Crump, Richard K.; Giannone, Domenico; Hundtofte, Sean

Working Paper

Changing risk-return profiles

Staff Report, No. 850

Provided in Cooperation with:
Federal Reserve Bank of New York

Federal Reserve Bank of New York
Staff Reports

Changing Risk-Return Profiles

Richard K. Crump
Domenico Giannone
Sean Hundtofte

Staff Report No. 850
June 2018

Changing Risk-Return Profiles

Richard K. Crump, Domenico Giannone, and Sean Hundtofte

Federal Reserve Bank of New York Staff Reports, no. 850
June 2018
JEL classification: C22, G17, G18

Abstract

We show that realized volatility, especially the realized volatility of financial sector stock returns, has strong predictive content for the future distribution of market returns. This is a robust feature of the last century of U.S. data and, most importantly, can be exploited in real time. Current realized volatility has the most information content on the uncertainty of future returns, whereas it has only limited content about the location of the future return distribution. When volatility is low, the predicted distribution of returns is less dispersed and probabilistic forecasts are sharper. Given this finding on the importance of financial sector volatility not just to financial equity return uncertainty but to the broader market, we test for changes in the realized volatility of banks over a $50 billion threshold associated with more stringent Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank) requirements. We find that the equity volatility of these large banks is differentially lower by 9 percentage points after Dodd-Frank compared to pre-crisis levels, controlling for changes over the same period for all banks and all large firms.

Key words: stock returns, realized volatility, density forecasts, optimal pools, Dodd-Frank, financial intermediation, financial conditions

---

Crump, Giannone, Hundtofte: Federal Reserve Bank of New York (emails: richard.crump@ny.frb.org, domenico.giannone@ny.frb.org, sean.hundtofte@ny.frb.org). The authors thank Nina Boyarchenko for helpful comments and discussions. Oliver Kim provided excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

To view the authors' disclosure statements, visit https://www.newyorkfed.org/research/staff_reports/sr850.html.

1 Introduction

Are stock returns predictable? This is a longstanding question in the finance literature. The dominant focus has been to identify variables, available in real time, which can provide information about the future conditional mean of stock returns. In an influential paper, Welch and Goyal (2008) argued that popular predictors in the literature showed limited evidence in out-of-sample performance relative to the naive model. These variables did not have sufficient strength and stability to be exploited in real time relative to a benchmark model where the conditional mean of returns is equal to the unconditional expectation. The literature on prediction has addressed these issues by modifying the estimation approach such as through model combination or by imposing economic constraints (as surveyed by Rapach and Zhou (2013)), focussing on other aspects of the distribution such as predicting tail events or volatility (e.g., Engle et al. (2008), Patton (2011)), and using information across assets to help inform predictions (e.g., Adrian et al. (2017b)).

In this paper, we evaluate whether the whole distribution of stock returns is predictable. We assess whether economic and financial variables can provide predictive content about the future distribution of stock returns. In this context, unpredictability would be concluded if it is impossible to improve over a benchmark model that forecasts the future distribution to be equal to the unconditional density of returns. We use the set of predictors assembled in Welch and Goyal (2008), adding an additional proxy for the conditions of financial intermediaries available over the entire time series (the realized volatility in financial sector equity returns), and we allow these predictors to shape the full conditional distribution as opposed to affecting only the conditional mean.

A natural metric to assess probabilistic forecasts is the average predictive score. According to which an accurate density forecast is one which assigns high ex-ante probability for events that are realized ex-post. To estimate the predictive density we use the methodology of Adrian et al. (2017a). Intuitively, the approach can be characterized as follows: In the first step quantile predictive regressions are performed for a set of fixed quantiles. In the second step, we fit a flexible parametric density function by choosing the parameter values to best approximate the fitted conditional quantiles.

Our main result is that realized volatility, especially the realized volatility of financial sector stock returns, has strong predictive content for the future distribution of returns. This is a robust feature of the data and, importantly, can be exploited in real time. Our estimation approach allows the predictors to affect the location, scale, skewness and kurtosis of the tails in a simple and flexible way. This allows us to easily construct conditional density forecasts and compare them to a benchmark of no predictability. This is a different way to judge the accuracy of a stock return prediction than is commonly used in the literature where the focus is generally on predicting the conditional mean or variance. This is an important distinction because it suggests that if we shift the focus away from very specific aspects of the distribution, we see meaningful and substantial gains in predictability.

We find that current realized volatility has significant information content for the uncertainty or spread of future returns, whereas it has only limited content about the location of the future return distribution. When volatility is low, our predictions of returns are sharper—in other words, returns

are more concentrated around the unconditional mean. Thus, in "normal" times, our predictive density forecasts will be quite accurate. When volatility is high, then the predictive distribution becomes flatter i.e., mass is spread over a more dispersed set of outcomes and so larger forecast errors are less surprising, which is reflected in the predictive score. The benchmark model with no predictability, which does not allow the variance of the predictive distribution to change with financial conditions, has a variance that is an average across regimes, and so is too flat in turbulent times and too large in "normal times". We also find that the default yield spread between BAA and AAA rated corporate bonds shows clear improvement over the benchmark model—albeit of a smaller magnitude and less robust than realized volatility.

After developing and testing our methodology, we then apply our findings to address a pressing policy (and open research) question. Is the banking system healthier or not today, and have recent regulatory reforms helped? Analysis addressing this question has, until recently, been limited, and part of the challenge has been what measure to examine as an indicator of risk or health in the banking system. Sarin and Summers (2016) look across a variety of measures that could arguably be relevant and argue the evidence is at best mixed. Chousakos and Gorton (2017) examine Market-to-Book ratios across advanced economies and find that the low post-crisis valuations of bank equities has been surprisingly persistent relative to previous financial crises.

Focusing on the volatility of bank equity as a financial condition most relevant to broad market risk as shown in our preceding analysis, we find that recent regulatory reforms are associated with an improvement in risk. Pursuing a differences-in-differences-in-differences empirical design, we compare the changes in market volatility of the largest banks (those over the Dodd-Frank \$50 Billion threshold for designation as a systemically important financial institution) to banks under the threshold and to other large non-bank firms. We find a material relative reduction in market volatility for the largest banks (-9%). A significant caveat to interpreting this result is if the structural relationship between financial conditions (as measured by bank equity volatility) and the real economy (as indicated by forward-looking aggregate stock market volatility) has changed. For example, regulations might have have chased financial intermediation to unregulated intermediaries. We check the underlying predictive relationship–between the volatility of financial sector stock returns to the dispersion and left tail of market returns established in the long time series of US data–and find no immediate evidence of a recent break in our predictive relationship.

Our paper is most closely related to Cenesizoglu and Timmermann (2008) who study whether economic and financial variables can help improve prediction of the quantiles of the return distribution. Cenesizoglu and Timmermann (2008) also find evidence of predictive power especially in the upper tail of the return distribution. Our paper is also related to Massacci (2015) who evaluates the accuracy of density forecasts but restricts the economic and financial variables to predict the location of the distribution only. Durham and Geweke (2014) predict higher-frequency, daily returns allowing for realized intraday volatility and option-implied volatility but restrict these variables to predict the scale of the distribution only.

Our paper is also related to those papers that investigate interactions between financial crises (high periods of market volatility) and equity returns. For example, Baron and Xiong (2017) ex-

amine changes in skew of bank equity returns in response to changes in economic leverage (bank credit/GDP). Moreira and Muir (2017) look at the returns of a market portfolio managed by lagged volatility, finding investors are not compensated for risk in periods of lagged high volatility. Similarly, when we examine the distribution of aggregate returns conditional on realized volatility we find no predictability in the average return, but do find predictability in the risk to holding the market portfolio (either in terms of variance at short horizon, or skew at longer horizons), indicating a breakdown in a risk-return tradeoff. Adrian et al. (2017b) document a strongly nonlinear dependence of stock and bond returns on past equity market volatility as measured by the VIX with expected stock returns increasing for stocks when volatility increases from moderate to high levels.

Our paper is organized as follows. Section 2 provides initial motivating evidence regarding future stock market returns and current realized volatility and credit spreads. Next, Section 3 describes our estimation approach and provides our main results. It is important to emphasize that all results in Section 3 are in real time with no look-ahead bias. Section 4 builds off the main results to assess the impact of post-crisis regulatory reforms in the US. Section 5 concludes.

2 Motivation

To motivate the results in the paper we start with a set of predictive quantile regressions. Let us denote by r_{T+h} the annualized CRSP value-weighted return between T and $T+h$ and by x_T a vector containing a set of predictors, including a constant. We work with a linear quantile regression model, i.e., we assume that

$$Q_{r_{T+h}|x_T}(p|x_T) = x_T'\beta_p, \tag{1}$$

where $Q_{r_{T+h}|x_T}(\cdot|x_T)$ is the quantile function of the distribution of h-period future returns conditional on the variables x_T. To develop intuition, consider the simple case when $x_t = 1$. Then, β_p is the unconditional quantile of the distribution of returns, and solving the minimization problem is equivalent to ordering the data and choosing the value of $\hat{\beta}_p$ so that a fraction p of observations take on smaller values. Equivalently, β_p is the Value-at-Risk for an investor holding the market portfolio.[1] Thus, in the general case we are assuming that the Value-at-Risk is, $x_t'\beta_p$, a linear function of the variables contained in x_t.

As a first step, we examine the realized volatility of financial sector equity returns (volFinancial). We view volFinancial as a proxy for financial conditions, i.e. on the state of financial intermediation, available throughout the long history of our sample, as compared to the National Financial Conditions Index (NFCI) which only starts in 1971.[2] In their review of systemic risk measures, Giglio et al. (2016) find realized financial volatility to be an important predictor, and the mean equity volatility of the financial sector can be interepreted as an approximation of the financial

[1] Value-at-Risk is a term commonly used in (bank) risk management to describe the maximum estimated loss of a portfolio within a given time period at a specified confidence level, or here $(1-p)$.

[2] NFCI is found by Adrian et al. (2017a) to be an important indicator in predicting shifts the distribution of aggregate output growth

fragility or soundness of financial intermediaries in an economy (Chousakos et al., 2016; Atkeson et al., 2017). We provide a full list of variable names and definitions on page 19, and include further description in Section 6 of the Appendix.

The top left panel of Figure 1 shows the results of both quantile and conventional OLS regressions of one-period ahead quarterly stock returns on volFinancial. Superimposed over the scatterplot of the data are four lines corresponding to the OLS fit (blue), median (red), and 5th (black) and 95th (green) percentiles. The estimated slopes for the median ($\beta_{.50}$) and the conditional mean (OLS) are essentially zero, while the extreme quantiles have steep positive and negative slopes. The flat slopes of the OLS and median lines illustrate the difficulty in predicting returns found in many studies that focus on point forecasts (e.g., Welch and Goyal (2008)). Essentially, realized volatility does not have any clear predictive power for the location of the distribution of future returns.

In contrast, the estimated slopes corresponding to the upper and lower tails are strongly positive and negative, respectively—as realized volatility rises, the gap between the quantiles grows. Consequently, when realized volatility is high (low) in the current quarter, then the central mass of the predicted distribution of returns for next quarter spreads (concentrates) over a wider (narrower) interval. Thus, although median predicted returns are largely unchanged, risk is higher (lower). The model captures what we can see in the scatterplot–realized returns are more variable when last quarter's realized volatility is higher. The right panel of Figure 1 shows results of the same analysis for the horizon $h = 4$. In this case, when realized volatility is high, the predicted distribution of returns for the subsequent period again spreads over a wider area but does so asymmetrically, with mass more concentrated on the left side of the distribution. We find the same relationships with stronger magnitudes when investigating the direct relationship to future financial sector-specific market returns.[3] We next investigate whether these results can be explained by pure statistical chance.

In the bottom panel of Figure 1 we report the estimated slope coefficients over a range of values for p. We also include the conditional mean and median estimates we would obtain with a Gaussian linear model of returns and lagged realized volatility along with confidence intervals corresponding to nominal coverage rates of 68%, 90%, and 95% for $\hat{\beta}_p$. The confidence intervals have been constructed as follows: first, we fit a bivariate vector autoregression (VAR) of returns and realized volatility with four lags. Second, we simulate datasets based on this model and the associated in-sample parameter estimates. Finally, for each simulation we run the appropriate quantile regression for each p. The confidence intervals reflect the distribution of the estimated slopes across simulations.[4] The OLS point estimate provides an appropriate benchmark for judging the deviation of quantile estimates from the mean relationship in the data, and the Gaussian model is a suitable benchmark since the quantiles as a function of lagged realized volatility will be parallel lines, i.e., the slopes are the same across quantiles.

Both charts in the bottom panel of Figure 1 show that, even if the benchmark model were

[3]For illustrations of the relationship of volFinancial to financial sector stock returns (rFinancial), see Appendix Figure 16.

[4]This can be interpreted as a parametric bootstrap exercise.

correct, there is no statistically significant evidence of predictability for the conditional mean as the OLS estimates are comfortably within the confidence bounds. This also true for the median and so, by properties of the Gaussian distribution, for all other quantiles. Notice that as p moves toward zero or one, sampling uncertainty increases, although not dramatically because the quantile regression uses *all* of the data to inform its estimates, not just extreme observations. In sharp contrast, the left and right tails of the conditional distribution are associated with slope coefficients that are well outside of our bootstrapped confidence intervals. This provides clear evidence against the linear model and strongly suggests that the conditional predictive distribution can be modeled as a function of lagged realized financial-sector volatility. The right column of Figure 1 shows the analogous results for the $h = 4$ horizon. In this case we see that realized volatility is strongly associated with the left tail of the predictive distribution of returns.

In Figure 2 we show the corresponding results using the default yield spread (DFY)–the difference between BAA and AAA rated yields. In general, the results are similar to that of Figure 1 except that we find some evidence that the location of the predictive distribution moves with DFY. This can be seen by the fact that zero is outside the confidence bands for $h = 1$ and especially for $h = 4$. Taking all this evidence in concert, we find suggestive evidence that the realized volatility of the financial sector and the default spread provide useful information about the predictive distribution of returns and especially about its dispersion.

3 Main Results

In the previous section, we provided preliminary evidence that features of the future distribution of returns change over time with respect to current values of realized volatility and credit spreads. In this section, we formally investigate this relationship across a range of variables common to the literature of return forecasting as tested by Welch and Goyal (2008), and show that our results provide new insights related to the debate of stock return forecastability.

It is important to emphasize that throughout this section, all results are based on real-time estimation of the associated model parameters, completely avoiding any look-ahead bias in our results. The only real-time aspect we do not account for are data revisions, as we use the data as available today, which might be different from the data available at the time the forecasts would have been made. That said, the predictors we find which forecast most accurately should be subject to minimal, if any, revisions.[5]

3.1 Forecasting Returns: From Quantile Regressions to Density Forecasts

In this subsection we formally describe our approach to producing conditional density forecasts. Further details may be found in Adrian et al. (2017a). We estimate the quantile of r_{T+h} conditional

[5]For example, our realized volatility variables are based on data from CRSP which is continually updated to correct data errors as they are identified on a monthly basis.

on x_T,

$$\hat{Q}_{r_{T+h}|x_T}(p|x_T) = x_T\hat{\beta}_p, \tag{2}$$

by choosing $\hat{\beta}_p$ to minimize the quantile weighted absolute value of the in-sample errors:

$$\hat{\beta}_p = \arg\min_{\beta_p \in \mathbb{R}^k} \sum_{t=1}^{T-h} \left(p \cdot \mathbb{1}_{(r_{T+h} \geq x_t\beta)} |r_{T+h} - x_t\beta_p| + (1-p) \cdot \mathbb{1}_{(r_{T+h} < x_t\beta)} |r_{T+h} - x_t\beta_p|\right) \tag{3}$$

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

The quantile regression provides us with approximate estimates of the quantile function, or equivalently, the inverse cumulative distribution function evaluated at a specific value.[6] A simple approach would be to estimate the conditional quantile function on a fine grid of different values of p and use this to recover the conditional CDF and PDF. In practice this is challenging, however, as approximation error and estimation noise would require local smoothing in concert with monotonicity restrictions (Cenesizoglu and Timmermann (2008), Schmidt and Zhu (2016)). A simpler approach is to instead choose a coarse grid and find the best approximation based on a flexible parametric family of distributions. Following Adrian et al. (2017a) we use the skewed t-distribution developed by Azzalini and Capitanio (2003) in order to smooth the quantile function and recover a valid probability density function:

$$f(y;\mu,\sigma,\alpha,\nu) = \frac{2}{\sigma} g\left(\frac{y-\mu}{\sigma};\nu\right) G\left(\alpha\frac{y-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+\frac{y-\mu}{\sigma}}};\nu+1\right). \tag{4}$$

Here $g(\cdot)$ and $G(\cdot)$ denote the PDF and CDF of the Student t-distribution. The skewed-t distribution has four parameters, (μ,σ,α,ν) which pin down the location μ, scale σ, skewness ν, and shape α. There are three useful special cases to gain intuition about this distribution:

1. if $\alpha = 0$ (and $\sigma = 1$) then $f(\cdot;\nu)$ is the PDF of the noncentral t-distribution with noncentrality parameter μ;
2. if $\alpha = 0$ and $\nu \to \infty$, then we obtain a normal distribution with mean μ and variance σ^2;
3. if $\alpha \neq 0$, and $\nu \to \infty$ then we obtain the skewed normal distribution.

Throughout, let $F(y;\mu,\sigma,\alpha,\nu)$ denote the CDF of the skewed t-distribution and $F^{-1}(y;\mu,\sigma,\alpha,\nu)$ its associated quantile function.

To implement our approach we choose our grid of quantiles as $S_p = \{.05,.25,.75,.95\}$. Adrian et al. (2017a) find that this choice performs well in practice. Then, for a fixed time period T we first run four quantile regressions to obtain $Q_{r_{T+h}|x_T}(p)$ for each $p \in S_p$ and then estimate

[6] Recall that the quantile function $Q(\cdot)$ is defined as $Q(p) = \inf\{x : F(x) \geq p\}$. For continuous distributions, we have that $Q(p) = F^{-1}(p)$.

$(\mu_T, \sigma_T, \alpha_T, \nu_T)$ via

$$(\hat{\mu}_{T+h}, \hat{\sigma}_{T+h}, \hat{\alpha}_{T+h}, \hat{\nu}_{T+h}) = \arg\min_{\mu,\sigma,\alpha,\nu} \sum_{p \in S_p} \left(\hat{Q}_{r_{T+h}|x_T}(p|x_T) - F^{-1}(p; \mu, \sigma, \alpha, \nu) \right)^2 \tag{5}$$

where $\hat{\mu}_{T+h} \in \mathbb{R}$, $\hat{\sigma}_{T+h} \in \mathbb{R}^+$, $\hat{\alpha}_{t+h} \in \mathbb{R}$, and $\hat{\nu}_{T+h} \in \mathbb{Z}^+$. It is important to emphasize that these parameter estimates are direct functions of the conditioning variable x_T although we drop the explicit dependence for notational convenience. Conceptually, as the value of x_t shifts over time the conditional distribution will also shift in a nonlinear way through the $(\hat{\mu}_{T+h}, \hat{\sigma}_{T+h}, \hat{\alpha}_{T+h}, \hat{\nu}_{T+h})$.

To obtain the forecast conditional density we can simply evaluate the PDF of the skewed-t distribution at our estimates: $\hat{p}_{r_{T+h}|x_T}(y) = f(y; \hat{\mu}_T, \hat{\sigma}_T, \hat{\alpha}_T, \hat{\nu}_T)$. Figures 3 and 4 provide two examples of the inputs and outputs of this two-step estimation procedure as a forecaster would have observed near the beginning of the last financial crisis, the fourth quarter of 2008. Figure 3 shows the conditional CDF and PDF based on volFinancial for both of our forecasting horizons. The charts in the top row show the raw quantile regression outputs (green) along with the fitted conditional quantile function (blue). For reference the fitted unconditional quantile function is also shown (red). The charts in the bottom row show the associated conditional and unconditional PDFs for this date. The conditional distribution is substantially flatter than the unconditional distribution implying that, conditional on the level of financial-sector volatility, tail outcomes were considered much more likely. In comparison, Figure 5 illustrates a quiet period at the end of 1993, where the conditional distribution is tighter than the unconditional.

The density evaluated at the ex-post realization of the return is referred to as the predictive score: $\hat{p}_{r_{T+h}|x_T}(r_{T+h})$ is the "probability" assigned ex-ante and in real-time by the model to the ex-post realized value. Using these scores, we can form a simple and intuitive measure of the out-of-sample accuracy of a model, the average log predictive score,

$$\frac{1}{|\mathbb{T}|_0} \sum_{T \in \mathbb{T}} \log \hat{p}_{r_{T+h}|x_T}(r_{T+h}),$$

where $\mathbb{T}$ is the sample (set of points in time) at which the forecasts are evaluated and $|\mathbb{T}|_0$ is the size of the evaluation sample. A desirable forecasting model should provide (ex-ante) high conditional probability to those (ex-post) realizations frequently observed and low probability to those infrequently observed. More formally, it can be shown that this criterion function is equivalent to minimizing the Kullback-Leibler loss function of the combined density relative to an unknown true density function. Clearly, a higher value of the average log predictive scores is preferred.

The right panel of Table 1 reports results for the out-of-sample performance of the density forecasts. The first row reports the average log predictive scores from the naive model which is produced by choosing the skewed-t distribution that best approximates the unconditional quantiles (up to time T). The subsequent rows show the difference in average log predictive scores relative to this benchmark; therefore, a *positive* number represents an improvement in accuracy relative to the naive model. The overall sample runs from 1926Q4 to 2016Q4. We use 80 observations (the first 20 years of data) as an initial window, then construct recursive forecasts, so that the out-of-sample

results begin in 1947Q1. Thus, $\mathbb{T} = \{1947Q1, ..., 2016Q4\}$ and $|\mathbb{T}| = 280$.

Table 1 clearly shows that realized volatility–either in the financial sector or the entire market– and the default yield spread produce consistent outperformance in predicting the distribution of future stock returns. The average log predictive scores are strongly positive at both the $h = 1$ and $h = 4$ horizon. In parentheses we include standard error estimates based on a HAC estimator as indicative measures of the variability of the scores. Realized volatility based on returns of the financial sector outperforms total market volatility at both horizons, but by only a modest amount. Both volatility measures, in turn, outperform the default yield spread especially at the horizon $h = 4$. Looking across all of our other candidate predictors, only NTIS at horizon $h = 1$ has a higher average log predictive score than the naive model but even there the improvement is marginal. Realized volatility and the default yield spread stand out sharply against the rest of the field.

As an additional comparison we also include mean square forecast errors (MSFE) from point forecasts obtained from OLS regressions in the left panel.[7] The first row reports the MSFE from the naive (constant only) model and the subsequent rows show the difference in MSFE relative to this benchmark; therefore, a *negative* number represents an improvement in accuracy relative to the naive model. The second column reports HAC standard errors for the MSFE differential series. It is clear from the table that there is little, if any, systematic improvement over the naive estimator when considering point forecasts based on the set of predictors we consider. Thus, we observe similar results in the out-of-sample analysis as in the full-sample analysis presented in Section 2.

Table 2 repeats the predictive exercises with a twenty-year fixed window to form forecasts, rather than estimating the model on an expanding window. We again observe little consistent predictive gains relative to the naive model in terms of point forecasts. We also again observe that volFinancial, volMarket, and DFY show strong evidence of density forecast outperformance relative to our benchmark at the $h = 1$ horizon. In fact, since the estimation windows are fixed we can formally test the null of equal predictive ability between the constant-only model and each univariate model using the test statistic of Diebold and Mariano (2002). For the $h = 1$ horizon, the average log predictive scores using either realized volatility measure or the default yield spread are more than twice their associated standard error. Interestingly, at the $h = 4$ horizon, this is no longer the case. It could be that the informational content of a previous financial crisis is required to make predictive gains on the downside of stock market returns, in the US' case the only systemic crisis prior to 2007 was the Great Depression, which occurs at the beginning of our dataset and hence is picked up with a recursive window formulation that "remembers" any previous crisis episodes in the time series.

In Figures 6-8, we plot the time series of the predictive scores (without applying the log transformation) for forecasts based on volFinancial, DFY along with the BM ratio relative to the naive benchmark model. The latter is included to show an example of poor performance in terms of log predictive scores. For both volFinancial (Figure 6) and DFY (Figure 7) it is clear that the outperformance relative to the benchmark model is not concentrated in specific time periods or outliers

[7] Similar results are obtained using the conditional median to construct point forecasts.

but is instead consistent over time. Furthermore, the outperformance does not come from better predictability for the tails of the return distribution (both conditional and unconditional forecasts are similar for these low probability events). Instead, the advantage comes from the ability to have sharper predictions in "normal" times. For example, when volatility is low we know that the probability of a tail event is lower and so we can attribute more mass to returns in the center of the distribution. Our results highlight an important distinction between point and density forecasts. We have shown that none of our candidate predictors are systematically informative about the conditional mean of the distribution or the median. In other words, the location does not change systematically with our predictors. Instead, what changes systematically is the likelihood that the model places on the event that future observations will be "close" to the central tendency of the distribution. The predictive score measures distance by the conditional density.[8]

To cement ideas, consider a simple case where returns are conditionally Gaussian,

$$y_{t+1} = \mu_{t+1|x_t} + \sigma_{t+1|x_t}\epsilon_{t+1}.$$

The associated log predictive score is,

$$\log(\hat{p}_{y_{T+1}|x_T}(y_{T+1})) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_{T+1|x_T}) - \frac{1}{2}\frac{(y_{T+1} - \mu_{T+1|x_T})^2}{\sigma_{T+1|x_T}}.$$

The benchmark case of non-predictability occurs when $\mu_{T+1|x_T}$ and $\sigma_{t+1|x_t}$ are constant. Accurate forecasting performance can be achieved by models which result in $\mu_{T+1|x_T}$ near y_{T+1}. This is the exclusive goal of point forecasting. However, even if a model is uninformative about the location, i.e., $\mu_{T+1|x_T}$ is constant, accurate predictions can be achieved if x_t is informative about the variance, i.e., $\sigma_{t+1|x_t}$ is non-constant and varies with x_t. $\sigma_{t+1|x_t}$ has two contrasting effects on the log predictive score. When $\sigma_{t+1|x_t}$ rises, predictions become less sharp and hence less informative. This is reflected in the second term. At the same time, the higher uncertainty means that the model is less surprised by outcomes which are far from the mean, and so forecast errors are down-weighted as reflected in the third term. In the general case, predictability of any aspect of the distribution is exploited. For example, if x_t predicts negative skewness for predictions of returns over the next four quarters, negative errors will be less surprising and then down-weighted in the calculation of the predictive score.

In light of these examples, one source of outperformance of the conditional model relative to the unconditional forecast is that larger forecast errors are typically preceded by high realized volatility or credit spreads, which presage a widening of the return distribution. The conditional density forecast accommodates this widening and so is less surprised by extreme realized returns.

[8] Anne et al. (2017) consider a framework where the focus is restricted to particular parts of the distribution to compare and combine models.

3.2 Combining density forecasts

A large literature in studies of point forecasting models have found that combining forecasts from multiple models outperforms that of any individual model (Timmermann (2006)). In a similar vein, we can also consider combining density forecasts from the individual models presented in Table 1. Moreover, analyzing the weights assigned to individual models provides an alternative metric to compare our candidate predictors' performance.

The problem of optimally combining density forecasts can be formulated as follows. Let us consider the N density forecasts $\hat{p}_{r_{T+h}|x_T^i}(y)$ made by using a set of predictors x_t^i, $(i = 1, \ldots, N)$ at time T over the horizon h. We may write the combined forecast density as

$$\sum_{i=1}^{N} w_i \hat{p}_{r_{T+h}|x_T^i}(y). \tag{6}$$

It is natural to impose that the weights, $(w_1, \ldots, w_N)$, are nonnegative and sum to one which ensures that the combined density is itself a probability density function (Hall and Mitchell, 2007; Geweke and Amisano, 2011). In practice, we will choose the weights at every point T so as to maximize real-time accuracy of the forecast combination over a given evaluation sample starting at a date T_0 and ending in T; accordingly, the forecasts are exactly the same as if constructed by a forecaster in real time.

As before, we measure accuracy by the average log score, and so for the combined forecast we may choose the optimal weights via:

$$\hat{w}_1, ..., \hat{w}_N = \arg \max_{w_i > 0; \sum_{i=1}^{n} w_i = 1} \frac{1}{T - T_0 - h + 1} \sum_{\mathcal{T}=T_0}^{T-h} \sum_{i=1}^{N} w_i \hat{p}_{r_{\mathcal{T}+h}|x_{\mathcal{T}}^i}(y), \tag{7}$$

where $\hat{p}_{r_{T+h}|x_T^i}(r_{T+h})$ is the "probability" assigned ex-ante and in real-time by the model based on the ith set of predictors to the ex-post realized value. The weights $\hat{w}_i$ depend on time T when they are computed, and are based only on information available up to time T. However, we drop the dependence on T for notational convenience.

The probability forecast of the optimal combination is given by:

$$\hat{p}_{r_{T+h}|x_T^1,\ldots,x_T^N}(y) = \sum_{i=1}^{N} \hat{w}_i \hat{p}_{r_{T+h}|x_T^i}(y). \tag{8}$$

Conflitti et al. (2015) show that optimally combining forecasts is a viable strategy even when the number of forecasts to be combined is large. Indeed, imposing the constraints that the weights should be nonnegative and sum to one enforces a form of shrinkage on the weights. This ensures a reasonable out-of-sample performance of the combined forecasts. More precisely, these two constraints imply that the sum of the absolute values of the weights sum to one, i.e., $w_i \geq 0$ and $\sum w_i = 1$ imply $\sum_{i=1}^{n} |w_i| = 1$. The Lagrangian dual of this problem is equivalent to the following

unconstrained optimization problem with a penalized objective:

$$\hat{w}_1, ..., \hat{w}_N = \arg\max_{w_i} \frac{1}{T - T_0 - h + 1} \sum_{\mathcal{T}=T_0}^{T-h} \sum_{i=1}^{N} w_i \hat{p}_{r_{\mathcal{T}+h}|x^i_{\mathcal{T}}}(y) + \lambda \left(\sum_{i=1}^{n} |w_i| - 1 \right) \tag{9}$$

It is important to emphasize the advantages of our approach of combining density forecasts compared to other popular alternatives such as Bayesian model averaging (BMA) that assign weights on the sole basis of the performance of each model, ignoring mutual dependence. Instead, the optimal combination takes into account mutual dependence among the component's forecasts. Consequently, predictions that, taken individually, are not accurate might have positive weights in optimal linear pools since they might tend to be more accurate when other forecasts perform poorly. Below, we will see that the mutual dependence is exploited in practice.

The solution to the maximization above is not available in closed form, but can be computed using the iterative algorithms developed by Conflitti et al. (2015), which can handle a large number of forecasts efficiently. At iteration $k+1$, the weights are given by:

$$w_i^{(k+1)} = w_i^{(k)} \frac{1}{T - T_0 - h + 1} \sum_{t=T_0}^{T-h} \frac{\hat{p}_{r_{T+h}|x^i_T}(r_{T+h})}{\sum_{i=1}^{N} w_i \hat{p}_{r_{T+h}|x^i_T}(y_{T+h})} \tag{10}$$

The non-negativity constraint for the weights is automatically satisfied at each iteration provided that the algorithm is initialized with positive weights summing to one, i.e. with $w_i^{(0)} = N^{-1}\ \forall i$. A primary advantage of this algorithm is that it is very simple to implement and scales well with the cross-sectional dimension of the problem, and hence we can consider combinations of all 15 models.

The out-of-sample score of the pooled distribution is given by

$$\hat{p}_{r_{T+h}|x^1_T,...,x^N_T}(r_{T+h}) = \sum_{i=1}^{N} \hat{w}_i \hat{p}_{r_{T+h}|x^i_T}(r_{T+h})$$

The rows at the bottom of Tables 1 and 2 report the average of the logarithmic score over the evaluation sample. It is evident that the predictive accuracy of the pooled forecasts at the one quarter ahead horizon is comparable to the best individual models based on volFinancial, volMarket, and DFY. At the four-quarter horizon, the combined forecast is superior to these individual models– especially when the forecasts are formed based on a rolling-window.

We also report the predictive accuracy of the optimal combination of point forecasts. The weights are computed by minimizing the MSFE of the pooled forecasts, instead of maximizing its average log predictive score. As for the combination of density forecasts, we impose the two constraints that the weights should be nonnegative and that they should sum to one. As shown by Conflitti et al. (2015), this problem displays a strong similarity to the problem of determining minimum-variance portfolios with short-selling constraints, with the component forecasts being replaced with the returns of individual assets. Moreover, this problem can also be viewed as

a special case of a so-called LASSO regression (Tibshirani, 1996) since, as described above, the constraints imply that the sum of the absolute values of the weights is equal to one. This type of constraint is known to enforce sparsity, namely the presence of zeros in the weight vector, which means that only a small number of forecasts will be selected (active) to form the combined forecast. The algorithm to compute the weights of the point forecasts is the constrained LASSO developed by Brodie et al. (2009) for portfolio optimization. The bottom row of the left panel of Tables 1 and 2 show that in half of our cases, this optimal combination is clearly preferred to the individual models whereas in the other half they are inferior to selected individual predictors. Recall that the optimal combination of weights for density forecasts, in contrast, showed consistent performance at least on par with the best individual models.

The real-time weights, $\{\hat{w}_i : i = 1, \ldots, n\}$, formed for the optimal combined conditional density forecast are reported in Figure 9. Strikingly, volFinancial is consistently chosen in the optimal density forecast at both the $h = 1$ and $h = 4$ horizons. Moreover, it is strictly preferred to volMarket suggesting that although the two series strongly co-move there is additional information in financial-sector volatility.[9] This is a surprising result: for predicting the dispersion or risk in aggregate market returns, it is information contained in financial intermediary equity returns that is more important than information from the aggregate market itself. For $h = 1$, only volFinancial and DFY contribute consistently to the optimal forecast whereas for $h = 4$ the log dividend-price ratio (logDP) also plays a role. We note that logDP demonstrated poor predictive density forecasting performance when evaluated individually, suggesting that the optimal combination is exploiting co-movement between the errors from this forecast and the other variables. In contrast to the density forecasting exercise where the weights are largely concentrated in a few variables, the optimal weights for the point forecasts are more dispersed–both across predictors and in the time series. For example, variables may receive a positive weight for short bursts suggesting instability in the optimal pooling of forecasts. This provides further evidence that the candidate predictors are more informative about the predictive distribution of returns more generally than about its central tendency.

We can also visualize the results in Table 1 by showing the realized returns along with the real-time conditional forecast confidence intervals for different levels of nominal coverage. This is shown in Figure 11. We first observe that realized returns are almost always contained in the 95% conditional confidence interval. Since all forecasts are done in a completely real-time manner this was not guaranteed. Second, we observe that the median of the predictive distribution shows very little movement whereas the conditional confidence intervals vary substantially over time. Third, at short forecast horizons, the confidence intervals are approximately symmetric whereas at longer forecast horizons there is a clear negative skewness in the predictive distribution.

We conclude this section by studying the calibration of the density forecasts. In Figure 13, we show the conditional and unconditional predictive distributions of the pooled forecast. We plot the fraction of realizations that are below any given quantile of the real time predictive distribution

[9] In the figure, black shading denotes a weight of one and white shading a weight of zero. Intermediate values range from light yellow (closer to zero) to dark red (closer to one).

as a function of the quantile.[10] In a perfectly calibrated model, this percentage would be exactly equal to the quantile itself and the statistics would be aligned on the 45-degree line. This would hold exactly if the density forecasts were based on the quantile of the empirical distribution based on the full sample. The calibration test evaluates whether the out-of-sample density forecasts have this natural and desirable property. Following Rossi and Sekhposyan (2017), we report confidence bands around the 45-degree line to account for sampling uncertainty. We observe that the fraction of ex-post realizations that are below any given quantile of the real time predictive distribution is not significantly different from the quantile itself. We conclude that the forecasts are not only sharp, at least sharper than the unconditional forecasts, but are also well calibrated. We report the PITs for individual components in the Appendix.

4 Case Study: After the Post-Crisis Regulatory Reforms

Our out-of-sample tests find realized volatility, particularly in financial sector returns, to be a consistent predictor of the risk in market returns, and of downside risk at longer horizons. Given this consistent relationship of realized volatility to the risk in returns, how does volatility appear to have changed after recent regulatory reforms? Do reforms appear to have helped reduce risk? To explore this question, we first examine if the underlying structural relationship of financial sector volatility to aggregate returns has changed. We then formally test for differential changes amongst banks most affected by regulatory reforms using less affected banks and large, non-banking firms as counterfactuals. Specifically, we test for changes associated with the enhanced regulatory and supervisory requirements of the Dodd-Frank Act passed on July 21 2010, which included stress testing and associated capital requirements for banks with above $50 billion in assets.[11,12]

We first examine the structural relationship between volatility and returns. Given the short time series of observations we have post-crisis for a statistical test of any change, we pursue a simple qualitative analysis of this relationship. Visually examining the spread graphs of Figure 1 does not indicate a large deviation from the relationship we found in the longer time series, either at one or four quarter forecasting horizons. Given the small number of observations in the time series, a statistical test for any change would be underpowered. Instead, we turn to the cross-section to test for differential changes associated with firms most impacted by Dodd-Frank regulations.

We next examine realized volatility of the large institutions specifically targeted by the Dodd-Frank Act, comparing banks with over $50B in assets to smaller banks and to other large, non-financial, firms. The results of a differences-in-differences-in-differences (DDD) analysis can be interpreted as a result of new regulations only if changes affecting these large banks are same

[10] Technically speaking, this fraction is a function of the quantiles of the empirical cumulative distribution of the Probability Integral Transform (PIT), which in turn is the cumulative predictive distribution evaluated at the realization.

[11] This asset cutoff was intended to apply more stringent capital requirements to about 30 bank holding companies and international holding companies (see Bank Systemic Risk Regulation: The $50 Billion Threshold in the Dodd-Frank Act

[12] In addition to firms legally designated as a 'bank' and which would have already fallen under a bank regulatory umbrella, Dodd-Frank also allowed for the designation of other bank-like or financially important entities as "Systemically Important Financial Institutions"

as those experienced by all banks, or all large non-banks over the same period, i.e. the parallel trend assumption for identification. Even if we cannot interpret the results causally, given our interpretation of realized volatility as a proxy for financial conditions, it is still informative to simply find out whether volatility is lower or higher for our largest banks today relative to pre-crisis levels. We therefore perform a Differences-in-Differences-in-Differences analysis, testing for changes in quarterly observations of realized volatility. We average realized volatility across firms from 2000-2016, distinguishing firms by their size ('large' indicating assets over \$50B as of 2010Q2), and industry ('bank' indicating firms defined as banking firms in their SIC codes using the Fama-French 48 industry classification.). We use time (quarterly) and industry/size fixed effects to allow for overall constant differences between "treated" and "untreated" firms, and allow for slopes to vary intra- and post-crisis as follows:

$$\begin{aligned}\sigma_{ist} = \alpha_t + \alpha_{i,s} &+ \beta \cdot \mathbb{1}\{t > 2010Q2\} \cdot \mathbb{1}\{s = large\} \cdot \mathbb{1}\{i = banking\} \\ &+ \delta_1 \cdot \mathbb{1}\{t > 2010Q2\} \cdot \mathbb{1}\{s = large\} \\ &+ \delta_2 \cdot \mathbb{1}\{t > 2010Q2\} \cdot \mathbb{1}\{i = banking\} \\ &+ \gamma_0 \cdot \mathbb{1}\{t >= 2008, t <= 2010Q2\} \cdot \mathbb{1}\{s = large\} \cdot \mathbb{1}\{i = banking\} \\ &+ \gamma_1 \cdot \mathbb{1}\{t >= 2008, t <= 2010Q2\} \cdot \mathbb{1}\{s = large\} \\ &+ \gamma_2 \cdot \mathbb{1}\{t >= 2008, t <= 2010Q2\} \cdot \mathbb{1}\{i = banking\} + \varepsilon_{ist}.\end{aligned} \tag{11}$$

Where "large" is defined as 2010Q2 assets > \$50B, and β represents the coefficient of interest: the marginal estimated change in the volatility of large banks relative to other firms, large firms, and banks in general.

Table 3 estimates this DDD. The specification in equation (11) allows for constant differences between types of firms, isolating an average change in volatility for large banks over and above those for large firms or other banks. The point estimate of the coefficient of interest is -9.1%, indicating a differentially lower volatility in the post-Dodd Frank period for large banks compared to the pre-crisis period. At the same time, banks and large non-bank firms have experienced increases in volatility. After factoring in relevant fixed effects, volatility for large banks is slightly lower (-3 percentage points) in the post-crisis period compared to pre-crisis. The mean annualized volatility in this sample is 41%, with a standard deviation of 23.5%. Results are of a similar magnitude if analysis is performed at the individual firm level of observation, and do not appear sensitive to the exact choice in quarter for crisis start date.

If all of this change in the realized volatility of banks over the Dodd-Frank size threshold could be ascribed to policy changes, we could interpret the policy as having reduced risk, not only for banks but for the broader market. A weakness of this evaluation exercise is if there are other coincident changes in the economy that particularly affect large banks. It could be that the market's perceptions of large banking stocks has changed in a way that is not reflected in the historical relationship inferred from the 1926 onwards time series. For instance, contrary to a stated intent to make it easier to fail larger firms, recent regulatory reforms may have dampened volatility for the largest banks through greater too-big-to-fail (TBTF) implicit insurance. So long

as market perceptions of this TBTF premium have remained steady (Afonso and Santos, 2015), however, smaller banks should provide a good counterfactual.

Our findings are consistent with a "banks as utilities" hypothesis of the impact of recent regulatory reforms, and stand in contrast to the findings of Sarin and Summers (2016). An alternative hypothesis of the effect of regulation, is that recent reforms have chased financial intermediary risk-taking activity outside of the publicly traded banking industry to the so-called "shadow banking" sector. Both of these hypotheses are consistent with the persistently low post-crisis bank Tobin's Qs found by Chousakos and Gorton (2017). While we interpret the decline in volatility as beneficial, this will hold only to the extent that the relationship does not alter going forward and activity does not shift outside of the traditional publicly traded banking sector, for example into new privately held financial intermediaries.

Based on a time series relationship that holds out-of-sample across various regulatory regimes since 1926, we conclude that the financial system looks healthier today. A key historical indicator of crisis-like events and the left tail of market returns has been volatility in financial sector returns. Examining the short time series of data we have post-crisis, we find no strong evidence to believe this relationship has changed. By pursuing cross-sectional analysis, we find that the realized volatility in the equities of banks most impacted by the signature regulatory reform of the last financial crisis have fallen when compared to banks in general or large firms in general.

5 Conclusion

Are stock returns predictable? We address this question by modeling the predictive density as a function of economic and financial variables and assessing the accuracy of the associated forecasts. We compare these forecasts in pure out-of-sample exercises relative to a naive benchmark of no predictability according to which the predictive distribution is equal to the historical unconditional distribution. Results show that stock returns are predictable because some financial variables provide useful information, especially about the uncertainty surrounding future returns.

The most informative variable we find is realized volatility, especially realized volatility of financial sector stock returns, followed by the default yield spread. This is a robust feature of the last century of US data and, importantly, can be exploited in real time. Realized volatility has the most information content for the uncertainty in future returns, whereas it has only limited information content about the location of the future return distribution. When volatility is low, the predicted distribution of returns is less disperse and probabilistic forecasts become sharper. That we find a proxy for financial conditions–the realized volatility of financial equity returns–consistently adds precision to forecasts of the distribution of aggregate market returns, mirrors recently documented relationships between financial conditions and the distribution of macroeconomic growth outcomes (Giglio et al., 2016; Adrian et al., 2017a).

Given this predictive relationship, we test for changes in the realized volatility of bank equities above and below a $50B threshold introduced by Dodd-Frank as a threshold for more stringent regulatory requirements. We find that the equity volatility of these large banks is differentially lower

by 9 percentage points after Dodd-Frank compared to pre-crisis levels (and compared to changes over the same period for smaller banks, or large non-bank firms). These findings are relevant to a current policy debate, suggesting that the more stringent requirements of Dodd-Frank may be associated with a reduction in a variable relevant to systemic market risk.

References

Adrian, T., Boyarchenko, N., Giannone, D., 2017a. Vulnerable growth. American Economic Review (forthcoming).

Adrian, T., Crump, R., Vogt, E., 2017b. Nonlinearity and flight to safety in the risk-return trade-off for stocks and bonds. Journal of Finance (forthcoming).

Afonso, G., Santos, J., 2015. What do rating agencies think about "too-big-to-fail" since dodd frank? Liberty Street Economics .

Anne, O., van Dijk, D., van der Wel, M., 2017. Combining density forecasts using focused scoring rules. Journal of Applied Econometrics 32, 1298–1313.

Atkeson, A. G., Eisfeldt, A. L., Weill, P.-O., 2017. Measuring the financial soundness of us firms, 1926–2012. Research in Economics 71, 613–635.

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society: Series B 65, 367–389.

Baron, M., Xiong, W., 2017. Credit expansion and neglected crash risk*. The Quarterly Journal of Economics 132, 713–764.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., Loris, I., 2009. Sparse and stable markowitz portfolios. Proceedings of the National Academy of Sciences 106, 12267–12272.

Cenesizoglu, T., Timmermann, A., 2008. Is the distribution of stock returns predictable?, working paper.

Chousakos, K., Gorton, G., 2017. Bank health post-crisis, working paper.

Chousakos, K., Gorton, G., Ordonez, G., 2016. Aggregate information dynamics. Tech. rep., Working Paper, Yale University.

Conflitti, C., De Mol, C., Giannone, D., 2015. Optimal combination of survey forecasts. International Journal of Forecasting 31, 1096–1103.

Diebold, F. X., Mariano, R. S., 2002. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Durham, G., Geweke, J., 2014. Improving asset price prediction when all models are false. Journal of Financial Econometrics 12, 278–306.

Engle, R. F., Ghysels, E., Sohn, B., 2008. On the economic sources of stock market volatility. AFA 2008 New Orleans Meeting Paper .

Geweke, J., Amisano, G., 2011. Optimal prediction pools. Journal of Econometrics 164, 130–141.

Giglio, S., Kelly, B., Pruitt, S., 2016. Systemic risk and the macroeconomy: An empirical evaluation. Journal of Financial Economics 119, 457–471.

Hall, S. G., Mitchell, J., 2007. Combining density forecasts. International Journal of Forecasting 23, 1–13.

Massacci, D., 2015. Predicting the distribution of stock returns: Model formulation, statistical evaluation, var analysis and economic significance. Journal of Forecasting 34, 191–208.

Moreira, A., Muir, T., 2017. Volatilitymanaged portfolios. The Journal of Finance 72, 1611–1644.

Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Rapach, D., Zhou, G., 2013. Chapter 6 - forecasting stock returns. In: Elliott, G., Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, pp. 328 – 383.

Rossi, B., Sekhposyan, T., 2017. Alternative tests for correct specification of conditional predictive densities, working paper.

Sarin, N., Summers, L. H., 2016. Have big banks gotten safer? Brookings Papers on Economic Activity .

Schmidt, L. D. W., Zhu, Y., 2016. Quantile spacings: A simple method for the joint estimation of multiple quantiles without crossing, working paper.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288.

Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger, C., Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Elsevier, vol. 1, chap. 04, pp. 135–196, first ed.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Variable Names and Definitions

All variable names, definitions and data are from Welch and Goyal (2008) other than realized volatilities (volFinancial and volMarket), which are constructed from CRSP data. For completeness we include primary details below (further information may be found in Welch and Goyal (2008)).

volFinancial Financial sector realized volatility is calculated as the standard deviation of CRSP value-weighted return daily returns for the financial sector within a quarter.

volMarket Market realized volatility is calculated using the same method as volFinancial. Note that this series is very similar to the volatility series, SVAR, used in Welch and Goyal (2008).

logDP The log dividend price ratio is the log-ratio of the 12-month moving sum of dividends paid on the S&P 500 index and the level of the index.

logDY The log dividend yield is the log-ratio of the 12-month moving sum of dividends paid on the S&P 500 index and the lagged level of the index.

logEP The log earnings price ratio is the log-ratio of the 12-month moving sum of earnings on the S&P 500 index and the level of the index.

logDE The log dividend earnings ratio is the log-ratio of the 12-month moving sum of dividends paid on the S&P 500 index and the 12-month moving sum of earnings on the index.

BM The book-to-market ratio is the ratio of book value to market value for the Dow Jones Industrial Average.

NTIS Net equity expansion is the ratio of a 12-month moving sum of net equity issues by New York Stock Exchange-listed (NYSE) stocks to the total end-of-year market capitalization of NYSE-listed stocks.

TBL The Treasury bill rate is the secondary-market interest rate on a three-month Treasury.

LTY The long-term government bond yield.

LTR The return on a long-term government bond.

TMS The term spread is the difference between the yields of the long-term government bond (LTY) and the three-month Treasury bill rate (TBL).

DFY The default yield spread is the difference between the yields of BAA and AAA-rated corporate bonds.

DFR The default return spread is the difference between returns on long-term corporate bonds and long-term government bonds.

INFL Inflation is calculated from the Consumer Price Index (CPI) for urban consumers. We use the one-period lag of inflation to account for the delay in CPI releases.

Table 1: **Diebold-Mariano Test: rMarket recursive**

The table below compares the out-of-sample market return performance of each predictor for point forecasts (left panel) and density forecasts (right panel) estimated with a recursive (expanding) window. In the left panel, the first row shows the mean-squared forecast error for the naive, or unconditional, model. The following rows show the MSFE differential over the naive model of conditioning on that predictor, with HAC standard errors shown in parentheses. In the right panel, the first row shows the sum of log scores for the naive, or unconditonal forecast. The following rows show the log score gain (or loss) from conditioning on that predictor, with HAC standard errors shown in parentheses. Note that for point forecasts, a negative differential indicates a more accurate forecast, while for density forecasts, a positive differential indicates a more accurate forecast.

	Point				Density			
	$h = 1$		$h = 4$		$h = 1$		$h = 4$	
Naive	1027.340		267.688		-4.937		-4.286	
volFinancial	29.027	(18.409)	14.738	(6.090)	0.089	(0.024)	0.070	(0.031)
volMarket	20.677	(7.345)	0.133	(2.173)	0.085	(0.018)	0.066	(0.024)
logDP	-6.948	(12.367)	-4.501	(21.850)	-1.020	(0.857)	-1.818	(1.238)
logDY	-13.872	(12.495)	-3.437	(23.681)	-0.396	(0.316)	-0.210	(0.155)
logEP	-4.782	(18.684)	0.798	(19.453)	-0.026	(0.024)	-0.163	(0.156)
logDE	99.862	(47.888)	11.920	(7.050)	-0.321	(0.270)	-1.074	(0.641)
BM	-4.196	(8.708)	22.188	(24.093)	-0.040	(0.034)	-0.049	(0.066)
NTIS	23.591	(18.907)	29.088	(31.353)	0.008	(0.019)	-0.067	(0.098)
TBL	15.055	(9.369)	20.070	(13.461)	-0.165	(0.125)	-0.271	(0.236)
LTY	26.378	(18.776)	33.346	(14.139)	-0.041	(0.016)	-0.508	(0.325)
LTR	25.410	(37.005)	7.795	(9.025)	-0.025	(0.020)	-0.066	(0.065)
TMS	16.970	(7.648)	11.734	(11.428)	-0.006	(0.012)	-0.134	(0.124)
DFY	21.412	(9.562)	2.533	(5.346)	0.076	(0.032)	0.047	(0.036)
DFR	33.348	(9.267)	10.032	(6.764)	-0.029	(0.007)	-0.223	(0.197)
INFL	-0.120	(1.710)	-3.253	(4.985)	-0.004	(0.015)	-0.008	(0.014)
Aggregation	1.628	(2.952)	-19.731	(14.193)	0.075	(0.020)	0.075	(0.030)

Table 2: **Diebold-Mariano Test: rMarket rolling**

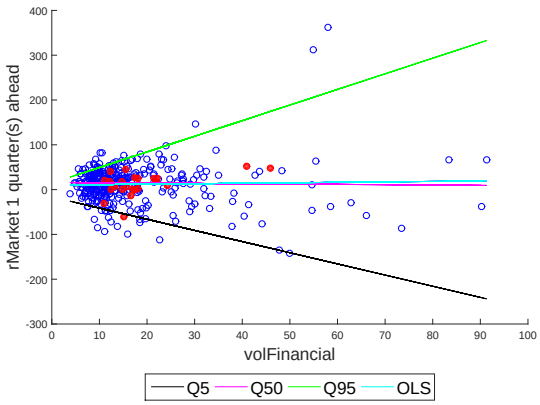
The table below compares the out-of-sample market return performance of each predictor for point forecasts (left panel) and density forecasts (right panel) estimated with a recursive (expanding) window. In the left panel, the first row shows the mean-squared forecast error for the naive, or unconditional, model. The following rows show the MSFE differential over the naive model of conditioning on that predictor, with HAC standard errors shown in parentheses. In the right panel, the first row shows the sum of log scores for the naive, or unconditonal forecast. The following rows show the log score gain (or loss) from conditioning on that predictor, with HAC standard errors shown in parentheses. Note that for point forecasts, a negative differential indicates a more accurate forecast, while for density forecasts, a positive differential indicates a more accurate forecast.

	Point				Density			
	$h = 1$		$h = 4$		$h = 1$		$h = 4$	
Naive	1042.598		280.792		-4.922		-4.272	
volFinancial	13.769	(21.513)	1.634	(7.695)	0.074	(0.026)	0.055	(0.052)
volMarket	5.419	(11.930)	-12.971	(4.697)	0.069	(0.023)	0.052	(0.040)
logDP	-22.206	(13.184)	-17.606	(25.077)	-1.035	(0.861)	-1.832	(1.232)
logDY	-29.130	(13.108)	-16.541	(26.746)	-0.411	(0.323)	-0.224	(0.169)
logEP	-20.040	(21.064)	-12.306	(21.601)	-0.041	(0.041)	-0.177	(0.163)
logDE	84.604	(48.854)	-1.184	(5.624)	-0.336	(0.251)	-1.088	(0.606)
BM	-19.454	(10.433)	9.084	(27.086)	-0.055	(0.048)	-0.064	(0.094)
NTIS	8.333	(24.316)	15.983	(31.242)	-0.007	(0.031)	-0.082	(0.096)
TBL	-0.203	(11.768)	6.966	(15.165)	-0.180	(0.128)	-0.285	(0.236)
LTY	11.120	(18.985)	20.242	(14.256)	-0.056	(0.034)	-0.523	(0.294)
LTR	10.152	(40.029)	-5.309	(10.860)	-0.040	(0.037)	-0.081	(0.079)
TMS	1.712	(13.080)	-1.370	(12.984)	-0.021	(0.026)	-0.148	(0.119)
DFY	6.153	(13.206)	-10.572	(6.216)	0.061	(0.027)	0.032	(0.053)
DFR	18.090	(13.335)	-3.073	(8.686)	-0.044	(0.033)	-0.237	(0.188)
INFL	-15.378	(9.368)	-16.357	(6.758)	-0.020	(0.023)	-0.023	(0.046)
Aggregation	-51.671	(31.894)	-4.528	(25.534)	0.056	(0.039)	0.118	(0.071)

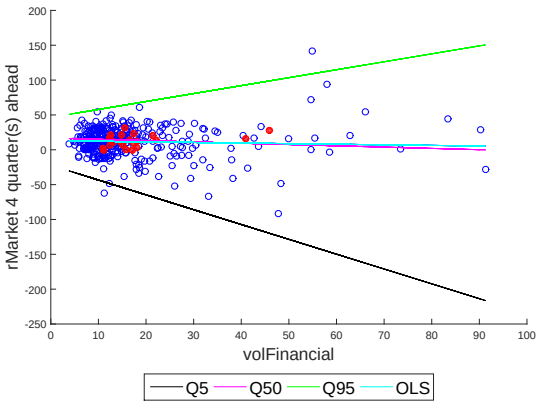
Figure 1: **rMarket vs. volFinancial**

The top panels in this figure show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on financial sector realized volatility. Four quarter ahead returns are an average of four quarterly observations. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

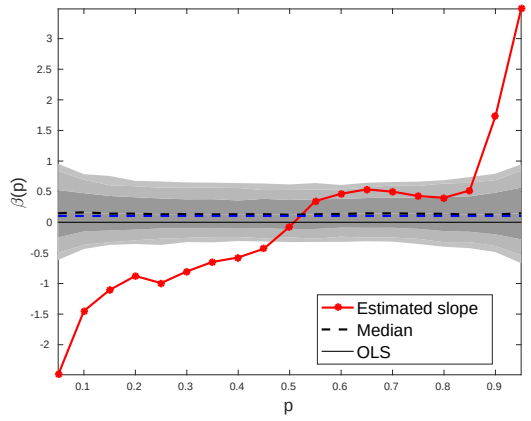
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

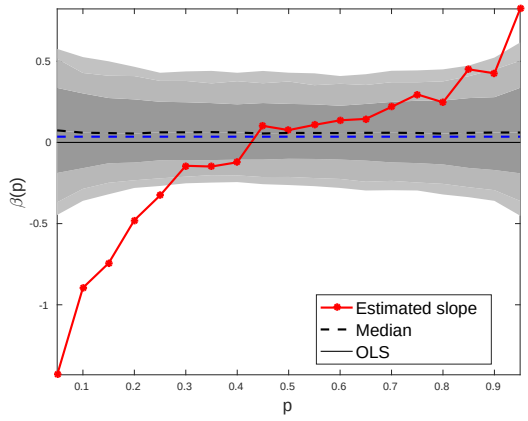
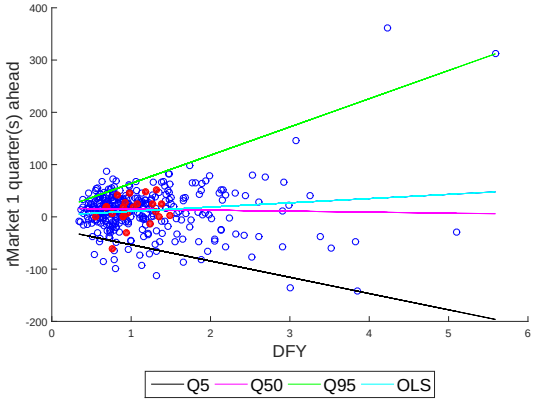


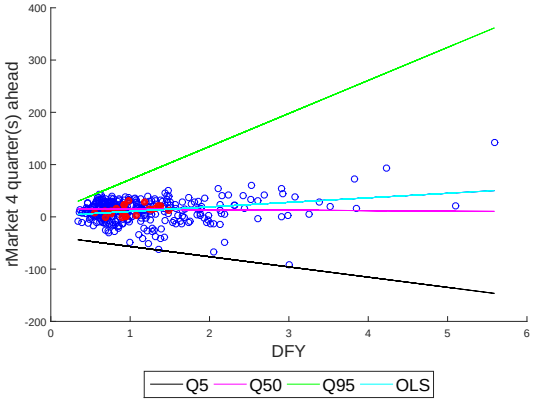
Figure 2: **rMarket vs. DFY**

The top panels in this figure show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on the default yield spread. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

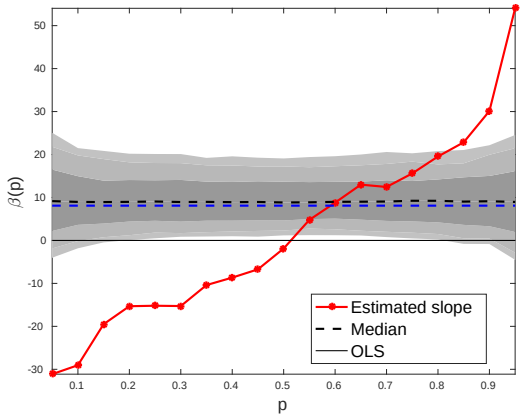
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

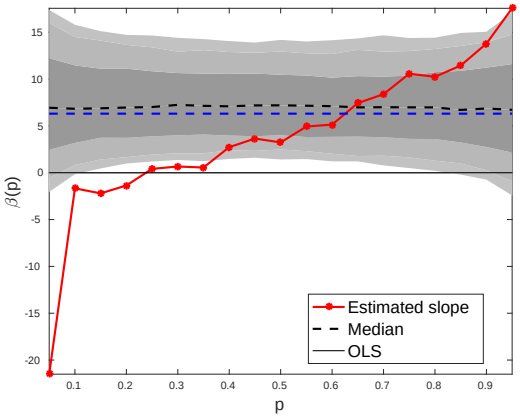
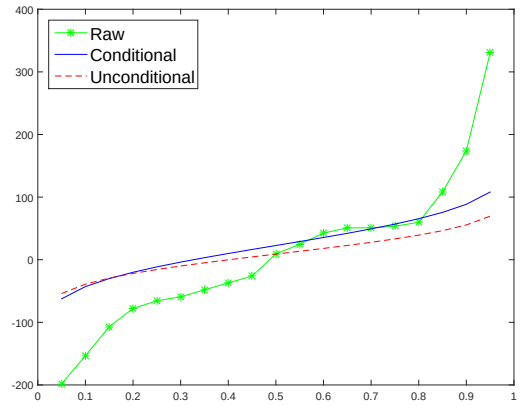


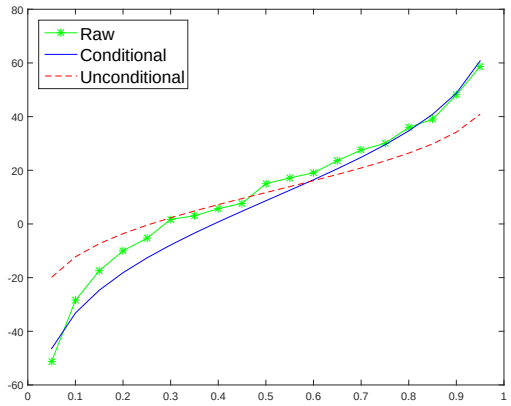
Figure 3: **Example of Fitted Distributions on 2008Q4: volFinancial**

The top panels in this figure show the conditional quantiles and fitted skewed t-inverse cumulative distribution functions of one quarter ahead (left) and four quarter ahead (right) market returns predicted by realized financial volatility in 2008Q4. The bottom panels show the corresponding density functions.

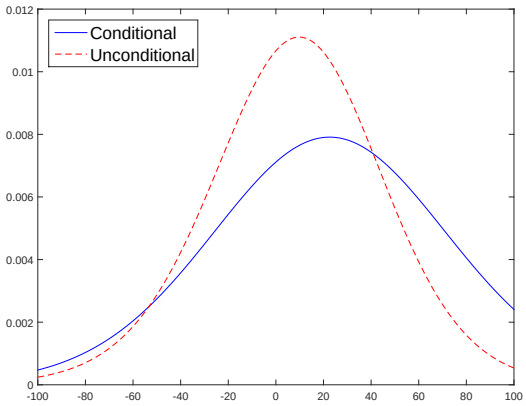
(a) **One quarter ahead**:
Conditional quantiles and skewed t-inverse CDF



(b) **Four quarters ahead**:
Conditional quantiles and skewed t-inverse CDF



(c) **One quarter ahead**:
Probability Densities



(d) **Four quarters ahead**:
Probability Densities

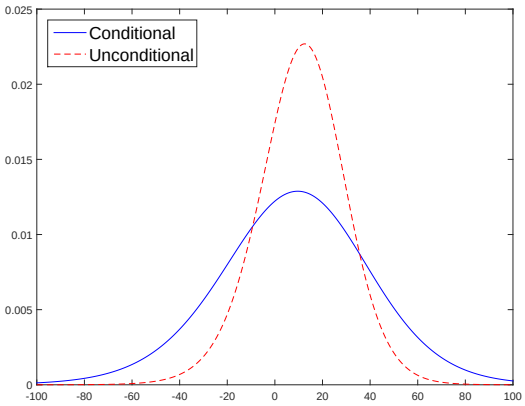
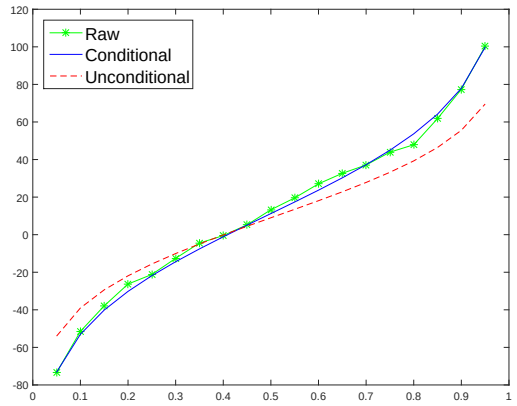


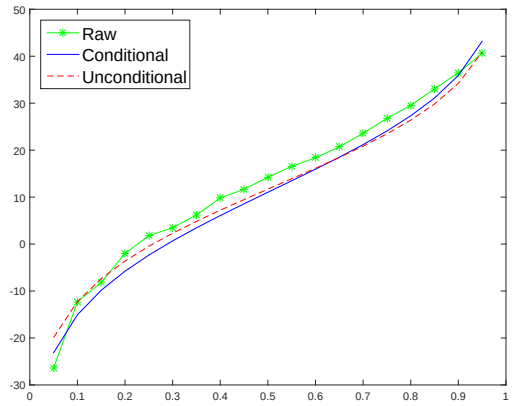
Figure 4: **Example of Fitted Distributions on 2008Q4: DFY**

The top panels in this figure show the conditional quantiles and fitted skewed t-inverse cumulative distribution functions of one quarter ahead (left) and four quarter ahead (right) market returns predicted by the default yield spread in 2008Q4. The bottom panels show the corresponding density functions.

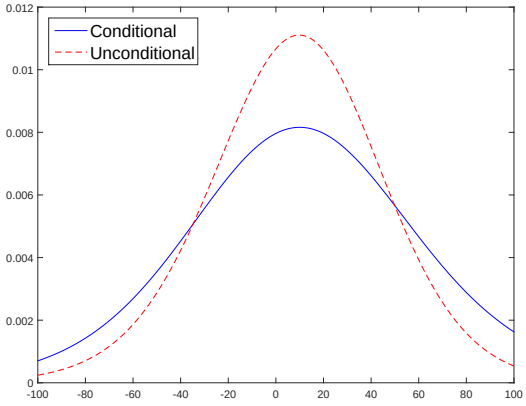
(a) **One quarter ahead**:
Conditional quantiles and skewed t-inverse CDF



(b) **Four quarters ahead**:
Conditional quantiles and skewed t-inverse CDF



(c) **One quarter ahead**:
Probability Densities



(d) **Four quarters ahead**:
Probability Densities

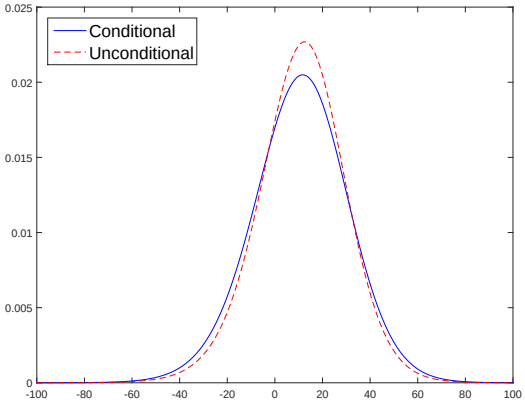
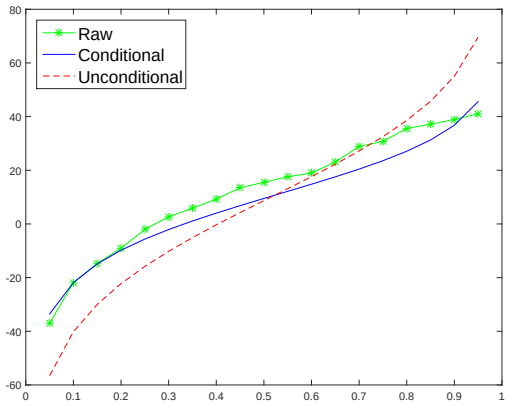


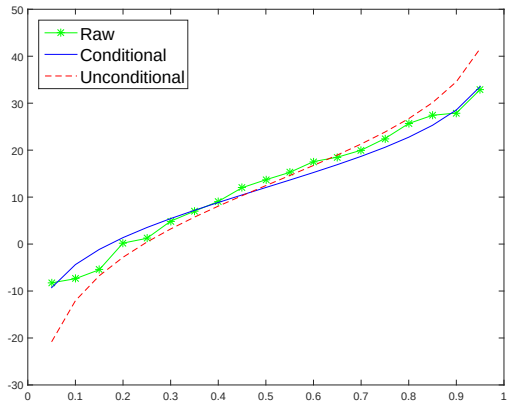
Figure 5: **Example of Fitted Distributions on 1993Q4: volFinancial**

The top panels in this figure show the conditional quantiles and fitted skewed t-inverse cumulative distribution functions of one quarter ahead (left) and four quarter ahead (right) market returns predicted by realized financial volatility in 1993Q4. The bottom panels show the corresponding density functions.

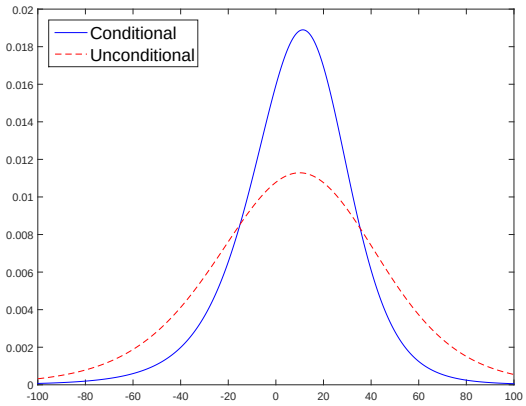
(a) **One quarter ahead**:
Conditional quantiles and skewed t-inverse CDF



(b) **Four quarters ahead**:
Conditional quantiles and skewed t-inverse CDF



(c) **One quarter ahead**:
Probability Densities



(d) **Four quarters ahead**:
Probability Densities

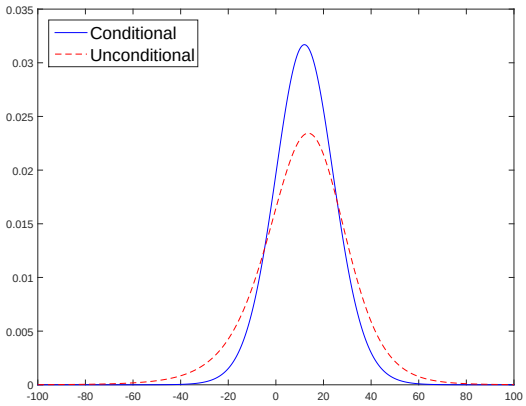
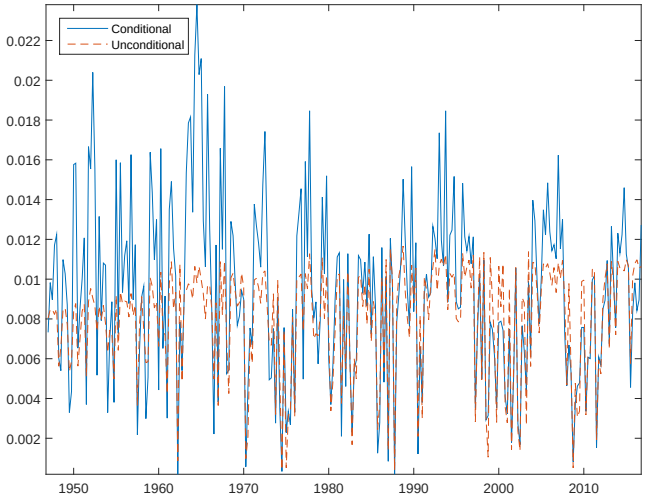


Figure 6: **Out-of-sample Scores, volFinancial**

The figure compares the out-of-sample predictive scores of the predicted distribution conditional on realized financial sector volatility (volFinancial) with the unconditional distribution. Scores are shown before log-transformation

(a) **One quarter ahead**



(b) **Four quarters ahead**

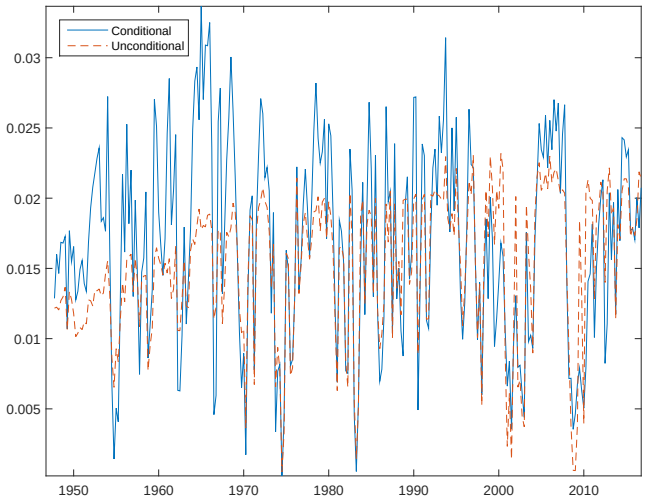
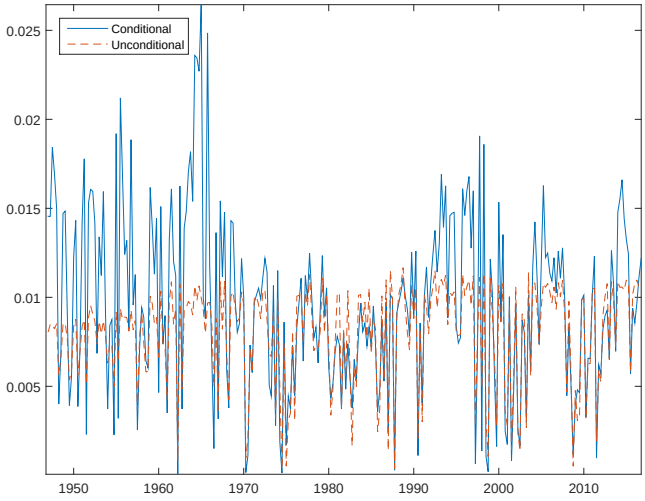


Figure 7: **Out-of-sample Scores, DFY**

The figure compares the out-of-sample predictive scores of the predicted distribution conditional on the default yield spread (DFY) with the unconditional distribution. Scores are shown before log-transformation

(a) **One quarter ahead**



(b) **Four quarters ahead**

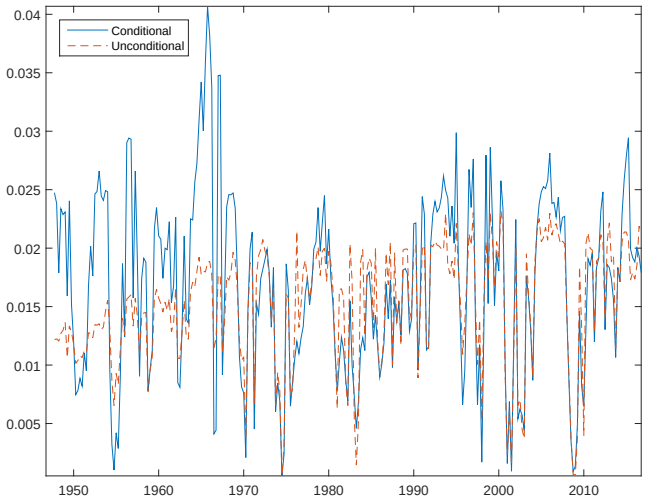
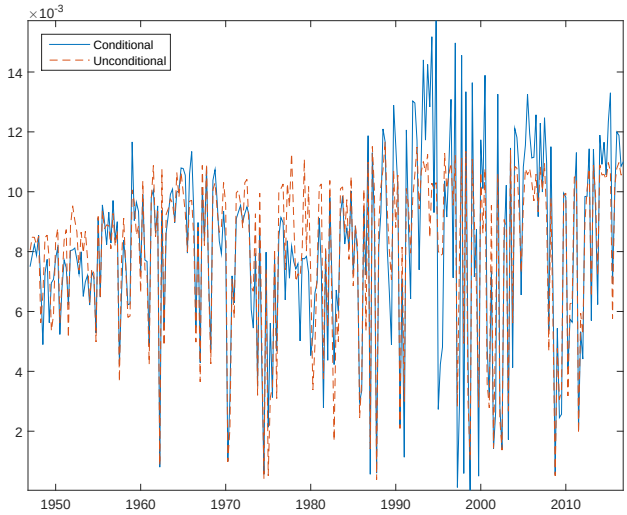


Figure 8: **Out-of-sample Scores, BM**

The figure compares the out-of-sample predictive scores of the predicted distribution conditional on the aggregate boolk-to-market ratio (BM) with the unconditional distribution. Scores are shown before log-transformation

(a) **One quarter ahead**



(b) **Four quarters ahead**

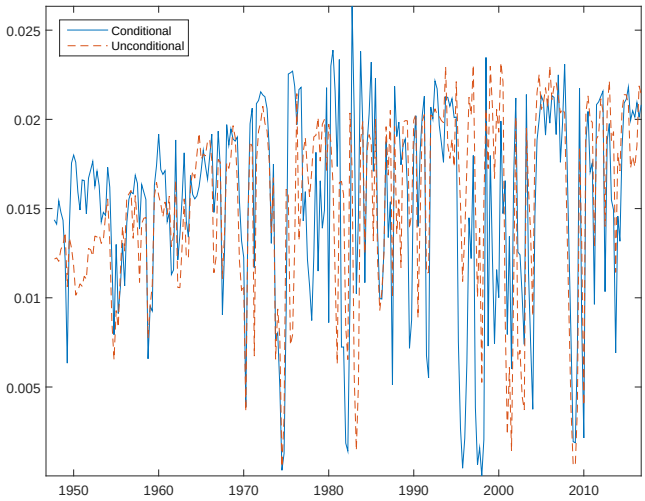
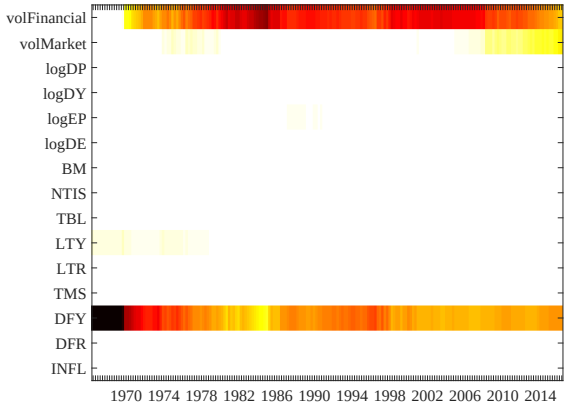


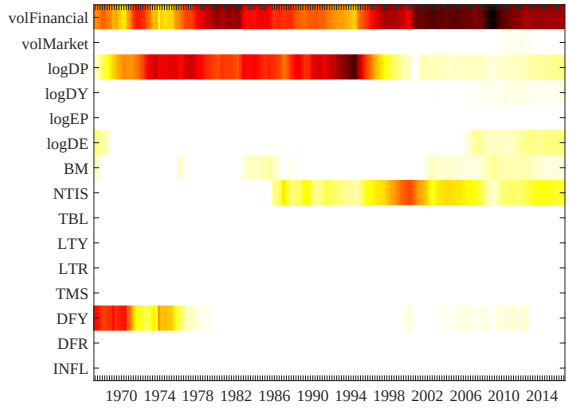
Figure 9: **Aggregation Weights**

The top panels of the figure show the weights, in time series, assigned to each of the forecasting variables when computing the optimal combination of forecasting densities. The bottom panels show the weights assigned when computing the optimal combination of *point* forecasts.

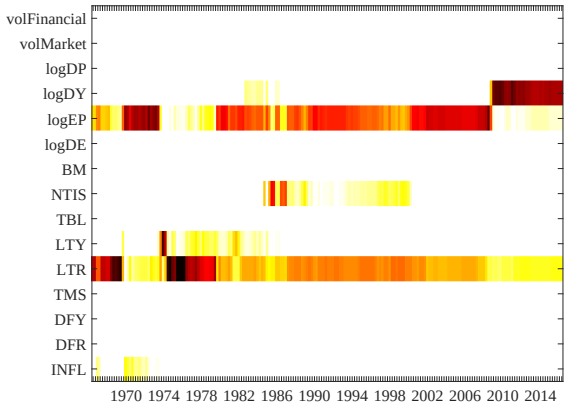
(a) **One quarter ahead**:
Density forecast weights



(b) **Four quarters ahead**:
Density forecast weights



(c) **One quarter ahead**:
Point forecast weights



(d) **Four quarters ahead**:
Point forecast weights

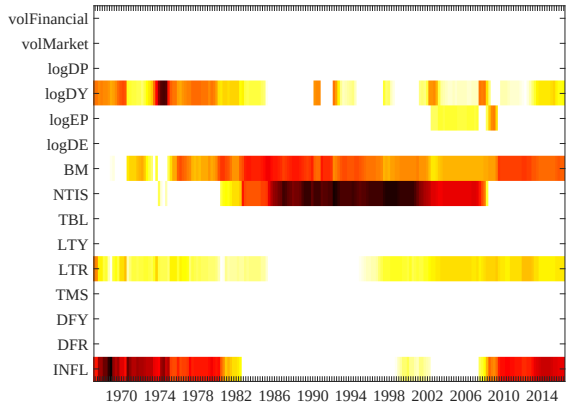
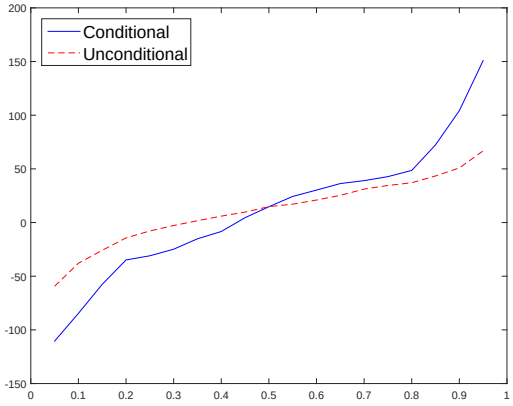


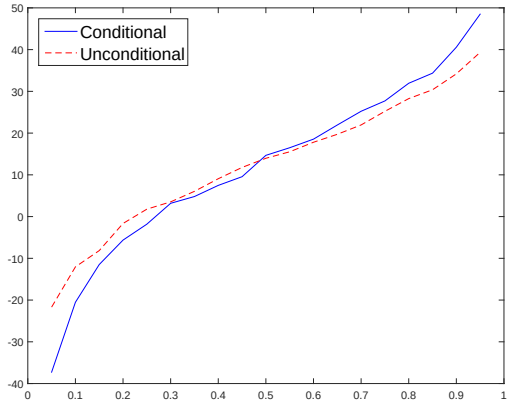
Figure 10: **Example of Fitted Distributions on 2008Q4: Optimal Forecast**

The top panels in this figure show the conditional quantiles and fitted skewed t-inverse cumulative distribution functions of one quarter ahead (left) and four quarter ahead (right) market returns predicted by the optimal combination of forecasting variables in 2008Q4. The bottom panels show the corresponding density functions.

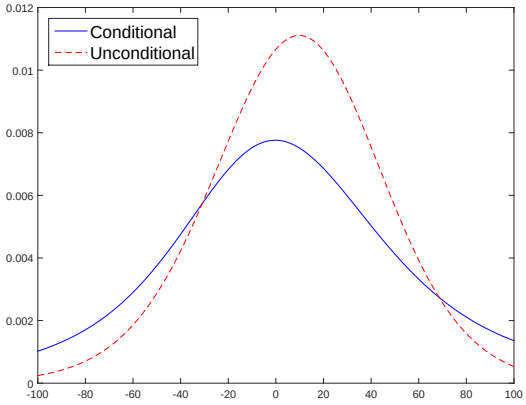
(a) **One quarter ahead**:
Conditional quantiles and skewed t-inverse CDF



(b) **Four quarters ahead**:
Conditional quantiles and skewed t-inverse CDF



(c) **One quarter ahead**:
Probability Densities



(d) **Four quarters ahead**:
Probability Densities

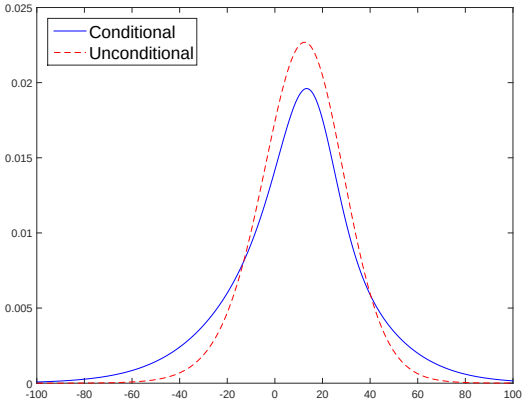
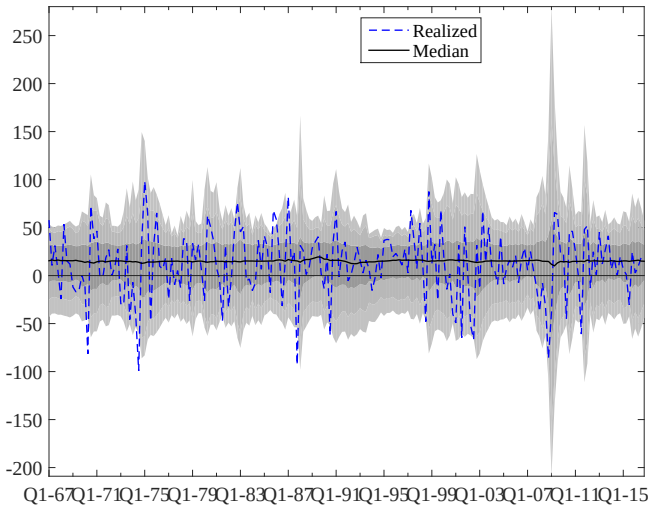


Figure 11: **Predicted Distributions**

The figure shows the time series of the predicted distribution of one quarter ahead (top) and four quarter ahead (bottom) market retuns using the optimal combination of forecasting variables. The conditional 5, 25, 75, and 95 percent quantiles are shaded in gray.

(a) **One quarter ahead**



(b) **Four quarters ahead**

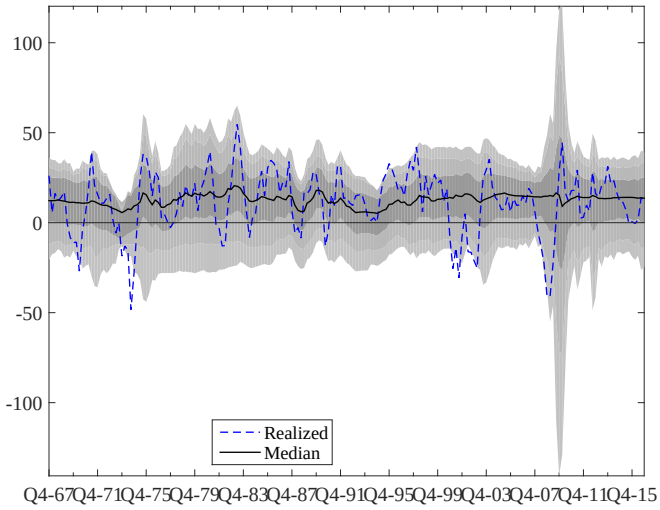
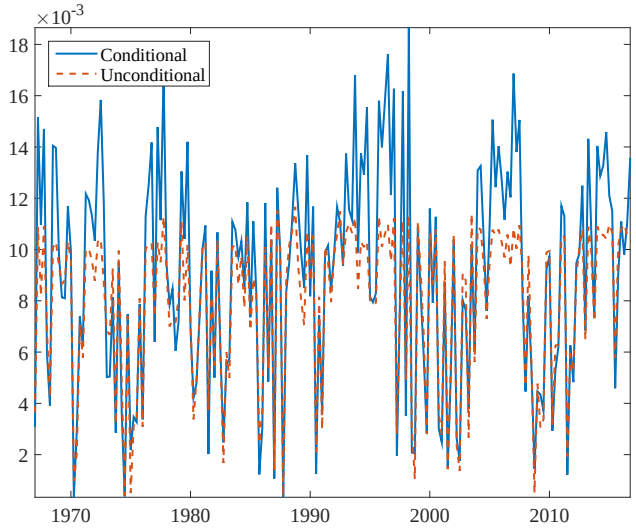


Figure 12: **Out-of-sample Scores, Aggregate Forecast**

The figure compares the out-of-sample predictive scores of the predicted distribution conditional on the optimal combination of predictors with the unconditional distribution.

(a) **One quarter ahead**



(b) **Four quarters ahead**

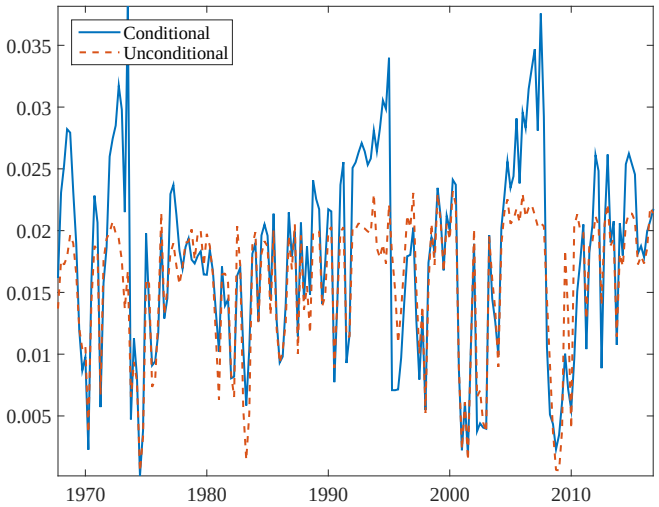
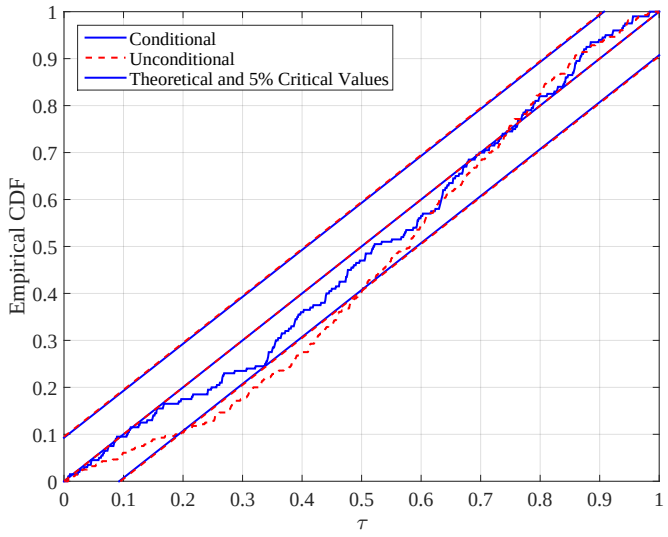


Figure 13: **PIT, Aggregate Forecast**

The figure compares the empirical cumulative distribution of the probability integral transform (PIT) for the optimal combination of predictors with the unconditional distribution.

(a) **One quarter ahead**



(b) **Four quarters ahead**

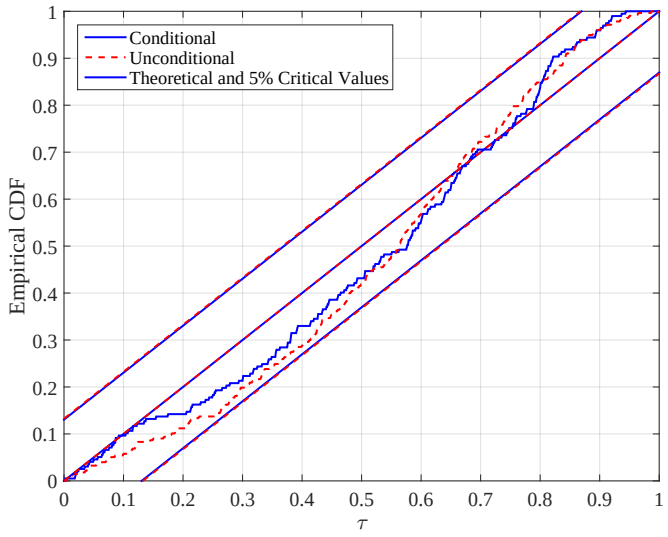


Table 3: Realised Volatility - a Differences-in-Differences-in-Differences Analysis

This table estimates equation 11. Observations are at the quarterly, industry x size level. "realised volatility" is quarterly realised volatility times 4. 'Large' indicates firms with assets greater than \$50B as of 2010Q2. 'Bank' is a binary indicator for firms belonging to industry 44 in the Fama-French 48 industry definitions. Standard errors are clustered by industry x size.

	(1) realised volatility
PostDodd-Frank x Large x Bank	-9.10***
	(0.00)
PostDodd-Frank x Large	5.86***
	(0.00)
PostDodd-Frank x Bank	16.35***
	(0.00)
Crisis x Large x Bank	6.91***
	(0.00)
Crisis x Large	9.90***
	(0.00)
Crisis x Bank	32.92***
	(0.00)
Year-quarter FEs	Y
Industry x Size FEs	Y
N	272
R^2	0.929

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

.

Supplementary Appendix for "Changing Risk-Return Profiles"

Richard K. Crump **Domenico Giannone** **Sean Hundtofte**

(Federal Reserve Bank of New York)

6 Data Illustrations

The data for many of our time series of candidate predictors begins in 1926 owing to stock market data availability. A full list of variable names and definitions we use, mostly but not entirely, from Welch and Goyal (2008) is available on page 19.

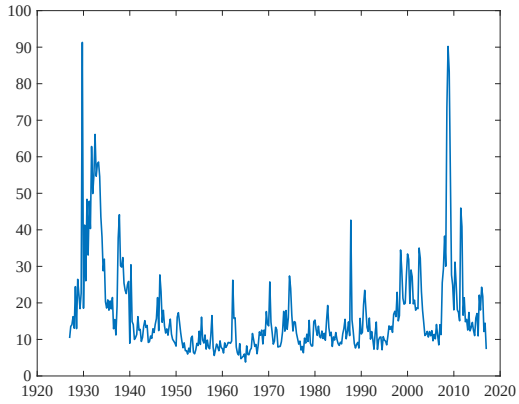
Figure 14 shows the times series of four important variables: realized volatility amongst financial stocks (volFinancial) and the aggregate market (volMarket), BBB-AAA credit spreads (DFY), and log dividend price-ratios (logDP). Figure 15 provides the quantile-quantile (QQ) plots of these predictors relative to market returns over the next year. The definition for the finance industry is taken from the Fama-French 12 industry definitions and refers to stocks associated with the SIC codes between 6000 and 6999.

The QQ-plots of Figure 15 show the empirical quantiles of aggregate stock market returns on the y-axis against the empirical quantiles of a sample of predictor variables on the x-axis. Examining these plots for deviations from linear relationships, volatility and DFY exhibit very pronounced nonlinearity with market returns compared to logDP. A nonlinearity, or systematic departure from the 45-degree line, indicates a difference in the conditional distribution function and foreshadows the findings of which variables are most useful indicators of stock market vulnerability. The highest realized stock market returns in the sample occur during the recovery from the Great Depression (1933) - if these quarters are excluded, the positive return outliers to the top right of each QQ-plot disappear.

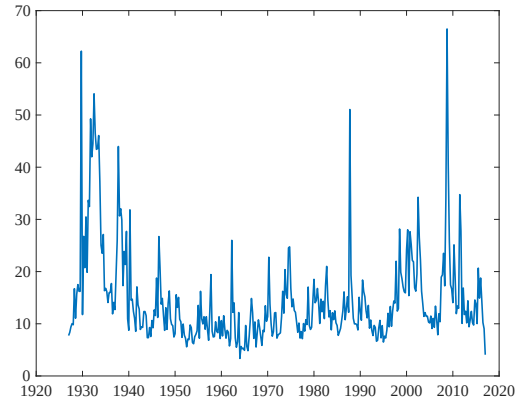
Figure 14: **Raw Data**

These figures show the time series of realized volatility of banking stocks ("volFinancial"), of the aggregate market ("volMarket"), the BBB-AAA credit spread ("DFY"), and the log of the dividend-price ratio ("logDP")

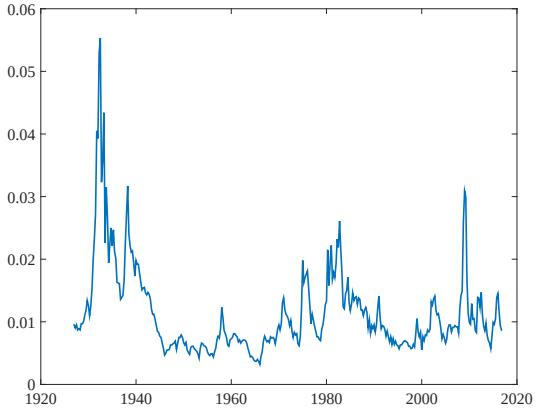
(a) **volFinancial**



(b) **volMarket**



(c) **DFY**



(d) **logDP**

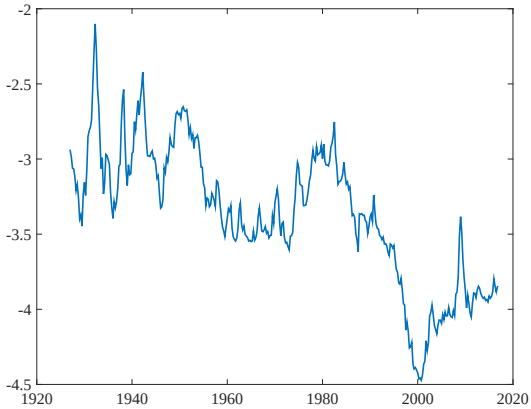
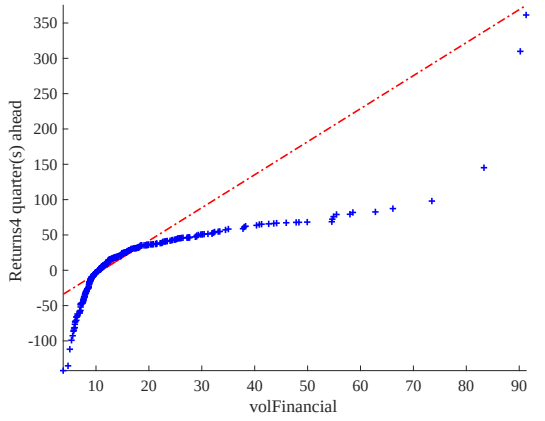


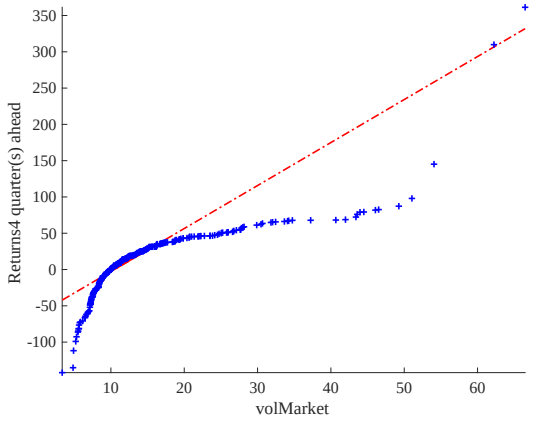
Figure 15: **Quantile-Quantile Plots**

These panels show the sample quantile-quantile relationship between 4 quarter horizon returns and various candidate predictors for the time series 1926-2017

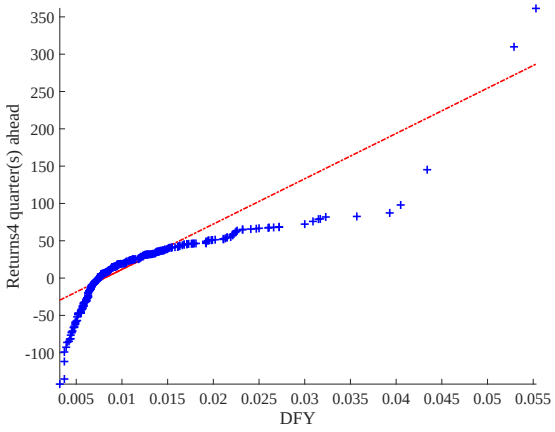
(a) **Market returns - volFinancial (4Q)**



(b) **Market returns - volMarket (4Q)**



(c) **Market returns - DFY (4Q)**



(d) **Market returns - logDP (4Q)**

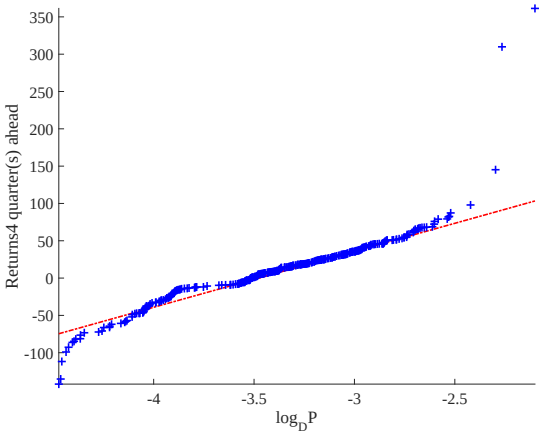
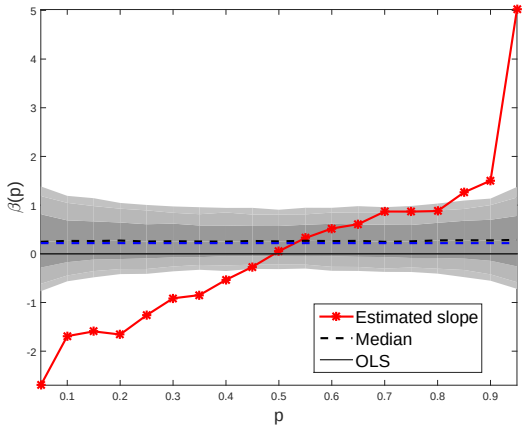


Figure 16: **rFinancial vs. volFinancial**

These panels show the estimated coefficients of quantile regressions with equity returns of the financial industry (as defined by Fama-French 11 industry definitions using SIC codes) as the outcome variable. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter

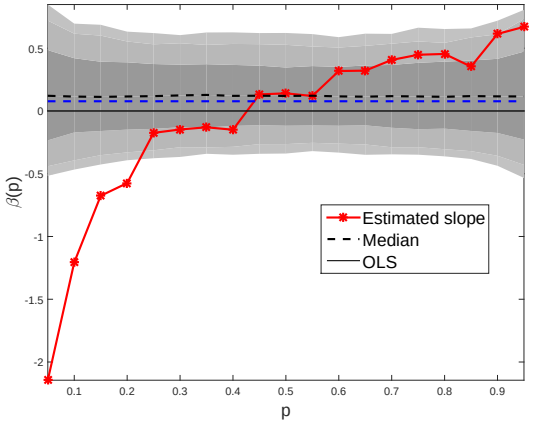
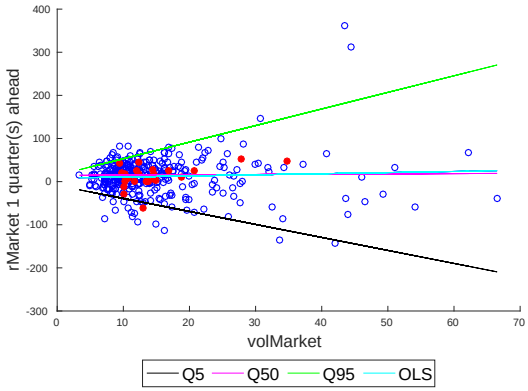


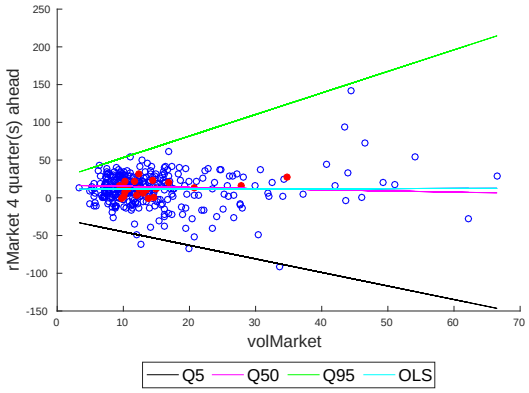
Figure 17: **Out-of-sample results: volMarket**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on volMarket. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

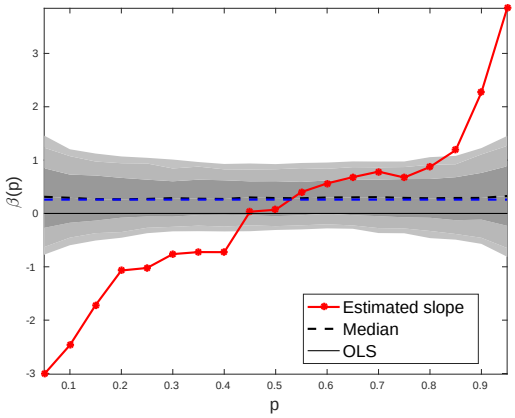
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

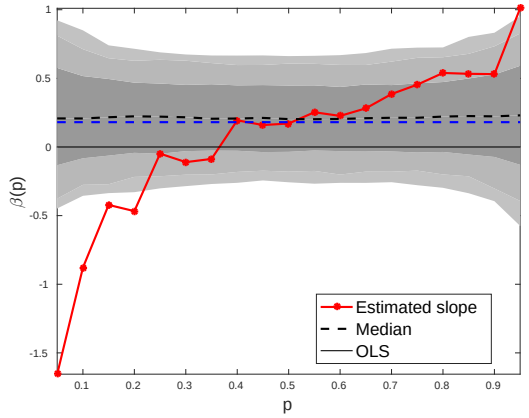
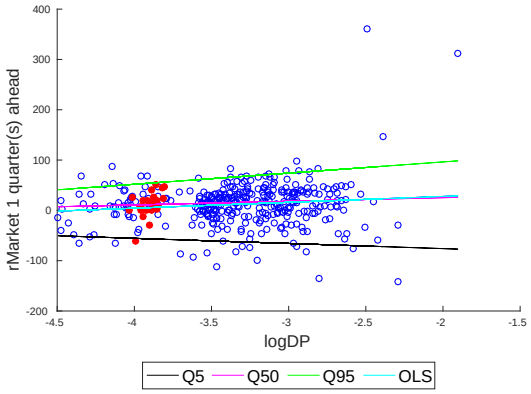


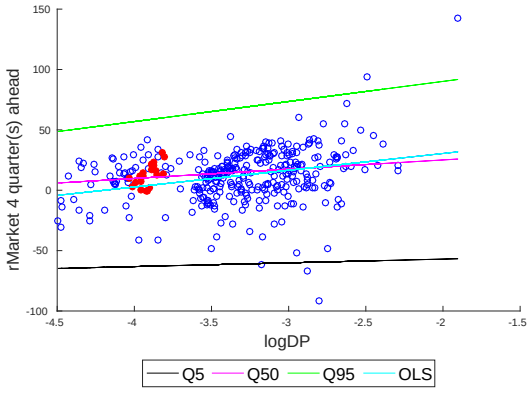
Figure 18: **Out-of-sample results: logDP**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on logDP. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

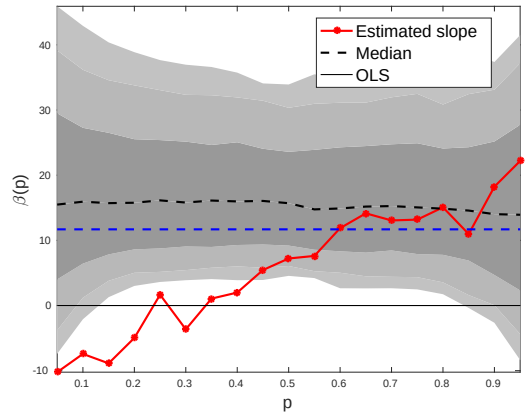
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

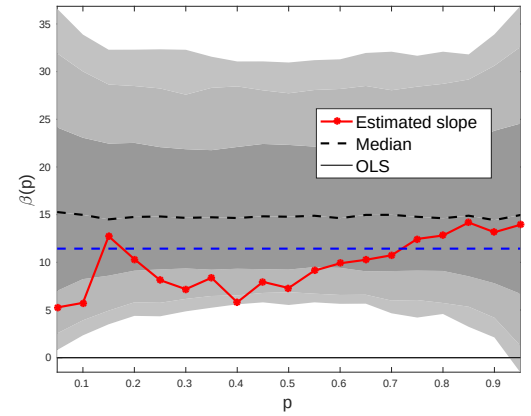
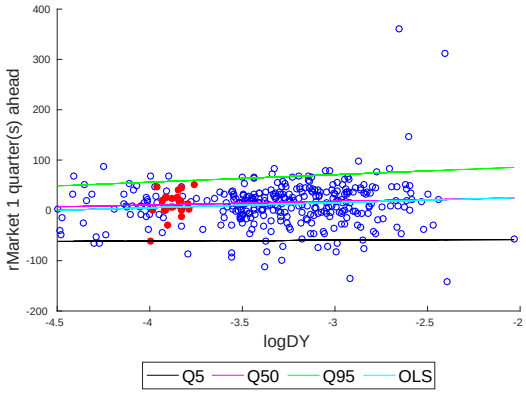


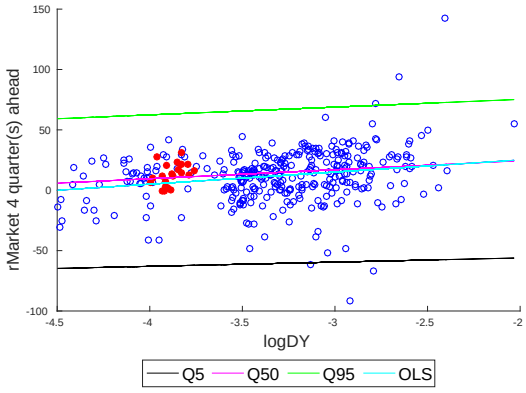
Figure 19: **Out-of-sample results: logDY**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on logDY. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

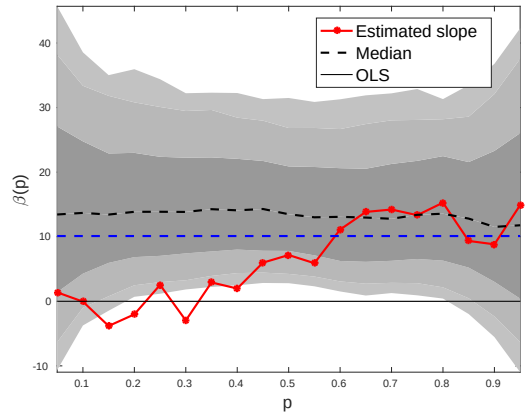
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

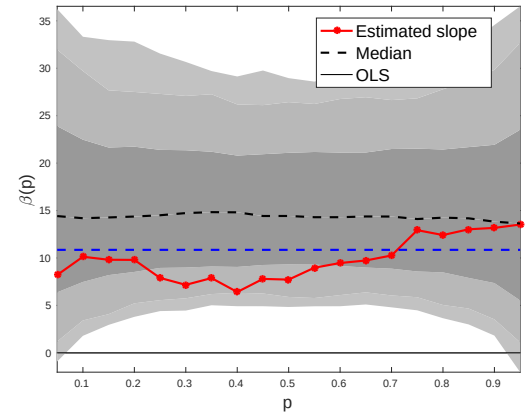
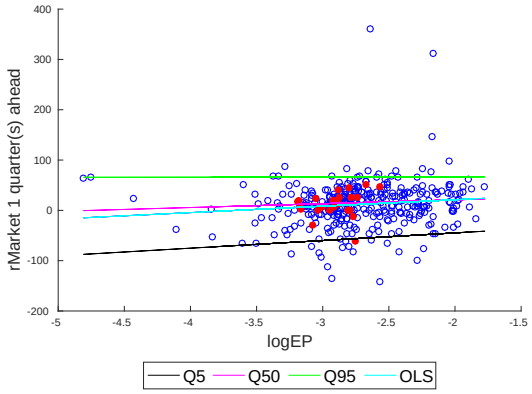


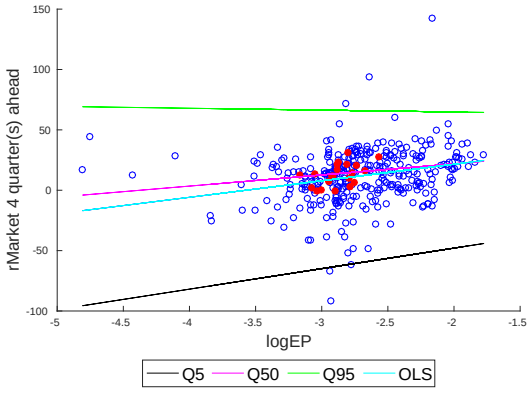
Figure 20: **Out-of-sample results: logEP**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on logEP. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

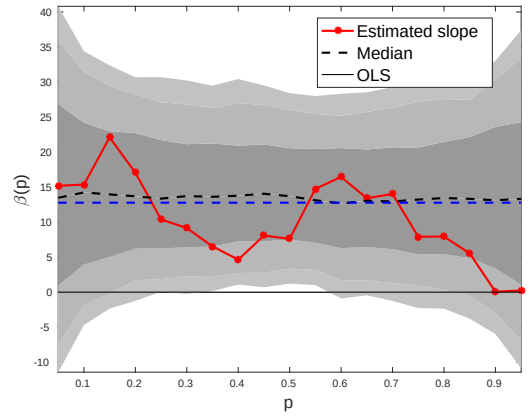
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

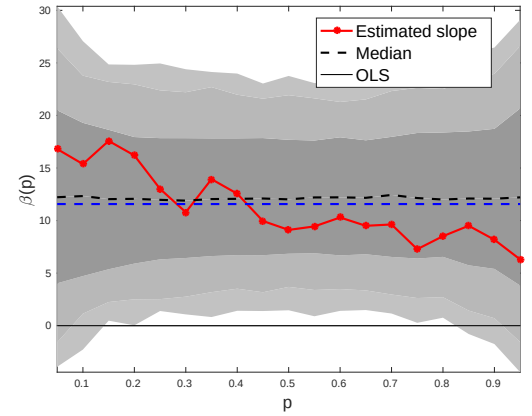
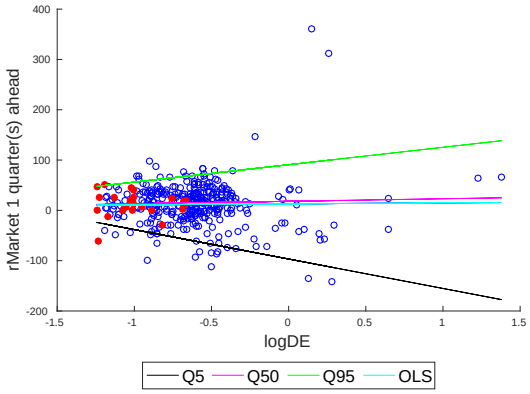


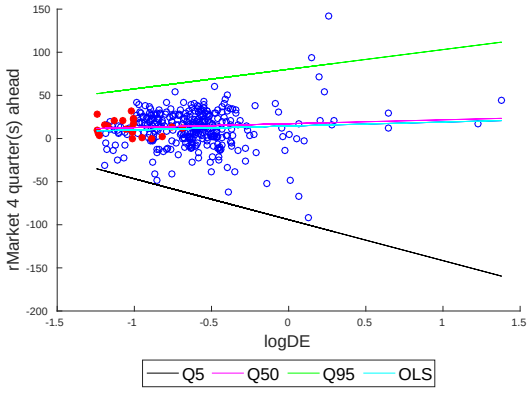
Figure 21: **Out-of-sample results: logDE**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on logDE. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

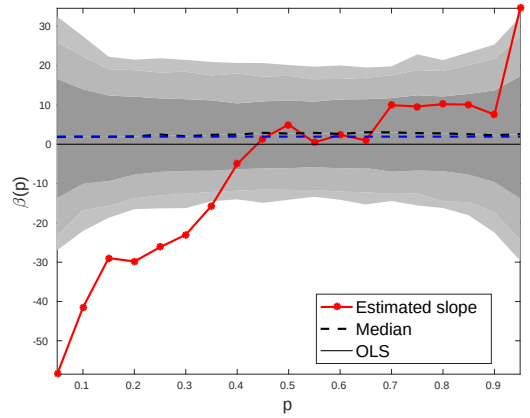
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

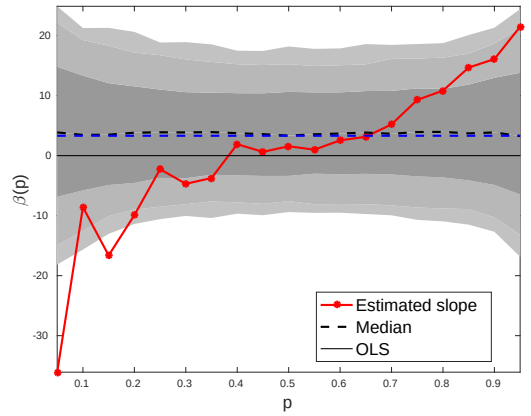
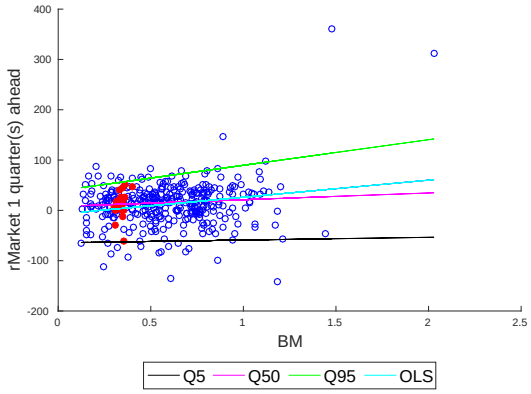


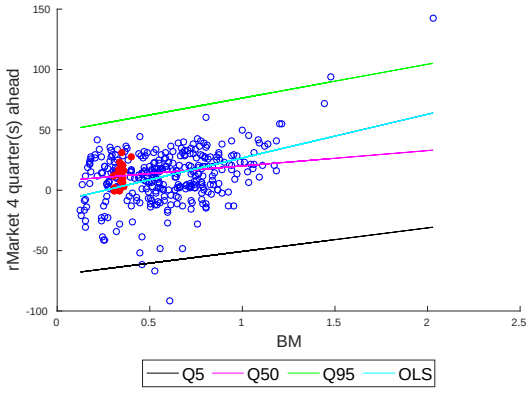
Figure 22: **Out-of-sample results: BM**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on BM. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

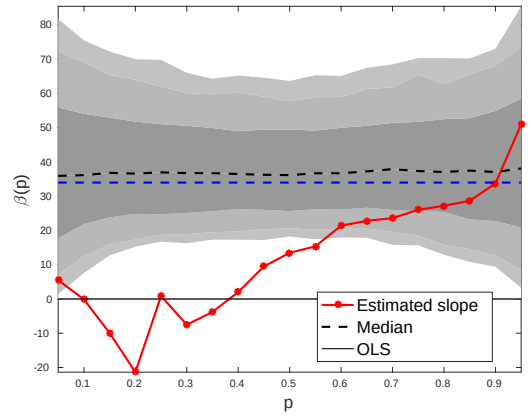
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

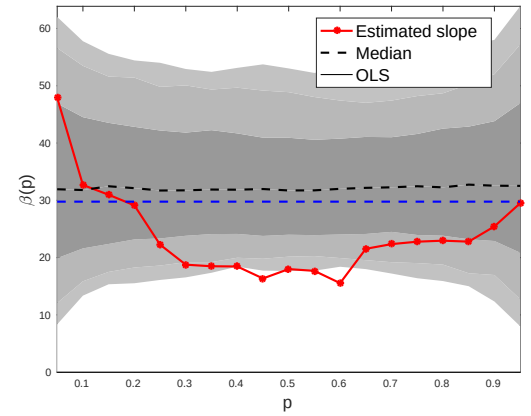
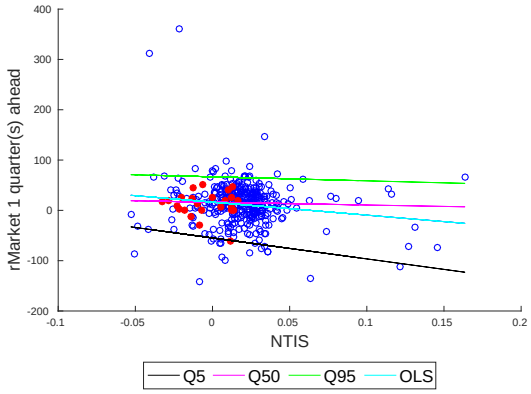


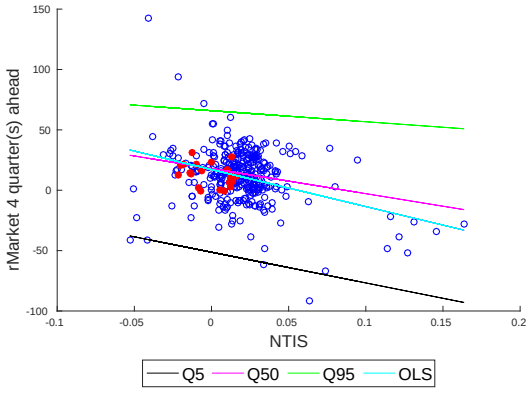
Figure 23: **Out-of-sample results: NTIS**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on NTIS. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

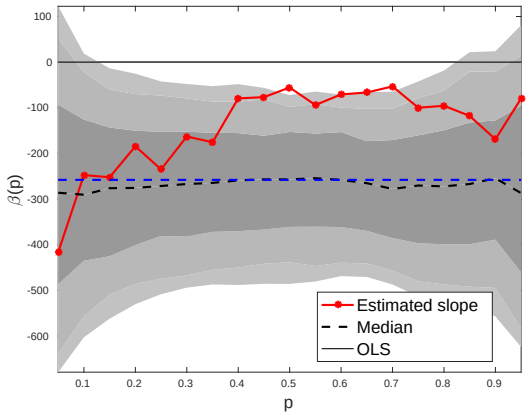
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

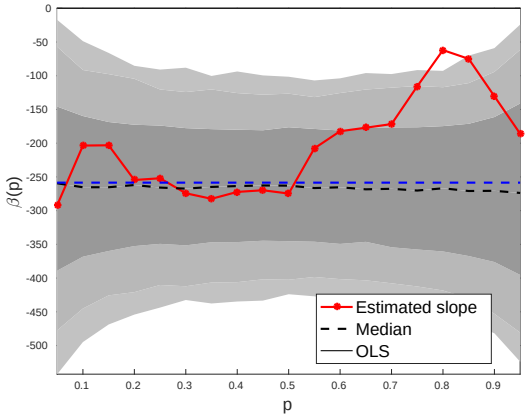
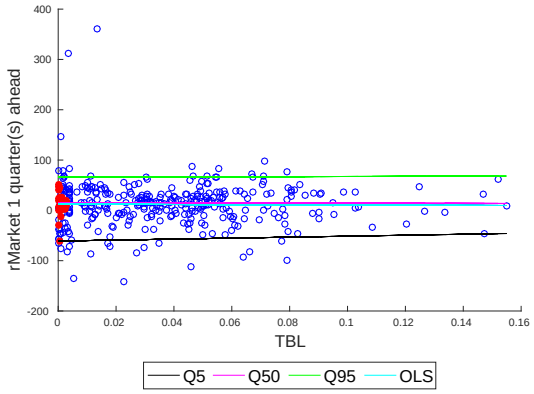


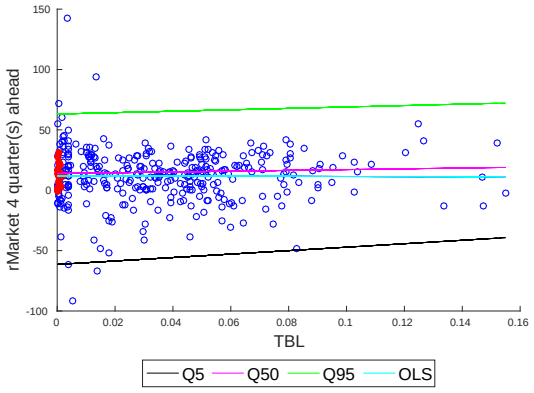
Figure 24: **Out-of-sample results: TBL**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on TBL. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

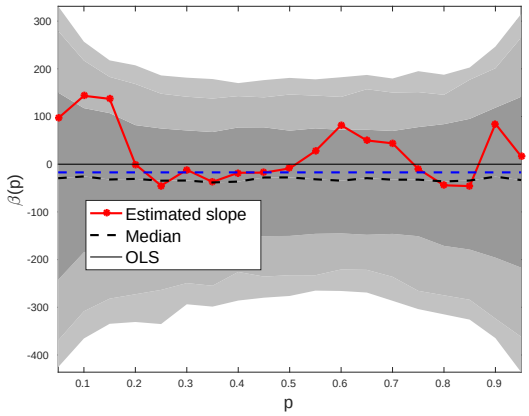
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

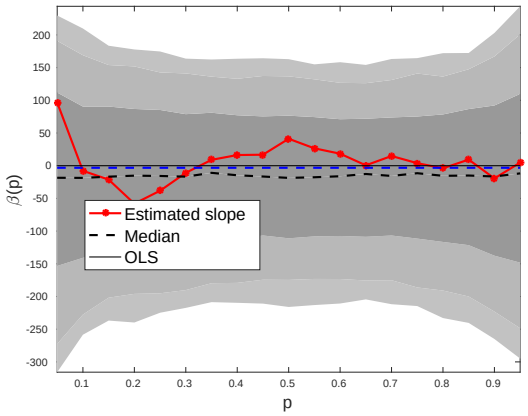
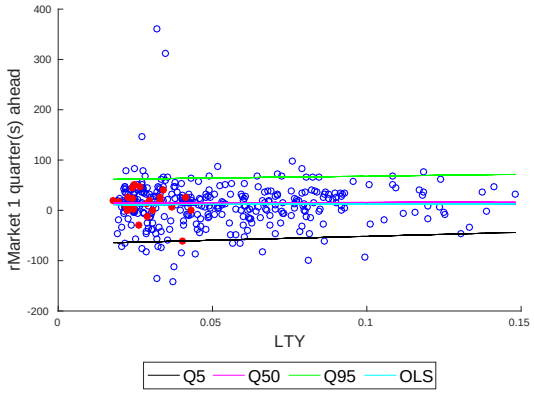


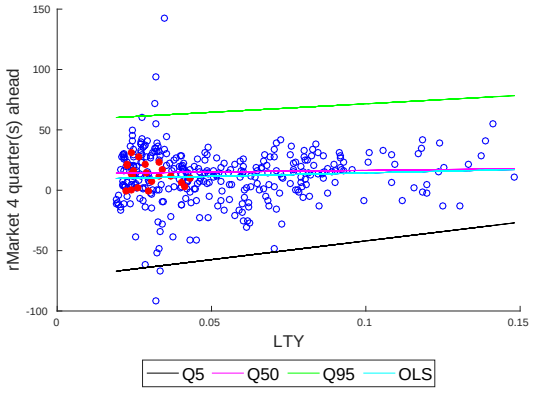
Figure 25: **Out-of-sample results: LTY**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on LTY. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

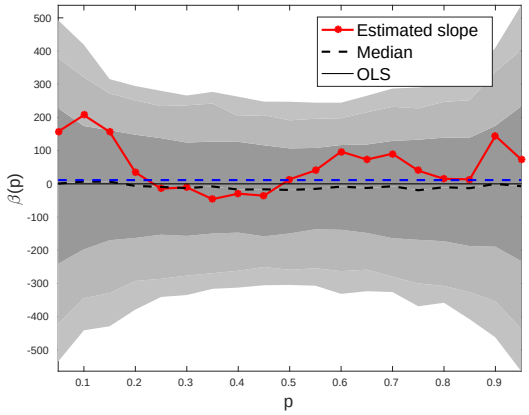
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

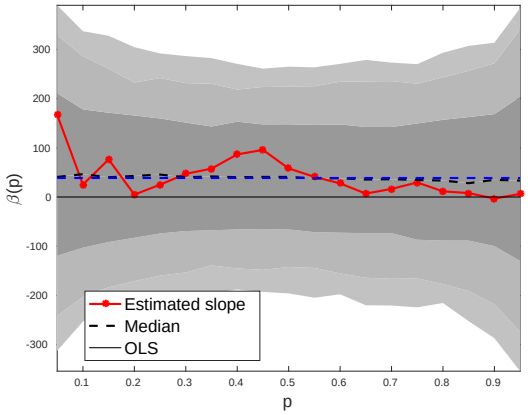
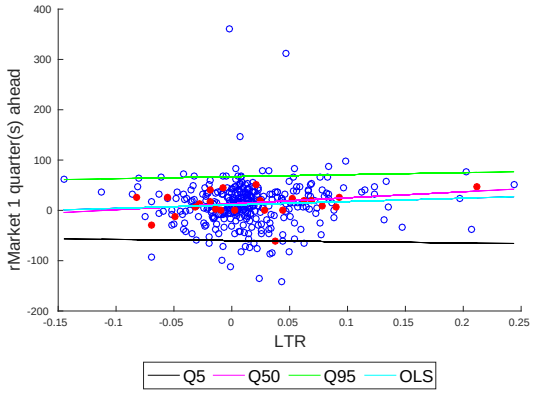


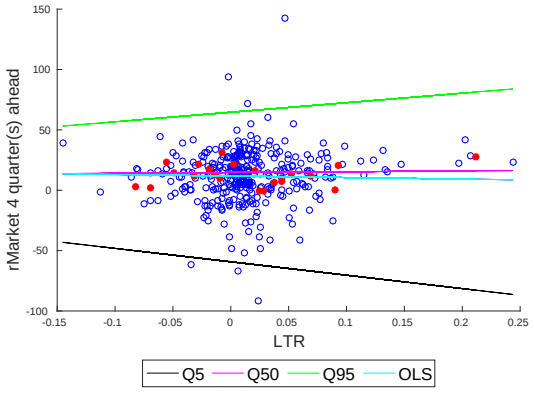
Figure 26: **Out-of-sample results: LTR**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on LTR. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

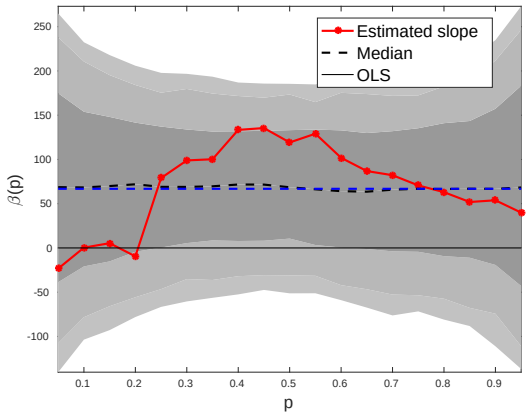
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

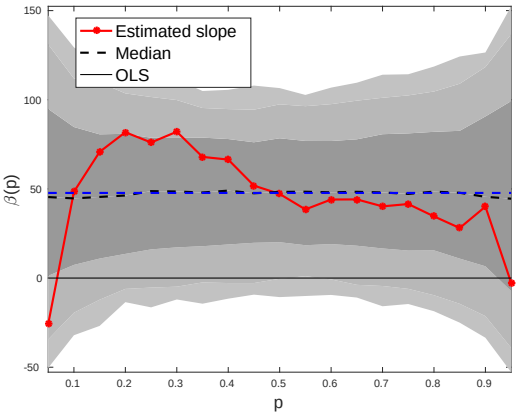
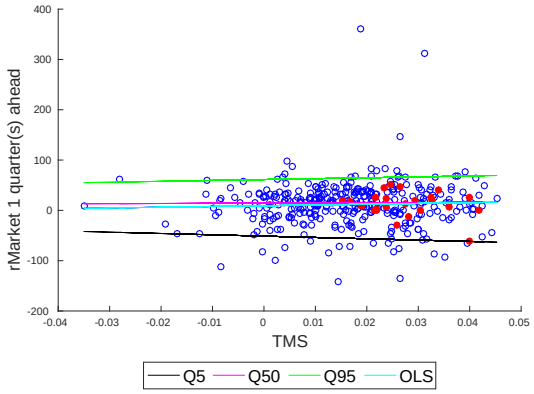


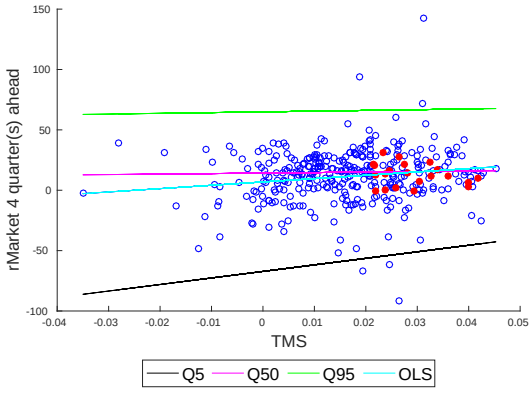
Figure 27: **Out-of-sample results: TMS**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on TMS. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

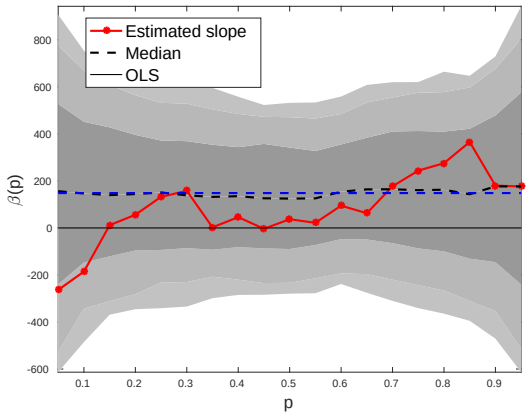
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

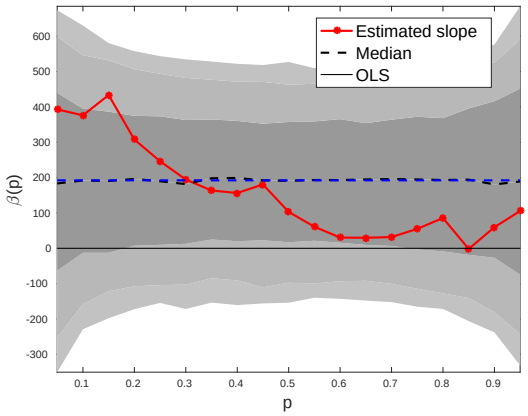
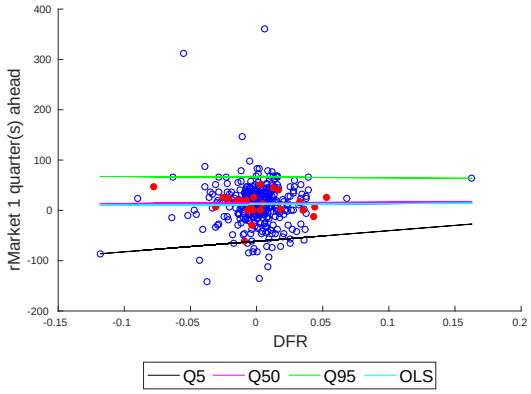


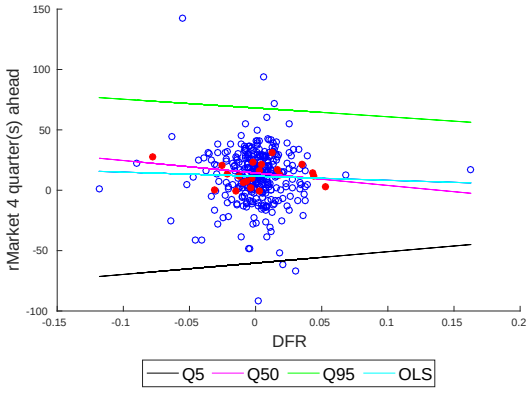
Figure 28: **Out-of-sample results: DFR**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on DFR. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

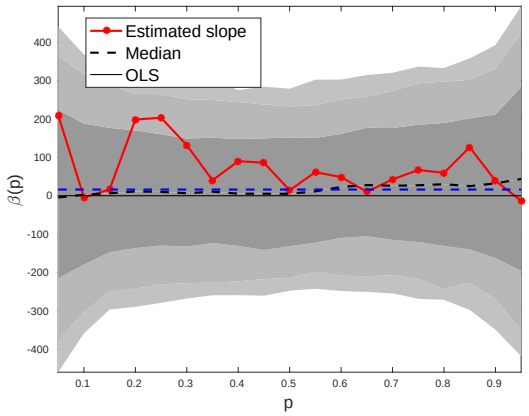
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

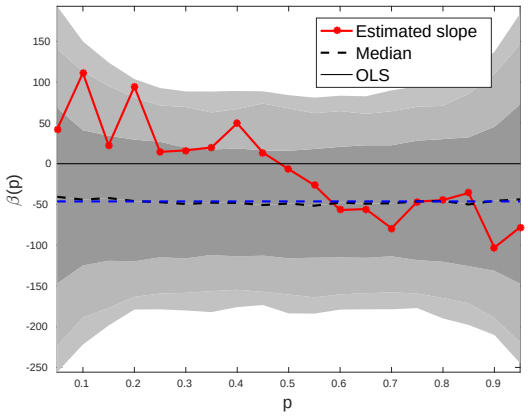
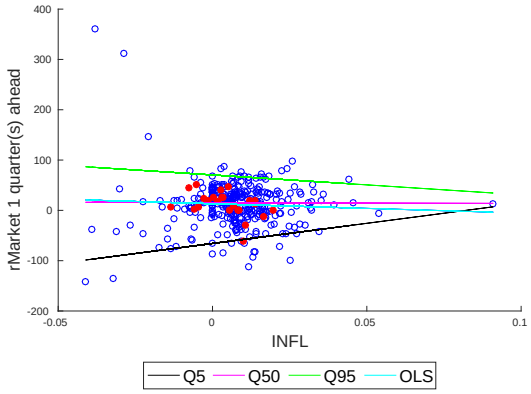


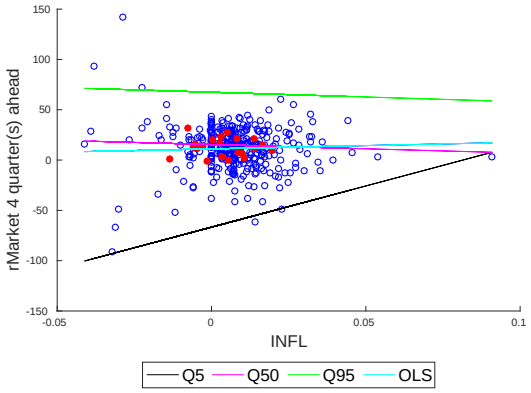
Figure 29: **Out-of-sample results: INFL**

The top panels show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on INFL. Data before 2010Q3 are shown as open blue circles; data after this date are shown as closed and red. The bottom panels show the estimated coefficients of these quantile regressions. We report confidence bounds for the null hypothesis that the true data-generating process is a VAR with 4 lags; bounds are computed using 1000 bootstrapped samples.

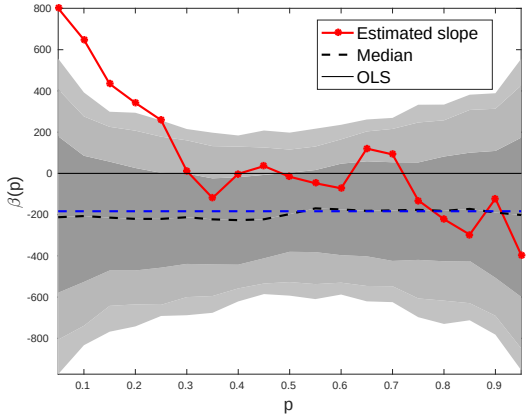
(a) **One quarter ahead**: Scatter



(b) **Four quarters ahead**: Scatter



(c) **One quarter ahead**:
Quantile Regression Coefficients



(d) **Four quarters ahead**:
Quantile Regression Coefficients

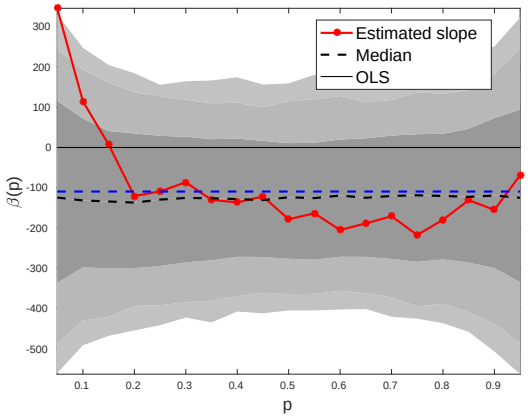
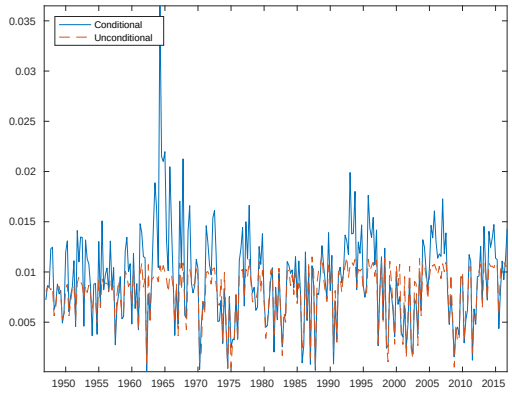


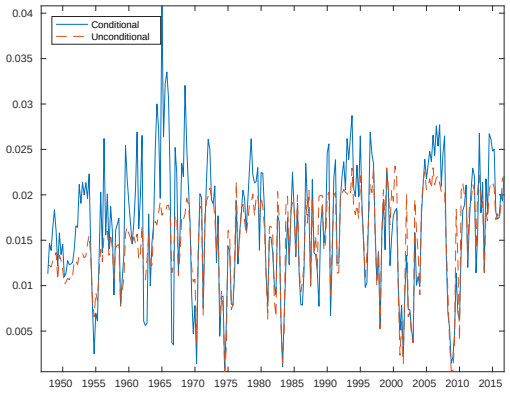
Figure 30: **Out-of-sample results: volMarket**

The top panels compare the out-of-sample scores of the predicted distribution conditional on volMarket with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the volMarket predictor with the unconditional distribution.

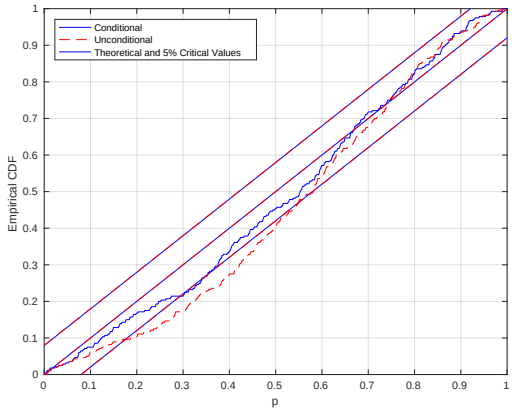
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

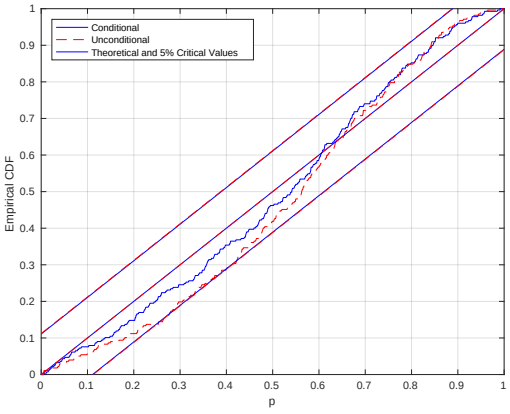
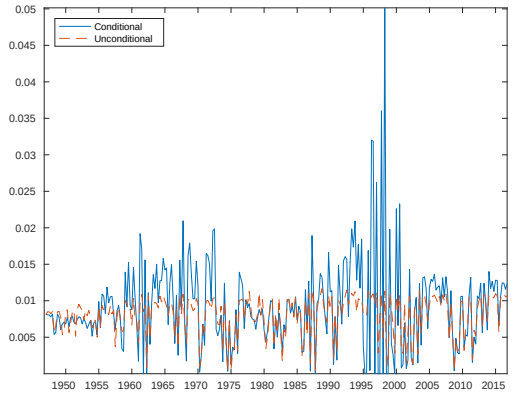


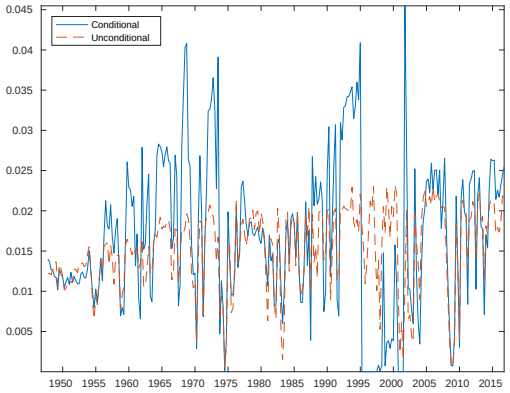
Figure 31: **Out-of-sample results: logDP**

The top panels compare the out-of-sample scores of the predicted distribution conditional on logDP with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the logDP predictor with the unconditional distribution.

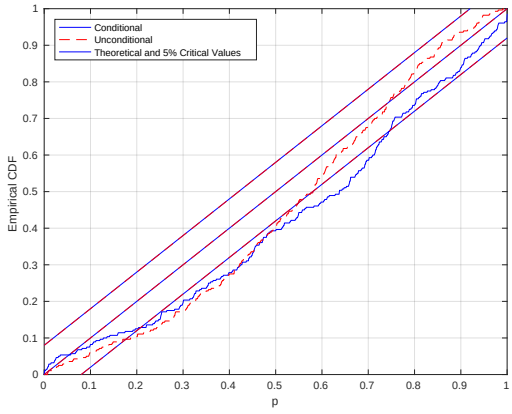
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

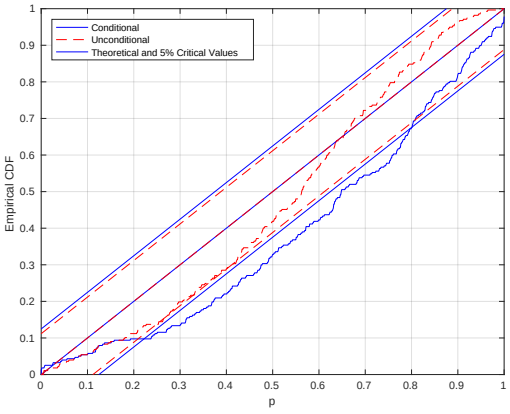
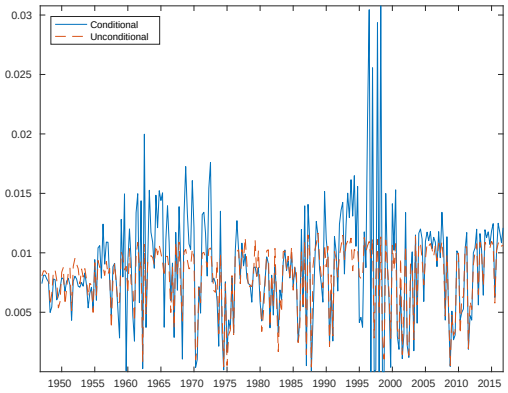


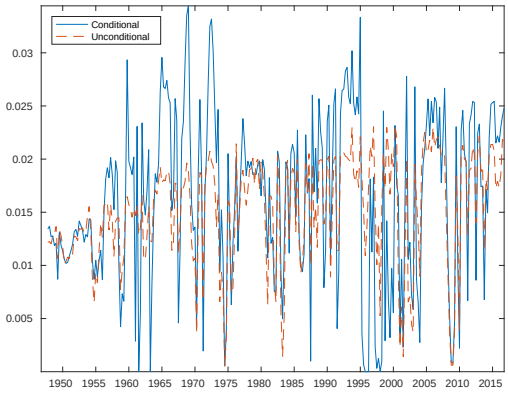
Figure 32: **Out-of-sample results: logDY**

The top panels compare the out-of-sample scores of the predicted distribution conditional on logDY with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the logDY predictor with the unconditional distribution.

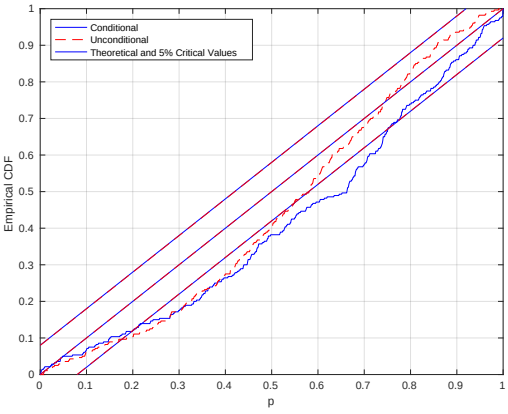
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

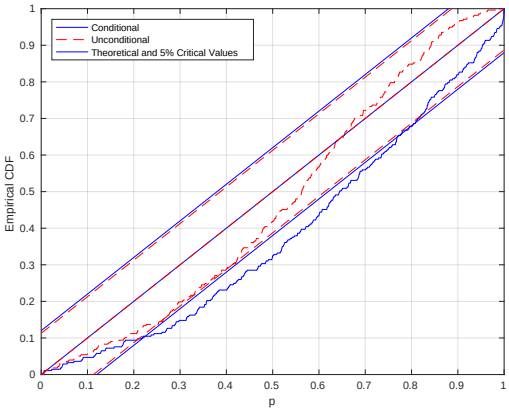
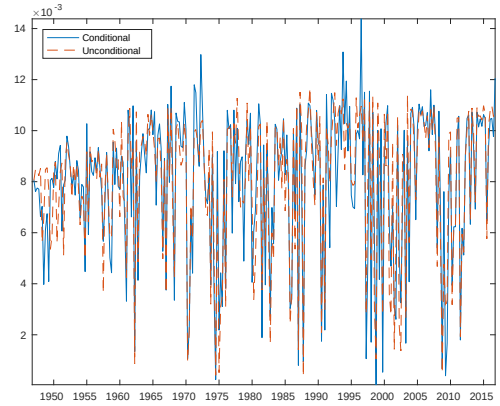


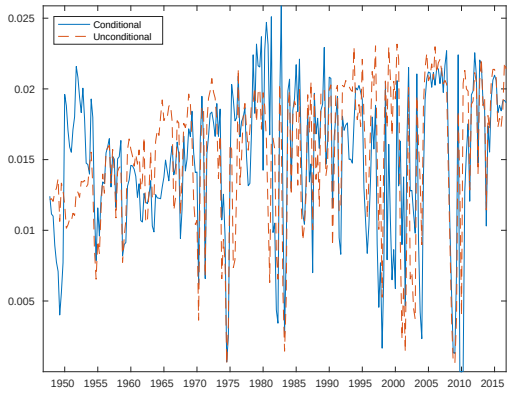
Figure 33: **Out-of-sample results: logEP**

The top panels compare the out-of-sample scores of the predicted distribution conditional on logEP with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the logEP predictor with the unconditional distribution.

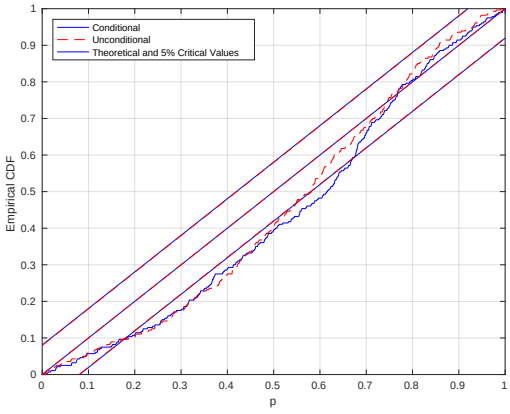
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

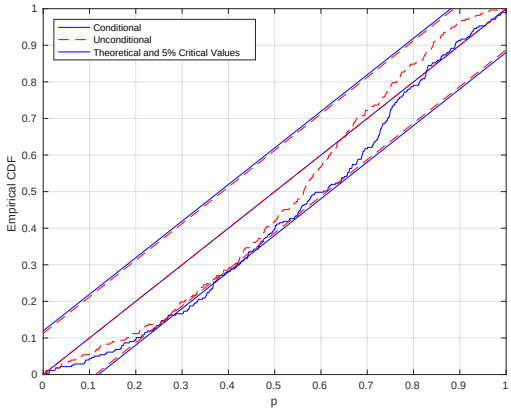
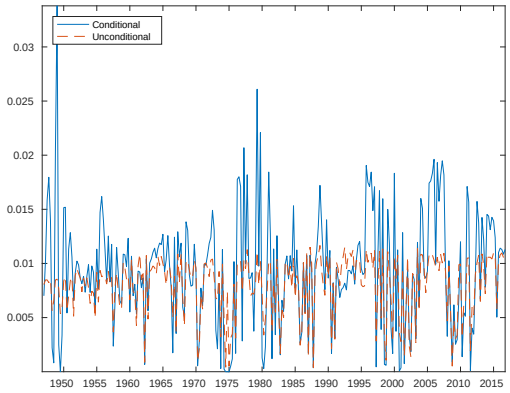


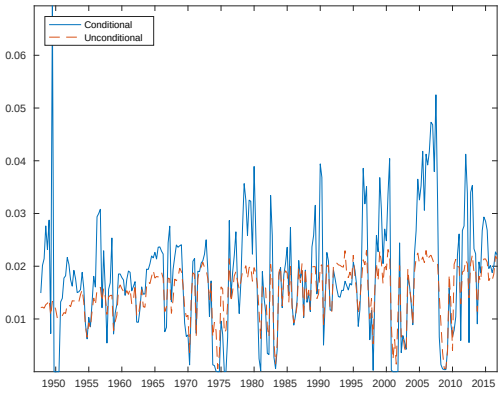
Figure 34: **Out-of-sample results: logDE**

The top panels compare the out-of-sample scores of the predicted distribution conditional on logDE with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the logDE predictor with the unconditional distribution.

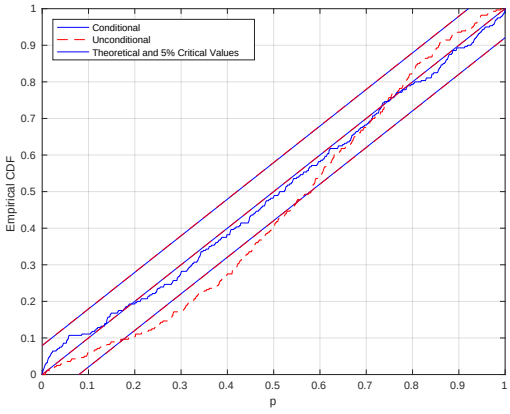
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

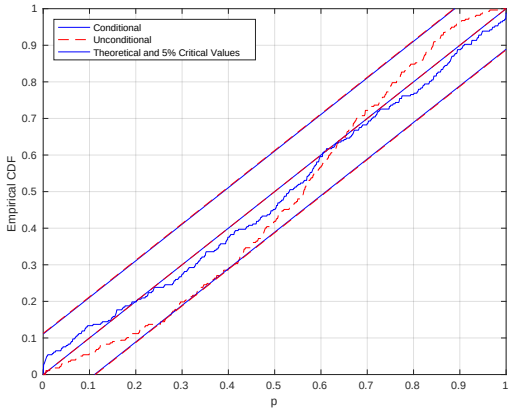
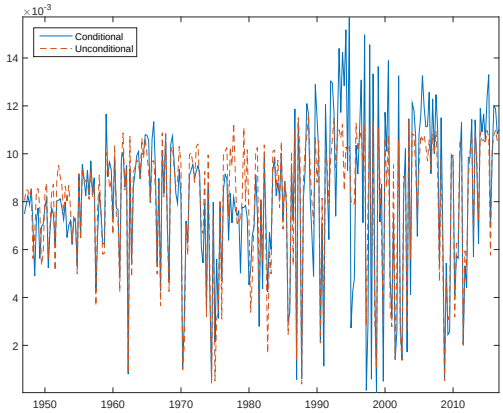


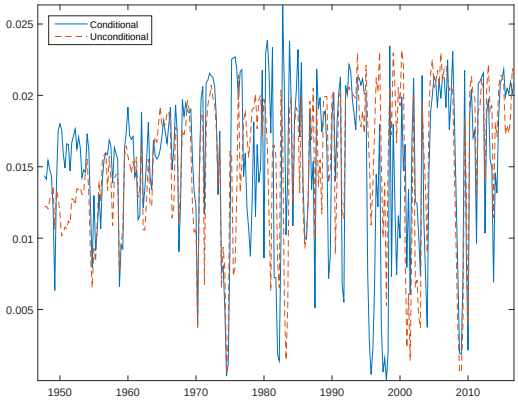
Figure 35: **Out-of-sample results: BM**

The top panels compare the out-of-sample scores of the predicted distribution conditional on BM with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the BM predictor with the unconditional distribution.

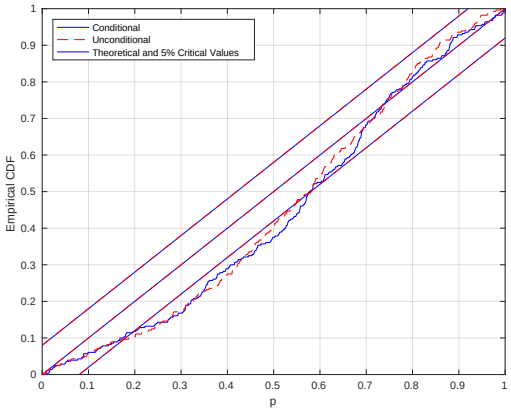
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

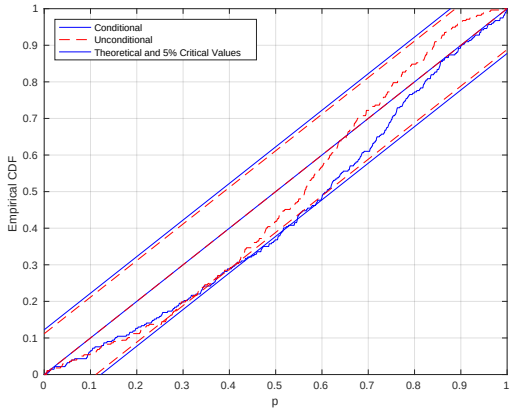
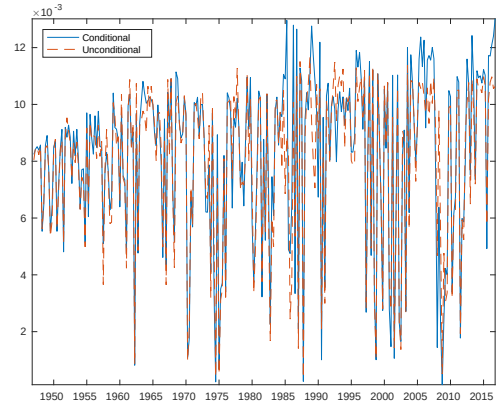


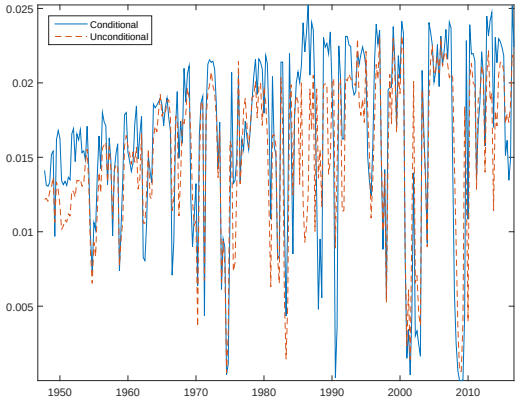
Figure 36: **Out-of-sample results: NTIS**

The top panels compare the out-of-sample scores of the predicted distribution conditional on NTIS with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the NTIS predictor with the unconditional distribution.

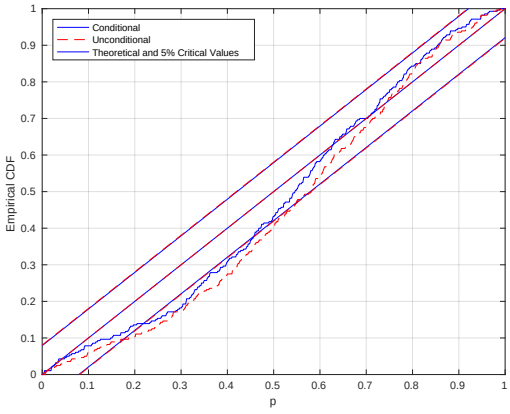
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

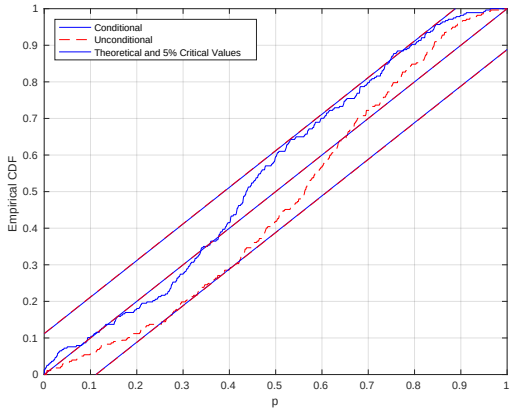
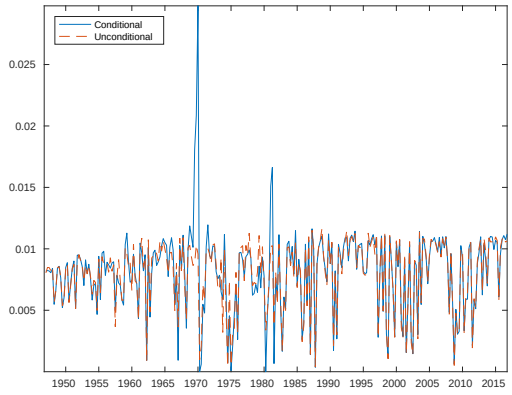


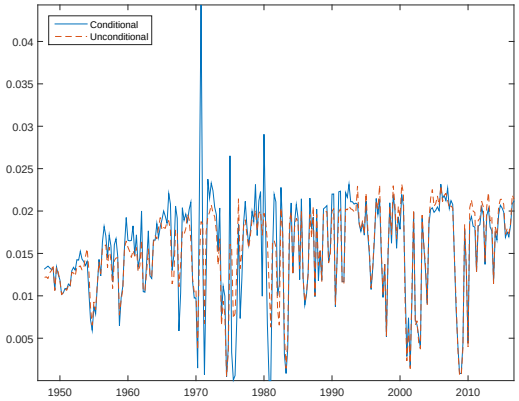
Figure 37: **Out-of-sample results: TBL**

The top panels compare the out-of-sample scores of the predicted distribution conditional on TBL with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the TBL predictor with the unconditional distribution.

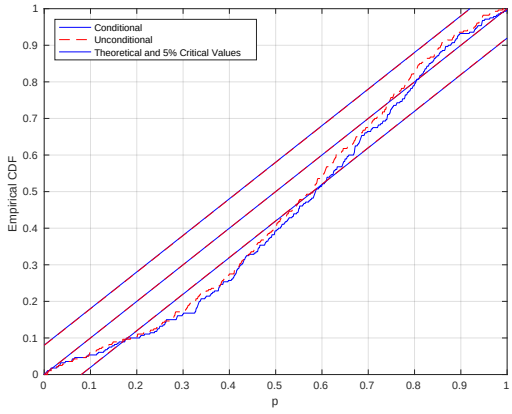
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

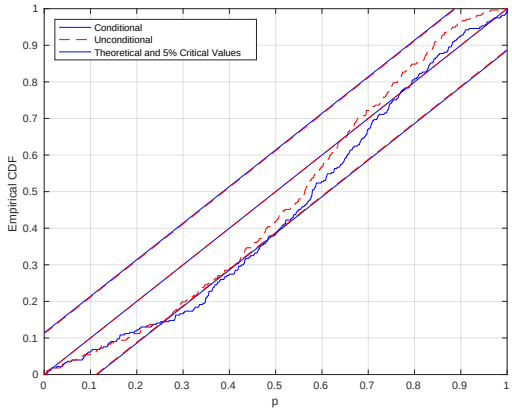
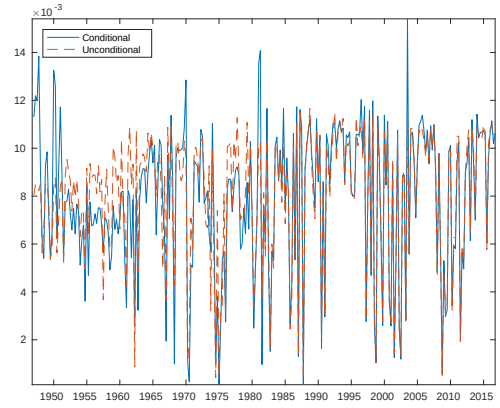


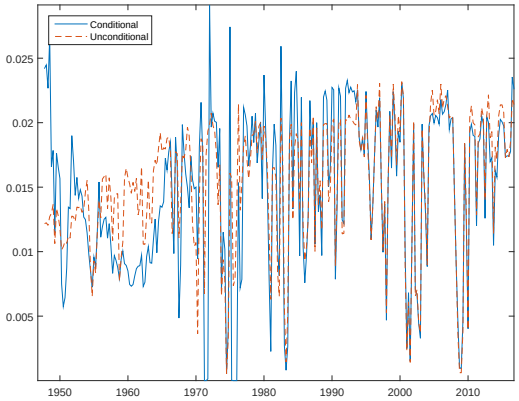
Figure 38: **Out-of-sample results: LTY**

The top panels compare the out-of-sample scores of the predicted distribution conditional on LTY with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the LTY predictor with the unconditional distribution.

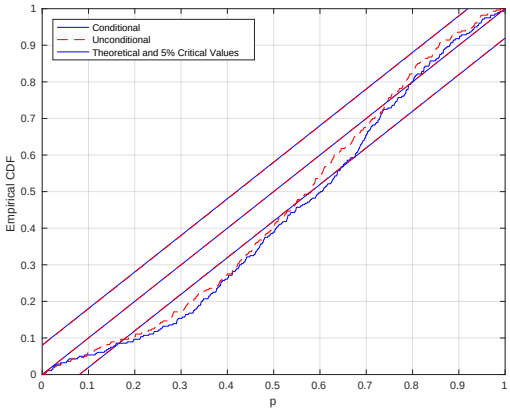
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

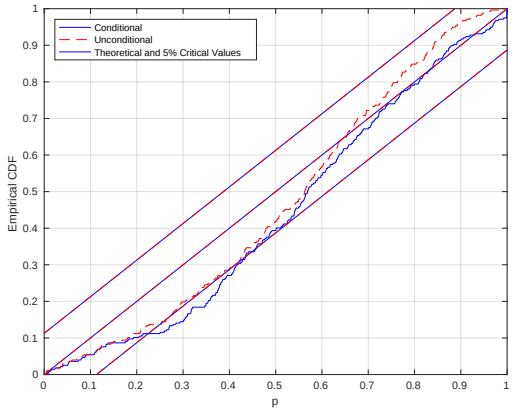
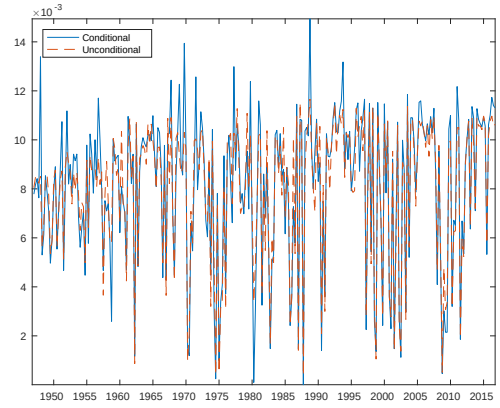


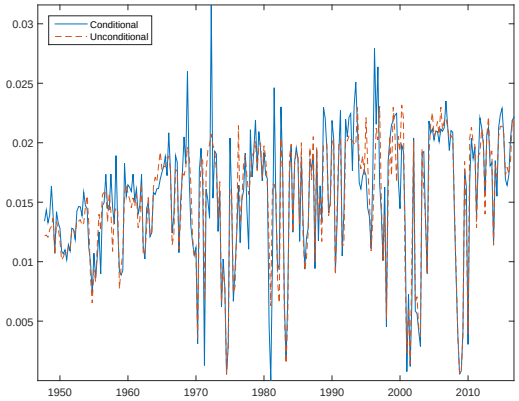
Figure 39: **Out-of-sample results: LTR**

The top panels compare the out-of-sample scores of the predicted distribution conditional on LTR with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the LTR predictor with the unconditional distribution.

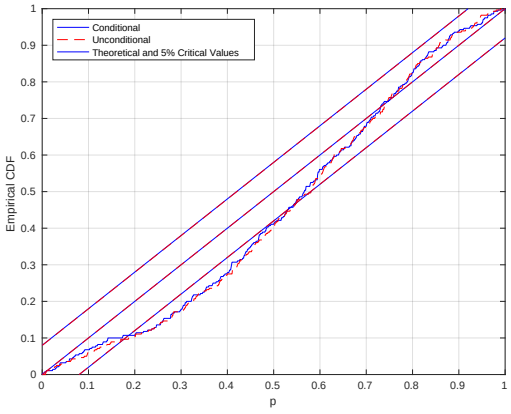
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

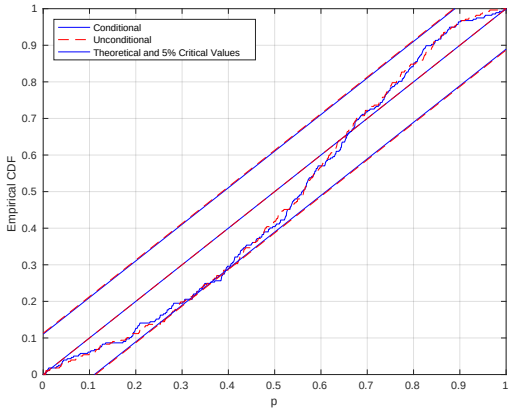
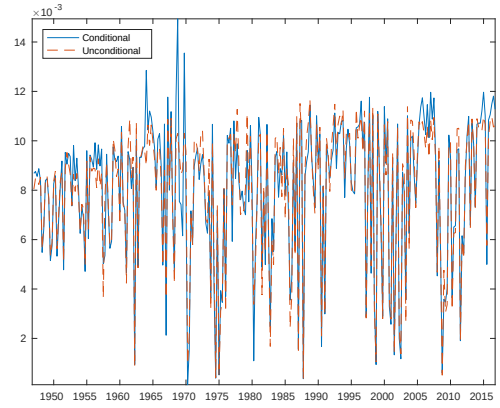


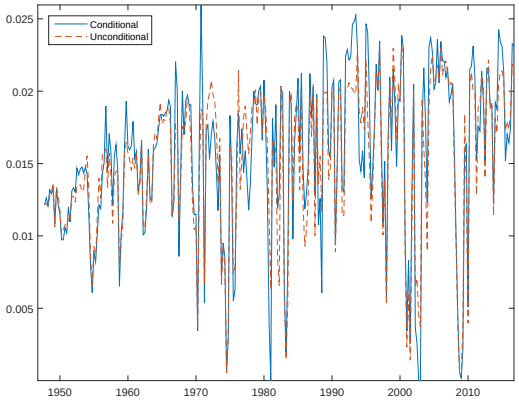
Figure 40: **Out-of-sample results: TMS**

The top panels compare the out-of-sample scores of the predicted distribution conditional on TMS with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the TMS predictor with the unconditional distribution.

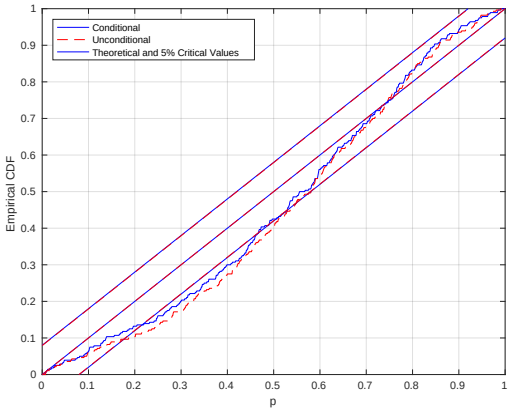
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

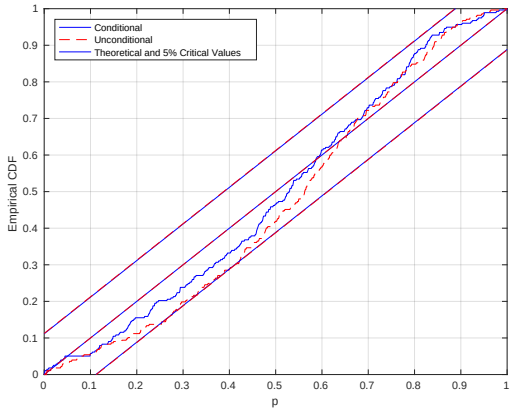
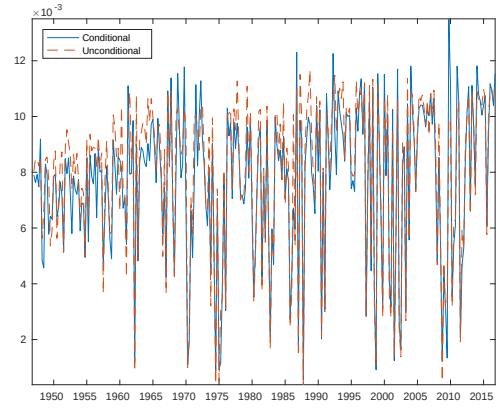


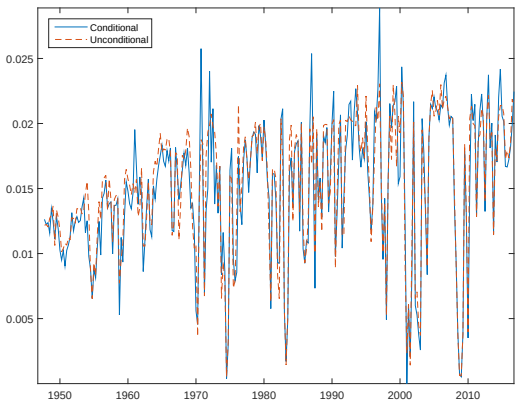
Figure 41: **Out-of-sample results: DFR**

The top panels compare the out-of-sample scores of the predicted distribution conditional on DFR with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the DFR predictor with the unconditional distribution.

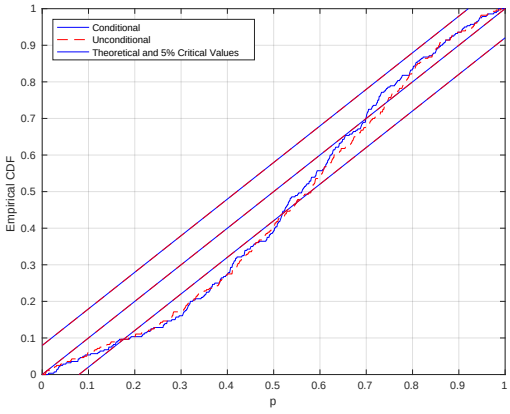
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform

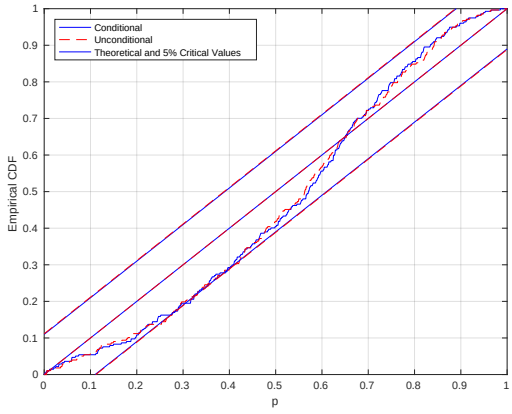
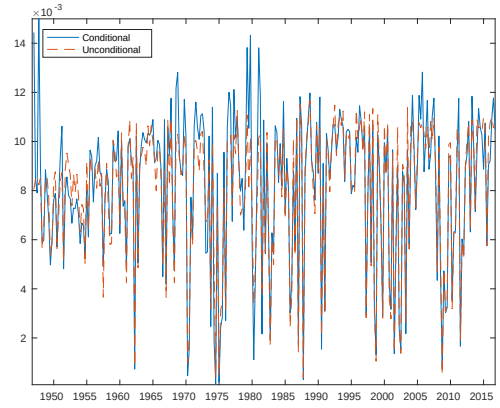


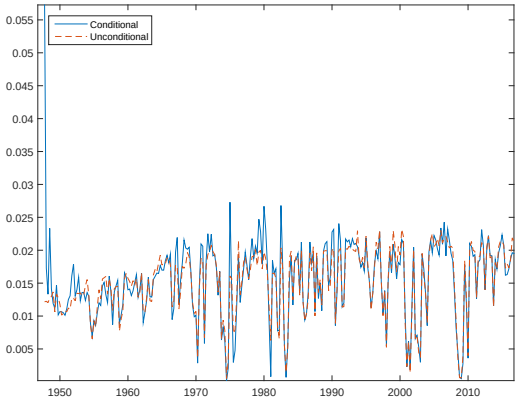
Figure 42: **Out-of-sample results: INFL**

The top panels compare the out-of-sample scores of the predicted distribution conditional on INFL with the unconditional distribution. The bottom panels compare the empirical cumulative distribution of the probability integral transform (PIT) for the INFL predictor with the unconditional distribution.

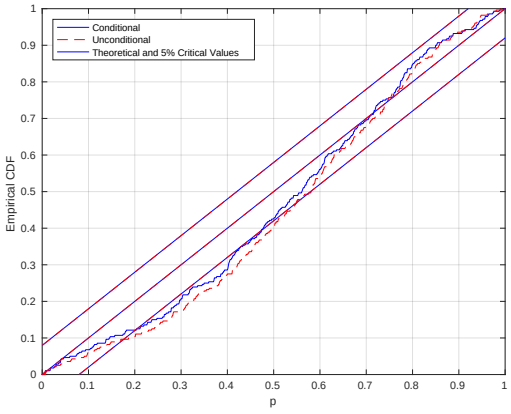
(a) **One quarter ahead**:
Out-of-sample scores



(b) **Four quarters ahead**:
Out-of-sample scores



(c) **One quarter ahead**:
Probability Integral Transform



(d) **Four quarters ahead**:
Probability Integral Transform