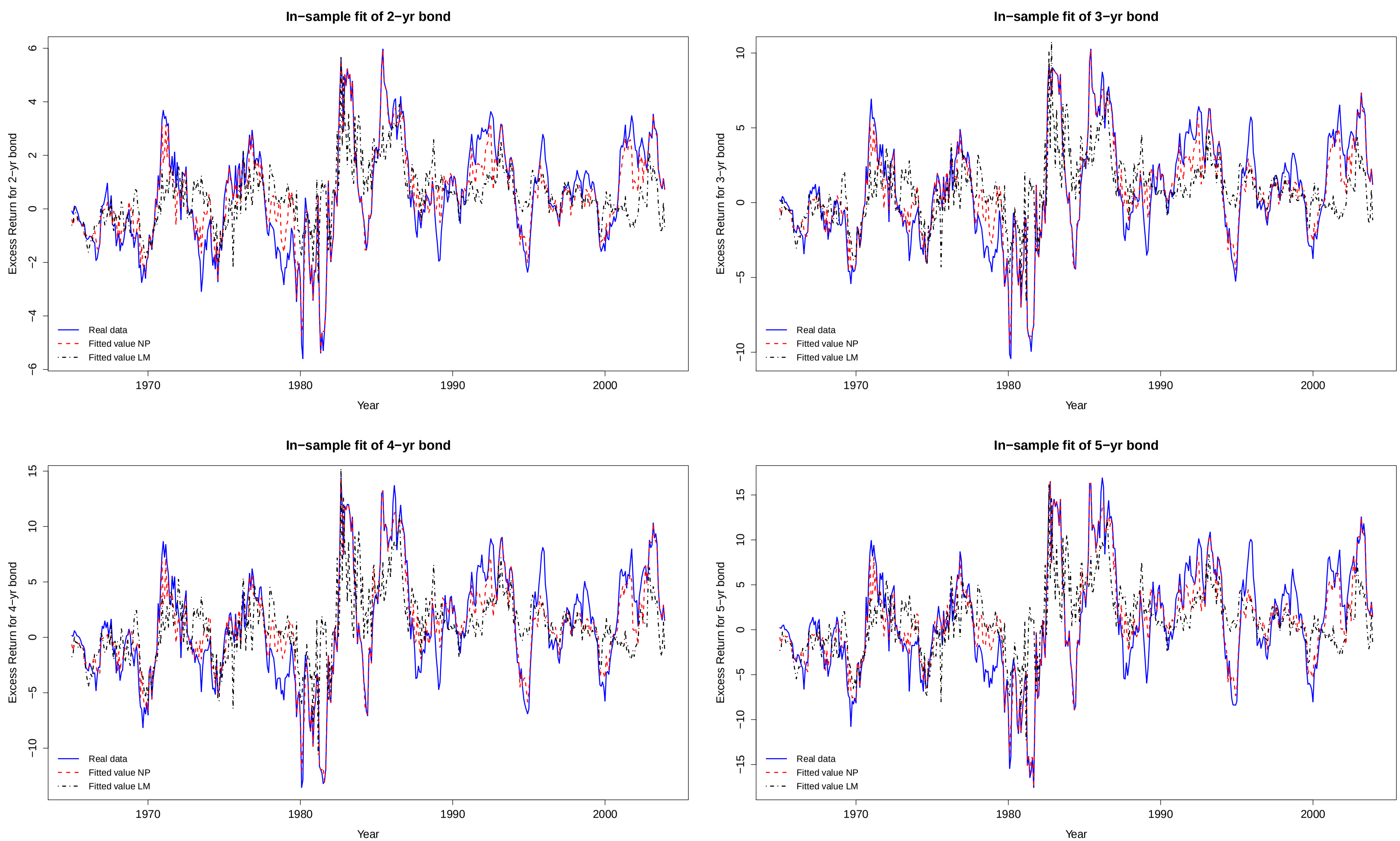
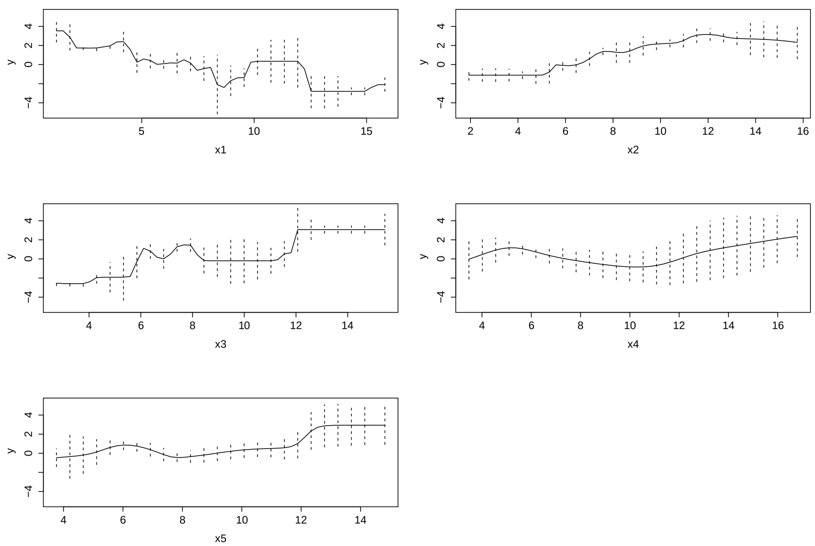
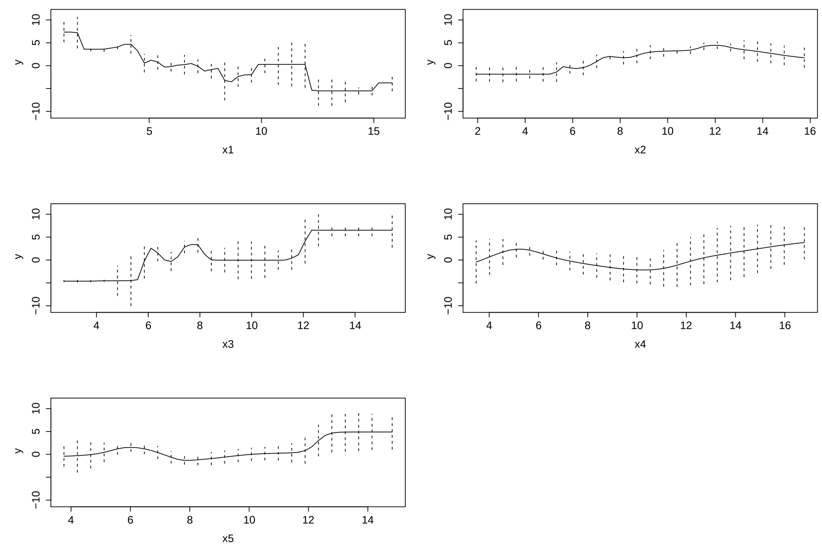


Figure 1.1: In-sample excess return forecast for 2-5 year Fama-Bliss bonds

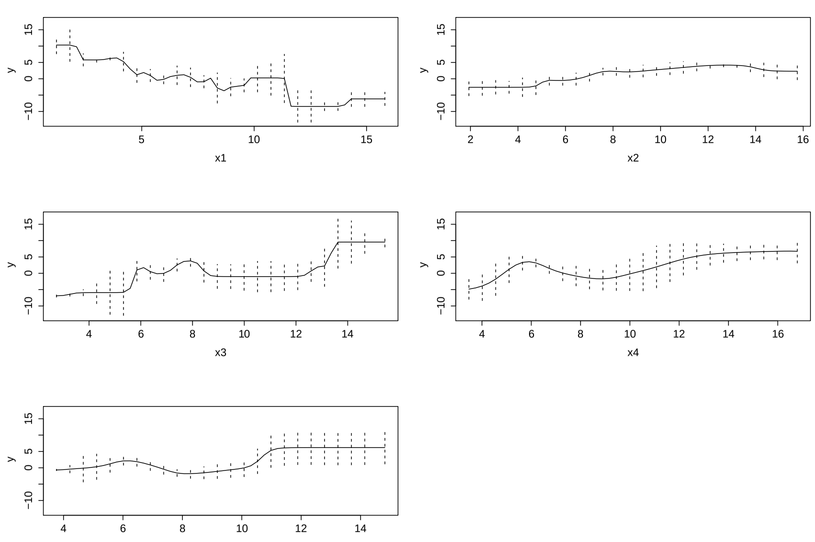
(a) Slope coefficients for $rx_{t+1}^{(2)}$.



(b) Slope coefficients for $rx_{t+1}^{(3)}$.



(c) Slope coefficients for $rx_{t+1}^{(4)}$.



(d) Slope coefficients for $rx_{t+1}^{(5)}$.

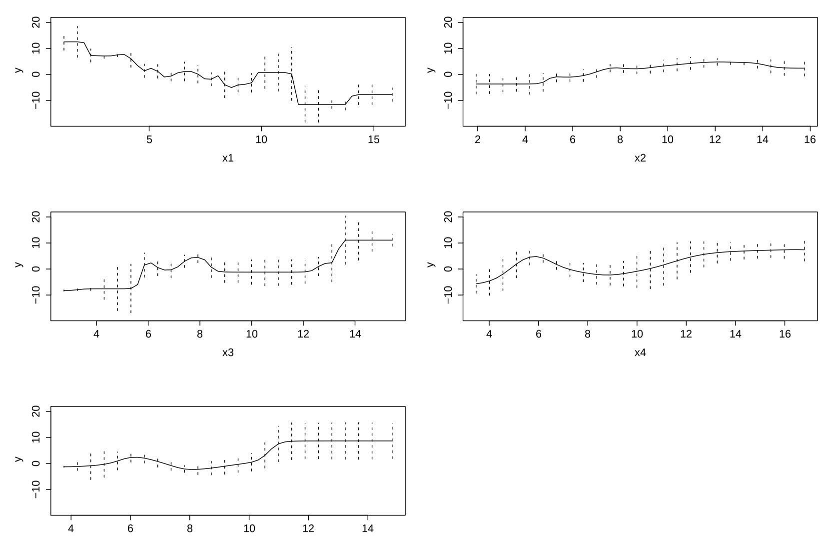


Figure 1.2: Slope coefficients evaluated at different values of independent variables for all bonds.

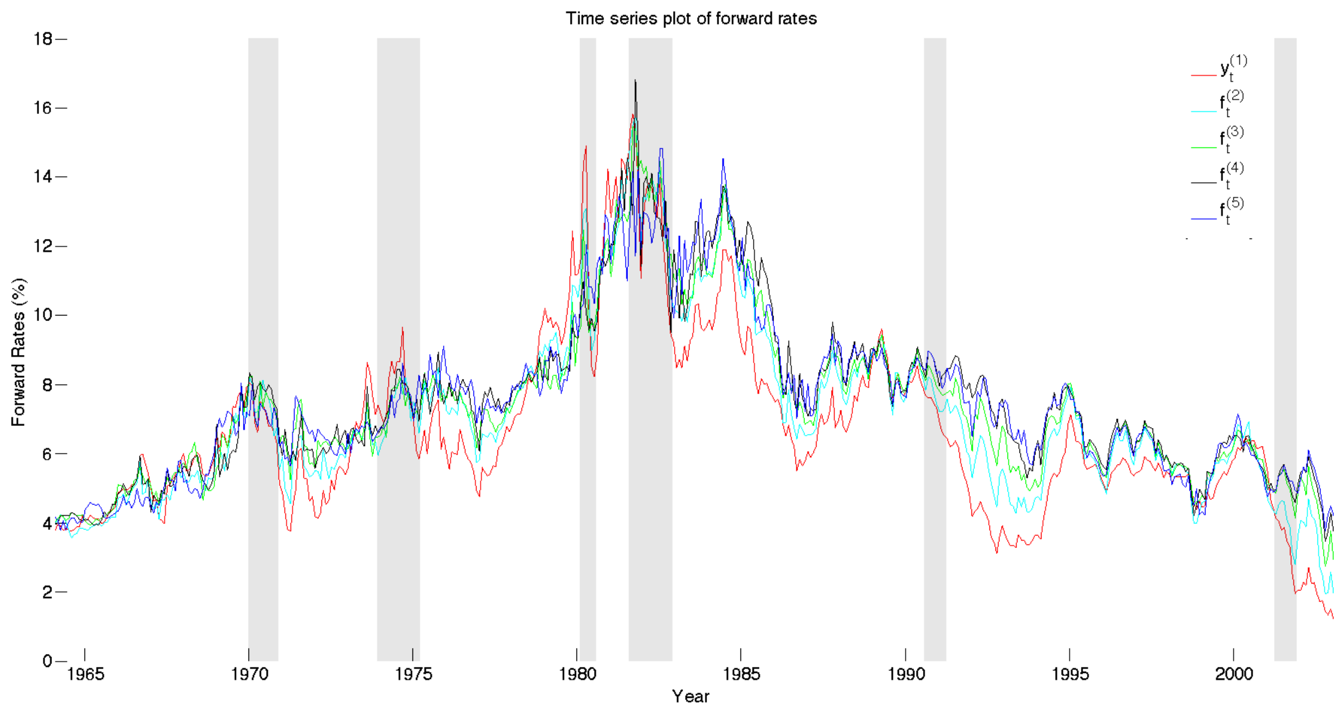
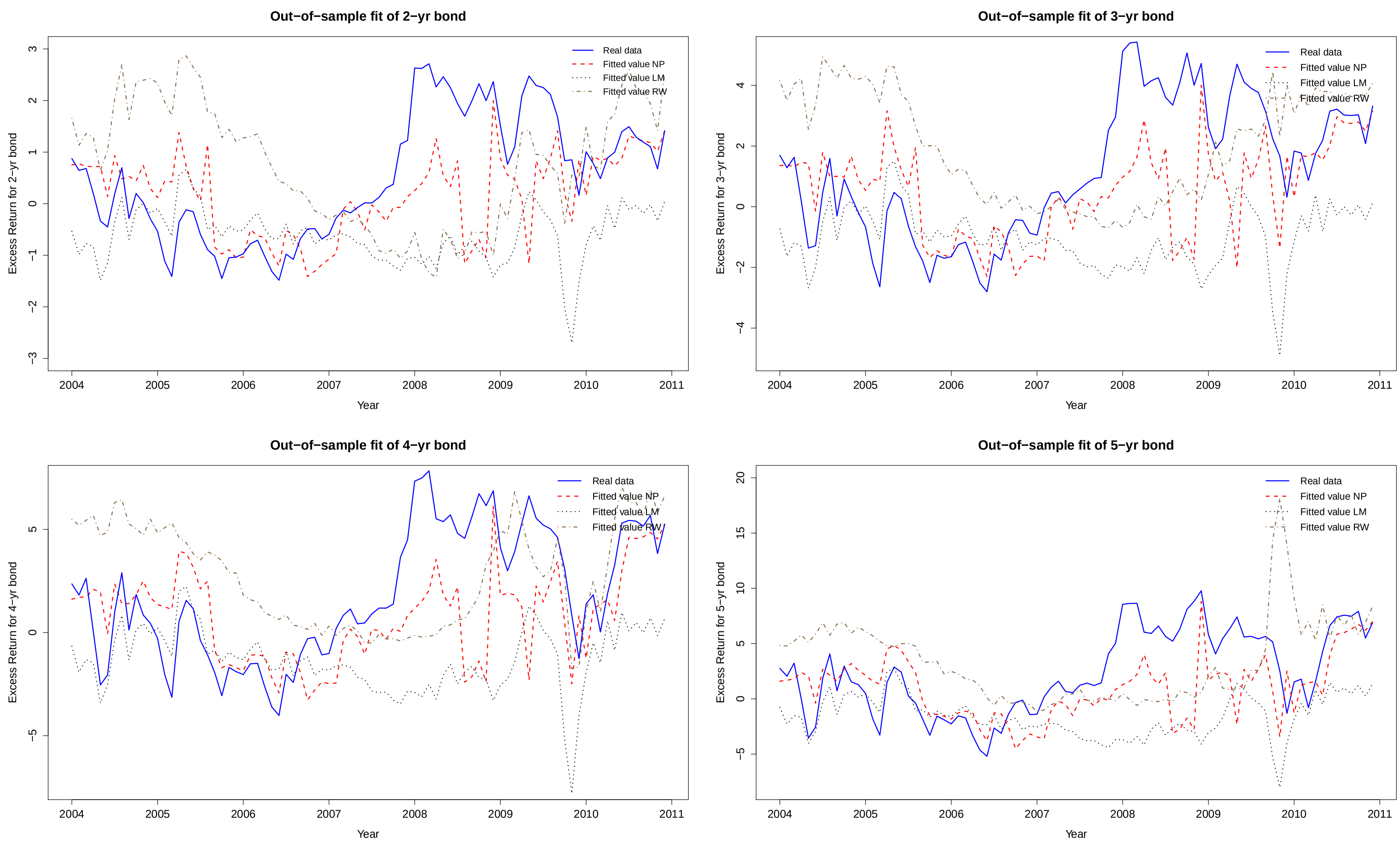


Figure 1.3: Time series plots of 2-5-yr forward rates.



Note: In all plots, solid blue line, red dashed line and brown dashed line represent real excess return, fitted excess return from non-parametric, linear and random walk model respectively.

Figure 1.4: Out-of-sample excess return forecast for 2-5 year Fama-Bliss bonds.

References

Aït-Sahalia, Y. 1996a. Nonparametric pricing of interest rate derivative securities. *Econometrica* 64:527–60.

———. 1996b. Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9:385–426.

Ang, A., and G. Bekaert. 2002. Short rate nonlinearities and regime switches. *Journal of Economic Dynamics & Control* 26:1243–74.

Ang, A., G. Bekaert, and M. Wei. 2008. The term structure of real rates and expected inflation. *Journal of Finance* 63:797–849.

Ang, A., and M. Piazzesi. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50:745–87.

Barillas, F. 2010. Macroeconomic releases and the term structure of interest rates. *Working Paper* Emory University.

Beran, R. 1988. Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83:687–97.

Campbell, J. Y., and R. J. Shiller. 1991. Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies* 58:495–514.

Cieslak, A., and P. Povala. 2013. Expected returns in treasury bonds. *Working Paper* Northwestern University.

Clark, T. E., and M. W. McCracken. 2013. Advances in forecast evaluation. *Handbook of Economic Forecasting* 2, Part B:1107–201.

Cochrane, J. H., and M. Piazzesi. 2005. Bond risk premia. *American Economic Review* 95:138–60.

Cooper, I., and R. Priestley. 2009. Time-varying risk premiums and the output gap. *Review of Financial Studies* 22:2801–33.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53:385–407.

Dai, Q., and K. J. Singleton. 2000. Specification analysis of affine term structure models. *Journal of Finance* 55:1943–78.

Dai, Q., K. J. Singleton, and W. Yang. 2007. Regime shifts in a dynamic term structure model of u.s. treasury bond yields. *Review of Financial Studies* 20:1669–706.

Diebold, F., and R. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13:253–65.

Diebold, F. X., and C. Li. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130:337–64.

Duffee, G. R. 2002. Term premia and interest rate forecasts in affine models. *Journal of Finance* 57:405–43.

———. 2011a. Forecasting with the term structure: The role of no-arbitrage restrictions. *Working Paper* Johns Hopkins University.

———. 2011b. Information in (and not in) the term structure. *Review of Financial Studies* 24:2895–934.

———. 2013. Forecasting interest rates. *Handbook of Economic Forecasting* 2, Part A:385–426.

Duffie, D., and R. Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6:379–406.

Epanechnikov, V. A. 1969. Nonparametric estimates of a multivariate probability density. *Theory of Probability and its Applications* 14:153–8.

Fama, E. F., and R. R. Bliss. 1987. The information in long-maturity forward rates. *American Economic Review* 77:680–92.

Feunou, B., and J.-S. Fontaine. 2014. Bond risk premia and gaussian term structure models. *Working Paper* Bank of Canada.

Fontaine, J.-S., and R. Garcia. 2012. Bond liquidity premia. *Review of Financial Studies* 25:1207–54.

Gray, S. F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42:27–62.

Hall, P. 1986. On the bootstrap and confidence intervals. *The Annals of Statistics* 14:1431–52.

Horowitz, J. L. 2001. The bootstrap. *Handbook of Econometrics* 5:3159–228.

Joslin, S., M. Priebsch, and K. J. Singleton. 2014. Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance* 69:1197–233.

Künsch, H. R. 1989. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17:1217–41.

Li, Q., and J. S. Racine. 2006. *Nonparametric econometrics: Theory and practice.* Princeton University Press.

Ludvigson, S., and S. Ng. 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22:5027–67.

Nadaraya, E. A. 1964. On estimating regression. *Theory of Probability and its Applications* 9:141–2.

Racine, J. S. 1997. Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics* 15:369–79.

Rilstone, P. 1991. Nonparametric hypothesis testing with parametric rates of convergence. *International Economic Review* 32:209–27.

Robinson, P. M. 1991. Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58:437–53.

———. 1994. The normal approximation for semiparametric averaged derivatives. *Econometrica* 63:667–80.

Stambaugh, R. F. 1988. The information in forward rates: Implications for models of the term structure. *Journal of Financial Economics* 21:41–70.

Stanton, R. 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52:1973–2002.

Stone, C. J. 1984. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics* 12:1285–97.

Vasicek, O. 1977. An equilibrium characterisation of the term structure. *Journal of Financial Economics* 5:177–88.

Watson, G. S. 1964. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* 26:359–72.

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* 14:1261–350.

Chapter 2

Do non-linearities matter in the yield curve?

Abstract

Do non-yield variables contain information beyond what is contained in the yield curve? This paper answers the question using a non-linear factor extracted non-parametrically from the yield curve. When the non-linear factor is included, non-yield factors, which are constructed from a large panel of macro-finance data, are no longer significant in predicting future bond excess returns both in- and out-of-sample. Moreover, my non-linear factor generates countercyclical and business cycle frequency bond risk premia. The findings underscore the importance of non-linearities embedded in the term structure, suggesting a fully spanned term structure model with non-linear state factors may be capable of matching features observed in the data.

JEL Classification: E0, E4, G0, G1.

Key words: Bond term structure, affine model, non-linearity, non-parametric methods, time-varying risk premia.

I Introduction

Empirical research in financial economics has uncovered significant forecastable variation in bond excess returns since the early work conducted in Fama and Bliss (1987), Stambaugh (1988) and Campbell and Shiller (1991). In most of the literature, factors constructed from the cross section of bond term structure are used as predictors. For example, Litterman and Scheinkman (1991) use principal component analysis (PCA) to construct the "level", "slope" and "curvature" factors, which are then used to forecast returns for holding bond portfolios. Cochrane and Piazzesi (2005) find that a linear combination of forward rates on average can explain 35% of future excess return variation. Therefore, a natural choice for the state vector in the standard affine term structure model[1] (ATSM) framework is to use yields or their linear combinations. However, more recent research shows that various measures of economic activity can provide additional predictability. Ang and Piazzesi (2003) combine bond yields with inflation and real productivity factors in a no-arbitrage vector autoregression (VAR) framework and find that models including macro factors can produce more precise forecasts. Cooper and Priestley (2009) report that the output gap has strong predictive power of bond excess returns both in- and out-of-sample. Ludvigson and Ng (2009) apply dynamic factor analysis (DFA) to a large panel of macro-finance data and find that factors closely related to real output and inflation can improve forecasting power for future excess returns. They also show that non-yield factors are the key ingredient in generating countercyclical risk premia. Fontaine and Garcia (2012) construct a liquidity factor to measure funding conditions confronted by financial intermediaries. They find that bond excess returns can be predicted by this factor, which suggests that liquidity risks might be an important source of bond excess return predictability.

These findings question the assumption made in the fully spanned ATSM that the yield curve contains all the information investors need to make future forecasts by suggesting that certain non-yield variables help predict bond yields and excess returns. Ludvigson and Ng (2009) and Joslin, Priebsch, and Singleton (2014), among many others, point out that the

[1] See for example, Duffie and Kan (1996), Dai and Singleton (2000) and Duffee (2002).

failure of the fully spanned ATSM is due to the invertibility between the state vector and yields. In other words, if the state vector is fully **linearly** spanned by the yield curve, so are the non-yield variables. Then, as Joslin, Priebsch, and Singleton (2014) point out, a direct implication of the fully spanned ATSM would be "macro [and other non-yield] variables are uninformative about the expected excess returns (risk premiums) [after conditioning on the current yield curve in a linear regression model]." Unfortunately, this statement is strongly rejected in the empirical research cited above. To resolve this conflict, Duffee (2011) and Joslin, Priebsch, and Singleton (2014) propose an ATSM in which part of the state vector is not spanned by the yield curve. Duffee (2011) assumes the existence of "hidden" factors that are undetectable from the cross section of yields but that have significant predictive power with respect to future bond excess returns. Joslin, Priebsch, and Singleton (2014) allow for macroeconomic risks that cannot be spanned by the yield curve. Two linear combinations of smoothed industrial production and inflation are chosen as the unspanned factors. They show that the unspanned factors can explain a substantial portion of the variation in forward terms premia. Barillas (2011) investigates the optimal portfolio choice problem when unspanned macro risks exist and shows significant utility gains for a risk averse agent who incorporates these unspanned macro risks.

Although the results of the unspanned ATSM look promising, the potential problem associated with this approach is that economic theory does not tell us which variables should be used as unspanned factors. For example, Cieslak and Povala (2013) argue the source of the unspanned risks is inflation. They construct a factor which is equal to the weighted average of past inflation, where the weights are estimated from survey data. However, it appears that the weights are chosen such that the dynamics of the factor follow the same pattern as the dynamics of the yields after smoothing. Therefore, Duffee (2013) argues that "there are good theoretical and empirical reasons to be skeptical of the evidence" and "the robustness of these results is not yet known." In this paper, I suggest that a term structure model with fully spanned but non-linear state variables may also be able to resolve the conflict. In fact, the non-linear features of yield dynamics are widely documented both empirically and theoretically.[2] By allowing for non-linearities, invertibility between state factors and yields generally will not exist. Therefore, in principle, it is possible that models with fully spanned non-linear state factors are compatible with the well-known empirical findings.

I empirically investigate this possibility with a non-linear factor extracted from the cross section of bond yields. The factor is constructed using non-parametric (NP) techniques.[3] I use this factor to simultaneously predict one-month ahead one-year excess returns for bonds with different maturities. Compared to the linear factor in Cochrane and Piazzesi (2005), the non-linear factor captures a greater portion of variation in bond excess returns. Results indicate that the information embedded non-linearly in the yield curve can provide substantial excess bond return forecastability. More importantly, the non-linear factor predicts excess returns beyond what is documented in the existing literature.

I then examine whether non-yield factors are informative about future bond excess returns when the non-linear factor is included. Using the same non-yield factors as in Ludvigson and Ng (2009), I find that non-yield factors fail to provide additional predictive power for future excess returns beyond what is already captured by the non-linear factor. The results are robust both in- and out-of-sample, implying that the information contained in non-yield factors is already incorporated non-linearly in the yield curve.

To answer the question of whether non-linear risks are related to macroeconomic shocks, I construct the model implied risk premia dynamics and analyze their cyclicality. The non-linear factor is able to generate countercyclical and business cycle frequency bond risk premia, while the risk premia generated by the linear Cochrane and Piazzesi (2005) factor are generally acyclical. I conclude that non-linearities in the yield curve add substantial predictability beyond what is contained in linear bond factors. Non-yield variables are not needed once non-linearities in the yield curve have been taken into account. Moreover, non-

[2]Non-linear yield dynamics are documented empirically in Gray (1996), Aït-Sahalia (1996a,b) and Stanton (1997). Theoretical approach to model non-linearities usually assumes regime switching (RS). Example includes Ang and Bekaert (2002), Dai, Singleton, and Yang (2007) and Ang, Bekaert, and Wei (2008).

[3]The NP estimation is conducted by evaluating the joint density of excess returns and yield cross section, hence non-linearities in the yield curve would be retained under the NP setting.

linearities may relate to fundamental macroeconomic shocks, which would allow us to build structure models with fully spanned non-linear state factors to jointly model the dynamics of bond yields and macro variables.

The remainder of this paper is organized as follows. In section 2, I introduce the factors used in this article and the econometric model to estimate. In section 3, I show the in- and out-of-sample performance of the non-linear factor, followed by the potential implications of the results on affine models. The risk premia decomposition and cyclicality pattern analysis are conducted in section 4. Section 5 concludes.

II The factors

II.1 Factors from the yield curve

The notation used in this paper generally follows that in Cochrane and Piazzesi (2005) (hereafter CP). The one-year bond excess return $rx_{t+1}^{(n)}$ for the n-year bond is defined as the net return for simultaneously taking a long position in the n-year bond and a short position in the one-year bond from time t to $t+1$,

$$rx_{t+1}^{(n)} = r_{t+1}^{(n)} - y_t^{(1)}.$$

Two factors extracted from the yield curve are used: the linear factor CP_t and the non-linear factor NP_t. In CP, the authors construct the linear factor CP_t by regressing the average excess returns on current forward rates in a linear model

$$\overline{rx}_{t+1} = \gamma_0 + \gamma^\top \mathbf{f}_t + \bar{\varepsilon}_{t+1}. \tag{2.1}$$

where $\overline{rx}_{t+1}$ is the simple average of excess returns for two- to five-year bonds

$$\overline{rx}_{t+1} = \frac{1}{4}\sum_{n=2}^{5} rx_{t+1}^{(n)},$$

and $\mathbf{f}_t$ represents the vector of forward rates at time t across maturities, i.e., $\mathbf{f}_t = [y_t^{(1)}, f_t^{(2)}, \cdots, f_t^{(5)}]^\top$.The vector of slope coefficients $\left[\gamma_0, \gamma^\top\right]^\top$ in (2.1) is estimated by OLS and CP_t is defined as the fitted value $\hat{\gamma}_0 + \hat{\gamma}^\top \mathbf{f}_t$.

To construct the non-linear factor, I assume that there exists a stationary function $g_n(\cdot)$ for the n-year bond such that

$$rx_{t+1}^{(n)} = g_n(\mathbf{f}_t) + \varepsilon_{t+1}^{(n)}, \quad n = 2, 3, 4, 5. \tag{2.2}$$

In order to retain non-linearities in the yield curve, I do not make additional assumptions on the format of $g_n(\cdot)$. (2.2) is estimated non-parametrically using the local constant kernel estimator (Nadaraya, 1964; Watson, 1964). Given the dependent variable $rx_{t+1}^{(n)}$ and the independent variables $\mathbf{f}_t$, the local constant kernel estimator of $\hat{g}_n(\mathbf{f})$ is written as

$$\hat{g}_n(\mathbf{f}) = \frac{\sum_{t=1}^{T} rx_{t+1}^{(n)} \cdot \mathbf{K}\left(\frac{\mathbf{f}-\mathbf{f}_t}{\mathbf{h}}\right)}{\sum_{t=1}^{T} \mathbf{K}\left(\frac{\mathbf{f}-\mathbf{f}_t}{\mathbf{h}}\right)}, \tag{2.3}$$

where $\mathbf{K}\left(\frac{\mathbf{f}-\mathbf{f}_t}{\mathbf{h}}\right) \equiv K\left(\frac{y^{(1)}-y_t^{(1)}}{h_1}\right) \cdot \left[\prod_{i=2}^{5} K\left(\frac{f^{(i)}-f_t^{(i)}}{h_i}\right)\right]$. $K(\cdot)$ is called a kernel function and assigns weight to each point based on how far it is from actual data points. The literature suggests that different $K(\cdot)$ have very little impact on estimation results (Epanechnikov, 1969). In this paper, I adopt the commonly used Gaussian kernel $K(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$. The bandwidth parameter h_i is also known as the smoothing parameter. When implementing the non-parametric estimation, one bandwidth is chosen for each explanatory variable to minimize some loss functions. Following the literature, the bandwidth vector $\mathbf{h} = (h_1, h_2, \cdots, h_5)^\top$ is selected via least square cross validation.[4]

By allowing for non-linearities, I essentially drop the constant-slope assumption made in the linear model. In a non-parametric model, the slope coefficient for an independent variable is defined as the partial derivative of the conditional mean function with respect to that variable, which is evaluated point-wisely over its domain. In this paper, I define β_j^n

[4] Section 2.2.2 of Li and Racine (2006) provides further discussion of least square cross validation and its properties.

as the slope coefficient of the forward rate $f^{(j)}$ for the n-year bond, i.e.,

$$\beta_j^n = \frac{\partial g_n(\mathbf{f})}{\partial f^{(j)}}. \tag{2.4}$$

Substituting $g_n(\mathbf{f})$ in (2.4) with (2.3), $\hat{\beta}_j^n = \frac{\partial \hat{g}_n(\mathbf{f})}{\partial f^{(j)}}$ is a consistent estimator of β_j^n.

Estimations of (2.2) show a substantial increase in bond excess return predictability across different maturities. On average, 90% of the variation in bond excess returns can be explained once non-linearities are incorporated.[5] The slope coefficient β_j^n is estimated point-wisely over the domain of each forward rate $f^{(j)}$ and plotted in different panels in Figure (1.2) for all bonds. Plots of $\hat{\beta}_j^n$ exhibit clear heterogeneous patterns for almost all forward rates over their domain. More importantly, the shape of the slope curves β_j^n are similar, differing only with respect to some maturity dependent constants.

Figure 1.2 motivates the following parsimonious model,

$$rx_{t+1}^{(n)} = \beta_n \cdot g(\mathbf{f}_t) + \varepsilon_{t+1}^{(n)}. \tag{2.5}$$

The same non-linear function of forward rates $g(\mathbf{f}_t)$ forecasts excess returns at all maturities. Different maturities differ only in loadings β_n on $g(\mathbf{f}_t)$. Model (2.5) is not identified unless an additional restriction is introduced. I use the same restriction as in CP,

$$\frac{1}{4}\sum_{n=2}^{5} \beta_n = 1. \tag{2.6}$$

Combining (2.5) and (2.6), we have

$$\overline{rx}_{t+1} = g(\mathbf{f}_t) + \bar{\varepsilon}_{t+1}, \tag{2.7}$$

I estimate (2.7) using the same non-parametric estimator as that which is discussed above. The non-linear factor NP_t is defined as the fitted value $\widehat{g(\mathbf{f}_t)}$ in (2.7).

[5]Detail results are available upon request.

II.2 The non-yield factors $\widehat{F}_t$

I use the same non-yield factors $\widehat{F}_t$ as those in Ludvigson and Ng (2009) (hereafter LN), which are summarized from a 132-variable panel of macro-finance data using the DFA discussed in Bai and Ng (2002, 2006). Under the information criteria developed in Bai and Ng (2002), the first 8 principal components (PCs) form a good proxy for this panel. The components of $\widehat{F}_t$ are then determined by assessing the predictive regression

$$rx_{t+1}^{(n)} = \gamma_n + \alpha_n' \widehat{F}_t + \beta_n Z_t + \epsilon_{t+1}^{(n)}. \tag{2.8}$$

Combinations of these eight PCs (including the current and lagged PCs and their simple power functions) are examined in LN. The preferred set of $\widehat{F}_t$ is the one that minimizes the Bayesian information criterion (Stock and Watson, 2002b). Depending on the choice of Z_t, two versions of (2.8) are estimated. In the restricted model

$$rx_{t+1}^{(n)} = \gamma_n + \alpha_n' \widehat{F}_t + \epsilon_{t+1}^{(n)},$$

the non-yield factors $\widehat{F}_t$ are given by $\widehat{F}_t = \overrightarrow{F6}_t = \left(\widehat{F}_{1t}, \widehat{F}_{1t}^3, \widehat{F}_{2t}, \widehat{F}_{3t}, \widehat{F}_{4t}, \widehat{F}_{8t}\right)^\top$. The second PC $\widehat{F}_{2t}$, which loads heavily on variables in the categories of money, credit and the financial sector, will lose its marginal predictive power once CP_t is included in the model. LN argue that the information contained in $\widehat{F}_{2t}$ is generally contained in CP_t as well. As a result, the five-factor vector $\overrightarrow{F5}_t = \left(\widehat{F}_{1t}, \widehat{F}_{1t}^3, \widehat{F}_{3t}, \widehat{F}_{4t}, \widehat{F}_{8t}\right)^\top$ is used in (2.8) instead of $\overrightarrow{F6}_t$ when Z_t=CP_t. Since the non-linear factor NP_t captures more information from the yield curve than the linear factor CP_t, following the same argument in LN, I expect $\widehat{F}_{2t}$ will become insignificant when Z_t=NP_t. In this paper, both non-yield factors $\overrightarrow{F6}_t$ and $\overrightarrow{F5}_t$ are examined.

II.3 The econometric model

If the non-linear factor NP_t is truly a part of the state vector, it should have unconditional predictive power for future bond excess returns. I first examine the validity of this statement by testing the significance of β_n's in (2.9),

$$rx_{t+1}^{(n)} = \beta_n \cdot NP_t + \epsilon_{t+1}^{(n)}. \tag{2.9}$$

for bonds with different maturities.

Then, I address the question of whether the information contained in the non-yield factor $\widehat{F}_t$ overlaps with that contained in NP_t. In other words, I investigate whether non-yield factors are informative about future excess returns when conditioning on the potential non-linear state factor NP_t. Two specifications of regression (2.8) are evaluated for all n, in which the vector $\widehat{F}_t$ denotes the six-factor vector $\overrightarrow{F6}_t$ and five-factor vector $\overrightarrow{F5}_t$ respectively. The factor Z_t in (2.8) is set to be NP_t for both specifications, i.e.,

$$rx_{t+1}^{(n)} = \gamma_n + \alpha_n' \widehat{F}_t + \beta_n \cdot NP_t + \epsilon_{t+1}^{(n)}. \tag{2.10}$$

The results are compared with the linear benchmark in LN, where Z_t=CP_t.

III Empirical results

III.1 Data

The forward rates and one-year bond excess returns are constructed using the monthly Fama-Bliss (F-B) bond prices data set obtained from the Center for Research in Securities Prices (CRSP). The F-B data set contains zero-coupon U.S. treasury prices interpolated from yields of coupon bonds. I use bonds maturing in one to five years to construct excess returns $rx_{t+1}^{(n)}$ (n=2,3,4 and 5) between 1964/01 and 2003/12 for the in-sample period. The large macro-finance panel, which is used to construct the non-yield factors $\overrightarrow{F6}_t$ and $\overrightarrow{F5}_t$, are provided by Stock and Watson (Stock and Watson (2002a, 2005, 2006)). It covers broad macro-finance categories including output and income, employment and hours, orders and housing, money, credit and prices of financial assets. For the details of the Stock and Watson macro-finance panel, readers can refer to Ludvigson and Ng (2009), in which the complete list is provided in their appendix.

The in-sample analysis covers four specifications of (2.8): (a) NP_t only , (b) $\overrightarrow{F6}_t$ only, (c) $\overrightarrow{F6}_t + NP_t$, (d) $\overrightarrow{F5}_t + NP_t$. Results are discussed in the following subsections.

III.2 In-sample analysis: NP_t only

If the non-linear factor NP_t is truly able to serve as a state factor, it should have unconditional predictive power of future bond excess returns, which is tested under the NP_t only specification. The multiple regressions in (2.9) are evaluated for different n. The point estimates $\hat{\beta}_n$, heteroskedasticity and serial-correlation adjusted t-statistics, together with the adjusted R^2, are reported in rows labeled (a) in Table 2.1. The non-linear factor NP_t is quite persistent with a first order autocorrelation of 0.92; hence, I adjust the estimates of $\hat{\beta}_n$ to take into account the small sample bias in Stambaugh (1999). Following the argument in LN, the standard errors are obtained using the Newey and West (1987) correction with 18 lags.

The estimated $\hat{\beta}_n$'s are positive for all bonds, hinting that the dynamics of NP_t are countercyclical. Figure 2.1 shows the time series plots of the 12-month moving average of industrial productivity growth (IP growth, IP_t) and the non-linear factor NP_t (in panel A) and the linear factor CP_t (in panel B) over the in-sample period. Shaded areas indicate dates designated by the National Bureau of Economic Research (NBER) as recession periods. The correlation coefficient between IP growth and the factor is reported in the upper left corner of each panel. The non-linear factor is negatively correlated with IP growth (with a correlation of -0.25) while the dynamics of the linear factor are almost acyclical (the correlation between the IP growth and CP_t is just -0.09). The countercyclicality of NP_t is also directly seen in the plot. For example, the factor NP_t rebounded from its low in 1981/06, which is just one month before the recession started. It continued to rise during the whole recession period until 1982/11, when the recession officially ended. Then NP_t

decreased dramatically as the economy recovered.

[Figure 2.1 about here.]

The non-linear factor NP_t is significant both statistically and economically for all bonds. On average, 85% of the variation in bonds excess returns can be captured by this factor. By comparison, the linear factor CP_t, on average, can explain only 35% of the next year's excess returns. Results indicate that the factor NP_t has unconditional predictive power, in which most of the predictability comes from non-linearities in the yield curve.

III.3 In-sample analysis: non-yield factors

The specification (b) replicates what has been done in LN. On average, the non-yield factors $\overrightarrow{F6}_t$ explain 22% of the variation in excess returns. Parameter estimates $\hat{\alpha}$ in (2.8) are statistically significant at the 5% level or better. As LN conclude in their paper, the non-yield factors $\overrightarrow{F6}_t$ contain substantial information about future bond excess returns unconditionally.

Specifications (c) and (d) investigate the extent to which the information contained in the non-yield factors $\overrightarrow{F6}_t$ and $\overrightarrow{F5}_t$ are orthogonal to that incorporated in NP_t, which is essentially equivalent to examining the possibility that NP_t could serve as the state factor of a fully spanned model. As discussed in the introduction, if NP_t is a plausible candidate for the state factor, other variables that are not obtained from the yield curve (such as non-yield factors) should be conditionally uninformative about future excess returns. Z_t in (2.8) is set as NP_t for both specifications. The point estimates of each parameter (except for the constants $\hat{\gamma}_n$), heteroskedasticity and serial correlation robust t-statistics, and adjusted R^2 are reported in the rows labeled (c) and (d) in Table 2.1. I also test and report the joint significance of the slope coefficients α_n using the χ^2 statistic and its corresponding p-value, where the standard errors are calculated after a Newy-West (18) correction.

[Table 1.1 about here.]

Table 2.1 shows that in specifications (c) and (d), the second PC $\widehat{F}_{2t}$ loses its marginal significance, as expected, in all predictive regressions both economically and statistically when the regression (2.8) is evaluated conditionally on the non-linear factor NP_t. Results also show that the non-linear factor NP_t possesses the largest marginal effect on future bond excess returns among all factors. For instance, a 1% increment in NP_t, on average, boosts the five-year excess return by 1.56%. Meanwhile, the predictability of non-yield factors deteriorates. The first PC $\widehat{F}_{1t}$— which is strongly correlated with employment and production (also named the "real factor" in LN)—has the largest marginal effect on future excess returns in specification (b). However, the marginal impact of $\widehat{F}_{1t}$ drops more than 60% for all bonds when NP_t is included as a predictor. A 1% increment in $\widehat{F}_{1t}$ used to cause the five-year excess return to decline by 2.24% on average in (b), but causes only a 0.29% decline after incorporating NP_t. Moreover, the slope coefficients associated with non-yield factors also lose their statistical significance as maturity increases. For example, the t-statistic of $\widehat{F}_{1t}$ decreases from -5.13 to -3.18 (in (c)) and -3.24 (in (d)) respectively for the two-year bond, but the same t-statistic for the five-year bond declines from -4.29 to -1.13 (in both (c) and (d)), in which case we fail to reject the null hypothesis of no predictive power. The same pattern can be found in other non-yield factors as well. For example, the eighth PC $\widehat{F}_{8t}$, which produces the second largest marginal effect on excess returns in (b), loses its predictive power even for the two-year bond (point estimation falls from 0.35 to 0.04; t-statistics from 4.23 to 1.2). The results of χ^2 tests reach the same conclusion. Although the null hypothesis that non-yield factors are jointly insignificant is rejected for the two-year bond, we fail to reject the same null hypothesis for bonds that will mature in three to five years at the 5% level or better.

III.4 Out-of-sample results

The in-sample findings support the idea of fully spanned non-linear factors. Following suggestions in Clark and McCracken (2013) and Duffee (2013), I examine the robustness of in-sample results using out-of-sample forecasts. The goal of this exercise is to determine whether the insignificance of non-yield factors in-sample is caused by non-linearities. Two comparisons on four models are conducted. In the first comparison, I simply follow LN by comparing the out-of-sample performance of the model (2.8) with $\overrightarrow{F5}_t$ and the linear factor CP_t (the unrestricted model) to the model with CP_t and a constant (the restricted model). Then, I replace the linear factor with the non-linear factor NP_t in both models and re-evaluate the performance in the out-of-sample period. Under this setting, the unrestricted model will always outperform the restricted model in-sample because the unrestricted model includes more explanatory variables. However, the out-of-sample results depend on whether the non-linear factor truly has predictive power.

Four different out-of-sample time periods are examined: 1. 1987/01-2003/12; 2. 1987/01-2008/12; 3. 1995/01-2003/12; 4. 1995/01-2008/12. Forecasts are conducted on a rolling window basis. Using the first sample period as an example, the forecast results are obtained as follows. The initial estimates are obtained from the regression over the period 1964/01 to 1986/12 (observations between 1965/01 and 1986/12 for the dependent variable; 1964/01 and 1985/12 for the independent variables). Next, I substitute the parameter estimations $\left(\widehat{\gamma}, \widehat{\alpha}, \widehat{\beta}\right)$ into (2.8) to calculate the expected excess returns for 1987/01 using the observation on 1986/01. The parameters (γ, α, β) are then re-estimated with data from 1964/02 to 1987/01 and a forecast is computed for 1987/02, and so on. A similar procedure is applied to the other sample periods. I use the adjusted mean squared predictive errors (MSPE-adj) test statistic proposed in Clark and West (2007) to compare the out-of-sample MSPE of the restricted model and the unrestricted model. The null hypothesis being tested is that the MSPE of the unrestricted model is the same as that of the restricted model (the restricted model encompasses the unrestricted model).

Table 2.2 shows the results of one-year excess returns for all bonds in the out-of-sample periods. The MSPE of the unrestricted (labeled MSE_u) and restricted models (labeled MSE_r), the value of the MSPE-adj test statistic and its corresponding p-value are reported in the table. Clark and West (2007) prove that MSPE-adj approximately follows the standard normal distribution.

[Table 1.2 about here]

At least two conclusions can be drawn from the results presented in the table. First, the out-of-sample results support the conclusion reached by LN,[6] Duffee (2011) and Joslin, Priebsch, and Singleton (2014) that the linear factor CP_t cannot be used as the state factor. I find from the first comparative study that the MSPE from the unrestricted model ($\overrightarrow{F5}_t$ + CP_t) is significantly (at the 5% level or better) smaller than that from the restricted model ($const$ + CP_t) for all bonds in all sample periods. The out-of-sample results imply that the non-yield factors truly provide additional information about future bond excess returns in-sample. In fact, the results indicate that the non-yield factors are informative about excess returns conditional on **any linear** function of bond prices. The second conclusion is much more interesting. The results show that if the information in the yield curve is summarized non-linearly (NP_t is just one of the candidates), the restricted model ($const$ + NP_t) will encompass the unrestricted model ($\overrightarrow{F5}_t$ + NP_t) in the out-of-sample periods. In the first two forecast samples, with a 22-year rolling window period, two out of eight comparisons favor the unrestricted model under the Clark and West criterion.[7] With a 30-year rolling window period, the MSE_r is smaller than MSE_u in all eight comparisons! The findings suggest that conditional on the non-linear factor NP_t, the non-yield factors are uninformative about future excess returns, which supports the idea that a fully spanned non-linear factor (such as NP_t) may be able to serve as the state factor in a dynamic term structure model (DTSM).

Another comparison can be made between the model with $\overrightarrow{F5}_t$ + CP_t and the model with $const$ + NP_t. The out-of-sample MSPE of the model with linear and non-yield factors is much larger than the MSPE of the model with the non-linear factor. I do not apply a

[6]Although the same predictors as in LN are used in this paper, LN's model updates under the recursive scheme.

[7]Although the unrestricted model produces smaller MSPE for the two- and three-year bonds in the shorter out-of-sample period between 1987/01 to 2003/12, the results reverse when the out-of-sample period extends to 2008/12.

formal test[8] to investigate whether the difference is significant, but the conclusion should be obvious.

III.5 Implications of the findings

The results in Tables 2.1 and 1.2 indicate that a properly constructed non-linear factor from the yield curve can produce good forecasts of excess bond returns both in- and out-of-sample. The non-yield factors in LN provide additional predictability when the information set only contains linear yield factors (such as CP_t).

These findings reveal the failure of the commonly employed fully spanned ATSM (as we expect), in which the state vector is a linear combination of bond yields. Under invertibility, it follows that bond yields can serve as state factors. The model produces counterfactual empirical results. The existing literature proposes several potential resolutions to this conflict, the main theme of which is to break the invertibility. Besides the unspanned ATSM framework in Duffee (2011) and Joslin, Priebsch, and Singleton (2014), Ang, Piazzesi, and Dong (2007) model all yields with measurement errors; Collin-Dufresne and Goldstein (2002) assume unspanned stochastic volatility. The findings in this paper show that a fully spanned non-linear state factor may also be able to resolve this conflict. In addition, the results indirectly show that the excess return forecasts based on this non-linear state variable (NP_t) are more accurate than those generated by the the unspanned framework ($\overrightarrow{F5}_t$ + CP_t) in Joslin, Priebsch, and Singleton (2014). Although the non-linear factor in this paper is constructed by a purely data driven method, there is hope that similar results can be obtained from a properly modeled non-linear pricing kernel.

IV Risk premia decomposition

Excess returns are compensation for a risk averse investor to bear various risks; hence, a natural question to ask is what is the source of the risk as related to non-linearities? The

[8]One can apply the Diebold and Mariano (1995) test to compare the out-of-sample performance of the two models.

answer to this question is crucial in both theory and policy. In much of the literature,[9] the macro risks, which are associated with business cycles and general macro activities, are modeled as the source of time varying risk premia. Therefore, for a fully spanned DTSM with non-linear state variables to be consistent with economic theory, macro risks should be captured by non-linear state factors. In other words, the model implied risk premia dynamics must be consistent with the consumption decision made by a risk averse agent. One direct implication is that the model implied risk premia dynamics should be countercyclical, i.e., long term bonds should offer higher expected excess returns in a bad state (when the economy is contracting), and lower expected excess returns in a good state (when the economy is expanding). As pointed out in LN, policy makers care about the answer because they are concerned about "the extent that fluctuations in long term yields reflect investor expectations of future short rates vs. changing risk premia".

To figure out whether the macro risks have been incorporated in the yield curve, I directly examine the cyclicality of the risk premia dynamics generated by different factors. Consider a bond that will mature in n years. The current yield $y_t^{(n)}$ can be written as sum of the average expected future short rates and *yield risk premium* $\varkappa_t^{(n)}$, i.e.,

$$y_t^{(n)} = \underbrace{\frac{1}{n} E_t \left(y_t^{(1)} + y_{t+1}^{(1)} + \cdots + y_{t+n-1}^{(1)} \right)}_{\text{average expected future short rates}} + \underbrace{\varkappa_t^{(n)}}_{\text{yield risk premium}} . \tag{2.11}$$

where $\varkappa_t^{(n)}$ is the average expected future *return risk premia* of declining maturity:

$$\varkappa_t^{(n)} = \frac{1}{n} E_t \left[rx_{t+1}^{(n)} + rx_{t+2}^{(n-1)} + \cdots + rx_{t+n-1}^{(2)} \right] . \tag{2.12}$$

The expectations in (2.12) are obtained based on the information set at time t. To form the multi-step ahead forecast of return risk premia, I follow LN by adopting the VAR(12)

[9] For example, Campbell and Cochrane (1999) introduce a slow-moving habit term to the investor's utility function. In this model, the equity risk premium varies with the difference between consumption and the habit term. Shocks to aggregate consumption will produce countercyclical risk premia, i.e., the risk premia increase in an economic down turn and vice versa. Wachter (2006) extends Campbell and Cochrane's work and show that similar countercyclical dynamics can also be found in bond risk premia.

model on monthly bond excess returns.

$$Z_{t+\frac{1}{12}} = c + \Phi_1 Z_t + \Phi_2 Z_{t-\frac{1}{12}} + \cdots + \Phi_{12} Z_{t-\frac{11}{12}} + \varepsilon_{t+\frac{1}{12}}. \tag{2.13}$$

The role non-linearities play in risk premia dynamics is examined by three Z_t specifications. All specifications include the current excess returns $rx_t^{(n)}$ as the common components. Differences are attributed to the choice of bond yield factors and whether the non-yield factors are included. The configurations of Z_t are listed as follows:

$$Z_{1t} = \left[rx_t^{(2)}, rx_t^{(3)}, rx_t^{(4)}, rx_t^{(5)}, CP_t\right]^\top, \quad (5 \times 1)$$

$$Z_{2t} = \left[rx_t^{(2)}, rx_t^{(3)}, rx_t^{(4)}, rx_t^{(5)}, CP_t, \overrightarrow{F5}_t\right]^\top, \quad (10 \times 1)$$

$$Z_{3t} = \left[rx_t^{(2)}, rx_t^{(3)}, rx_t^{(4)}, rx_t^{(5)}, NP_t\right]^\top. \quad (5 \times 1)$$

The differences in risk premia dynamics are caused by the unique components in Z_t, which are the linear factor CP_t, the linear factor CP_t and the non-yield factors $\overrightarrow{F5}_t$, and the nonlinear factor NP_t, respectively. $\left(\widehat{c}, \hat{\Phi}_1, \hat{\Phi}_2, \cdots, \hat{\Phi}_{12}\right)$ is estimated from (2.13) via OLS. Under the VAR setting, multi-step ahead forecasts can be obtained by iterating the one-step ahead forecasts. Hence, I re-define the variables in Z_t as their deviations from the unconditional mean $\hat{\mu}$,

$$\widetilde{Z}_t = Z_t - \hat{\mu},$$

where $\hat{\mu} = (I - \widehat{\Phi}_1 - \widehat{\Phi}_2 - \cdots - \widehat{\Phi}_{12})^{-1}\hat{c}$. I can then write the one-step ahead forecasts $E_t\widetilde{Z}_{t+\frac{1}{12}}$ as

$$E_t\widetilde{Z}_{t+\frac{1}{12}} = \sum_{i=0}^{11} \hat{\Phi}_{i+1}\widetilde{Z}_{t-\frac{i}{12}} \tag{2.14}$$

Defining the stack vector $\xi_t = \left[\widetilde{Z}_t, \widetilde{Z}_{t-\frac{1}{12}}, \cdots, \widetilde{Z}_{t-\frac{11}{12}}\right]^\top$, (2.14) can be rewritten as

$$E_t\xi_{t+\frac{1}{12}} = A\xi_t,$$

where

$$A = \begin{bmatrix} \hat{\Phi}_1 & \hat{\Phi}_2 & \hat{\Phi}_3 & \cdots & \hat{\Phi}_{11} & \hat{\Phi}_{12} \\ I & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & I & \mathbf{0} \end{bmatrix}.$$

Hence, the h-step ahead forecasting is simply

$$E_t \xi_{t+\frac{h}{12}} = A^h \xi_t.$$

Defining the vector e_i such that $e_i{}'\xi_t$ picks out the ith element of ξ_t, the $\frac{h}{12}$-year return risk premia of the n-year bond $E_t rx^{(n)}_{t+\frac{h}{12}} = e'_{n-1} A^h \xi_t$.

The cyclical properties of the yield risk premia $\varkappa^{(n)}_t$ and the return risk premia $E_t rx^{(n-i)}_{t+i+1}$ from different VAR vectors are illustrated in Figures 2.2 to 2.4. Figure 2.2 plots the time series of the 12-month moving average of the model implied return risk premia for the five-year bond, $E_t rx^{(5)}_{t+1}$, along with the 12-month moving average of the IP growth in three panels. Results under specifications Z_{1t}, Z_{2t} and Z_{3t} are displayed in panels A, B and C, respectively. At first glance, all specifications successfully generate business cycle frequency risk premia dynamics. However, if we take a closer look at the correlation between the return risk premia and the IP growth (correlation -0.09), results indicate that the risk premia in Z_{1t} specification are almost acyclical. If the non-yield factors $\overrightarrow{F5}_t$ are also included in the VAR vector, as LN suggest (corresponding to Z_{2t}), the return risk premia dynamics will show stronger countercyclical patterns with a correlation of -0.16. Interestingly, the model implied return risk premia from Z_{3t} exhibit much stronger countercyclical patterns (correlation -0.23). Results also show that the return risk premia from the non-linear factor (Z_{3t}) are largest in magnitude and volatility. Panel A of Figure 2.4 plots the time series of 12-month moving average of return risk premia for all three Z_t. The mean/standard deviation of the expected return risk premia per annum are 1.03%/4.86%, 1.08%/4.94% and 1.13%/5.77% for the linear, linear and macro, and non-linear factor models, respectively. Although these numbers appear almost the same in magnitude, the figure shows that a major difference happens during the recession period. For example, during the 2002 recession, the figure shows that the maximum difference between the specifications Z_{3t}/Z_{1t} (Z_{3t}/Z_{2t}) was 6.29% (4.77%) per annum, which is about five (four) times greater than the mean value.

[Figure 2.2 about here.]

[Figure 2.4 about here.]

Similar patterns can be found in the estimated yield risk premia. Figure 2.3 shows the time series plots of the 12-month moving average of the model implied $\varkappa_t^{(5)}$ for the five-year bond, together with the IP growth. When I only use the linear factor to estimate (2.13), the yield risk premia $\varkappa_t^{(5)}$ is almost acyclical (panel A, correlation -0.03). The dynamics generated by the non-yield factors (panel B, correlation -0.38) and the non-linear factor (panel C, correlation -0.22) show stronger countercyclical patterns. Properties of the mean and standard deviation of the expected yield risk premia are similar to those of the expected return risk premia. The time series of $\varkappa_t^{(5)}$ from different models are displayed in panel B of Figure 2.4. By incorporating non-linearities, the mean yield risk premia increases from 0.7% per annum (produced by Z_{1t}) to 0.74% per annum (produced by Z_{3t}). The standard deviations for the three specifications are 1.06%, 1.14% and 1.21% respectively.

[Figure 2.12 about here.]

All figures share at least two common aspects. First, with all specifications of Z_t, the model implied risk premia generally increase during the recession period and decline during recovery. However, the linear factor constructed from the yield curve fails to generate noticeable countercyclical dynamics for the risk premia, which is inconsistent with economic theories. Second, in contrast to LN, introducing non-yield factors is not the only way to resolve the conflict. By including the non-linear factor, which only uses the information embedded in the yield curve, risk premia dynamics also exhibits strong countercyclicality. Moreover, the risk premia produced by the non-linear factor is much larger than those from other specifications in the recession period.

The results show that fully spanned DTSMs with non-linear state variables may be able to generate risk premia compatible with economic theories, which provides indirect evidence about the source of the risks associated with the non-linear factor. As opposed to the linear channel assumption in standard macro finance models (Ang and Piazzesi, 2003), results in this paper suggest that bond prices may be affected by fundamental shocks through a non-linear channel. The impact of the non-linear channel on macro finance models are generally unknown and model dependent. How to incorporate such a channel into macro finance models to match features observed in the data remains an open question, which can lead to potentially interesting theoretical research in the future.

V Concluding remarks

In the standard fully spanned ATSM, when conditioning on factors constructed from yields, non-yield variables should not have predictive power on future bond excess returns. However, a conflict arises as the data shows the opposite. In the existing literature, researchers seek to resolve the conflict by amending the model to include non-yield variables as the unspanned factors. I provide empirical evidence that questions the necessity of this approach. This paper contributes to the literature by showing that this conflict can be resolved by allowing for fully spanned but non-linear state variables. To show this, I construct a non-linear factor from the yield curve non-parametrically and use it to simultaneously predict bond excess returns.

The following aspects of this paper can be emphasized. First, I show that information contained in the non-yield factors is also contained in this non-linear factor. When conditioning on the non-linear factor, non-yield factors that used to be informative about future bond excess returns in the fully spanned ATSM no longer provide extra predictive power. Second, results suggest that fully spanned DTSMs with non-linear state variables may be able to bypass the difficulties we have in standard ATSMs. The dynamics of estimated risk premia show that the non-linear factor successfully produces business cycle frequency and countercyclical risk premia, indicating that non-linear risks may originate from fundamental macroeconomic shocks.

The main goal of this paper is to discuss the potential of the non-linear mechanism. The non-parametric regression method used in this paper is not the only way to introduce non-linearities. The research question remains open on how to incorporate the non-linear feature into a structure model, as it can provide a deeper understanding of the macro economy. Christensen, Diebold, and Rudebusch (2011) extend the dynamic Nelson and Siegel (1987) model in Diebold and Li (2006) by introducing the no-arbitrage restriction; Feldhütter, Heyerdahl-Larson, and Illditsch (2013) assume explicit non-linear dynamics in the stochastic discount factor. Both models generate promising forecasting results on bond yields, but the extent to which we can rule out the non-yield factors in their models is still unknown. Interesting extensions also include jointly predicting stock and bond returns Koijen, Lustig, and Nieuwerburgh (2014) and optimal portfolio choice Sangvinatsos and Wachter (2005); Barillas (2011), which I leave for future investigation.

Table 2.1: Regression of bond excess returns on lagged factors.

Model : $rx_{t+1}^{(n)} = \gamma + \alpha' \hat{F}_t + \beta NP_t + \varepsilon_{t+1}$.

		$\hat{F}_{1t}$	$\hat{F}_{1t}^3$	$\hat{F}_{2t}$	$\hat{F}_{3t}$	$\hat{F}_{4t}$	$\hat{F}_{8t}$	NP_t	$\bar{R}^2$	χ^2
$rx_{t+1}^{(2)}$	(a)							0.50	0.82	
								(28.66)		
	(b)	-0.93	0.06	-0.38	0.19	-0.33	0.35		0.25	
		(-5.13)	(2.75)	(-2.96)	(2.42)	(-2.95)	(4.23)			
	(c)	-0.34	0.01	0.07	0.14	-0.13	0.04	0.47	0.86	40.24
		(-3.18)	(2.06)	(1.75)	(4.85)	(-3.03)	(1.20)	(27.37)		(0)
	(d)	-0.35	0.01		0.14	-0.13	0.05	0.46	0.86	29.60
		(-3.24)	(2.14)		(5.10)	(-3.25)	(1.30)	(28.74)		(0)
$rx_{t+1}^{(3)}$	(a)							0.93	0.87	
								(38.26)		
	(b)	-1.57	0.11	-0.80	0.21	-0.53	0.63		0.23	
		(-4.91)	(2.98)	(-3.20)	(1.54)	(-2.51)	(4.28)			
	(c)	-0.45	0.02	0.05	0.12	-0.14	0.06	0.89	0.89	10.38
		(-2.68)	(1.94)	(0.85)	(2.41)	(-2.05)	(1.04)	(38.05)		(0.11)
	(d)	-0.46	0.02		0.12	-0.14	0.06	0.89	0.89	10.18
		(-2.72)	(2.02)		(2.50)	(-2.13)	(1.10)	(43.60)		(0.07)
$rx_{t+1}^{(4)}$	(a)							1.30	0.89	
								(40.69)		
	(b)	-2.02	0.15	-1.19	0.22	-0.62	0.94		0.22	
		(-4.71)	(3.05)	(-3.24)	(1.09)	(-2.02)	(4.27)			
	(c)	-0.44	0.02	0.01	0.09	-0.07	0.12	1.26	0.90	7.39
		(-2.08)	(1.77)	(0.14)	(1.39)	(-0.83)	(1.77)	(37.72)		(0.29)
	(d)	-0.44	0.02		0.09	-0.07	0.12	1.26	0.90	7.08
		(-2.09)	(1.78)		(1.40)	(-0.83)	(1.82)	(41.86)		(0.21)
$rx_{t+1}^{(5)}$	(a)							1.57	0.89	
								(42.59)		
	(b)	-2.24	0.18	-1.50	0.23	-0.77	1.12		0.20	
		(-4.29)	(2.90)	(-3.33)	(0.92)	(-2.04)	(4.27)			
	(c)	-0.29	0.02	-0.01	0.07	-0.10	0.12	1.56	0.89	3.44
		(-1.13)	(0.84)	(-0.15)	(0.87)	(-0.99)	(1.47)	(37.36)		(0.75)
	(d)	-0.29	0.02		0.07	-0.10	0.12	1.56	0.89	3.43
		(-1.13)	(0.83)		(0.86)	(-0.98)	(1.48)	(40.81)		(0.63)

Note: The table reports estimates from the OLS regressions of four specifications of (2.8). Point estimates are reported, along with the value of t-statistics in the parentheses. As in LN, t-statistics are carried out based on a corrected variance-covariance matrix following the Newy and West (1987) correction with 18 lags. The $\overline{R}^2$ in the table stands for adjusted R^2. For specifications (c) and (d), I also conduct a χ^2 test for the joint significance of non-yield factors $\overrightarrow{F6}_t$ and $\overrightarrow{F5}_t$. The value of test statistics and p-values are reported below (in the parentheses). Estimates of the constant γ are not reported in this table. The in-sample period spans from January 1964 to December 2003.

Table 2.2: Out-of-sample predictive power of the NP factor

Forecast Sample	Comparison	MSE_u	MSE_r	Test Statistic	p-value
	$rx_{t+1}^{(2)}$				
1987:01-2003:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	1.58	1.78	4.09	0.000
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	0.59	0.61	2.49	0.006
1987:01-2008:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	1.67	1.78	3.45	0.000
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	0.62	0.61	1.58	0.060
1995:01-2003:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	2.07	2.12	2.13	0.020
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	0.69	0.63	0.17	0.430
1995:01-2008:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	2.17	2.11	1.85	0.030
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	0.69	0.62	0.01	0.500
	$rx_{t+1}^{(3)}$				
1987:01-2003:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	6.11	6.78	4.17	0.000
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	1.88	1.91	1.74	0.040
1987:01-2008:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	6.13	6.48	3.53	0.000
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	1.91	1.85	0.57	0.280
1995:01-2003:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	7.75	8.27	2.58	0.005
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	2.21	2.14	0.48	0.320
1995:01-2008:12	$\overrightarrow{F5}_t$+CP vs. $const$ +CP	8.04	8.16	2.30	0.011
	$\overrightarrow{F5}_t$+NP vs. $const$ +NP	2.15	2.01	0.21	0.420

Notes: The table reports out-of-sample forecast comparisons of n-year excess bond returns $rx_{t+1}^{(n)}$. $\overrightarrow{F5}_t$ is defined as the vector of PCs of the macro-finance panel $\left(\widehat{F}_{1t}, \widehat{F}_{1t}^3, \widehat{F}_{2t}, \widehat{F}_{3t}, \widehat{F}_{4t}, \widehat{F}_{8t}\right)^\top$. "$\overrightarrow{F5}_t$ + CP" refers to (2.8) with the specification that $\widehat{F}_t = \overrightarrow{F5}_t$ and $Z_t = CP_t$. "$\overrightarrow{F5}_t$ + NP" refers to (2.8) with the specification that $\widehat{F}_t = \overrightarrow{F5}_t$ and $Z_t = NP_t$. "$const$ + CP" and "$const$ + NP" denote two restricted specifications of (2.8) with $\alpha = \mathbf{0}$. In each row, an unrestricted model is compared with a restricted model. The corresponding MSE (labeled "MSE_u" and "MSE_r" respectively) are used to test whether the restricted model encompasses the unrestricted model. The test statistic reports $f = MSE_r - (MSE_u - adj.)$ as in Clark and West (2007), the p-value is calculated under the standard normal curve and reported in the last column.

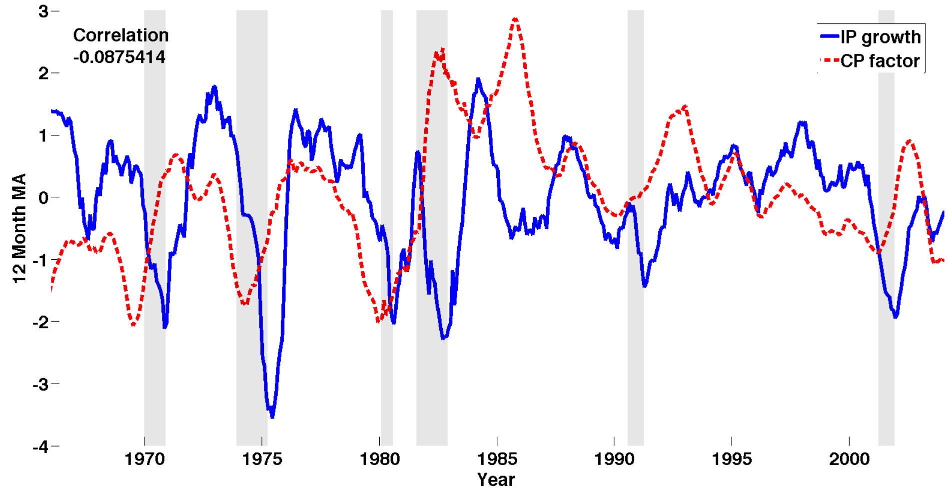
Table 2.2: (Cont'd) Out-of-sample predictive power of the NP factor

Forecast Sample	Comparison	MSE_u	MSE_r	Test Statistic	p-value
	$rx_{t+1}^{(4)}$				
1987:01-2003:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	11.73	12.89	4.19	0.000
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	3.55	3.56	1.02	0.150
1987:01-2008:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	11.38	12.05	3.66	0.000
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	3.45	3.33	-0.14	0.560
1995:01-2003:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	14.15	15.13	2.68	0.004
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	4.02	3.94	0.32	0.380
1995:01-2008:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	14.92	15.34	2.47	0.007
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	3.74	3.61	0.04	0.490
	$rx_{t+1}^{(5)}$				
1987:01-2003:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	18.11	19.56	4.02	0.000
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	5.61	5.56	0.25	0.400
1987:01-2008:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	16.83	17.76	3.72	0.000
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	5.14	4.97	-0.63	0.740
1995:01-2003:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	21.91	23.37	2.63	0.004
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	6.84	6.82	0.54	0.290
1995:01-2008:12	$\overrightarrow{F5}_t+CP$ vs. $const+CP$	22.43	23.17	2.52	0.006
	$\overrightarrow{F5}_t+NP$ vs. $const+NP$	5.79	5.71	0.33	0.370

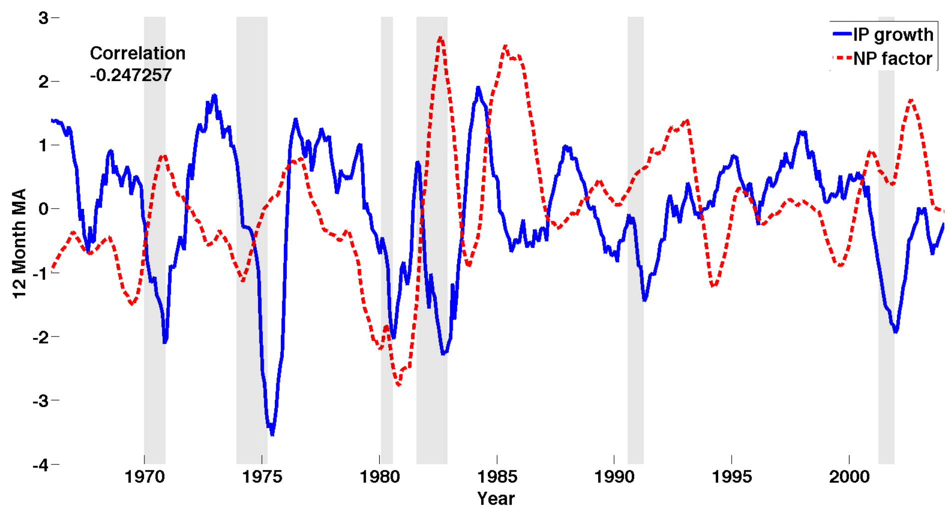
Notes: The table reports out-of-sample forecast comparisons of n-year excess bond returns $rx_{t+1}^{(n)}$. $\overrightarrow{F5}_t$ is defined as the vector of PCs of the macro-finance panel $\left(\widehat{F}_{1t}, \widehat{F}_{1t}^3, \widehat{F}_{2t}, \widehat{F}_{3t}, \widehat{F}_{4t}, \widehat{F}_{8t}\right)^\top$. "$\overrightarrow{F5}_t$ + *CP*" refers to (2.8) with the specification that $\widehat{F}_t = \overrightarrow{F5}_t$ and $Z_t = CP_t$. "$\overrightarrow{F5}_t$ + *NP*" refers to (2.8) with the specification that $\widehat{F}_t = \overrightarrow{F5}_t$ and $Z_t = NP_t$. "*const* + *CP*" and "*const* + *NP*" denote two restricted specifications of (2.8) with $\alpha = \mathbf{0}$. In each row, an unrestricted model is compared with a restricted model. The corresponding *MSE* (labeled "MSE_u" and "MSE_r" respectively) are used to test whether the restricted model encompasses the unrestricted model. The test statistic reports $f = MSE_r - (MSE_u - adj.)$ as in Clark and West (2007), the p-value is calculated under the standard normal curve and reported in the last column.

Figure 2.1: Time series plot of the 12-lagged moving average of CP and NP factor vs. IP growth.

A: CP factor and IP growth.

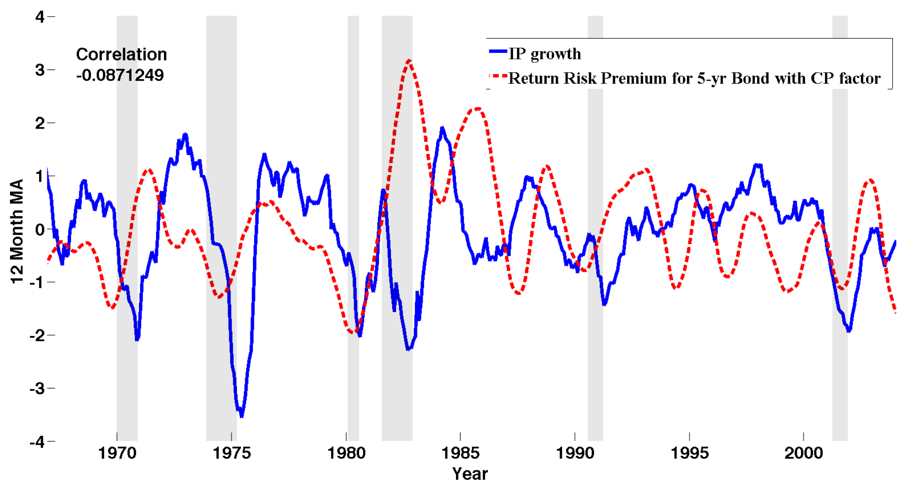


B: NP factor and IP growth.

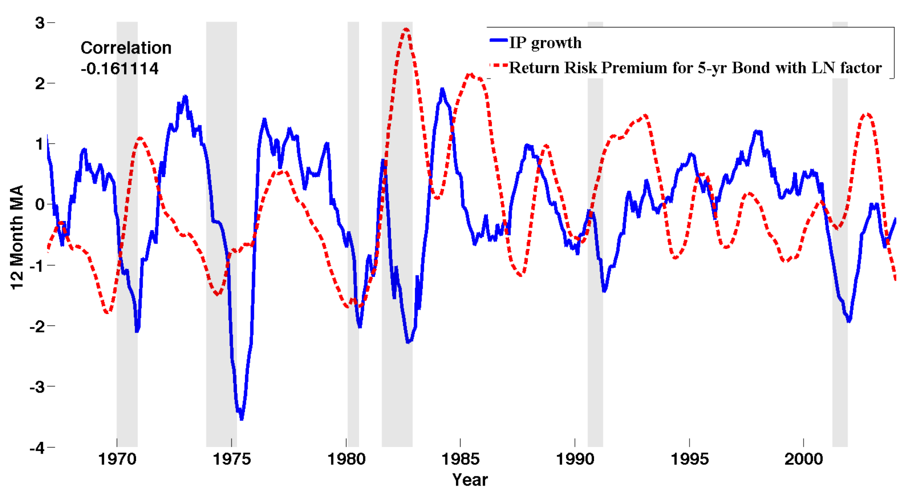


Note: Variables are reported under standardized units. Shadings denote periods designated as recessions by the National Bureau of Economic Research. The correlation between variables is displayed on upper left corner.

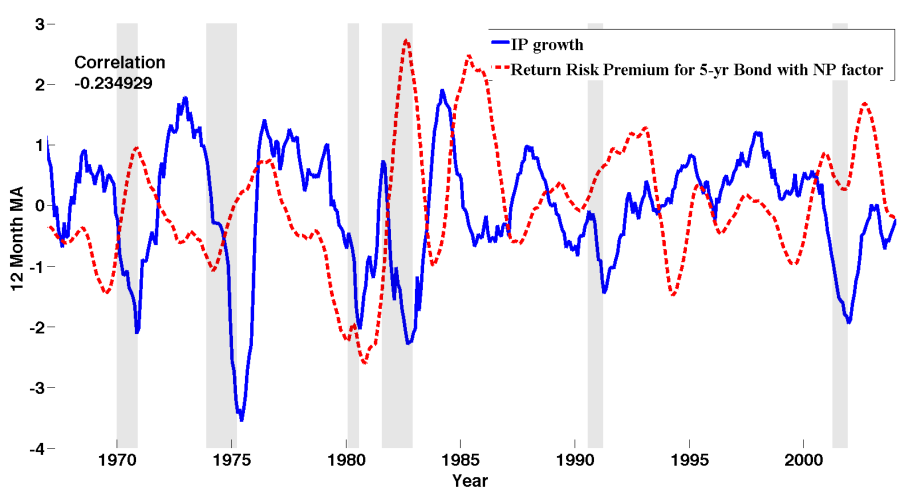
Figure 2.2: Implied return risk premia constructed by different factors.
A: Return risk premium constructed from the CP factor.



B: Return risk premium constructed from the non-yield factors.

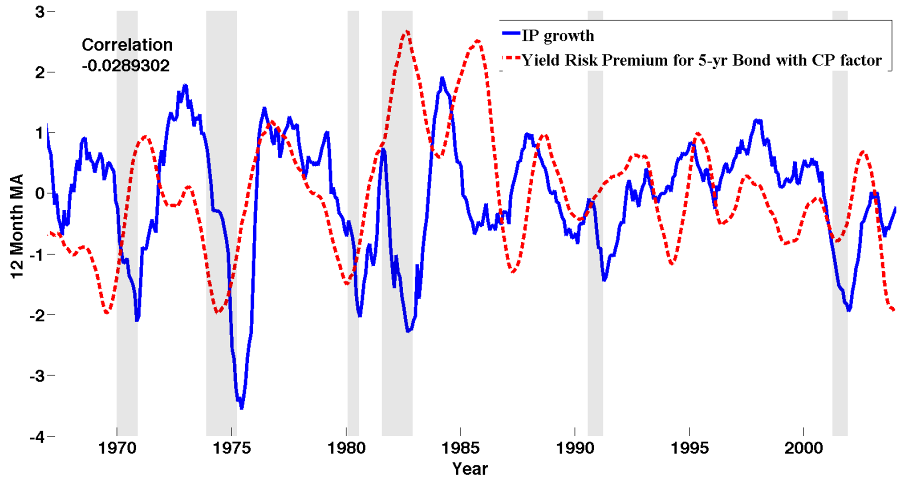


C: Return risk premium constructed from the NP factor.

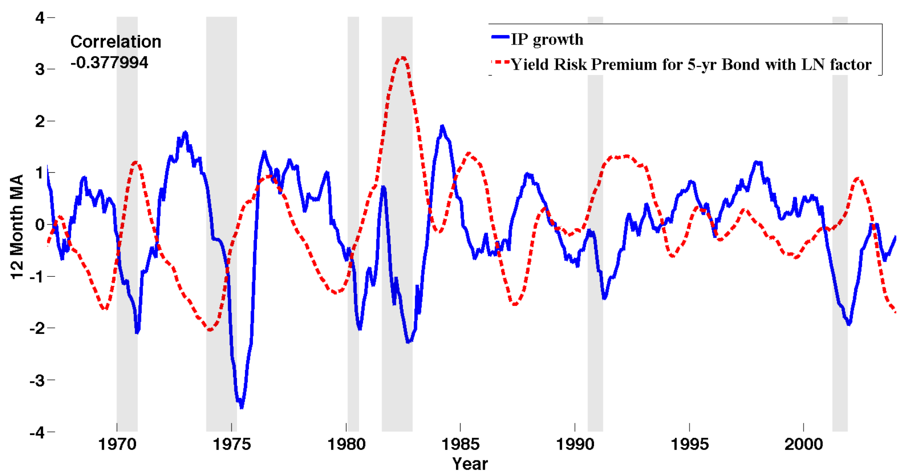


Note: Variables are reported under standardized units. Shadings denote periods designated as recessions by the National Bureau of Economic Research. The correlation between variables is displayed on upper left corner.

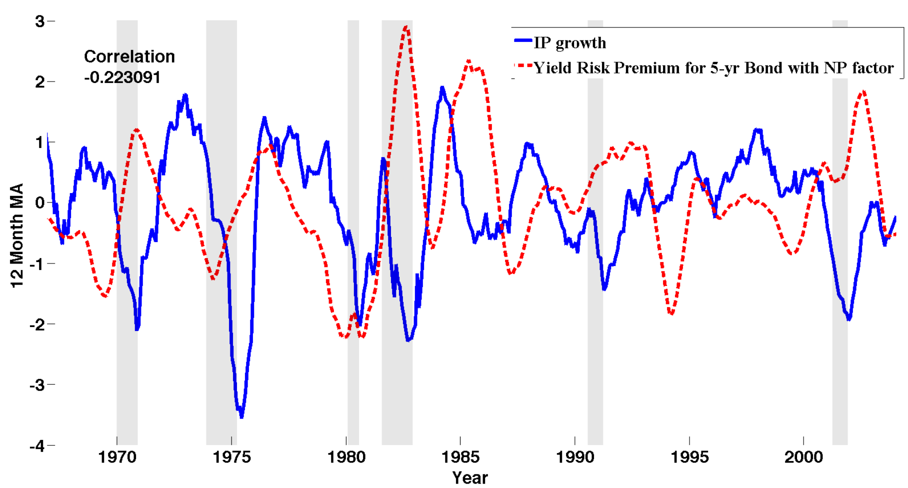
Figure 2.3: Implied yield risk premia constructed by different factors.
A: Yield risk premium constructed from the CP factor.



B: Yield risk premium constructed from the non-yield factors.



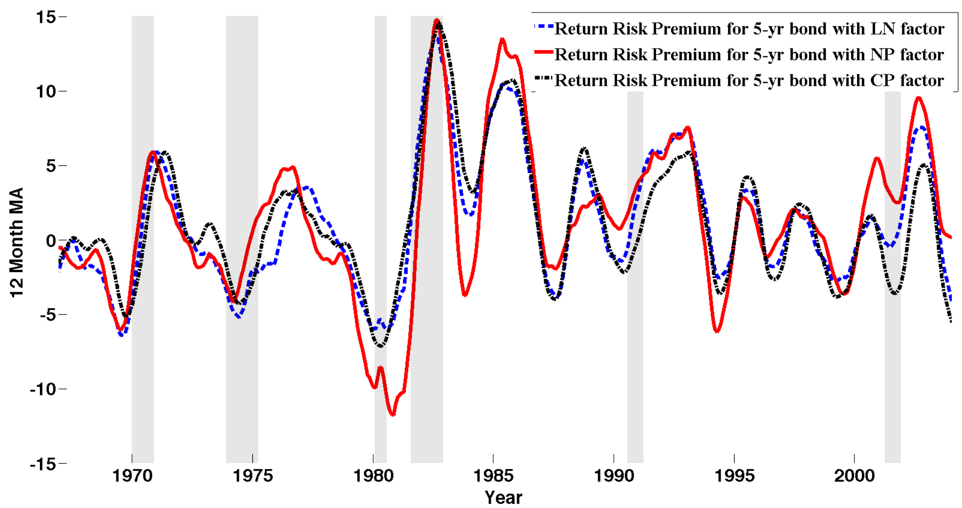
C: Yield risk premium constructed from the NP factor.



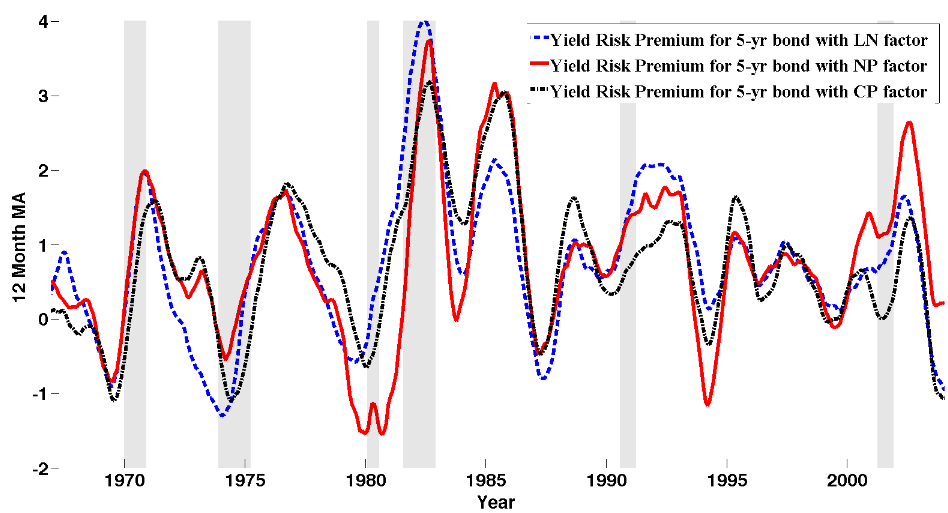
Note: Variables are reported under standardized units. Shadings denote periods designated as recessions by the National Bureau of Economic Research. The correlation between variables is displayed on upper left corner.

Figure 2.4: Yield and return risk premia from different VAR estimation.

A: Fitted return risk premium from different factors.



B: Fitted yield risk premium from different factors.



Note: Variables are reported under standardized units. Shadings denote periods designated as recessions by the National Bureau of Economic Research. The correlation between variables is displayed on upper left corner.

References

Aït-Sahalia, Y. 1996a. Nonparametric pricing of interest rate derivative securities. *Econometrica* 64:527–60.

———. 1996b. Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9:385–426.

Ang, A., and G. Bekaert. 2002. Short rate nonlinearities and regime switches. *Journal of Economic Dynamics & Control* 26:1243–74.

Ang, A., G. Bekaert, and M. Wei. 2008. The term structure of real rates and expected inflation. *Journal of Finance* 63:797–849.

Ang, A., and M. Piazzesi. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50:745–87.

Ang, A., M. Piazzesi, and S. Dong. 2007. No-arbitrage taylor rules. *Working Paper* Columbia University.

Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70:191–221.

———. 2006. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74:1133–50.

Barillas, F. 2011. Can we exploit predictability in bond markets? *Working Paper* Emory University.

Campbell, J. Y., and J. H. Cochrane. 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107:205–51.

Campbell, J. Y., and R. J. Shiller. 1991. Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies* 58:495–514.

Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch. 2011. The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics* 164:4–20.

Cieslak, A., and P. Povala. 2013. Expected returns in treasury bonds. *Working Paper* Northwestern University.

Clark, T. E., and M. W. McCracken. 2013. Advances in forecast evaluation. *Handbook of Economic Forecasting* 2, Part B:1107–201.

Clark, T. E., and K. D. West. 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138:291–311.

Cochrane, J. H., and M. Piazzesi. 2005. Bond risk premia. *American Economic Review* 95:138–60.

Collin-Dufresne, P., and R. S. Goldstein. 2002. Do bonds span the fixed incomemarkets? theory and evidence for unspanned stochastic volatility. *Journal of Finance* 57:1685–730.

Cooper, I., and R. Priestley. 2009. Time-varying risk premiums and the output gap. *Review of Financial Studies* 22:2801–33.

Dai, Q., and K. J. Singleton. 2000. Specification analysis of affine term structure models. *Journal of Finance* 55:1943–78.

Dai, Q., K. J. Singleton, and W. Yang. 2007. Regime shifts in a dynamic term structure model of u.s. treasury bond yields. *Review of Financial Studies* 20:1669–706.

Diebold, F., and R. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13:253–65.

Diebold, F. X., and C. Li. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130:337–64.

Duffee, G. R. 2002. Term premia and interest rate forecasts in affine models. *Journal of Finance* 57:405–43.

———. 2011. Information in (and not in) the term structure. *Review of Financial Studies* 24:2895–934.

———. 2013. Forecasting interest rates. *Handbook of Economic Forecasting* 2, Part A:385–426.

Duffie, D., and R. Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6:379–406.

Epanechnikov, V. A. 1969. Nonparametric estimates of a multivariate probability density. *Theory of Probability and its Applications* 14:153–8.

Fama, E. F., and R. R. Bliss. 1987. The information in long-maturity forward rates. *American Economic Review* 77:680–92.

Feldhütter, P., C. Heyerdahl-Larson, and P. Illditsch. 2013. Risk premia, volatilities and sharpe ratios in a nonlinear term structure model. *Working Paper* London Business School.

Fontaine, J.-S., and R. Garcia. 2012. Bond liquidity premia. *Review of Financial Studies* 25:1207–54.

Gray, S. F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42:27–62.

Joslin, S., M. Priebsch, and K. J. Singleton. 2014. Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance* 69:1197–233.

Koijen, R. S., H. Lustig, and S. V. Nieuwerburgh. 2014. The cross-section and time-series of stock and bond returns. *NBER Working Paper 15688* .

Li, Q., and J. S. Racine. 2006. *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Litterman, R. B., and J. Scheinkman. 1991. Common factors affecting bond returns. *Journal of Fixed Income* 1:54–61.

Ludvigson, S., and S. Ng. 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22:5027–67.

Nadaraya, E. A. 1964. On estimating regression. *Theory of Probability and its Applications* 9:141–2.

Nelson, C. R., and A. F. Siegel. 1987. Parsimonious modeling of yield curves. *Journal of Business* 60.

Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–08.

Sangvinatsos, A., and J. Wachter. 2005. Does the failure of the expectations hypothesis matter for long-term investors? *Journal of Finance* 60:179–230.

Stambaugh, R. F. 1988. The information in forward rates: Implications for models of the term structure. *Journal of Financial Economics* 21:41–70.

———. 1999. Predictive regressions. *Journal of Financial Economics* 54:375–421.

Stanton, R. 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52:1973–2002.

Stock, J. H., and M. W. Watson. 2002a. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97:1167–79.

———. 2002b. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20:147–62.

———. 2005. Implications of dynamic factor models for var analysis. *NBER Working Paper 11467* .

———. 2006. Macroeconomic forecasting using many predictors. *Handbook of Economic Forecasting* .

Wachter, J. 2006. A consumption based model of the term structure of interest rates. *Journal of Financial Economics* 79:365–99.

Watson, G. S. 1964. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* 26:359–72.

Chapter 3

A Test on Asymmetric Dependence

(Joint with Lei Jiang, Esfandiar Maasoumi, and Ke Wu)

Abstract

We provide a model-free test for asymmetric dependence between stock and market returns, based on the Kullback-Leibler mutual information measure. Our test has greater power in small samples than previous tests of asymmetric correlation proposed by Hong, Tu and Zhou (2007). Empirically, we find that asymmetric comovment is a prevailing phenomenon in most commonly used portfolios.

JEL Classification: C12, C15, C32, G12.

Key words: Asymmetric dependence, Kullback-Leibler entropy, mutual information, copulas, GARCH, Monte-Carlo simulation.

I Introduction

For the individual stock returns and market return in the U.S. stock market, joint normal distribution and symmetric assumption during market upturns and downturns have prevailed since the Capital Asset Pricing Model (CAPM) (see, e.g. Sharpe, 1964; Lintner, 1965, and many others). However, empirical evidence indicates that during market downturns, the cross-sectional average of three conceptually similar measures, the Betas (Ball and Kothari, 1989, Braun, Nelson, and Sunier, 1995, and Ang, Chen, and Xing, 2006), correlations (Ang and Chen, 2002 and Longin and Solnik, 2001) and covariance (Kroner and Ng, 1998, Cho and Engle, 1999 and Conrad, Gultekin, and Kaul, 1991) are much higher than during market upturns, especially for size portfolios (Kroner and Ng, 1998 and Conrad, Gultekin, and Kaul, 1991). Therefore, two important questions are still under debate: whether the observed difference in comovement is statistically significant, and whether the difference is a common phenomenon for the U.S. stock market. A positive answer to these questions may call for different methods of portfolio management when the market collapses.

In order to answer the questions mentioned above, Ang and Chen (2002) first come up with a test for asymmetric correlation and conclude that asymmetric comovements are significant, and stocks with certain characteristics (such as small size, high book to market ratio, and losers) have higher asymmetries than otherwise. However, more recent work by Hong, Tu, and Zhou (2007) point out that the evidence provided by Ang and Chen (2002) may reflect the fact that the joint densities of individual stock returns and market return are different from normal rather than asymmetry. That is why they propose a model-free test and conclude that asymmetric comovement in fact rarely occurs. Many commonly used portfolios such as book to market portfolios and momentum portfolios do not exhibit a significant difference in comovement conditional on market downturns and upturns.

In this paper, we interpret the empirical evidence in Hong, Tu, and Zhou (2007) differently. We argue that the insignificant difference in the correlation does not necessarily imply that the comovements, which should be more precisely measured by dependence, are the same. The reason is based on the stylized fact documented in various empirical research that stock returns are not normally distributed (Embrechts, McNeil, and Straumann, 2002), where the dependence between two random variables are fully captured by the linear correlation. Therefore, we propose a model-free test for asymmetric dependence (a direct nonlinear extension of correlation) of individual stock returns and market return based on the Kullback-Leibler mutual information measure. The new method directly tests the equality of the dependence of returns at a given exceedance level of the return distribution and generates greater power than Hong, Tu, and Zhou (2007) in small samples. The empirical evidence indicates that the comovements of market return and individual stock return are significantly higher during market downturns for many portfolios. Portfolio managers must pay special attention to diversification during market downturns because of the excessive comovements.

Ang, Chen, and Xing (2006) also notice the correlation is higher in downturns. But they did not extend the correlation into dependence as we do to capture the nonlinear statistical relation between stock returns and market return. Kelly and Jiang (2014) propose a time-dynamic tail risk factor to capture the adverse effect of extremely small return on stocks. In contrast, we consider the asymmetric dependence of individual stock returns and market return, which also differs from Chabi-Yo, Ruenzi, and Weigert (2014), who propose a left tail dependence measure to capture crash sensitivities.

The rest of the paper is organized as follows. Section 2 introduces the test for asymmetric dependence based on the relative entropy measure. Section 3 examines the asymptotic size and finite sample performance of the test statistic. Section 4 applies the test to investigate asymmetry dependence in common portfolios sorted by size, book-to-market and momentum. Section 5 concludes.

II A Relative Entropy Based Test on Asymmetric Dependence

In this section, we first present an entropy based measure of exceedance dependence which is motivated by the Kullback-Leibler mutual information measure. A test for asymmetric dependence based on this measure is developed. The bootstrap algorithm for obtaining the sampling distribution of the test statistic is also discussed in detail.

II.1 A relative entropy based measure of exceedance dependence

Let R_{1t}, R_{2t} be the returns on two portfolios in period t, and both of them are assumed to be stationary with $E(R_i) = \mu_i,\ \text{Var}(R_i) = \sigma_i^2,\ i = 1, 2$. For any exceedance level c, we consider how R_1 and R_2 co-move with each other when both of them exceed c standard deviations away from their means, respectively. In most of the existing literature, (see, e.g. Longin and Solnik, 2001; Ang and Chen, 2002; Hong, Tu, and Zhou, 2007, and many others), researchers use exceedance correlation to measure the co-movements. However, sample moment-based dependence measures such as linear correlation and co-skewness only capture dependence up to the order of that moment. Therefore, any possible higher order dependence would be ignored, as pointed out by Jiang, Wu, and Zhou (2014), which implies huge information loss when the underlying distribution is not jointly normal. Meanwhile, non-normality of financial time series is widely documented in the literature (Embrechts, McNeil, and Straumann, 2002). Therefore, an ideal dependence measure should be able to summarize general dependence in certain areas of a given distribution.

Originating in physics and information theory, entropy has a long history of use as an aggregate measure of information contained in a distribution. During recent years, Kullback-Leibler relative entropy (Kullback and Leibler, 1951) has been employed more frequently in finance and economics research (see, for example Backus, Chernov, and Martin, 2011; Hansen, 2012; Backus, Chernov, and Zin, 2014, among others). In particular, Kullback-Leibler relative entropy has also been used to construct a widely used mutual information (MI) measure that can measure the mutual dependence between two random variables from the distributional perspective.

The MI for random variables R_1 and R_2 is defined as the Kullback-Leibler relative entropy between the joint density $g(R_1, R_2)$ and product of their marginals $g_1(R_1) \cdot g_2(R_2)$:

$$I(R_1; R_2) \equiv E(\log \frac{g(R_1, R_2)}{g_1(R_1) \cdot g_2(R_2)}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(R_1, R_2) \log \frac{g(R_1, R_2)}{g_1(R_1) \cdot g_2(R_2)} dR_1 dR_2. \tag{3.1}$$

Essentially, MI measures the expected difference between the log likelihood of $g(R_1, R_2)$ and $g_1(R_1) \cdot g_2(R_2)$. To serve as a measure for dependence, the MI measure yields the following desirable properties. First, MI is always non-negative, i.e., $I(R_1; R_2) \geq 0$. $I = 0$ if and only if R_1 and R_2 are independent, and it increases as the dependence between them grows.[1] Second, the measure is obtained by comparing the whole distribution $g(R_1, R_2)$ and $g_1(R_1) \cdot g_2(R_2)$. Hence, it captures all higher order dependencies between R_1 and R_2 beyond commonly employed moment-based dependence measures.

Motivated by the fact that $I(R_1, R_2)$ measures the dependence between R_1 and R_2 in the whole sample space $\mathbb{R}^2$, we propose a partial MI measure defined on subspaces to measure variables' exceedance dependence. For a given exceedance level c, we define left and right tail exceedance dependence $\rho^-_{c,o}$ and $\rho^+_{c,o}$ as

$$\rho^-_{c,o} = \int_{-\infty}^{\mu_2 - c\sigma_2} \int_{-\infty}^{\mu_1 - c\sigma_1} g(R_1, R_2) \log \frac{g(R_1, R_2)}{g_1(R_1) \cdot g_2(R_2)} dR_1 dR_2, \tag{3.2}$$

$$\rho^+_{c,o} = \int_{\mu_2 + c\sigma_2}^{+\infty} \int_{\mu_1 + c\sigma_1}^{+\infty} g(R_1, R_2) \log \frac{g(R_1, R_2)}{g_1(R_1) \cdot g_2(R_2)} dR_1 dR_2. \tag{3.3}$$

$\rho^+_{c,o}$ and $\rho^-_{c,o}$ measure the general dependence between R_1 and R_2 in the upper (in the subspace $(\mu_1 + c\sigma_1, +\infty) \times (\mu_2 + c\sigma_2, +\infty)$) and lower tail (in the subspace $(-\infty, \mu_1 - c\sigma_1) \times (-\infty, \mu_2 - c\sigma_2)$), respectively. Testing for asymmetric dependence simply requires testing

[1] See, for example, Cover and Thomas (2006) pp.42 for reference.

hypothesis

$$H_0: \ \rho_{c,o}^{+} = \rho_{c,o}^{-}. \tag{3.4}$$

Evaluating $\rho_{c,o}^{+}$ and $\rho_{c,o}^{-}$ involves integration with respect to the unknown parameters μ_i and σ_i. Following Hong, Tu, and Zhou (2007) and other finance literature, we focus on the standardized returns X and Y with zero mean and unit variance. Similar to (3.2) and (3.3), we define the partial MI measure for X and Y at level c by

$$\rho_c^{-} = \int_{-\infty}^{-c}\int_{-\infty}^{-c} f(X,Y)\log\frac{f(X,Y)}{f_1(X)f_2(Y)}dXdY, \tag{3.5}$$

$$\rho_c^{+} = \int_{c}^{+\infty}\int_{c}^{+\infty} f(X,Y)\log\frac{f(X,Y)}{f_1(X)f_2(Y)}dXdY, \tag{3.6}$$

where $f(X,Y)$, $f_1(X)$ and $f_2(Y)$ denote the joint and marginal densities for the standardized returns. Although the MI measure is not invariant under general linear transformation, the following theorem shows that invariability holds under simple standardization.

Theorem II.1. *Under the assumptions made in Section II.1, the partial MI measure is invariant under simple standardization, i.e.,* $\forall c,\ \rho_{c,o}^{+} = \rho_c^{+}$ *and* $\rho_{c,o}^{-} = \rho_c^{-}$.

Proof. See Appendix. □

To compare the exceedance dependence of R_1 and R_2 at any given level c, we can simply test the equivalence of the partial MI measure on standardized returns X and Y,

$$H_0: \ \rho_c^{+} = \rho_c^{-} \quad \text{for a given exceedance level } c. \tag{3.7}$$

If the null hypothesis is rejected, the dependence measure between the positive returns of the two portfolios is different from that between their negative returns. In other words, dependence measures are asymmetric. Hence, the alternative hypothesis is $H_1: \ \rho_c^{+} \neq \rho_c^{-}$.

II.2 The non-parametric estimator

Now we consider how to estimate the partial MI measure given the data. Similar to as the MI in (3.1), ρ_c^- and ρ_c^+ also can be interpreted as expectations,

$$\begin{aligned}\rho_c^- &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(X,Y)\log\tfrac{f(X,Y)}{f_1(X)f_2(Y)}\cdot \mathbf{1}(X<-c,Y<-c)dXdY\\ &= E(\log\tfrac{f(X,Y)}{f_1(X)f_2(Y)}\cdot \mathbf{1}(X<-c,Y<-c)),\end{aligned}$$

$$\begin{aligned}\rho_c^+ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(X,Y)\log\tfrac{f(X,Y)}{f_1(X)f_2(Y)}\cdot \mathbf{1}(X>c,Y>c)dXdY\\ &= E(\log\tfrac{f(X,Y)}{f_1(X)f_2(Y)}\cdot \mathbf{1}(X>c,Y>c)),\end{aligned}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function.

ρ_c^- and ρ_c^+ can be estimated by their sample analogues. For a random sample of returns which consists of T observations $\{X_t,Y_t\}_{t=1}^T$, let T_c^+ and T_c^- be the number of observations in which both X_t and Y_t are simultaneously larger and smaller than c, respectively. We can express the sample exceedance dependence as

$$\hat{\rho}_c^- = \frac{1}{T_c^-}\sum_{t=1}^{T}\log\frac{\hat{f}(X_t,Y_t)}{\hat{f}_1(X_t)\hat{f}_2(Y_t)}\mathbf{1}(X_t<-c,Y_t<-c), \tag{3.8}$$

$$\hat{\rho}_c^+ = \frac{1}{T_c^+}\sum_{t=1}^{T}\log\frac{\hat{f}(X_t,Y_t)}{\hat{f}_1(X_t)\hat{f}_2(Y_t)}\mathbf{1}(X_t>c,Y_t>c), \tag{3.9}$$

where the probability density functions $\hat{f}(X_t,Y_t)$, $\hat{f}_1(X_t)$ and $\hat{f}_2(Y_t)$ are estimated by robust non-parametric kernel estimators as proposed in Rosenblatt (1956) and Parzen (1962). Kernel estimation provides consistent estimators for estimating the joint density of a set of random variables. Given a series of m-dimensional random vectors Z that consists of T observations $z_1, z_2, \ldots, z_T$, the Parzen-Rosenblatt kernel density estimator of $f(z)$ is

$$\hat{f}(z) = \frac{1}{Th_1h_2\cdots h_m}\cdot\sum_{t=1}^{T}K\left(\frac{z_t-z}{h}\right), \tag{3.10}$$

where $K\left(\frac{z_t-z}{h}\right)\equiv\prod_{i=1}^{m}k\left(\frac{z_{i,t}-z_i}{h_i}\right)$. $k(\cdot)$ is a symmetric nonnegative bounded function and

h_i is the bandwidth (or smooth parameter). The density at any point z is estimated based on its distance to the observations z_t, scaled by the bandwidth h. The kernel density estimator is a generalization of a multidimensional histogram, as rectangles are chosen for $k(\cdot)$ in a histogram. Various studies (see, for example, Epanechnikov, 1969) suggest that different kernel functions have very little impact on estimations. In this paper, we use the popular Gaussian kernel $k(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. On selecting the bandwidth, we choose to use the likelihood cross-validation method. The method is also known as the Kullback-Leibler cross-validation (see Li and Racine, 2006, for details), since it minimizes the Kullback-Leibler relative entropy measure between the actual density and the estimated one. Specifically, it solves the following maximum likelihood problem,

$$\max_{h_1, h_2, \cdots, h_m} \mathcal{L} = \sum_{t=1}^{T} \ln \left[\hat{f}_{-t}(z) \right], \tag{3.11}$$

where

$$\hat{f}_{-t}(z) = \frac{1}{T h_1 h_2 \cdots h_m} \cdot \sum_{s \neq t}^{T} K \left(\frac{z_t - z}{h} \right), \tag{3.12}$$

which is equal to $\hat{f}(z)$ without the t-th realization. Based on the efficient market hypothesis (Fama, 1970), stock returns can be seen as *i.i.d.*, or weakly dependent series, and under such assumptions, the estimated density (3.10) converges to the actual density at a fairly fast speed (see, e.g., Li and Racine, 2006, for technical details).

II.3 Test statistic and its sampling distribution

Let $\hat{\theta} = \hat{\rho}_c^+ - \hat{\rho}_c^-$. (3.7) can be tested using an intuitive t-type test statistic

$$\hat{t} = \frac{\hat{\theta}}{\hat{\sigma}_\theta}. \tag{3.13}$$

Although asymptotic theory for the MI measure under the null hypothesis of independence has been developed in previous research (see, e.g., Robinson (1991) and Hong and White (2005)), the asymptotic distribution for the partial MI measure ($\hat{\rho}_c^+$ and $\hat{\rho}_c^-$, as in the numerator of (3.13)) when allowing for general dependence is unavailable. Moreover, while the asymptotic normality looks appealing at first glance, various studies, including Rilstone (1991) and Robinson (1991), report that inferences based on the asymptotic distribution are not reliable in finite samples. Part of the reason is that the test statistic does not depend on bandwidth asymptotically, as h vanishes when the number of observations $T \to \infty$. However, in finite samples, the test statistic is highly sensitive to $\hat{h}$, which varies across different approaches for bandwidth selection. Following the suggestion of Racine (1997); Hong and White (2005), and many others, we construct the sampling distribution for $\hat{t}$ using the pivotal bootstrap resampling approach.[2]

Stock returns are known to be stationary and weakly dependent across time. Following Künsch (1989), a natural choice to take into account the dependent structure would be bootstrap resampling with overlapping blocks. Stationarity is ensured by letting the length of each block be randomly sampled from the geometric distribution (Politis and Romano, 1994), whose mean is determined by the algorithm proposed in Politis and White (2004) and Patton, Politis, and White (2009). We use their method because it was designed to minimize the mean squared error of the estimated long-run variance of the time series. To achieve asymptotic refinement (Horowitz, 2001), $\hat{\theta}$ in (3.13) is pivotalized by its standard error $\hat{\sigma}_\theta$, which is obtained by the nested resampling method (Hinkley and Shi, 1989; Efron and Tibshirani, 1993). The details of the bootstrap resampling procedure are described below.

Obtaining the test statistic $\hat{t}_0$ for the original data sample is straightforward. After calculating $\hat{\theta}_0$ for the original data sample using (3.8) and (3.9), we use the stationary geometric bootstrap to create a sequence of B_1 nested samples of the original data sample. For each of the B_1 nested samples, we calculate its sample estimates for $\hat{\theta}$ and form a sequence $\{\hat{\theta}_0^{(i)}\}_{i=1}^{B_1}$. The standard error for the original data sample is simply the sample

[2]General properties of the bootstrap resampling approach can be found in Efron (1982). Horowitz (2001) provides excellent reviews of the literature.

standard deviation of these B_1 nested samples.

$$\hat{\sigma}_{\theta_0} = \frac{1}{B_1 - 1} \sum_{i=1}^{B_1} \left(\hat{\theta}_0^{(i)} - \overline{\hat{\theta}_0^{(i)}} \right)^2 .$$

Given both $\hat{\theta}_0$ and $\hat{\sigma}_{\theta_0}$, $\hat{t}_0$ can be directly computed by (3.13).

For the sampling distribution of the t-statistic, we first generate B bootstrap samples from the original data set using the stationary block bootstrap. For the jth bootstrap sample, we calculate $\hat{\theta}_j$ following (3.8) and (3.9). $\hat{\sigma}_{\theta_j}$ is also constructed using nested block bootstrap resampling with the same mean in the underlying geometric distribution: we create B_1 nested bootstrap samples by resampling from the given bootstrap sample and calculate $\hat{\theta}_j^{(i)}$ for each nested sample. $\hat{\sigma}_{\theta_j}$ is then simply computed by the sample standard deviation of $\{\hat{\theta}_j^{(i)}\}_{i=1}^{B_1}$. Following Horowitz (2001), the t-statistic from the bootstrap samples are adjusted for sampling bias,

$$\hat{t}_j = \frac{\hat{\theta}_j - \hat{\theta}_0}{\hat{\sigma}_{\theta_j}}.$$

We then estimate the empirical distribution F for $\{\hat{t}_j\}_{j=1}^{B}$ and report the percentile of t_0 under F. For a given level of significance α, the null hypothesis of symmetric dependence will be rejected if t_0 is located in the upper $1 - \alpha/2$ or lower $\alpha/2$ percentile.

II.4 Asymmetric Dependence vs. Asymmetry in Distribution

Jiang, Wu, and Zhou (2014) propose an entropy-based test on asymmetry of stock return distributions, i.e., to determine whether the joint distribution of market and individual stock returns is symmetric around its mean. Under the null hypothesis of symmetry, a rotation of the return distribution around the mean would have no impact on its shape. In their paper, a normalized Hellinger measure S_ρ proposed by Granger, Maasoumi, and Racine (2004) is used to measure the distance between return distributions before and after rotation. $S_\rho = 0$ under the null hypothesis that the distribution is symmetry.

Although there are similarities shared by symmetry and symmetric dependence, the difference between them is essentially crucial. The exceedance dependence measure in this paper describes the average co-movement of two random variables in subspaces of the sample space. Symmetric exceedance dependence means that two random variables on average have the same co-movement in two tails of the distribution while symmetry in distribution requires distributions before and after rotation to be equal point-wisely. Symmetry in distribution is sufficient but not necessary for a joint distribution to have symmetric dependence. The main reason for an investor to be concerned about asymmetry in exceedance dependence is that different co-movement behaviors in different return regimes may cause failure of a portfolio hedge. Therefore, the exceedance dependence measure proposed in this paper is more suitable.

III Simulation results

A valid test should possess the following asymptotic properties. First, as the sample size increases, the probability of falsely rejecting a true null hypothesis should converge to its nominal size. Second, the power of the test increases monotonically with the sample size and converge to 1 when the sample size tends to infinity. In this section, we present both asymptotic size and finite sample performance (size and power) of our test using copula-GARCH Monte Carlo simulation. To mitigate the concern that different bandwidth selection methods may have an impact on the final result, we also examine the robustness of our test results with respect to different bandwidth selection approaches.

III.1 Simulation setup

Our test focuses on detecting potential asymmetries in exceedance dependence of two random variables, thus data with different dependence structures are needed. In statistics and risk management, dependence structure is usually modeled by copulas. Therefore, a natural choice for our simulation is to generate random samples using parametric copulas with different tail behaviors. Among commonly used parametric copulas, Gaussian and Student's t copulas are known to have symmetric tail dependencies while Clayton and Gumbel copulas have strong asymmetric dependence in their tails. To study the empirical size of our test, a copula with symmetric exceedance dependence, i.e., $\rho_c^- = \rho_c^+$, is needed. In our paper, we use Student's t copula

$$C_d(u, v; \rho) = t_{d,\rho}(t_d^{-1}(u), t_d^{-1}(v)),$$

where $\rho \in (-1, 1)$ is the correlation coefficient between the marginal distributions and d is the degree of freedom for the Student t distribution. Compared with the Gaussian copula, the t copula has fatter tails and thus is an ideal candidate for examining the empirical size of our test in large samples.

In order to study the power of our test, we use distribution with different levels of asymmetric dependence. Following Hong, Tu, and Zhou (2007) and many others, the random samples are generated using the mixture copula. Different levels of asymmetric dependence are achieved by mixing the Gaussian copula, which has symmetric tail dependence, with the Clayton copula, which exhibits stronger left tail dependence. The mixture Gaussian-Clayton copula has the following specification,

$$C_{mix}(u, v; \rho, \tau, \kappa) = \kappa C_{nor}(u, v; \rho) + (1 - \kappa) C_{clay}(u, v; \tau), \quad \kappa \in [0, 1] \tag{3.14}$$

The parameter ρ in the Gaussian copula is the correlation coefficient, and the parameter τ governs the dependence between the marginal distributions in the Clayton copula. A higher τ indicates stronger left tail dependence. The parameter κ represents the weight we put on the Gaussian copula. Different levels of asymmetric dependence can be achieved by adjusting κ. When $\kappa = 1$, equation (3.14) reduces to the Gaussian copula with symmetric tail dependence. Asymmetries in tail dependence gradually increase as κ decreases. When $\kappa = 0$, equation (3.14) reduces to the Clayton copula, which shows strongest asymmetric tail dependence.

The mixture copula only determines the dependence structure between two random variables. In order to obtain the joint density, the marginal density of each random variable is needed. Since our ultimate goal is to investigate whether asymmetric dependence exists between stock portfolio and market returns, marginal distributions that mimic portfolio return distributions are used in our simulation. As in Jiang, Wu, and Zhou (2014), the value-weighted size 5 portfolio is selected as the benchmark. Following the finance literature, we model the marginal distributions of stock return with a GARCH (1,1) specification with no ARMA components.

When conducting the simulation, we first fit copula-GARCH model to the portfolio and market returns and obtain Maximum Likelihood estimates of parameters in copula and GARCH specification. The true data generating process (DGP) is assumed to follow the copula-GARCH model with all parameters set as the ML estimates. For each sample size T, 1,000 simulated random samples are generated under the same DGP, on which the properties of our test are examined.

III.2 Asymptotic size

To examine the asymptotic size of our bootstrap test, random samples with large sample size are needed. In the simulations, we set the sample size T to be 1,000 and 1,500, respectively. Table 3.1 shows the probabilities of rejecting the null hypothesis of symmetric dependence at the exceedance level $c = 0$ under the nominal sizes of 10%, 5% and 1%. The rejecting probabilities are computed as the portion of rejection decisions made in 1,000 simulated random samples. For each random sample, the inference is based on 199 stationary bootstraps. Since the random samples are generated by the t-copula, the null hypothesis of symmetric exceedance depedence is correct under the true DGP. The rejecting probabilities are thus empirical sizes. Patterns in Table 3.1 clearly suggest that our test possesses the correct asymptotic size. As the sample size increases from 1,000 to 1,500, the empirical size at the 10% level increases from 8.2% to 9.7%, which converges to the nominal size of 10%. For the 5% nominal level, the empirical sizes of both samples stay at 4.4%, which are also very close to the nominal level. Moreover, the empirical size of the 1,500 sample exactly hits the nominal size at the 1% level.

[Insert Table 3.1 about here]

III.3 Finite sample performance

The empirical size and power of our test in finite samples are examined for three sample sizes $T = 240, 420$ and 600, which are equivalent to 20, 35 and 50 years' monthly data, respectively. In this paper, the empirical size and power are computed as the portion of rejecting the null hypothesis of symmetric dependence at the exceedance level $c = 0$ in 1,000 simulated random samples. For each given random sample, inference is made based on 199 stationary bootstraps. The DGPs of our simulation follow equation (3.14) where the parameter κ governs the level of asymmetries in data exceedance dependence. In our simulation, we consider the case that $\kappa = 0, 0.25, 0.375, 0.5,$ and 1, representing five levels of dependence asymmetries, from the highest to the lowest.

Panel A of Table 3.2 reports the empirical size and power of our asymmetric dependence test for nominal size of 10%, 5% and 1%. For any given nominal size, when an asymmetric dependence exists in the true DGP (which corresponds to $\kappa = 0, 0.25, 0.375$ and 0.5 in our simulation), the empirical power of our test increases monotonically with the sample size. When distributions show stronger asymmetric dependence ($\kappa = 0$ and 0.25), we are able to reject the null hypothesis in all simulated samples under the nominal size of 5%. These findings suggest that our test is consistent. Our test suffers slight size distortion in small samples (reflected as under rejecting the null hypothesis under $\kappa = 1$). However, this is not a problem because we already show that our test possesses correct asymptotic size, and under rejection in the small sample is mainly due to the weak correlation embedded in the DGP.

[Insert Panel A of Table 3.2 about here]

To investigate whether our test is able to capture the non-linear dependence that is associated with higher order moments in data, the asymmetric correlation test proposed by Hong, Tu, and Zhou (2007) (HTZ test thereafter) is selected as the benchmark. In addition to using the sampling distribution derived from asymptotic theory, we also examine the performance of the HTZ test when its sampling distribution is obtained by stationary bootstrap resampling. The HTZ test is conducted with the same simulated random samples at the same exceedance level. Its empirical size and power are presented in Panel B of Table 3.2. By comparing the Panel B results with the results in Panel A, several interesting patterns are discovered. First, the HTZ test based on asymptotic distribution does not provide reliable inferences in small samples, especially when the simulated data has very tiny asymmetries in its dependence structure. For example, when $\kappa = 0.5$, the HTZ test based on asymptotic distribution fails to make any rejection at the 1% nominal level even in the sample with $T = 600$. Second, on average, our asymmetric dependence test is more powerful. In the table, we report the difference in power at all nominal levels for all DGPs in which dependence asymmetries exist. The difference in power is defined as the rejecting probability of the asymmetric dependence test minus that of the HTZ test. Except for very rare cases, e.g., $\kappa = 50\%$ with stationary bootstrap, our asymmetric dependence test consistently beats both versions of the HTZ test. Moreover, the average power increment becomes greater as the nominal size decreases. In the sample with $T = 600$ of DGP where $\kappa = 25\%$, the minimum power difference is 5.9% when the nominal size is set at 10%, but this difference jumps to 22.4% when the nominal size is 1%. Another reason that our test is favorable is that the power difference is most significant when the dependence asymmetries are not very strong in the underlying DGP. When the DGP is a 37.5% Gaussian-62.5% Clayton mixture copula, our test on average rejects 28.68% more than the HTZ test.

It is clear that our relative entropy based asymmetric dependence test performs better than the HTZ test in finite samples, but it is more critical to realize the role of information in the power improvement. The HTZ and other asymmetric correlation tests only use information up to the second moment. On the other hand, by using the kernel density estimation, our test essentially uses all information embedded in the distribution.

[Insert Panel B of Table 3.2 about here]

III.4 Robustness of results

Although the asymmetric dependence test performs well in finite samples, robustness of all results requires further investigation. As we discussed in section (II.3), the main reason for not using the asymptotic distribution is that the value of our kernel based test statistic relies on the choice of bandwidth parameters in finite sample. Hence, it is worthwhile to investigate the impact of different bandwidths on the test. Moreover, the block length used in the bootstrap resampling procedure does not have any meaning because the algorithm in Patton, Politis, and White (2009) is a purely data-driven approach to minimize the variable's long-run variance. For stock returns, we must also examine how the test results are affected when fixed block lengths that reflect investor's beliefs are used in resampling.

The impact of bandwidth is determined using fixed bandwidths for all simulated random samples in each DGP. Given a specific DGP, we first conduct non-parametric kernel density estimation and obtain the least square cross-validated bandwidth for all 1,000 simulated random samples. All random samples are generated under the same DGP; thus, differences in bandwidth across simulated random samples are the results of sampling variation. In our simulation, we fix the bandwidth parameters as the simple average of all individual bandwidths and use this fixed bandwidth to re-conduct the test.

In Table 3.3 we report the asymptotic size of the asymmetric dependence test using the fixed average bandwidth. Compared with the size presented in Table 3.1, where a sample specific bandwidth used in the test, we find that the asymptotic size of the asymmetric test shows consistent pattern as the sample size increases. As the sample size increases from 1,000 to 1,500, the empirical size increases from 8.7% to 9.2% at the 10% nominal level. When the nominal size is set at 5% or 1% level, the test's empirical sizes are quite close to the nominal sizes in both samples.

[Insert Table 3.3 about here]

Table 3.4 provides the empirical size and power of the test under fixed average bandwidth. When asymmetric dependence exists in the sample ($\kappa \neq 1$), the power of test using fixed average bandwidth consistently beats the test using individual bandwidth (as in Panel A of Table 3.1). Moreover, the size of the test also shows a promising pattern. When the sample size equals 600, the empirical size equals 9.6% at the 10% level. At the 5% nominal level, the empirical size increases from 2.8% to 3.7% as the sample size increases from 240 to 600. All results indicate that changing bandwidth has almost no impact on the testing results.

[Insert Table 3.4 about here]

To examine the robustness of the test with respect to block length, we use fixed block lengths of 6 and 12 when conducting simulations. The block lengths are selected to reflect 6- and 12-month's memory in monthly return data, which is enough to take care of the weak serial correlation in stock returns. The test is re-conducted on the samples with size 600.

[Insert Table 3.5 about here]

We present the robustness check of the asymmetric dependence test with respect to fixed block length in Table 3.5. Comparing the results of fixed block length to those obtained using Patton, Politis, and White's algorithm, different approaches to selecting block length are not likely to affect the size and power of the asymmetric dependence test. For example, when the block length is fixed at 6, the empirical size of the asymmetric test is 8.6%, 4.3% and 0.8% for the nominal level of 10%, 5% and 1%, respectively, which is almost identical to the results when the block length is obtained using the algorithm from Patton, Politis, and White (8.7%, 3.7%, 0.7%). When we use a fixed block length of 6, the average difference in testing power (for $\kappa = 0, 0.25, 0.375$ and 0.5) between two block length selection approaches is just 0.55% at 5% nominal level. Similar results can also be found when the block length is fixed at 12, which shows the robustness of the asymmetric dependence test with respect to different block lengths.

IV Asymmetric dependence in stock returns

In this section, we apply our test to equity portfolios sorted by size, book-to-market and momentum to investigate whether asymmetric dependence is a common phenomenon in stock returns.

IV.1 Data

Following the existing literature on asymmetric correlation, equity portfolios sorted by size, book-to-market ratio and momentum are used in this paper. As in Ang and Chen (2002) and Hong, Tu, and Zhou (2007), among others, we consider excess returns for value-weighted size and book-to-market decile portfolios and equal-weighted decile momentum portfolios which are formed based on cumulative returns from 12 to 2 months prior to formation. The CRSP (Center for Research in Security Prices) value-weighted index return, which includes all stocks listed in NYSE/AMEX/NASDAQ, is used as a proxy for the market return. All returns are recorded in excess of the rate on one-month T-bill. The entire data set is available on Kenneth French's site. The sample is formed in monthly frequency from January 1965 to December 2013, which includes 588 observations in total.

IV.2 Empirical results

We apply the asymmetric dependence test and the asymmetric correlation test in Hong, Tu, and Zhou (2007) to the equity portfolios. The exceedance level is set at 0 for both tests. P-values of the asymmetric dependence test are obtained based on 399 stationary bootstrap resamplings and we use the asymptotic Chi-square distribution to compute the p-values for the asymmetric correlation test. Table 3.6 provides the results on all three sets of portfolios. The asymmetric dependence test is able to reject the null hypothesis of symmetric dependence for the 1st to the 8th smallest size portfolios (Panel A) at the 5% level. Meanwhile, the HTZ test, which considers dependence up to the second order, can only reject symmetry for the smallest size portfolio. This finding is quite intuitive. Since our proxy for the market portfolio is a value-weighted index, larger firms will co-move more closely with the market. Therefore the asymmetries in dependence vanish as the firm size increases. For value-weighted book-to-market portfolios (Panel B), the asymmetric dependence test rejects symmetry for all except the first and the fourth smallest book-to-market portfolios. Our findings are consistent with what has been documented in past studies. For example, Ang and Chen (2002) and Jondeau (2015) report that value stocks exhibit more asymmetric co-movement with the market. On the other hand, the HTZ test fail to detect any asymmetry for all book-to-market portfolios. In Panel C, we present the results for equal-weighted momentum portfolios. Both tests find significant asymmetries for the highest (the past winner) and lowest (the past loser) momentum portfolio. In addition, we also find a statistically significant difference in the remaining equal-weighted momentum portfolios at the 1% level under our asymmetric dependence test. This finding suggest that the exceedance dependence of the middle decile momentum portfolios is associated with higher order moment beyond the second.

[Insert Table 3.6 about here]

V Conclusion

Whether individual stock return comoves with market return significantly more during market downturns than during upturns and whether the phenomenon is prevailed enough to be considered in asset management are questions under debate. Ang and Chen (2002) give a positive answer, but their test concerns joint normality rather than asymmetry. Hong, Tu, and Zhou (2007) is model-free but lacks sufficient power because it focuses on correlation rather than dependence.

In the paper, we propose a test of asymmetric dependence between stock returns and market return. The test is motivated by the Kullback-Leibler mutual information measure. We find a better power in finite samples than the previous model-free test by Hong, Tu, and Zhou (2007). The power comes from the fact that we test the difference in dependence rather than correlation. Furthermore, dependence is a more relavant question based on the fact that empirically, stocks return are not normally distributed (Embrechts, McNeil, and Straumann, 2002). Using the new measure of asymmetry, we test the portfolios sorted by size, book to market ratio and momentum. We find that most of the time, although symmetric correlation cannot be rejected, those portfolios are in fact asymmetric in dependence. Portfolio managers should pay more attention to risk-hedging when the market is down.

3.A Appendix: Proofs

3.A.1 Proofs of Theorem II.1

Proof. We want to show that the exceedance dependence measure is invariant under simple standardization. Without loss of generality, we will prove the equality for upper tail dependence measures $\rho_{c,o}^{+}$ and ρ_{c}^{+} in detail. The same proof also works for lower tail dependence measures $\rho_{c,o}^{-}$ and ρ_{c}^{-}.

Under simple standardization

$$X = \frac{R_1 - \mu_1}{\sigma_1} \text{ and } Y = \frac{R_2 - \mu_2}{\sigma_2}. \tag{3.15}$$

we have the following equalities hold for the marginal densities g_1, g_2 and f_1, f_2:

$$\begin{aligned} g_1(R_1) &= \frac{1}{\sigma_1} f_1\left(\frac{R_1 - \mu_1}{\sigma_1}\right) = \frac{1}{\sigma_1} f_1(X), \\ g_2(R_2) &= \frac{1}{\sigma_2} f_2\left(\frac{R_2 - \mu_2}{\sigma_2}\right) = \frac{1}{\sigma_2} f_2(Y). \end{aligned}$$

For the joint density g and f, we have

$$g(R_1, R_2) = f\left(\frac{R_1 - \mu_1}{\sigma_1}, \frac{R_2 - \mu_2}{\sigma_2}\right) \cdot |J|,$$

where J is the Jacobian of the transformation, which is defined as

$$J = \begin{bmatrix} \frac{\partial X}{\partial R_1} & \frac{\partial X}{\partial R_2} \\ \frac{\partial Y}{\partial R_1} & \frac{\partial Y}{\partial R_2} \end{bmatrix}.$$

Particularly under the simple standardization (3.15),

$$\det(J) = \begin{vmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{vmatrix} = \frac{1}{\sigma_1 \sigma_2}.$$

Hence

$$\begin{aligned}
\rho_{c,o}^{+} &= \int_{\mu_2+c\sigma_2}^{+\infty}\int_{\mu_1+c\sigma_1}^{+\infty} g(R_1,R_2)\log\frac{g(R_1,R_2)}{g_1(R_1)g_2(R_2)}dR_1dR_2 \\
&= \int_{c}^{+\infty}\int_{c}^{+\infty} |J|\cdot f(X,Y)\log\frac{|J|\cdot f(X,Y)}{\frac{1}{\sigma_1\sigma_2}\cdot f_1(X)f_2(Y)}\cdot\frac{1}{|J|}dXdY \\
&= \rho_c^{+}.
\end{aligned}$$

Similarly, we also have

$$\rho_{c,o}^{-} = \rho_c^{-}.$$

□

Table 3.1: Asymptotic Size of the Asymmetric Dependence Test

Sample size (T)	1000			1500		
Nominal size	10%	5%	1%	10%	5%	1%
Empirical size	0.082	0.044	0.009	0.097	0.044	0.01

Note:The table reports the probabilities of rejecting the null hypothesis of symmetric exceedance dependence under different nominal sizes, which are estimated based on the statistical inferences made in 1,000 simulated random samples. The random samples are generated by t copula, which exhibits symmetric tail dependence. For each random sample, the inferences are made based on 199 stationary bootstrap resamplings and the exceedance level $c = 0$ in all scenarios.

Table 3.2: Size and Power Comparison: Asymmetric Dependence test and HTZ test.

Panel A. Asymmetric dependence test at the exceedance dependence level c=0

		Weight on Normal Copula (κ %)				
Sample size (T)	Nominal size	100% (Size)	50%	37.50%	25%	0%
	10%	0.080	0.277	0.658	0.893	0.972
240	5%	0.035	0.170	0.503	0.800	0.942
	1%	0.006	0.052	0.204	0.497	0.706
	10%	0.075	0.44	0.880	0.993	1.000
420	5%	0.029	0.288	0.790	0.977	0.998
	1%	0.004	0.092	0.448	0.816	0.958
	10%	0.087	0.563	0.966	1.000	1.000
600	5%	0.037	0.418	0.929	1.000	1.000
	1%	0.007	0.173	0.702	0.964	0.996

Note: The table reports the probabilities of rejecting the null hypothesis of symmetric exceedance dependence under different nominal sizes, which are estimated based on the statistical inferences made in 1000 simulated random samples. All random samples are generated by the mixture copula in equation (3.14), whose degree of asymmetry in exceedance dependence is governed by the parameter κ. When $\kappa = 1$, equation (3.14) reduces to Gaussian copula with symmetric tail dependence. In all other cases, equation (3.14) produces distributions with asymmetric tail dependence. For each random sample, the inferences are made based on 199 stationary bootstrap resamplings and the exceedance level $c = 0$ in all scenarios.

Table 3.2: (Cont'd) Size and Power Comparison: Asymmetric Dependence test and HTZ test.

Panel B. Asymmetric correlation test as in Hong et al. with c=0

		Asymptotic Theory					Stationary Bootstrap				
Sample	Nominal	Weight on Normal Copula (κ %)					Weight on Normal Copula (κ %)				
size (T)	size	100% (Size)	50%	37.50%	25%	0%	100% (Size)	50%	37.50%	25%	0%
240	10%	0.000	0.016	0.200	0.660	0.944	0.085	0.386	0.608	0.748	0.807
	5%	0.000	0.005	0.096	0.478	0.857	0.039	0.233	0.423	0.549	0.614
	1%	0.000	0.003	0.010	0.159	0.559	0.000	0.074	0.145	0.221	0.273
Power Difference	10%	N/A	0.261	0.458	0.233	0.028	N/A	-0.109	0.050	0.145	0.165
	5%	N/A	0.165	0.417	0.322	0.085	N/A	-0.063	0.080	0.251	0.328
	1%	N/A	0.049	0.194	0.338	0.147	N/A	-0.022	0.059	0.276	0.433
420	10%	0.000	0.033	0.428	0.932	0.996	0.093	0.513	0.754	0.871	0.862
	5%	0.000	0.003	0.215	0.791	0.983	0.045	0.323	0.582	0.725	0.766
	1%	0.000	0.000	0.032	0.415	0.865	0.007	0.092	0.241	0.325	0.437
Power Difference	10%	N/A	0.371	0.452	0.061	0.004	N/A	-0.109	0.126	0.122	0.138
	5%	N/A	0.285	0.575	0.186	0.015	N/A	-0.035	0.208	0.252	0.232
	1%	N/A	0.092	0.416	0.401	0.093	N/A	0	0.207	0.491	0.521
600	10%	0.000	0.068	0.731	0.998	0.998	0.100	0.682	0.907	0.947	0.926
	5%	0.000	0.014	0.426	0.969	0.993	0.043	0.469	0.758	0.854	0.859
	1%	0.000	0.000	0.084	0.740	0.960	0.007	0.133	0.367	0.507	0.659
Power Difference	10%	N/A	0.495	0.235	0.002	0.002	N/A	-0.119	0.059	0.053	0.074
	5%	N/A	0.404	0.503	0.031	0.007	N/A	-0.051	0.171	0.146	0.141
	1%	N/A	0.173	0.618	0.224	0.036	N/A	0.040	0.335	0.457	0.337

Note: The table reports the probabilities of rejecting the null hypothesis of symmetric exceedance correlation under different nominal sizes, which are estimated based on the statistical inferences made in 1000 simulated random samples. All random samples are generated by the mixture copula in equation (3.14), whose degree of asymmetry in exceedance dependence is governed by the parameter κ. When $\kappa = 1$, equation (3.14) reduces to Gaussian copula with symmetric tail dependence. In all other cases, (3.14) produces distributions with asymmetric tail dependence. For each random sample, the sampling distribution are obtained using both asymptotic theory and 199 stationary bootstrap resamplings. The exceedance level $c = 0$ in all scenarios.

Table 3.3: Asymptotic Size of the Asymmetric Dependence Test using Fixed Average Bandwidth

Sample size (T)		1000			1500	
Nominal size	10%	5%	1%	10%	5%	1%
Empirical size	0.087	0.044	0.012	0.092	0.043	0.013

Note: The table reports the probabilities of rejecting the null hypothesis of symmetric exceedance dependence under different nominal sizes, which are estimated based on the statistical inferences made in 1,000 simulated random samples. The random samples are generated by t copula, which exhibits symmetric tail dependence. Fixed bandwidth, which equals the average least square cross-validated bandwidth for individual random samples, is used in the test. For each random sample, the inferences are made based on 199 stationary bootstrap resamplings and the exceedance level $c = 0$ in all scenarios.

Table 3.4: Size and Power of the Asymmetric Dependence Test using Fixed Average Bandwidth

Asymmetric dependence test with fixed bandwidth at the exceedance dependence level c=0

		Weight on Normal Copula (κ %)				
Sample size (T)	Nominal size	100% (Size)	50%	37.50%	25%	0%
	10%	0.079	0.288	0.680	0.904	0.976
240	5%	0.028	0.182	0.493	0.818	0.952
	1%	0.007	0.054	0.212	0.510	0.704
	10%	0.076	0.417	0.878	0.994	1.000
420	5%	0.033	0.299	0.780	0.980	1.000
	1%	0.008	0.096	0.467	0.819	0.967
	10%	0.096	0.571	0.967	1.000	1.000
600	5%	0.037	0.423	0.922	0.998	1.000
	1%	0.005	0.175	0.723	0.969	0.997

Note: The table reports the probabilities of rejecting the null hypothesis of symmetric exceedance dependence under different nominal sizes, which are estimated based on the statistical inferences made in 1000 simulated random samples. All random samples are generated by the mixture copula in equation (3.14), whose degree of asymmetry in exceedance dependence is governed by the parameter κ. When $\kappa = 1$, equation (3.14) reduces to Gaussian copula with symmetric tail dependence. In all other cases, equation (3.14) produces distributions with asymmetric tail dependence. Fixed bandwidth, which equals the average least square cross-validated bandwidth for individual random samples, is used in the test. For each random sample, the inferences are made based on 199 stationary bootstrap resamplings and the exceedance level $c = 0$ in all scenarios.

Table 3.5: Robustness of finite sample performance with fixed block length

Asymmetric dependence test at the exceedance dependence level c=0

		Weight on Normal Copula (κ %)				
Block Length	Nominal size	100% (Size)	50%	37.5%	25%	0%
	10%	0.086	0.548	0.960	1.000	1.000
6	5%	0.043	0.413	0.914	0.998	1.000
	1%	0.008	0.157	0.691	0.955	0.991
	10%	0.077	0.559	0.962	1.000	1.000
12	5%	0.036	0.416	0.912	1.000	1.000
	1%	0.007	0.182	0.690	0.951	0.996

Note: The table reports the probabilities of rejecting the null hypothesis of symmetric exceedance dependence under different nominal sizes, which are estimated based on the statistical inferences made in 1000 simulated random samples of size 600. All random samples are generated by the mixture copula in (3.14), whose degree of asymmetry in exceedance dependence is governed by the parameter κ. When $\kappa = 1$, equation (3.14) reduces to Gaussian copula with symmetric tail dependence. In all other cases, equation (3.14) produces distributions with asymmetric tail dependence. In stationary bootstrap resampling, fixed block length is used. We examine the cases where the block length is equal to 6 and 12, which stands for 6- and 12-month's memory in stock return distribution. For each random sample, the inferences are made based on 199 stationary bootstrap resamplings and the exceedance level $c = 0$ in all scenarios.

Table 3.6: Asymmetric Dependence in Returns of Commonly used Portfolios

Panel A. Value-Weighted Size Portfolios						Panel B. Value-Weighted Book-to-Market Portfolios						Panel C. Equal-Weighted Momentum Portfolios					
Port.	ρ_c^+	ρ_c^-	$\rho_c^- - \rho_c^+$	Asym. Dep.	Asym. Corr.	Port.	ρ_c^+	ρ_c^-	$\rho_c^- - \rho_c^+$	Asym. Dep.	Asym. Corr.	Port.	ρ_c^+	ρ_c^-	$\rho_c^- - \rho_c^+$	Asym. Dep.	Asym. Corr.
Size 1	0.507	0.816	0.309	0.000	0.040	BE/ME 1	0.953	1.023	0.070	0.175	0.880	L	0.442	0.628	0.186	0.000	0.023
Size 2	0.681	0.985	0.304	0.000	0.152	BE/ME 2	1.174	1.289	0.115	0.043	0.876	2	0.519	0.713	0.194	0.000	0.125
Size 3	0.766	1.020	0.254	0.005	0.333	BE/ME 3	1.159	1.350	0.191	0.013	0.807	3	0.581	0.784	0.203	0.000	0.228
Size 4	0.786	1.045	0.259	0.003	0.434	BE/ME 4	1.042	1.218	0.176	0.090	0.800	4	0.663	0.907	0.244	0.000	0.315
Size 5	0.884	1.136	0.252	0.003	0.512	BE/ME 5	0.927	1.149	0.222	0.023	0.703	5	0.669	0.944	0.275	0.000	0.339
Size 6	0.984	1.175	0.191	0.028	0.629	BE/ME 6	0.911	1.096	0.185	0.003	0.817	6	0.711	0.988	0.277	0.000	0.379
Size 7	1.162	1.286	0.124	0.005	0.761	BE/ME 7	0.872	1.046	0.174	0.020	0.774	7	0.708	1.028	0.320	0.000	0.397
Size 8	1.244	1.400	0.156	0.000	0.753	BE/ME 8	0.829	1.028	0.199	0.003	0.634	8	0.714	1.015	0.301	0.000	0.311
Size 9	1.420	1.508	0.088	0.068	0.945	BE/ME 9	0.764	1.027	0.263	0.000	0.504	9	0.655	1.005	0.350	0.000	0.174
Size 10	1.363	1.430	0.067	0.090	0.930	BE/ME 10	0.623	0.828	0.205	0.013	0.370	W	0.584	0.916	0.332	0.003	0.068

Note: The table reports the p-values of asymmetric dependence and asymmetric correlation test (as in Hong, Tu, and Zhou (2007)) on commonly used portfolios sorted by size, book-to-market and momentum at the exceedance level 0. The exceedance dependence measure ρ_c^+ and ρ_c^- and the difference $\rho_c^- - \rho_c^+$ are reported as well. The sample period is from January 1965 to December 2013. P-values of asymmetric dependence test are computed based on 399 stationary bootstrap resampling. P-values of asymmetric test are computed based on the asymptotic Chi-square distribution.

References

Ang, A., and J. Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63:443–94.

Ang, A., J. Chen, and Y. Xing. 2006. Downside risk. *Review of Financial Studies* 19:1191–239.

Backus, D., M. Chernov, and I. Martin. 2011. Disasters implied by equity index options. *The Journal of Finance* 66:1969–2012.

Backus, D., M. Chernov, and S. Zin. 2014. Sources of entropy in representative agent models. *The Journal of Finance* 69:51–99.

Ball, R., and S. P. Kothari. 1989. Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics* 25:51–74.

Braun, P. A., D. B. Nelson, and A. M. Sunier. 1995. Good news, bad news, volatility, and betas. *Journal of Finance* 50:1575–603.

Chabi-Yo, F., S. Ruenzi, and F. Weigert. 2014. Crash sensitivity and the cross-section of expected stock returns. *Working Paper* Available at SSRN 2011746.

Cho, Y.-H., and R. F. Engle. 1999. Time-varying betas and asymmetric effect of news: Empirical analysis of blue chip stocks. *NBER Working Paper* No. 7330.

Conrad, J., M. N. Gultekin, and G. Kaul. 1991. Asymmetric predictability of conditional variances. *Review of Financial Studies* 4:597–622.

Cover, T. M., and J. A. Thomas. 2006. *Elements of information theory.* 2 ed. Wiley-Interscience.

Efron, B. 1982. *The jackknife, the bootstrap, and other resampling plans*, vol. 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM.

Efron, B., and R. J. Tibshirani. 1993. *An introduction to the bootstrap.* New York: Chapman and Hall.

Embrechts, P., A. McNeil, and D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond* 176–223.

Epanechnikov, V. A. 1969. Nonparametric estimates of a multivariate probability density. *Theory of Probability and its Applications* 14:153–8.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25:383–417.

Granger, C. W., E. Maasoumi, and J. C. Racine. 2004. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25:649–69.

Hansen, L. P. 2012. Dynamic valuation decomposition within stochastic economies. *Econometrica* 80:911–67.

Hinkley, D. V., and S. Shi. 1989. Importance sampling and the nested bootstrap. *Biometrika* 76:435–46.

Hong, Y., J. Tu, and G. Zhou. 2007. Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20:1547–81.

Hong, Y., and H. White. 2005. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* 73:837–901.

Horowitz, J. L. 2001. The bootstrap. *Handbook of Econometrics* 5:3159–228.

Jiang, L., K. Wu, and G. Zhou. 2014. Asymmetry in stock returns: An entropy measure. *Working Paper* Emory University.

Jondeau, E. 2015. Asymmetry in tail dependence of equity portfolios. *Computational Statistics & Data Analysis* forthcoming.

Kelly, B., and H. Jiang. 2014. Tail risk and asset prices. *Review of Financial Studies* 27:2841–71.

Kroner, K. F., and V. K. Ng. 1998. Modeling asymmetric comovements of asset returns. *Review of Financial Studies* 11:817–44.

Kullback, S., and R. Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 79–86.

Künsch, H. R. 1989. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17:1217–41.

Li, Q., and J. S. Racine. 2006. *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Lintner, J. 1965. Security prices, risk and maximal gains from diversification. *Journal of Finance* 20:587–616.

Longin, F., and B. Solnik. 2001. Extreme correlation of international equity markets. *Journal of Finance* 56:649–76.

Parzen, E. 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33:1065–76.

Patton, A., D. N. Politis, and H. White. 2009. Correction to "automatic block-length selection for the dependent bootstrap" by d. politis and h. white. *Econometric Reviews* 28:372–5.

Politis, D. N., and J. P. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89:1303–13.

Politis, D. N., and H. White. 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23:53–70.

Racine, J. S. 1997. Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics* 15:369–79.

Rilstone, P. 1991. Nonparametric hypothesis testing with parametric rates of convergence. *International Economic Review* 32:209–27.

Robinson, P. M. 1991. Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58:437–53.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27:832–7.

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:425–42.