Market Returns Dormant in Option Panels*

Yoosoon Chang† Youngmin Choi‡ Soohun Kim§ Joon Park¶

April 30, 2019

Abstract

This paper offers a novel approach in identifying the relationship between the option prices and market risk premium using functional predictive regression. We provide an evidence that the predictability of the aggregate market return can be greatly improved by utilizing the linkage between the cross-section of option prices and the underlying asset returns. Applying our framework into the option panel data on S&P500 and the realized returns of S&P500 over our sample period from January 1996 to December 2015, we achieve a remarkable performance in predicting S&P500 index monthly returns, yielding 4.702% (6.198%) of in-sample (out-of-sample) R^2. We examine the relation between the information in risk-neutral density dynamics and that in the popular return predictors.

JEL classification codes: G12, G17

Keywords: functional predictive regression, return predictability, risk neutral measure, option market, market risk premium

*We are thankful to seminar participants at KAIST. Any remaining errors are solely ours. The usual disclaimer applies.

†Department of Economics, Indiana University, Bloomington, IN. E-mail: yoosoon@iu.edu.

‡Department of Economics and Finance, Zicklin School of Business at Baruch College, the City University of New York, New York, NY. E-mail: youngmin.choi@baruch.cuny.edu.

§Scheller College of Business, Georgia Institute of Technology, Atlanta, GA. E-mail: soohun.kim@scheller.gatech.edu.

¶Department of Economics, Indiana University, Bloomington, IN. E-mail: joon@iu.edu.

1 Introduction

According to a tail in Aristotle's 'Politics', Thales monetized his forecast on the olive price through forward contracts on the exclusive use of the olive-presses. Since the adoption of modern financial market where firms' ownerships are publicly traded, many traders as well as academics have been fascinated by the topic of forecasting stock returns. From practitioners' viewpoint, it is necessary to exploit real-time forecasts of stock returns for successful investment performance. Hence, it is quite natural that finance practitioners eagerly employ various variables and adopt novel methodologies for the purpose of forecasting stock returns. From academics' perspective, by analyzing the nature of stock return forecastability, we can deepen our understanding on the market participants' assessment of risks and their aversion towards the risks.

This paper proposes a novel methodology of predicting market risk premium by relating the risk-neutral density extracted from a cross section of option prices to the physical density observed from realized market returns. Our approach stands on two theoretical claims in finance: (i) the cross-sectional option prices contain the information on the risk-neutral density of the underlying asset (See Ross (1976), Banz and Miller (1978) and Breeden and Litzenberger (1978)) and (ii) the risk-neutral density and the physical density are equivalent (See Harrison and Kreps (1979)) and an asset pricing model can be interpreted as the change of measure between the two measures (See Hansen and Richard (1987)). For exploiting (i), many researchers have proposed various methods to find risk-neutral density from the cross-section of option prices.[1] We adopt the method by Ait-Sahalia and Duarte (2003) which imposes no-arbitrage restrictions nonparametrically and measure the risk-neutral density of S&P500 index after a month from a cross-section of SPX prices which expire after a month. Resorting to (ii), we attempt to identify the relation between the physical and risk-neutral densities. To this end, we do not take a stance on a particular asset pricing model but utilize the functional regression developed by Park and Qian (2012). In particular, we construct the physical density of S&P500 index monthly returns from bootstrapping S&P500 index daily returns in a given month and regress the physical density on the risk-neutral density estimated as of the previous month end. From this, we identify the relation between the two pairs of distribution and predict the physical density over the

[1] See Bliss and Panigirtzoglou (2002) and Jackwerth (2004) for comprehensive review on this topic.

following month through the observed risk-neutral density.

Given the theoretical association between risk-neutral and physical densities, we find that the risk-neutral density shows strong predictive power in explaining the physical density of the S&P500 index. The functional predictive regression of the risk-neutral density at time $t-1$ on the physical density at time t exhibits significant predictability with in-sample R^2statistics ranging from 4.375% to 4.720%, which outperforms the performance of conventional predictors such as dividend yield, earnings-price ratio, and other variables by Welch and Goyal (2008). While out-of-sample estimation of Welch and Goyal (2008) casts doubt about predictability, it is noteworthy that our approach delivers even stronger predictive power in out-of-sample prediction than in-sample prediction. Following out-of-sample forecast assessment of Campbell and Thompson (2008), our prediction model achieves 6.198% of the out-of-sample R^2 statistics. We believe that the aforementioned strong out-of-sample performance is due to the theoretical linkage between risk-neutral and physical densities.

This paper lies at the intersection of two literatures: return predictability and options. The academic history of predicting stock returns goes back to Cowles (1933) and Cowles and Jones (1937). In the early literature, the return predictability was interpreted against the market efficiency (Fama (1965), Fama (1970) and Samuelson (1965)). However, Fama (1991) harmonizes the empirical findings of return predictability with the market efficiency. Over the past decades, researchers have proposed various models featuring return predictability: external habits (Campbell and Cochrane (1999)), dynamic risk-sharing opportunities among heterogeneous agents (Lustig and Nieuwerburgh (2005)), long-run risks (Bansal and Yaron (2004), Bansal et al. (2010)), time-varying disaster risks (Gabaix (2012)). Furthermore, advanced methodologies are widely pursued: structural VAR (Cochrane (2008), Van Binsbergen and Koijen (2010)), model combination (Rapach, Strauss, and Zhou (2010) and Dangl and Halling (2012)), structural breaks (Guidolin and Timmermann (2007), Henkel, Martin, and Nardari (2011)). Also, given the inherent kinship between Q-density and option prices (Ross (1976), Banz and Miller (1978) and Breeden and Litzenberger (1978)), researchers have been proposing various methods to recover such relation (Jackwerth and Rubinstein (1996), Ait-Sahalia and Lo (1998), Ait-Sahalia and Duarte (2003)). Furthermore, Rosenberg and Engle (2002) and Jackwerth (2000, 2004) show how we can learn about the risk aversion of investors by jointly observing option markets and returns dynamics

of the underlying market. As recent endeavors in this line of research, Ross (2015) and Carr and Yu (2012) propose how to recover both of Q- and P- densities only from option panels under certain restrictions.

This paper contributes to the literature in two-fold. First, we highlight the valuable information dormant in option panels for predicting market returns. Recall that the transition between risk-neutral density and physical density is pinned down by the risk preference of pricing agents (Hansen and Jagannathan (1991)). Then, a rather mild assumption of enduring preference naturally implies a stable relation between between risk-neutral density and physical densities. As far as we know, this is the first paper aiming to predict the market return exploiting such relation. Second, we introduce a novel prediction method which handle predictors in a high-dimensional space, such as risk-neutral density function. Given that the size of relevant data such as SNS postings, household-level consumption or demographic changes is exceedingly growing, the ability to extract the relevant information from Big Data becomes crucial. The methodology that we use to connect two densities can be easily applied to connect any two objects.

This paper is organized as follows. In Section 2, we explain the data that we use for our empirical analysis. Section 3 describes our prediction methodology. Empirical results are reported in Section 4. Section 5 concludes.

2 Data

We explain the data source and filters. In extracting a risk-neutral density from option prices, we obtain the data on S&P500 index options over January 1996 to December 2015 from Option Metrics Database. In particular, we collect data on implied volatility, strike prices, expiration dates, dividend yields, the price of underlying asset (S&P500 index) of the options, and the risk-free rate data. We filter out each option contract with zero trading volume, zero open interest, or zero or missing implied volatility data. We also eliminate options with the average of the bid and ask quotes less than \$3/8. We work with only put option data.

We describe how we select the observation date and the corresponding time-to-maturity. From the filtered option data, we use the options whose remaining days until their expiration are close to 30 days so that a horizon of return predictability analysis is a monthly basis. Figure 2 shows on which date we collect option data with

certain time-to-maturity and how we aggregate market return data for our analysis. For example, when the expiration date of the index options is January 17th 2016, we collect the option data observed on 18th December 2015 whose time to expiration date are 30 days.[2] In this example, the *observation date* of the option data is December 18th 2015, and the *expiration date* of the options is January 17th 2016. Similar adjustments are made to the dividend yield and risk-free rate data. That is, we interpolate the dividend yields and risk-free rate data provided by Option Metrics to obtain 30-day dividend yield and risk-free rate on the *observation date*.

When *observation* and *expiration dates* of options do not correspond to the first or last date of a month, we make an analogous adjustment constructing the data on market return. Recall that our main objective is to investigate the predictability of a risk-neutral density obtained from options data on the physical density of stock market returns. Hence, we examine whether the risk-neutral density extracted from options data on the *observation date* has the predictability in explaining the stock market return realized over the *observation date* to the *expiration date* of the options used in prediction. To this end, we collect daily returns of S&P index from the *observation* to *expiration dates* and use bootstrap method to construct the physical density of monthly returns. Subsection 3.1 and 3.2 provide detailed descriptions on P- and Q-density construction, respectively.

Table 1 provides descriptive statistics for option data used in our main analysis. The second column of Table 1 provides annual and overall averages of the S&P500 index, and the next column shows the number of put options used in our analysis exhibiting a dramatic increase in recent years. The next four and last four columns display information on strike prices and implied volatility of the put options, respectively. Similarly, the range of strike prices and implied volatility exhibits wide dispersion in the second half of the sample period, compared to the first half.

[2]If there are no options whose days to expiration are exactly 30 days, we collect options whose time to expiration is closest to 30 days.

3 Methodology

3.1 Q-density construction

This subsection describes our approach to extract a risk-neutral distribution from panel data of option prices. The value of an option contract is the expected payoff on the expiration date discounted back to the present. Under risk-neutrality, the value of a call option at time t can be written as

$$C_t = \int_K^\infty e^{-r_f(T-t)}(S_T - K)q(T)dS_T, \tag{1}$$

where K is a strike price, r_f is a risk-free rate, T is date of expiration, $S_T$is the price of an underlying asset, and $q(\cdot)$ and risk-neutral density, respectively. Breeden and Litzenberger (1978) and Banz and Miller (1978)show that, from Equation (1), the risk-neutral density can be obtained by taking a second order derivative with respect to strike price. That is,

$$q(S_T) = e^{r_f(T-t)}\frac{\partial^2 C_t}{\partial K^2}. \tag{2}$$

A practical application of the above approach to extract a risk-neutral density has several empirical challenges. First, we observe only limited number of option contracts with discrete prices. Second, an option contract is traded based on bid and ask prices with microstructure noise. Third, there is a limited range of available strike prices. These issues make data on option prices coarse and noisy. Furthermore, the problem gets worsen as we take the second order derivative, which is our main objective of interest.

In this paper, we obtain a risk-neutral density by applying monotonicity and convexity of call option prices following Ait-Sahalia and Duarte (2003). In particular, from the positivity of the density and its integrability to one, two constraints can be written as follows.

$$-e^{-r(T-t)} \leq \frac{\partial C_t}{\partial K} \leq 0$$

$$\frac{\partial^2 C_t}{\partial K^2} \geq 0.$$

In particular, following Ait-Sahalia and Duarte (2003), a risk-neutral density is obtained using constrained least square regression with locally polynomial kernel smoothing approach. Using option panel data described in Section 1, we measure the risk-neutral density of S&P500 index after a month from a cross-section of prices of options on S&P500 (SPX) which expire in a month. The top two plots in Figure 1 display our estimated Q-density (left) and demeaned Q-density (right). As we apply theory-motivated constraints in extracting risk-neutral density from option panel data, the estimated Q-density function is well-behaved.

3.2 P-density construction

In this subsection, we describe a bootstrap method to construct the physical density of S&P500 index monthly returns. Recall that we aim to predict a physical density of S&P500 index return of a given month using a risk-neutral density of option prices observed at the end of a previous month. In particular, we collect daily returns of S&P500 index over the period between *observation date* to the *expiration date* described in Figure 2, which corresponds with the lifetime of option contracts used to construct the risk-neutral density.

Suppose, as of *observation date* t, that there are N number of days until expiration date $t+1$. That is, a cross-section of prices of options on S&P500 index is observed at month t, and these options expire at month $t+1$. N denotes a number of days between time t and $t+1$. Then, a realized monthly return of a given month t is written as $r_t = \prod_{i=1}^{N}(1+r_{i,t}) - 1$, where $r_{i,t}$is a daily return in day i of month t. So, we collect daily returns in a given month, $\{r_{1,t}, r_{2,t}, \cdots r_{i,t}, \cdots, r_{N,t}\}$. From this set of daily returns, we construct a series of bootstrapped $\left\{r^j_{1,t}, r^j_{2,t}, \cdots r^j_{i,t}, \cdots, r^j_{N,t}\right\}$, where $j = 1, \cdots, B$ and B is the number of bootstrapped samples. Using bootstrapped samples, monthly returns are simulated as $r^j_t = \prod_{i=1}^{N}(1+r^j_{i,t}) - 1$ for $j = 1, \cdots, B$. Thus, as we set $B = 10,000$ in our main analysis, 10,000 simulated monthly returns are generated for a given month and used to construct a physical density function of S&P500 index. The bottom two plots in Figure 1 show our physical density of S&P500 index (left) estimated using the bootstrap method and its demeaned density (right).

3.3 Projection of P-density on Q-density

Finally, we explain how we can relate the two densities described in the previous two subsections. Recall that $\mathbf{X}_R$ is the set of 1024 potential one-month horizon S&P500 returns used for Q-density and P-density estimation. Let $\mathbf{q}_t : \mathbf{X}_R \to \mathbb{R}$ denote the Q-density function mapping the one-month horizon S&P500 returns from the end of month t to the end of month $t+1$ to a real number, which is constructed using one-month horizon option prices observed at time t as in Section 3.1. Let $\mathbf{p}_{t+1} : \mathbf{X}_R \to \mathbb{R}$ denote the P-density function mapping the one-month horizon S&P500 returns from month t to $t+1$ to a real number, which is constructed using realized daily returns from the end of month t to the end of month $t+1$ as in Section 3.2. For practical convenience, we work with the demeaned versions of the density functions, $\mathbf{d}_{\mathbf{p},t+1}$ and $\mathbf{d}_{\mathbf{q},t}$ corresponding to $\mathbf{p}_{t+1}$ and $\mathbf{q}_t$, respectively. For in-sample (out of sample) estimation, we use the densities over the whole (previous) time-series to compute the mean densities.

We consider the function regression as follows:

$$\mathbf{d}_{\mathbf{p},t+1} = \mathbf{A}\mathbf{d}_{\mathbf{q},t} + \boldsymbol{\varepsilon}_{t+1}, \tag{3}$$

where $\mathbf{A}$ is a mapping from a space of real-valued functions to itself. We assume all technical conditions in Park and Qian (2012). Because our main interest lies in out of sample prediction, we will describe the estimation procedures as we perform out of sample prediction.

Step 1. We transform the vectors of $\mathbf{d}_{\mathbf{p},s+1}$, $\mathbf{d}_{\mathbf{q},s}$ into the vectors of $\mathbf{w}_{\mathbf{p},s+1}$, $\mathbf{w}_{\mathbf{q},s}$ through wavelet for $s \leq t$.

Step 2. Find the first K eigenvalues and eigenvectors from the matrix $\mathbf{W}_\mathbf{q}\mathbf{W}'_\mathbf{q}$, where $\mathbf{W}_\mathbf{q} = [\mathbf{w}_{\mathbf{q},1}\ \mathbf{w}_{\mathbf{q},2}\ \cdots\ \mathbf{w}_{\mathbf{q},t-1}]$. Let λ_k and $\mathbf{e}_k$ denote the k-th eigenvalue and eigenvectors, respectively, for $k \leq K$.

Step 3. Using the regularized regressors from Step 2, we find

$$\widehat{\mathbf{A}}_w = \left(\sum_{s\leq t} \mathbf{w}_{\mathbf{p},s+1}\mathbf{w}'_{\mathbf{q},s}\right)\left(\sum_{k\leq K} \lambda_k^{-1}\mathbf{e}_k\mathbf{e}'_k\right).$$

Step 4. From the estimated mapping in Step 3, we make a prediction $\widehat{\mathbf{w}}_{\mathbf{p},t+1} = \widehat{\mathbf{A}}_w\mathbf{w}_{\mathbf{q},t}$ in the wavelet space

Step 5. Transform the estimated function $\widehat{\mathbf{w}}_{\mathbf{p},t+1}$ in the wavelet space to $\widehat{\mathbf{d}}_{\mathbf{p},t+1}$ by reversing the procedures in Step 1.

Step 6. Add back the historical mean to $\widehat{\mathbf{d}}_{\mathbf{p},t+1}$ to obtain $\widehat{\mathbf{p}}_{t+1} = \widehat{\mathbf{d}}_{\mathbf{p},t+1} + \frac{1}{t}\sum_{s=0}^{t-1}\mathbf{p}_{s+1}$.

The intuition of the suggested procedures follows. Steps 1-3 stabilize the estimation outcomes. Recall that our target of $\mathbf{A}$ is a high-dimensional object. Hence, a brute-force regression approach is highly unstable. Instead, we propose an alternative route. We transform a real-value vector to a vector of elements corresponding to wavelet-basis in Step 1 and summarize the information in the regressor by the most important K components in Step 2. As a result of these regularization, the estimator in Step 3 does not suffer from the ill-conditioning problem. Steps 4 and 5 simply reverse the pre-treatments.

4 Empirical Results

Over the sample period from January 1996 to December 2015, we use the functional regression framework of Park and Qian (2012) to predict the realized returns of the S&P500 index. Panel A of Table 2 reports R^2 statistics, which is measured as following:

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \bar{r})^2}, \tag{4}$$

where $\hat{r_t}$is the predicted mean of our physical density obtained using our functional predictive regression, and $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$. We provide the R^2 statistics of the predictive functional regression for different numbers of eigenvalues (and corresponding eigenvectors) used in the estimation. The in-sample estimation exhibits significant predictability of the risk-neutral distribution extracted option prices on S&P500 index return, ranging from 4.375% to 4.720% of R^2 statistics.

Existing literature has documented, that even in the in-sample prediction, most of well-known predictors have poor predictability in explaining the market risk premium (see, among many others, Welch and Goyal (2008), Campbell and Thompson (2008), and Rapach et al. (2010)). To compare the performance of our approach to existing predictors in the literature, we also compute the in-sample R^2 statistics of the well-

known predictors used in Welch and Goyal (2008).[3] In particular, we obtain the R^2 statistics from the following time-series regression:

$$r_t = \alpha + \beta X_{t-1} + \varepsilon_t, \tag{5}$$

where r_t is the excess return on the S&P500 index of period t, X_{t-1} is a set of predictors observed in period $t-1$, and ε_t is an error term. Panel B of Table 2 provides the in-sample R^2 from using 13 monthly variables by Welch and Goyal (2008) as the predictor. The estimated R^2 statistics range from the lowest 0.001% of Treasury Bill rate to the highest 2.084% of stock variance. As already well known, when we include all 13 variables in the regression ("Kitchen sink"), the in-sample R^2 obviously increases as much as 11.55%. However, the stability of this finding will be investigated in the next subsection. Overall, the result provided in Table 2 suggests that our approach using information embedded in option prices to predict market return shows impressive performance in the prediction compared to the existing and well-known predictors in the financial market.

4.1 Out-of-sample prediction

Even though numerous economic variables have been proposed as predictors of stock returns, including valuation ratios and other variables as in Welch and Goyal (2008), the existence of out-of-sample predictability has still been controversial. In this subsection, we provide the performance of our approach in predicting stock return using the functional predictive regression with the Q-measure extracted from options data.

In out-of-sample prediction, we run the functional predictive regression (Equation(x.x)) using the expanding window. For each estimation, we generate out-of-sample forecast of stock market return and compare the forecast with a realized stock market return. Following Campbell and Thompson (2008) and Welch and Goyal (2008), we use the historical average of S&P500 returns ($\bar{r}_t$) as a natural benchmark model.

We use the out-of-sample R^2 suggested by Campbell and Thompson (2008) to compare the forecast from the functional regression ($\hat{r}_{t+1}$) and the forecast using the historical average of stock market returns ($\bar{r}_{t+1}$). The out-of-sample R^2 of Campbell and Thompson (2008) is computed as follows:

[3]The data are available from Amit Goyal's homepage: http://www.hec.unil.ch/agoyal/

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \bar{r}_t)^2}, \tag{6}$$

where $\hat{r}_t$ is the fitted value from the functional regression estimated through period t in an out-of-sample manner, and $\bar{r}_{t+1}$ is the historical average of stock market returns through period t. To estimate the historical average of stock returns, we use the long historical data on S&P500 index returns starting from 1927, giving the historical average advantage of data availability.

Table 3 provides the out-of-sample R^2 statistics of our proposed approach (Panel A) and predictors used in Welch and Goyal (2008) (Panel B). Our prediction model achieves 6.198% (6.102%) of the out-of-sample R^2 when using five (three) eigenvalues and corresponding eigenvectors in the predictive functional regression. As well documented in the existing literature, 13 predictors exhibit very poor performance in predicting stock market return in out-of-sample, even with negative out-of-sample R^2 statistics. Consistent with other papers, the kitchen sink regression using all 13 variables delivers a significantly negative out-of-sample R^2 of -3.422%. Overall, our approach using the functional regression in predicting stock market returns with option data provides unprecedentedly high predictability, even in out-of-sample analysis.

4.2 LASSO analysis

In this subsection, we examine the relationship between information embedded in the option panel data that we used to predict stock market return and conventional predictors which have been frequently used in the return predictability literature. In doing so, we use LASSO method (Tibshirani (1996)). Aiming to identify a linkage between risk-neutral distribution and widely used predictors in the literature, we start from 13 variables used in Welch and Goyal (2008), stacked in X_t. Using risk-neutral density we extracted from option panel, we construct the factor, f_t^k, on k-th principal component in the dynamics of risk-neutral density. With X_t and f_t^k, we estimate the following LASSO problem to identify variables which have significant effects on f_t^k:

$$\min_{\beta_0,\beta} \left(\sum_{t=1}^{T} (f_t^k - \beta_0 - \beta X_t)^2 + \lambda \left\| \beta \right\| \right), \tag{7}$$

where λ is a non-negative regularization parameter, and $\|\cdot\|$ is the standard ℓ^1-norm.

Table 4 provides the result for the LASSO analysis. Among well-known predictors used in Welch and Goyal (2008), the result suggest that dividend yield spread and stock variance are most strongly associated with all three factors of the risk-neutral density dynamics. In addition, inflation, net equity expansion, and book-to-market ratio are the next three important variables in explaining the first factor, while term spread, long-term yield, and net equity expansion have significant association with the second and the third factor of the risk-neutral density.

5 Conclusion

We This paper offers a novel approach in identifying the relationship between the option prices and market risk premium using functional predictive regression. We provide an evidence that the predictability of the aggregate market return can be greatly improved by utilizing the linkage between the cross-section of option prices and the underlying asset returns. Applying our framework into the option panel data on S&P500 and the realized returns of S&P500 over our sample period from January 1996 to December 2015, we achieve a remarkable performance in predicting S&P500 index monthly returns, yielding 4.702% (6.198%) of in-sample (out-of-sample) R^2. We examine that the information in risk-neutral density which contributes to this stark improvement in the predictability is not spanned by the information in the popular return predictors.

We propose new methodology to predict the market returns in a real time, exploiting the cross-section of option prices. Our methodology combines the risk-neutral density extraction by Ait-Sahalia and Duarte (2003) and the functional regression by Park and Qian (2012). Applying the proposed method to a large panel of option data, we find a stark improvement in predicting S&P500 index monthly returns, 6.198% of out-of-sample R^2, over existing studies.

We see a number of avenues for future research. A natural next step is to examine the predictability of other macro variables such as interest rate or exchange rate using the data of the option markets, the underlyings of which are those macro variables. Moreover, investigating the qualitative feature in the risk-neutral density, such as investor sentiment or slow price reaction, is also a possible direction for future research.

Reference

Ait-Sahalia, Yacine, and Jefferson Duarte, 2003, Nonparametric option pricing under shape restrictions, *Journal of Econometrics* 116, 9–47.

Ait-Sahalia, Yacine, and Andrew W. Lo, 1998, Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance* 53, 499–547.

Bansal, Ravi, Dana Kiku, and Amir Yaron, 2010, Long Run Risks, the Macroeconomy, and Asset Prices, *American Economic Review* 100, 542–546.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles, *Journal of Finance* 59, 1481–1509.

Banz, Rolf W, and Merton H Miller, 1978, Prices for State-contingent Claims: Some Estimates and Applications, *The Journal of Business* 51, 653–672.

Bliss, Robert R., and Nikolaos Panigirtzoglou, 2002, Testing the stability of implied probability density functions, *Journal of Banking & Finance* 26, 381–422.

Breeden, Douglas T, and Robert H Litzenberger, 1978, Prices of State-contingent Claims Implicit in Option Prices, *The Journal of Business* 51, 621–651.

Campbell, John Y., and John Cochrane, 1999, Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *Review of Financial Studies* 21, 1509–1531.

Carr, Peter, and Jiming Yu, 2012, Risk, return, and ross recovery, *Journal of Derivatives* 20, 38.

Cochrane, John H., 2008, The Dog That Did Not Bark: A Defense of Return Predictability, *Review of Financial Studies* 21, 1533–1575.

Cowles, Alfred, 1933, Can stock market forecasters forecast?, *Econometrica* 1, 309–324.

Cowles, Alfred, and Herbert E Jones, 1937, Some a posteriori probabilities in stock market action, *Econometrica* 5, 280–294.

Dangl, Thomas, and Michael Halling, 2012, Predictive regressions with time-varying coefficients, *Journal of Financial Economics* 106, 157–181.

Fama, Eugene F., 1965, The behavior of stock-market prices, *The Journal of Business* 38, 34–105.

Fama, Eugene F, 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* 25, 383–417.

Fama, Eugene F, 1991, Efficient Capital Markets: II, *Journal of Finance* 46, 1575–1617.

Gabaix, Xavier, 2012, Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance, *The Quarterly Journal of Economics* 127, 645–700.

Guidolin, Massimo, and Allan Timmermann, 2007, Asset allocation under multivariate regime switching, *Journal of Economic Dynamics and Control* 31, 3503–3544.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of Security Market Data for Models of Dynamic Economies, *Journal of Political Economy* 99, 225–262.

Hansen, Lars Peter, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Harrison, J. Michael, and David M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.

Henkel, Sam James, J. Spencer Martin, and Federico Nardari, 2011, Time-varying short-horizon predictability, *Journal of Financial Economics* 99, 560–580.

Jackwerth, Jens Carsten, 2000, Recovering Risk Aversion from Option Prices and Realized Returns, *Review of Financial Studies* 13, 433–451.

Jackwerth, Jens Carsten, 2004, *Option-Implied Risk-Neutral Distributions and Risk Aversion* (Charlotteville : Research Foundation of AIMR).

Jackwerth, Jens Carsten, and Mark Rubinstein, 1996, Recovering Probability Distributions from Option Prices, *Journal of Finance* 51, 1611–1632.

Lustig, Hanno N., and Stijn G. Van Nieuwerburgh, 2005, Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective, *Journal of Finance* 60, 1167–1219.

Park, Joon Y., and Junhui Qian, 2012, Functional regression of continuous state distributions, *Journal of Econometrics* 167, 397–412.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy, *Review of Financial Studies* 23, 821–862.

Rosenberg, Joshua V., and Robert F. Engle, 2002, Empirical pricing kernels, *Journal of Financial Economics* 64, 341–372.

Ross, Stephen A, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Ross, Steve, 2015, The Recovery Theorem, *Journal of Finance* 70, 615–648.

Samuelson, Paul A, 1965, Proof that properly anticipated prices fluctuate randomly, *IMR; Industrial Management Review (pre-1986)* 6, 41.

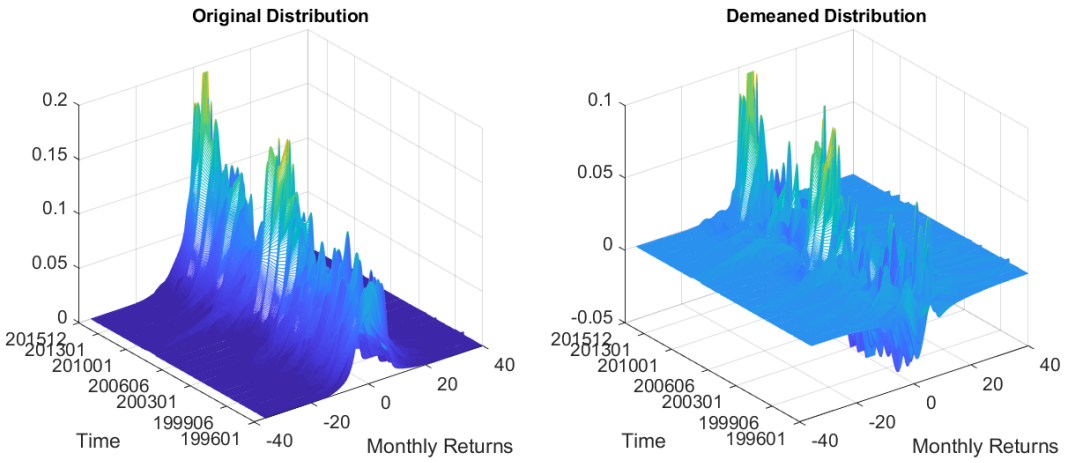
Tibshirani, Robert, 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society* 58, 267–288.

Van Binsbergen, Jules H, and Ralph SJ Koijen, 2010, Predictive regressions: A present-value approach, *Journal of Finance* 65, 1439–1471.

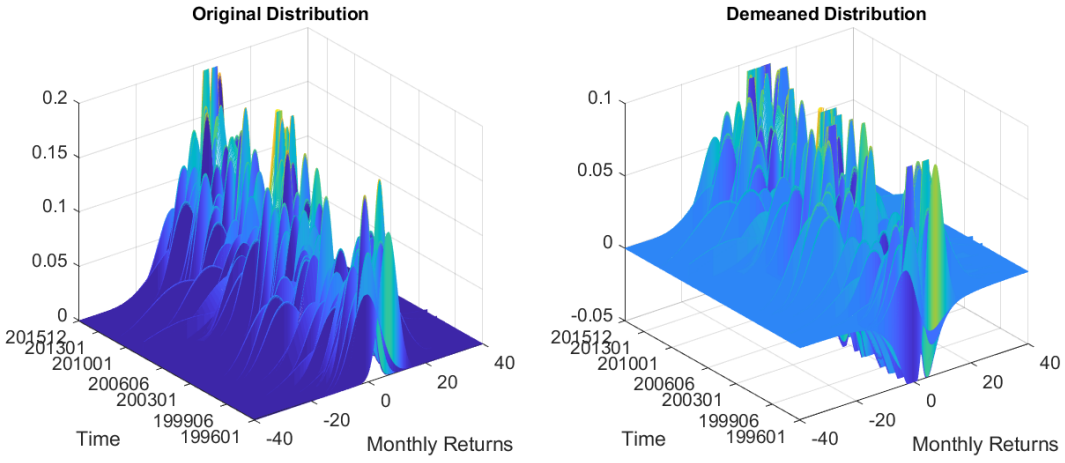
Welch, Ivo, and Amit Goyal, 2008, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* 21, 1455–1508.

Figure 1: Estimated Q and P densities from sample data

(1) Q-density

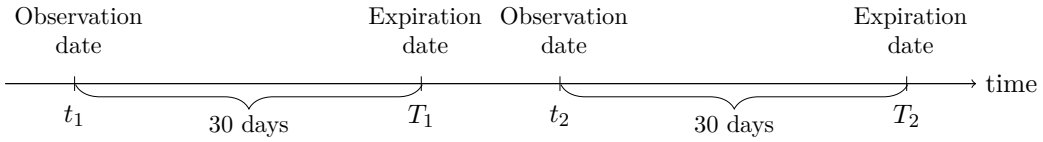


(2) P-density



The plots above display estimated Q density (top) and P density (bottom) from our sample data. Q density is estimated by following Ait-Sahalia and Duarte (2003) which is described in Subsection 3.1. P density is obtained using daily returns of S&P500 index as described in Subsection 3.2. In each top and bottom sections, we provided estimated Q and P densities along with their demeaned densities which are used in our main predictive analysis in Section 4.

Figure 2: Time Matching and Aggregation of Option and Market Return Data



The timeline above displays how observation date and expiration date of option data are coordinated and how market return data are aggregated accordingly. Option data are collected on *observation dates*, which are 30 days before the option *expiration dates*. That is, the collected S&P500 index options have 30-days of time-to-maturity. Once these *observation* and *expiration* dates are specified, daily returns on S&P500 from the *observation* and *expiration* dates are collected.

Table 1: Descriptive Statistics

			Strike Price Range				Implied Volatility			
Year	S&P 500 Index Avg.	No. of Put Options	Min	Median	Mean	Max	Min	Median	Mean	Max
1996	674.85	287	375	645.0	641.59	775	0.0865	0.1778	0.1974	0.7993
1997	875.86	370	400	830.0	831.68	1075	0.1500	0.2354	0.2620	0.9156
1998	1087.86	386	400	1025.0	1005.47	1275	0.1096	0.2600	0.3019	1.3293
1999	1330.58	353	600	1275.0	1235.42	1550	0.1009	0.2598	0.2916	0.9613
2000	1419.73	297	850	1395.0	1365.74	1800	0.0813	0.2470	0.2692	0.7016
2001	1185.75	305	700	1130.0	1137.72	1800	0.1350	0.2900	0.3209	1.0501
2002	988.59	331	600	975.0	977.05	1800	0.1487	0.3064	0.3289	0.9350
2003	967.93	328	500	930.0	920.70	1300	0.1139	0.2280	0.2535	0.6290
2004	1133.97	365	700	1090.0	1074.12	1300	0.0848	0.1611	0.1860	0.5351
2005	1207.77	410	800	1170.0	1152.28	1400	0.0741	0.1443	0.1632	0.5458
2006	1318.31	505	800	1255.0	1244.95	1500	0.0575	0.1505	0.1662	0.5459
2007	1478.10	660	900	1395.0	1383.64	1700	0.0782	0.2098	0.2179	0.5905
2008	1215.22	870	200	1162.5	1107.11	1900	0.1393	0.3282	0.4147	1.6948
2009	948.52	902	300	835.0	829.77	1500	0.1392	0.3819	0.4103	1.1949
2010	1130.68	1016	400	1005.0	989.55	1500	0.1104	0.3082	0.3393	1.0780
2011	1280.76	1014	400	1130.0	1099.66	1700	0.0997	0.3364	0.3699	1.2981
2012	1386.51	902	500	1250.0	1223.90	1550	0.0907	0.2444	0.2709	1.1321
2013	1652.29	1114	500	1460.0	1453.11	2000	0.0606	0.2329	0.2453	1.1972
2014	1944.41	1256	1000	1725.0	1700.35	2175	0.0450	0.2305	0.2540	0.7937
2015	2051.93	2016	300	1720.0	1707.86	2250	0.0556	0.3003	0.3236	1.9824
All	1263.98	13687	200	1225	1259.15	2250	0.0450	0.2579	0.2956	1.9824

The table shows descriptive statistics of put options used in main analyses. Each row represents annual averages, and the last row provides statistics of the data for the full sample period from 1996-2015. The second and third columns show an average of S&P500 index and a total number of put options on the index used to extract risk-neutral probability distribution, respectively. The next four columns (the last four columns) provide minimum, median, mean, and maximum of strike prices (implied volatility) of the put options.

Table 2: In-Sample Prediction Results

Panel A. Functional Regression	
Number of Eigenvalues	In-Sample R^2
$K = 3$	4.375%
$K = 4$	4.487%
$K = 5$	4.720%

Panel B. Predictors in Goyal and Welch (2008)	
Variable	In-Sample R^2
Dividend-Price Ratio	1.113%
Dividend Yield	1.437%
Earnings-Price Ratio	0.246%
Dividend Payout Ratio	0.004%
Stock Variance	2.084%
Book-to-Market Ratio	0.174%
Net Equity Expansion	1.840%
Treasury Bill Rate	0.001%
Long-Term Yield	0.068%
Long-Term Return	0.134%
Term Spread	0.098%
Default Yield Spread	0.606%
Inflation	0.126%
All (Kitchen sink)	11.552%

The table reports the R^2 statistics from the functional predictive regression provided in Section 3 and R^2 statistics of variables used in Welch and Goyal (2008). In Panel A, the in-sample R^2 statistics from the functional predictive regression is computed using Equation (HERE Functional Regression Eq) and Equation (4) in Subsection **??**. The value of K in the first column represents the number of eigenvalues and corresponding eigenvectors used in the estimation of the functional regression. In Panel B, the in-sample R^2 statistics for variables of Welch and Goyal (2008) are computed from the predictive regression of Equation (5). The sample period of estimation spans from January 1996 to December 2015.

Table 3: Out-of-Sample Prediction Results

Panel A. Functional Regression	
Number of Eigenvalues	Out-of-Sample R^2
$K = 3$	6.012%
$K = 4$	5.749%
$K = 5$	6.198%

Panel B. Predictors in Goyal and Welch (2008)	
Variable	Out-of-Sample R^2
Dividend-Price Ratio	2.431%
Dividend Yield	2.273%
Earnings-Price Ratio	0.269%
Dividend Payout Ratio	-0.723%
Stock Variance	1.910%
Book-to-Market Ratio	-0.249%
Net Equity Expansion	-4.420%
Treasury Bill Rate	-2.050%
Long-Term Yield	-2.456%
Long-Term Return	-1.297%
Term Spread	-0.554%
Default Yield Spread	-1.828%
Inflation	-1.585%
All (Kitchen sink)	-3.422%

The table reports the out-of-sample R^2 statistics from the functional predictive regression approach provided in Section 3. The out-of-sample R^2 statistics is computed following by Campbell and Thompson (2008)as Equation (6). The period of the out-of-sample prediction is over the last 5 years of our sample period, starting from January 2010. The value of K in the first column represents the number of eigenvalues and corresponding eigenvectors used in the estimation of the functional regression. The sample period spans from January 1996 to December 2015.

Table 4: Selected Variables for the First Factor of Risk-Neutral Density

1st Factor	2nd Factor	3th Factor
Default Yield Spread	Stock Variance	Stock Variance
Stock Variance	Default Yield Spread	Default Yield Spread
Inflation	Term Spread	Term Spread
Net Equity Expansion	Long Term Yield	Long Term Yield
Book to Market Ratio	Net Equity Expansion	Net Equity Expansion

This table represents selected variables in explaining the first three principal components extracted from the dynamics of risk-neutral density. A complete set of predictors used in the LASSO analysis includes 13 variables used in Welch and Goyal (2008). Among all 13 predictors, the table reports the most significant five variables for three principal components in each column.