STAFF
REPORTS

NO. 850
JUNE 2018

REVISED
AUGUST 2023

Changing Risk-Return Profiles

Richard K. Crump | Miro Everaert | Domenico Giannone | Sean Hundtofte

FEDERAL RESERVE BANK of NEW YORK

Changing Risk-Return Profiles

Richard K. Crump, Miro Everaert, Domenico Giannone, and Sean Hundtofte

Federal Reserve Bank of New York Staff Reports, no. 850
June 2018; revised August 2023
JEL classification: C22, G17, G18

Abstract

We show that realized volatility in market returns and financial sector stock returns have 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.

Key words: stock returns, realized volatility, density forecasts, optimal pools

---

Crump: Federal Reserve Bank of New York (email: richard.crump@ny.frb.org). Everaert: Freddie Mac. Giannone: University of Washington, Amazon.com, CEPR. Hundtofte: Solve Finance. The authors thank Nina Boyarchenko and two anonymous referees for helpful comments and discussions. Oliver Kim provided excellent research assistance.

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)), focusing 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, Crump, and Vogt (2019b)).

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). Furthermore, 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 a high ex-ante probability for events that are realized ex-post. To estimate the predictive density, we use the methodology of Adrian, Boyarchenko, and Giannone (2019a). 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 in overall market returns and financial sector stock returns have 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 of judging 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 informational 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 it 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.

We also investigate the performance of optimal-weighted density forecasts based on a linear combination of density forecasts based on individual predictors. We find that the optimal combination is only modestly worse than the best individual models based on realized volatility or the default yield spread and comfortably outperform the vast majority of the candidate predictors. Furthermore, we show that the pooled density forecast based on optimal weights strongly outperforms an equal-weighted alternative. This stands in contrast to the point forecasting literature that generally finds that equal weighting is a challenging benchmark to beat. To our knowledge, we are the first to introduce such a result.

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.

Finally, our paper is related to the literature on the econometrics of high-dimensional data and its use in forecasting. Stock and Watson (2002), Forni, Hallin, Lippi, and Reichlin (2005) and Doz, Giannone, and Reichlin (2012) focused on compressing *predictors*. Here instead, we focus on compressing *predictions*. We also depart from standard forecasting exercises that focus on point (return) forecasts and also predict the risk.

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 concludes. Additional results are available in an online Supplementary Appendix (hereafter, "SA").

2 Motivation

To motivate the results in the paper we start with a set of predictive quantile regressions. Let us denote 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 variable 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).[2] 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, for example, the National Financial Conditions Index (NFCI) which only starts in 1971.[3] 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 interpreted as an approximation of the financial fragility or soundness of financial intermediaries in an economy (Chousakos et al., 2018; Atkeson et al., 2017).

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

[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] Section SA.1 in the SA provides a full list of variable names and definitions along with key properties of volFinancial and the other three variables which play a prominent role in our results: realized volatility of the aggregate market (volMarket), BBB-AAA credit spreads (DFY), and log dividend price-ratios (logDP).

[3] NFCI is found by Adrian et al. (2019a) to be an important indicator in predicting shifts in the distribution of aggregate output growth

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 (see Figure SA-3 in the SA). 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 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 SA-4 in the SA we show the corresponding results using the default yield spread (DFY). 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 both the overall market and

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

the financial sector, along with the default yield spread, provides 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 is 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 identify as providing the most accurate forecasts 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. (2019a). We estimate the quantile of r_{T+h} conditional 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:

$$\hat{\beta}_p = \arg\min_{\beta_p \in \mathbb{R}^k} \sum_{t=1}^{T-h} \left(p \cdot \mathbb{1}_{(r_{T+h} \geq x_t'\beta)} \left| r_{T+h} - x_t'\beta_p \right| + (1-p) \cdot \mathbb{1}_{(r_{T+h} < x_t'\beta)} \left| r_{T+h} - x_t'\beta_p \right| \right) \tag{3}$$

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. 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. (2019a) we use the skewed t-distribution developed by Azzalini and Capitanio (2003) in order to smooth the quantile function and recover

[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.

a valid probability density function:

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

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 $F\left(y;\mu,\sigma,\alpha,\nu\right)$ denote the CDF of the skewed t-distribution and $F^{-1}\left(y;\mu,\sigma,\alpha,\nu\right)$ its associated quantile function. To implement our approach we choose our grid of quantiles as $S_p=\{.05,.25,.75,.95\}$. Adrian et al. (2019a) find that this choice performs well in practice. In our implementation we also restrict μ to the interval $[-20,20]$. Then, for a fixed time period T we first run four quantile regressions to obtain $Q_{r_{T+h}|x_T}\left(p\right)$ for each $p\in S_p$ and then estimate $(\mu_T,\sigma_T,\alpha_T,\nu_T)$ via

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

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: $\hat{p}_{r_{T+h}|x_T}(y)=f\left(y;\hat{\mu}_T,\hat{\sigma}_T,\hat{\alpha}_T,\hat{\nu}_T\right)$. Figure SA-5 in the SA provides two examples of the inputs and outputs of this two-step estimation procedure as a forecaster would have observed in an environment with mild financial conditions (1993Q4) and one with tight financial conditions (2008Q4). 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 the corresponding date. In the bottom left chart, the conditional distribution as of 2008Q4 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, the bottom right chart illustrates a "quiet" period at the end of 1993, where the conditional distribution is more concentrated than the unconditional distribution.

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. 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 2021Q4. 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, $\mathbb{T} = \{1947Q1, ..., 2021Q4\}$ and $|\mathbb{T}| = 300$.

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. For market and financial sector volatility and the default yield spread, 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. Both volatility measures outperform the default yield spread at both horizons. Between the two volatility measures, we prefer volFinancial for its more consistent performance. 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. 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.

We repeat the predictive exercises with a twenty-year fixed window to form forecasts, rather than estimating the model on an expanding window. We observe that the fixed window specification does not appear to significantly improve point or density forecasts relative to the recursive window specification. 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 accommodated by a recursive window formulation that "remembers" any previous crisis episodes in the historical time series. This result provides suggestive evidence that for density forecasts the dominant concern is the paucity of extreme events rather than changing features such as structural breaks. This stands in contrast to empirical findings when the focus is the conditional mean of future stock returns (e.g., Rapach et al. (2010)).

In Figures 2 and SA-6 in the SA we plot the time series of the predictive scores (without applying the log transformation) for forecasts based on volFinancial and DFY relative to the naive benchmark model, respectively. For both volFinancial and DFY it is clear that the outperformance relative to the benchmark model is not concentrated in specific time periods or outliers 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.[6]

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 that are far from the mean, and so forecast errors are down-weighted, as reflected in the third term. In the general case, the predictability of any aspect

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

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. As both higher realized volatility and higher credit spreads tend to occur during economic downturns, this suggests there is an underlying business-cycle component which relates to the density forecasting results we present; likewise, this has been shown to be an important feature for the equity risk premium (e.g., Rapach et al. (2010)).

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 (e.g., Timmermann (2006), Rapach et al. (2010)). 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 the 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.

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 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:

$$w_i^{(k+1)} = w_i^{(k)} \frac{1}{T - T_0 - h + 1} \sum_{t=T_0}^{T-h} \frac{\hat{p}_{r_{T+h}|x_T^i}(r_{T+h})}{\sum_{i=1}^{N} w_i^{(k)} \hat{p}_{r_{T+h}|x_T^i}(y_{T+h})} \tag{9}$$

De Mol (2023) derives the convergence properties of this iterative algorithm. Note that the non-negativity and sum-to-one constraints for the weights are 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 $\hat{p}_{r_{T+h}|x_T^1,\ldots,x_T^N}(r_{T+h})$. The row at the bottom of Table 1 reports the average of the logarithmic score over the evaluation sample. It is evident that the predictive accuracy of the pooled forecasts are modestly worse than the best individual models based on volFinancial, volMarket, and DFY. However, the pooled forecasts comfortably outperform the vast majority of the candidate predictors. Figure SA-7 in the SA provides an example of the fitted distribution based on the pooled forecasts for the end of 2008 and Figure SA-8 in the SA shows the out-of-sample predictive scores.

Table 1 also reports the results from an equal weighting of the density forecasts based on each individual predictor. The broad consensus from the literature on point estimation is that equal weighting is a challenging benchmark to outperform (e.g., DeMiguel et al. (2009)). However, it does not appear that this result transfers to the case of density forecasting. Instead, the predictor

using optimal weights strongly outperforms the equal weighted benchmark (see also Figure SA-9 in the SA). To our knowledge, we are the first to introduce such a result.

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 Table 1 shows that this optimal combination is clearly preferred to the individual models when $h = 4$ but is middle of the pack for $h = 1$. We also report the results for the equal weighted combination. Consistent with existing results in the literature, this equal-weighted approach outperforms most individual predictors and the optimally-weighted combination. Recall that the optimal combination of weights for density forecasts, in contrast, showed consistent performance only modestly worse than the best individual models and far better than the equal-weighted alternative.

The real-time weights, $\{\hat{w}_i : i = 1, \ldots, n\}$, formed for the optimal combined conditional density forecast are reported in Figure 3. 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. 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. The optimal weighting places zero weight on the majority of the predictors and yet still performs competitively with the best individual density forecasting models. This stands in stark contrast to the results in the point forecasting literature which find that (approximately) equal-weighted forecasts have superior out-of-sample performance (Rapach et al. (2010)). We can also see this in the bottom charts as the 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 4. 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 5, 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 as a function of the quantile.[7] 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 (2019), 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 Online Appendix.

4 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. Our 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, both of financial sector stock returns and market 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. 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 (e.g., Giglio, Kelly, and Pruitt 2016; Adrian, Boyarchenko, and Giannone 2019a).

We also investigate the performance of optimal-weighted density forecasts and find that the optimal combination is only modestly worse than the best individual models based on realized volatility or the default yield spread. Furthermore, we show that the pooled density forecast based on optimal weights strongly outperforms an equal-weighted alternative. This suggests that existing results for point forecasting may not transfer directly to density forecasting. Further work is warranted to better understand the commonalities and differences between forecast performance in these two different settings.

[7] 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.

In this paper, we have focused on the prediction of a single outcome, the market return. A natural extension is to study the joint predictions of several outcomes, for example, equity returns across different industries or countries. This would be useful for studying the risks of different portfolios and identifying the locations of the strongest co-dependencies.

References

Adrian, T., Boyarchenko, N., Giannone, D., 2019a. Vulnerable growth. American Economic Review 109, 1263–1289.

Adrian, T., Crump, R., Vogt, E., 2019b. Nonlinearity and flight to safety in the risk-return trade-off for stocks and bonds. Journal of Finance 74, 1931–1973.

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.

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., Ordonez, G., 2018. Aggregate information dynamics. 2018 Meeting Papers No. 167, Society for Economic Dynamics.

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

De Mol, C., 2023. Multiplicative algorithms for density combination and deconvolution, working paper.

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? Review of Financial Studies 22, 1915–1953.

Doz, C., Giannone, D., Reichlin, L., 2012. A quasi maximum likelihood approach for large, approximate dynamic factor models. Review of Economics and Statistics 94, 1014–1024.

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.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: One-sided estimation and forecasting. Journal of the American Statistical Association 100, 830–840.

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.

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

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

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Rossi, B., Sekhposyan, T., 2019. Alternative tests for correct specification of conditional predictive densities. Journal of Econometrics 208, 638–657.

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

Stock, J., Watson, M., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288.

Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger, C., Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Elsevier, vol. 1, chap. 4, 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.

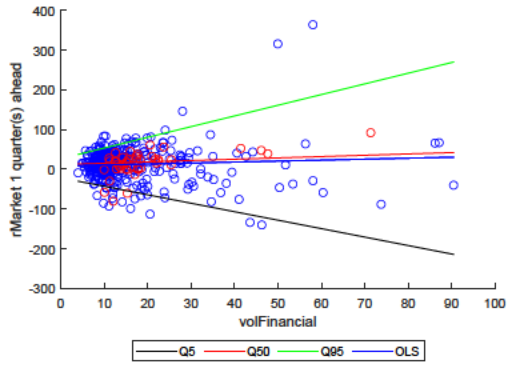
Table 1: **Diebold-Mariano Test: rMarket recursive** The table below compares the out-of-sample market return performance of each predictor for point forecasts (left panel) and density forecasts (right panel) estimated with a recursive (expanding) window. In the left panel, the first row shows the mean-squared 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 unconditional 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. "Equal-Weighted" refers to the forecast combination based on equal weights across all predictors. "Aggregation" refers to the forecast combination based on the optimal weights as described in Section 3.2.

	Point				Density			
	$h = 1$		$h = 4$		$h = 1$		$h = 4$	
Naive	1065.072		268.049		-4.955		-4.294	
volFinancial	31.317	(20.163)	10.699	(6.054)	0.079	(0.024)	0.075	(0.030)
volMarket	4.762	(10.131)	5.310	(5.076)	0.081	(0.024)	0.060	(0.022)
logDP	-5.826	(12.132)	-2.902	(21.071)	-1.315	(1.116)	-2.329	(1.456)
logDY	-7.057	(11.978)	-0.045	(22.934)	-0.469	(0.366)	-0.282	(0.173)
logEP	-5.586	(17.754)	5.572	(18.995)	-0.032	(0.024)	-0.197	(0.175)
logDE	92.327	(44.746)	12.320	(7.121)	-0.481	(0.347)	-1.512	(0.913)
BM	-0.372	(6.579)	25.598	(23.005)	-0.044	(0.033)	-0.078	(0.069)
NTIS	20.322	(18.860)	29.355	(29.961)	0.002	(0.018)	-0.076	(0.111)
TBL	11.013	(8.307)	19.076	(13.262)	-0.124	(0.079)	-0.428	(0.372)
LTY	23.767	(16.969)	35.205	(13.789)	-0.042	(0.015)	-0.638	(0.412)
LTR	20.948	(36.753)	8.203	(10.000)	-0.029	(0.019)	-0.294	(0.283)
TMS	14.281	(6.210)	12.816	(11.182)	-0.014	(0.012)	-0.332	(0.302)
DFY	18.285	(7.816)	2.610	(5.061)	0.073	(0.031)	0.058	(0.034)
DFR	27.567	(7.742)	9.750	(6.740)	-0.299	(0.263)	-0.266	(0.243)
INFLlag	-0.508	(0.989)	-3.448	(4.896)	-0.004	(0.016)	-0.012	(0.013)
Equal-Weighted	-1.433	(5.124)	-12.216	(6.449)	0.026	(0.007)	0.026	(0.015)
Aggregation	14.274	(13.167)	-4.813	(8.879)	0.054	(0.018)	0.055	(0.029)

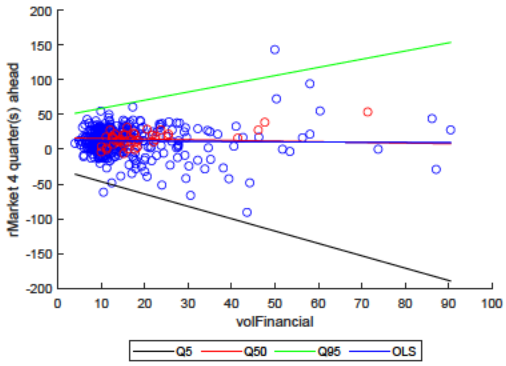
Figure 1: rMarket vs. volFinancial

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 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.

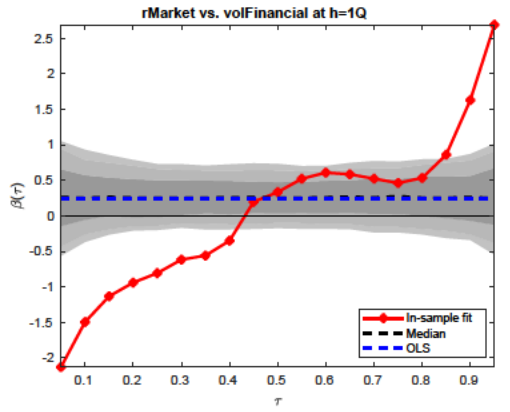
(a) **One quarter ahead**: Scatter



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



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



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

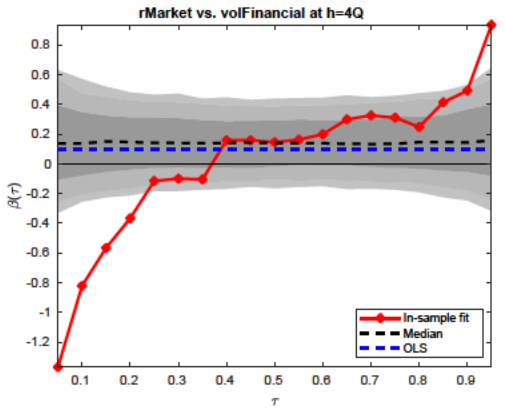
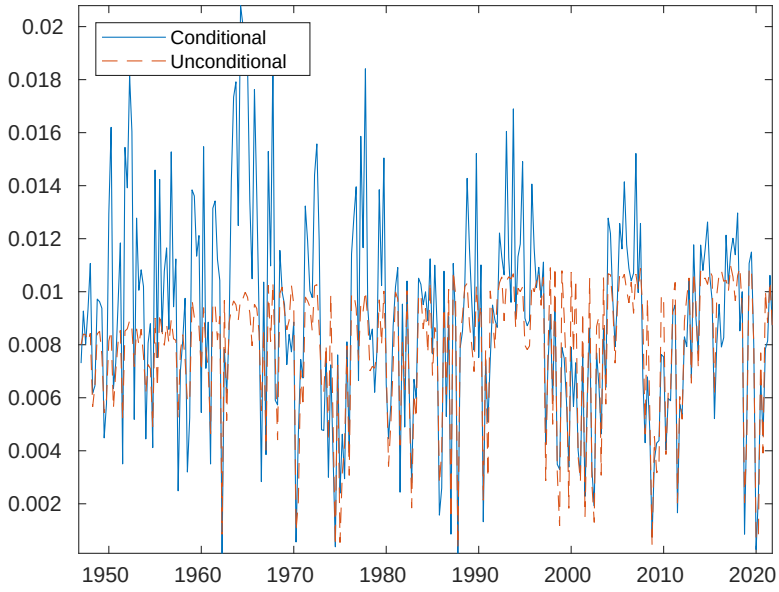


Figure 2: **Out-of-sample Scores, volFinancial**

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**

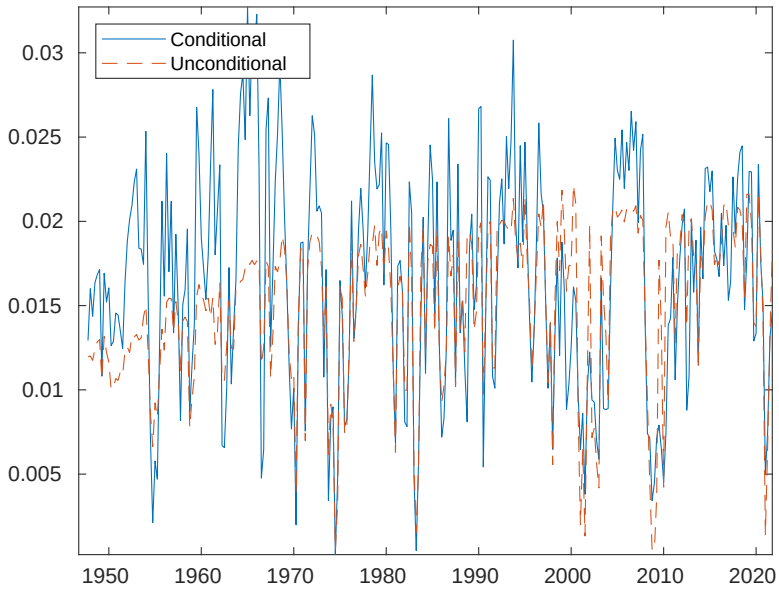
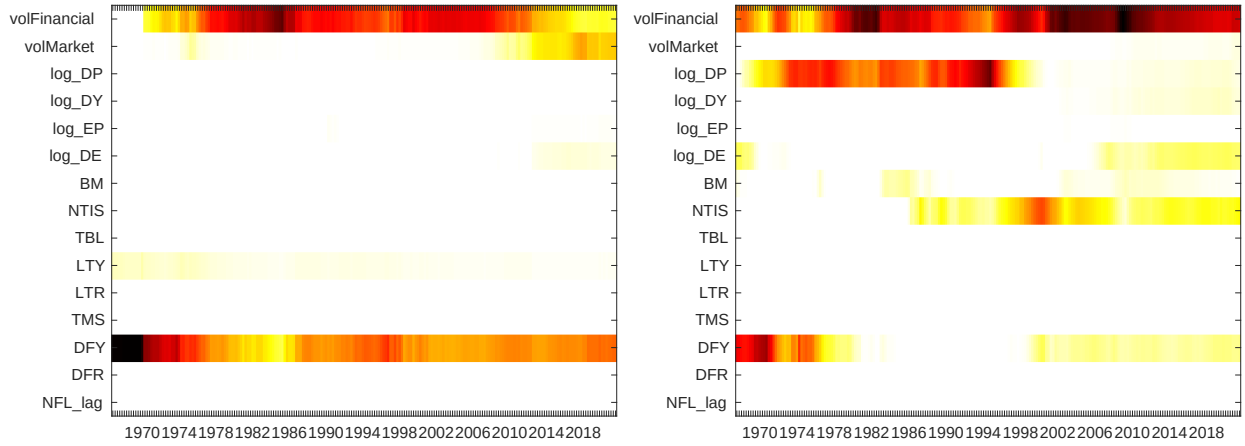


Figure 3: **Aggregation Weights**

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. 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).

(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

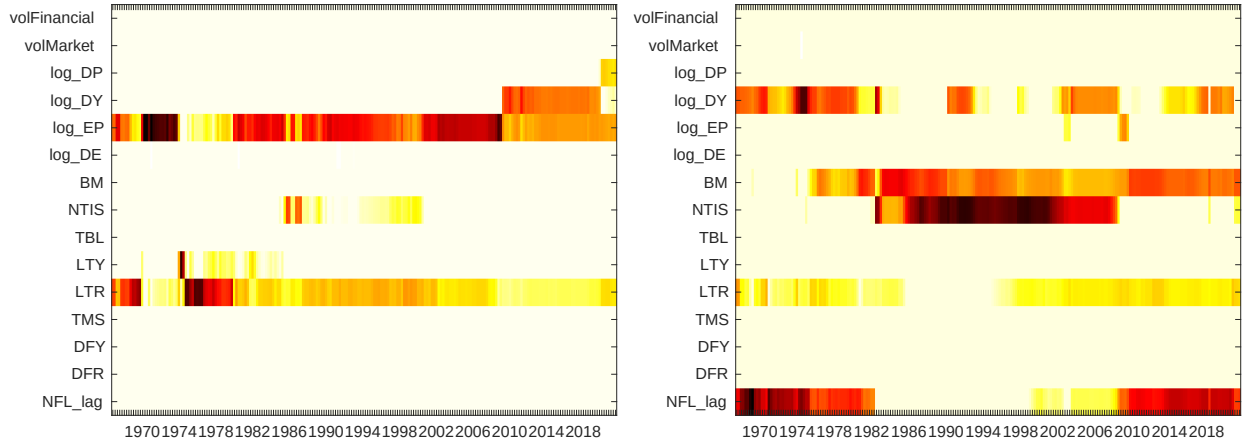
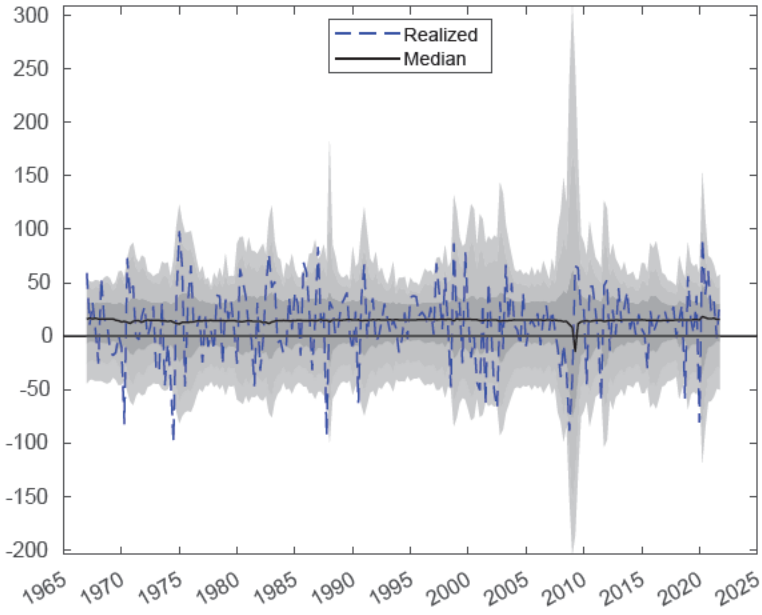


Figure 4: **Predicted Distributions**

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**

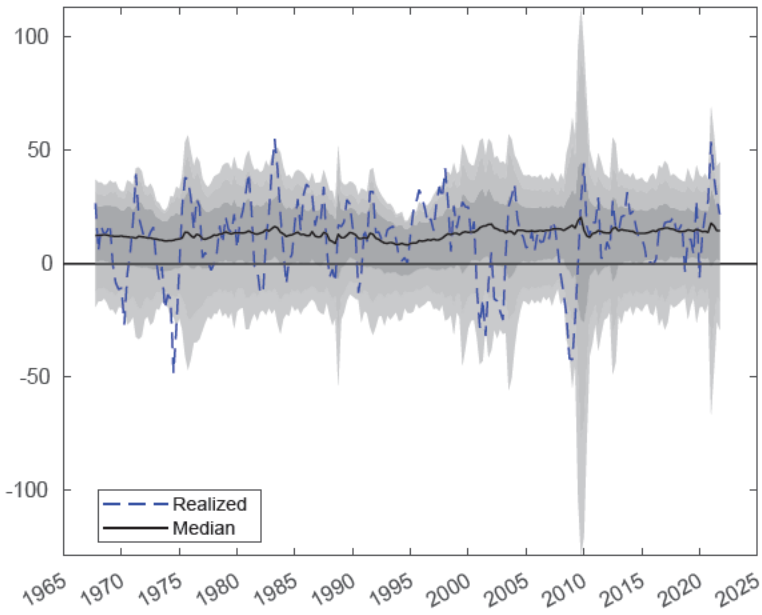
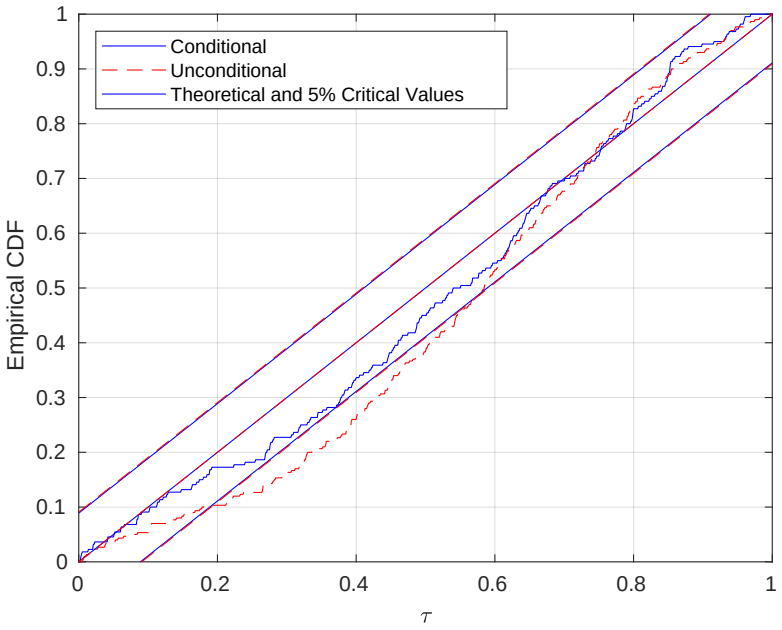


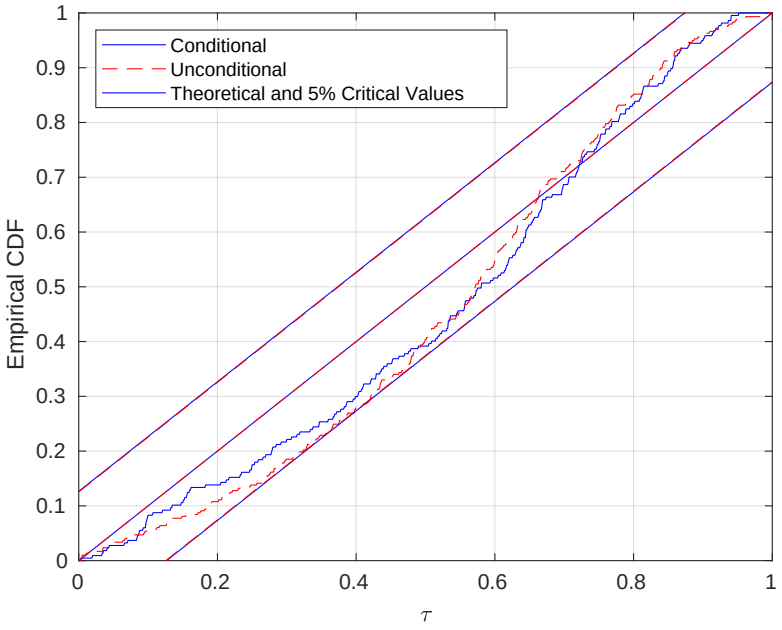
Figure 5: **PIT, Aggregate Forecast**

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**



.

Supplementary Appendix for "Changing Risk-Return Profiles"

Richard K. Crump **Miro Everaert**

Domenico Giannone **Sean Hundtofte**

This Supplemental Appendix provides additional supporting results for the main paper. Section SA.1 summarizes the data used and discusses some key features. Section SA.2 provides figures highlighted in the main text whereas Section SA.3 provides additional figures as a reference.

SA.1 Data

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 in the next subsection.

Figure SA-1 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 SA-2 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 SA-2 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.

SA.1.1 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.

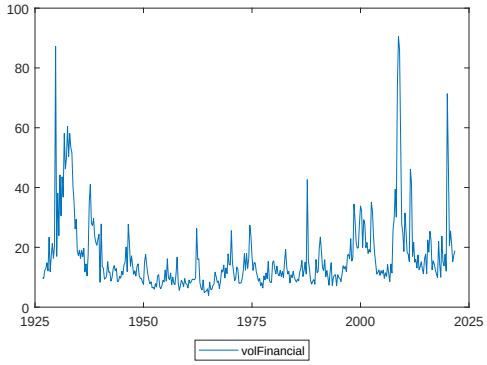
DFR The default return spread is the difference between returns on long-term corporate bonds and long-term government bonds.

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.

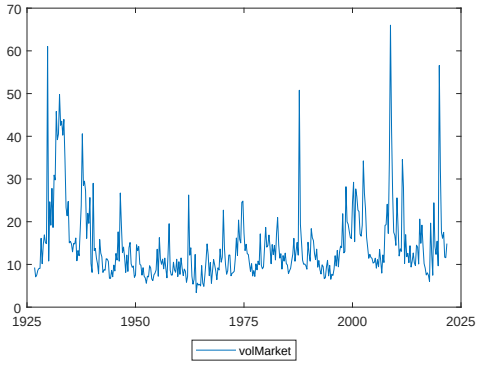
Figure SA-1: **Raw Data**

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")

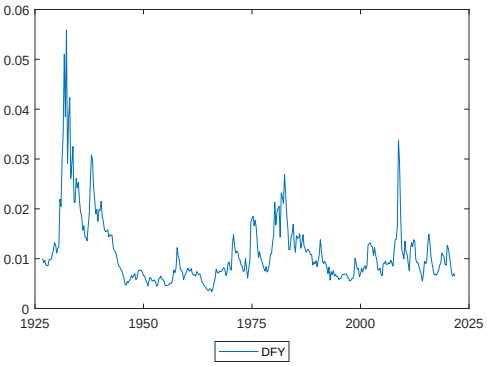
(a) **volFinancial**



(b) **volMarket**



(c) **DFY**



(d) **logDP**

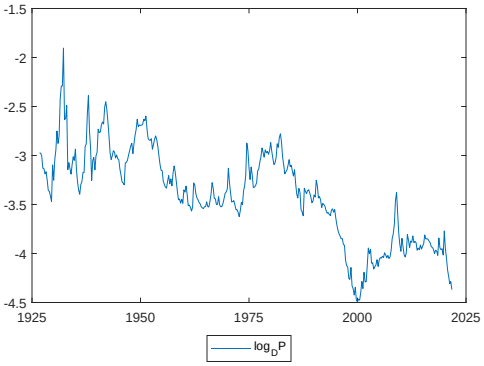
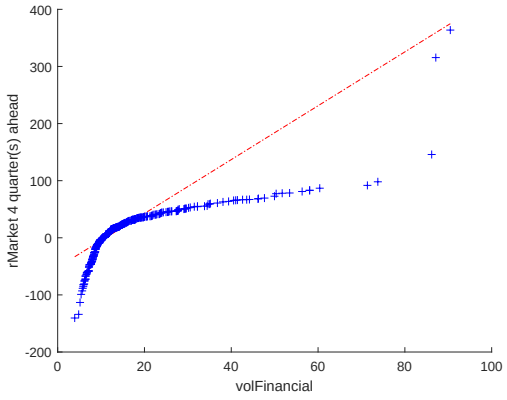


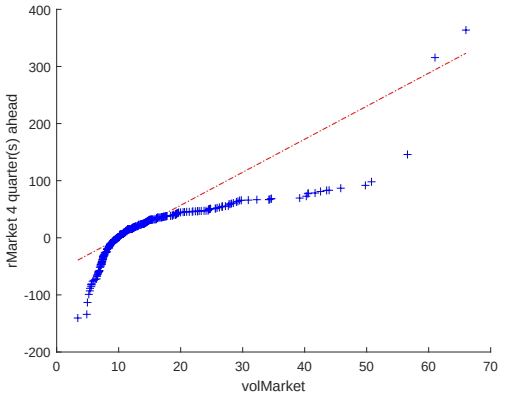
Figure SA-2: **Quantile-Quantile Plots**

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

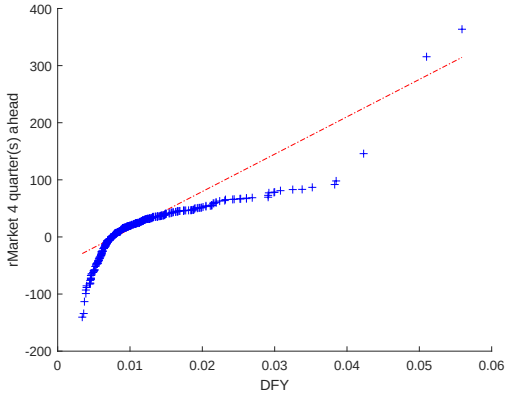
(a) **Market returns - volFinancial (4Q)**



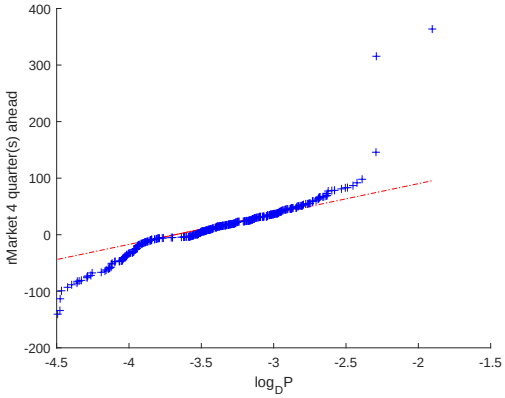
(b) **Market returns - volMarket (4Q)**



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



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

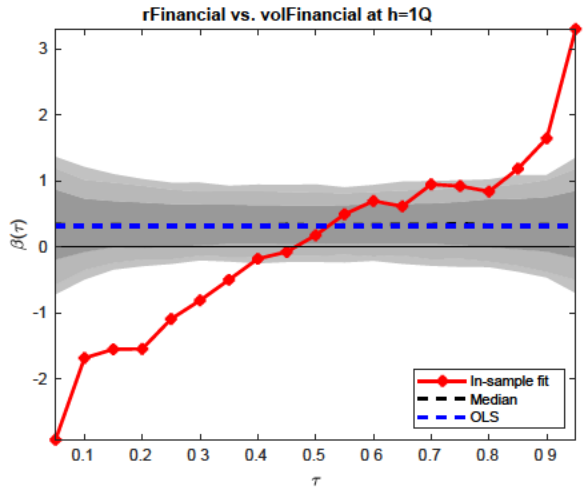


SA.2 Additional Figures for Main Results

Figure SA-3: **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

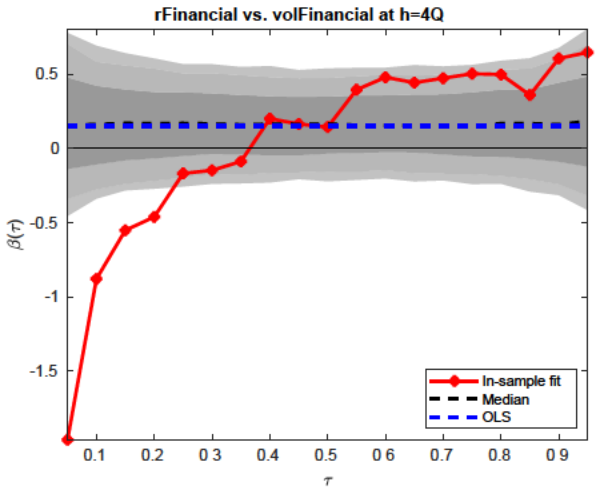
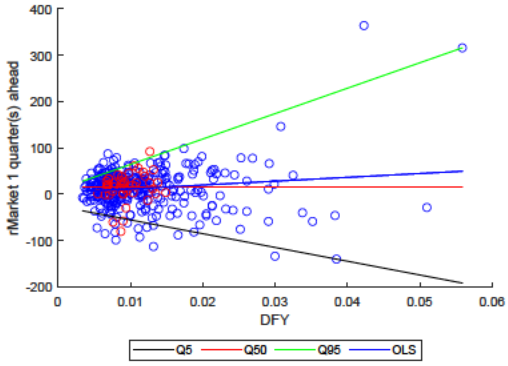


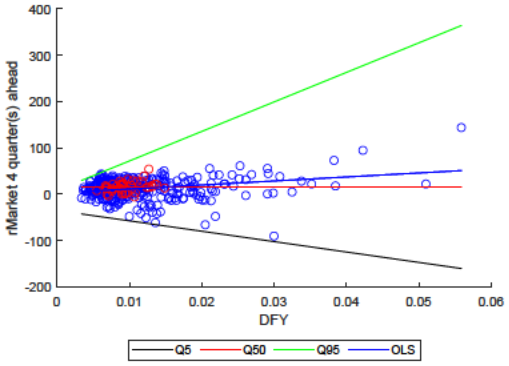
Figure SA-4: rMarket vs. DFY

The top panels in this figure show the univariate quantile regressions of one quarter ahead (left) and four quarter ahead (right) market returns on DFY. Data before 2010Q3 are shown as open blue circles; data after this date are shown as 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.

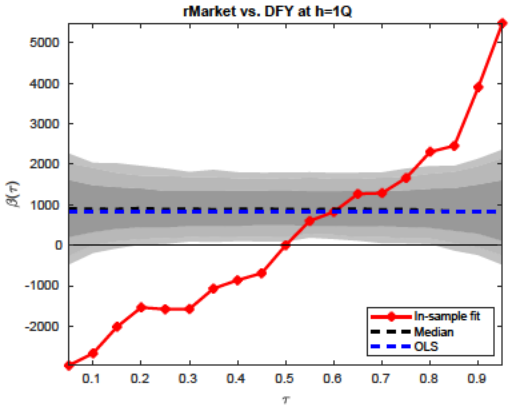
(a) **One quarter ahead**: Scatter



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



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



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

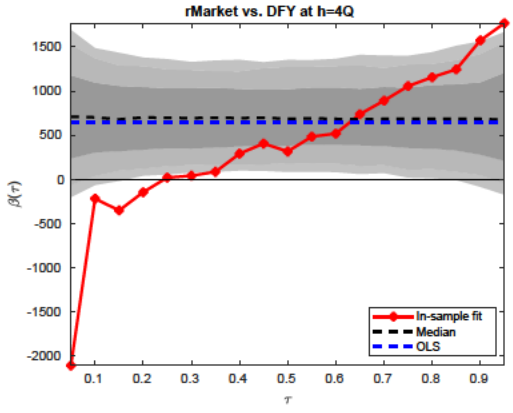
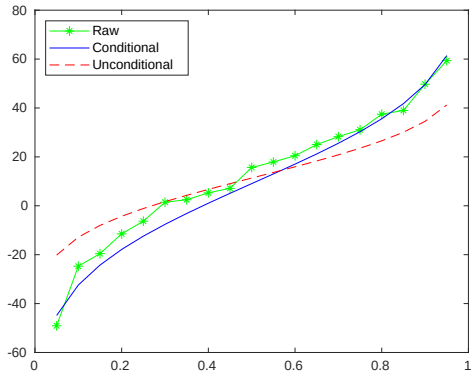


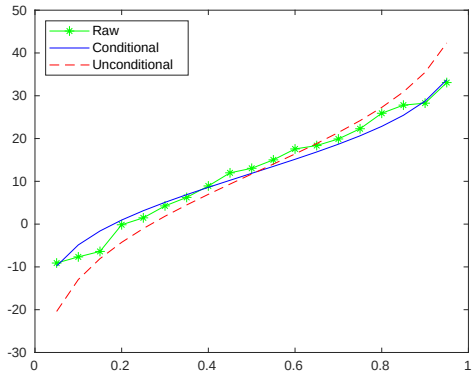
Figure SA-5: **Example of Fitted Distributions: volFinancial**

The top panels in this figure show the conditional quantiles and fitted skewed t-inverse cumulative distribution functions of four quarter ahead market returns predicted by realized financial volatility in 2008Q4 (left) and 1993Q4 (right). The bottom panels show the corresponding density functions.

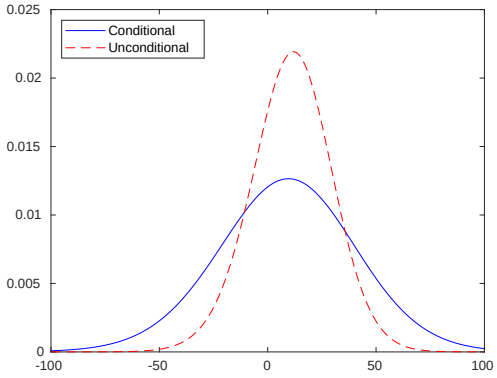
(a) **2008Q4**
Conditional quantiles and skewed t-inverse CDF



(b) **1993Q4**
Conditional quantiles and skewed t-inverse CDF



(c) **2008Q4**
Probability Densities



(d) **1993Q4**
Probability Densities

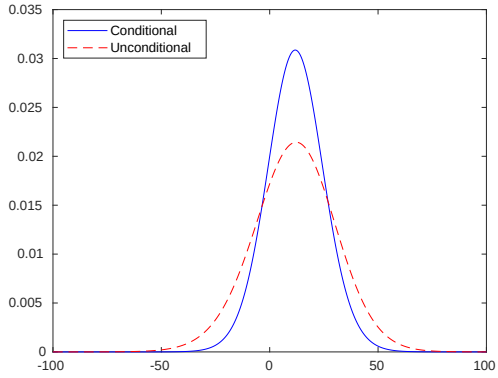
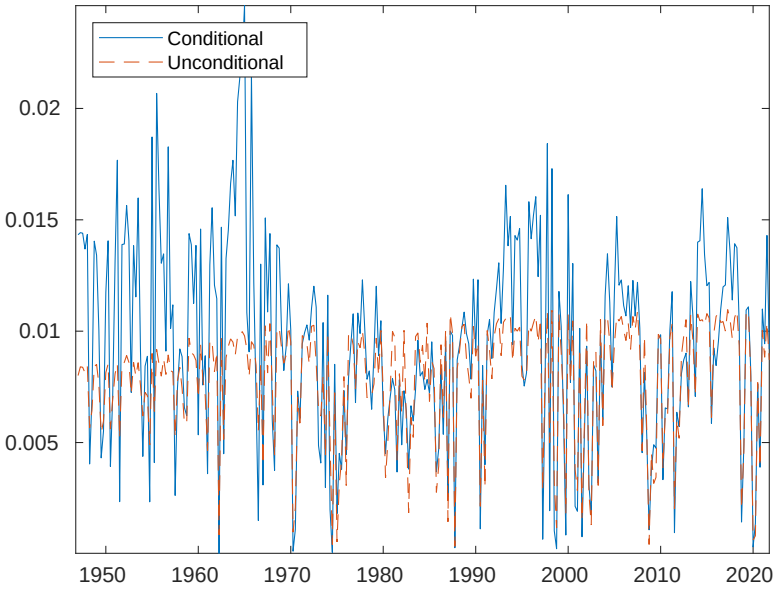


Figure SA-6: **Out-of-sample Scores, DFY**

The figure compares the out-of-sample predictive scores of the predicted distribution conditional on the default yield spread (DFY) with the unconditional distribution.

(a) **One quarter ahead**



(b) **Four quarters ahead**

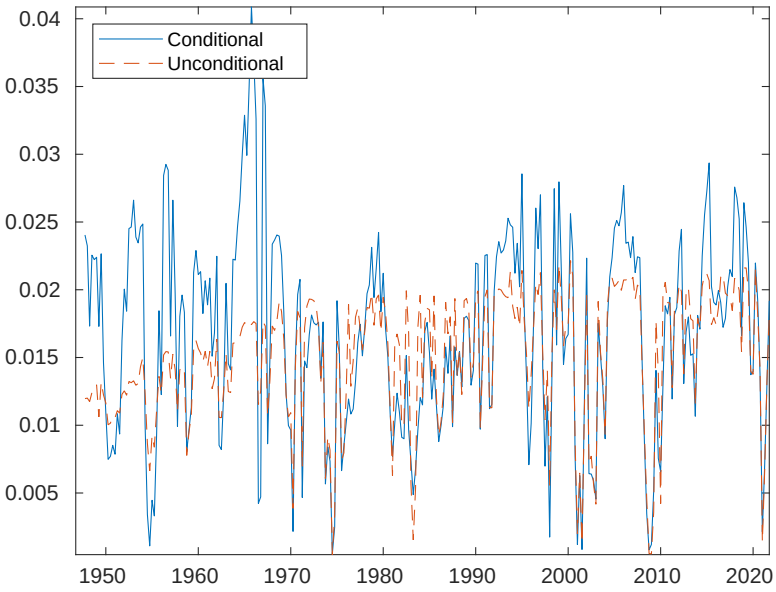
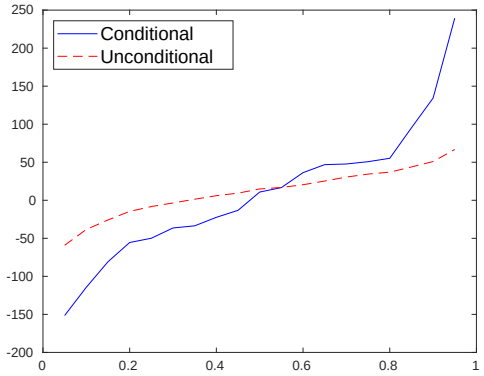


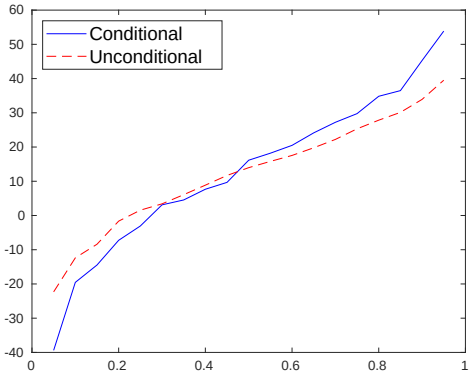
Figure SA-7: **Example of Fitted Distributions on 2008Q4: Optimal Forecast**

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.

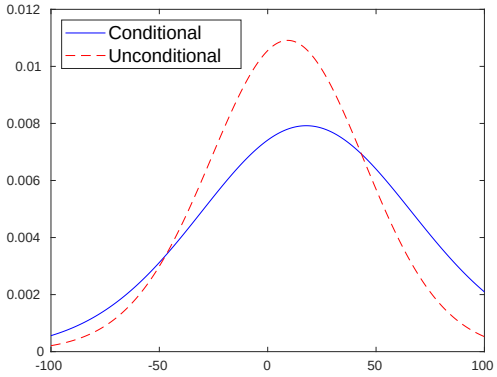
(a) **One quarter ahead**:
Conditional quantiles and skewed t-inverse CDF



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



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



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

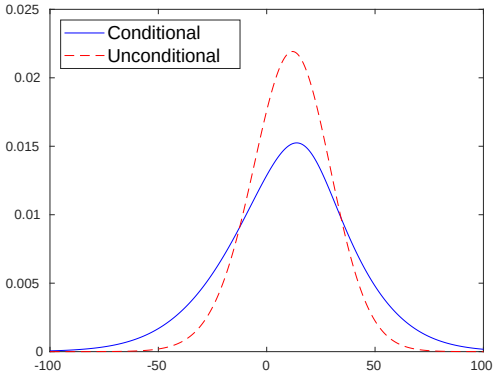
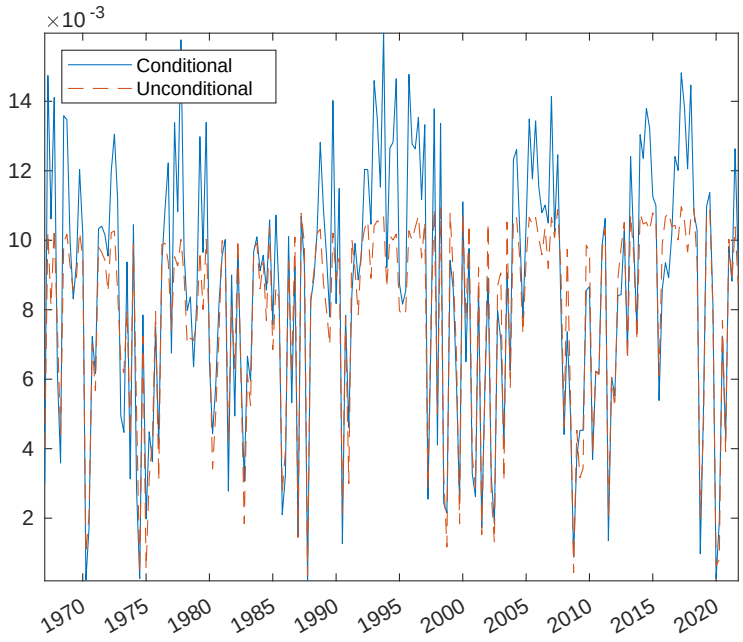


Figure SA-8: **Out-of-sample Scores, Aggregate Forecast**

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**

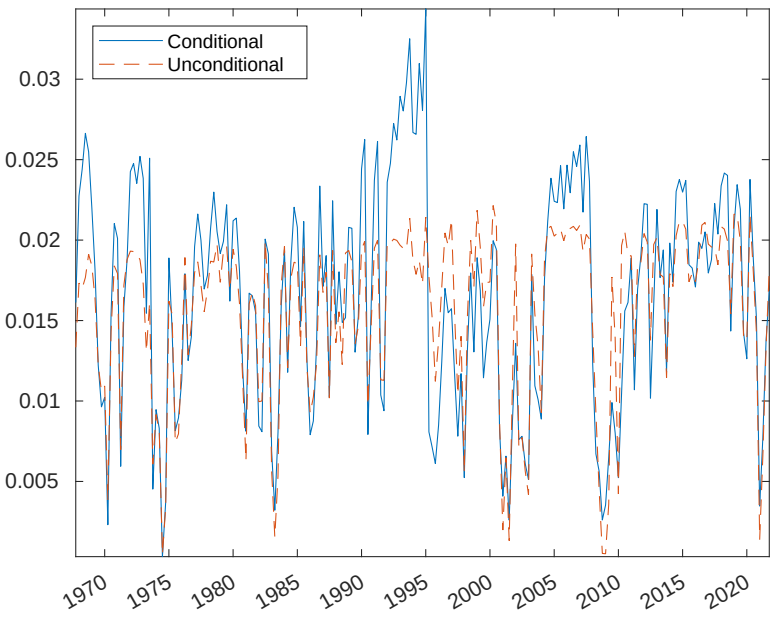
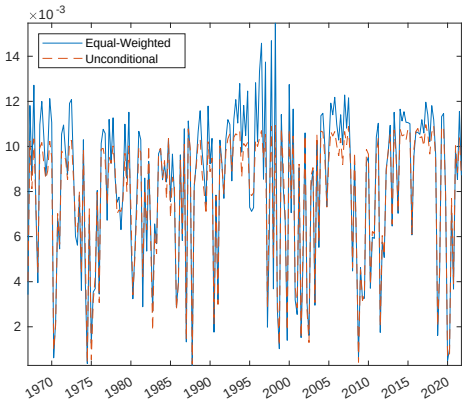


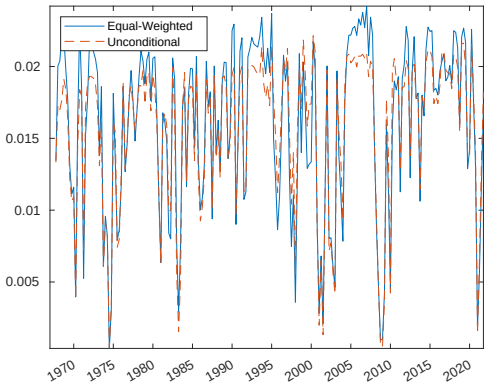
Figure SA-9: **Out-of-sample Scores, Equal-Weighted**

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

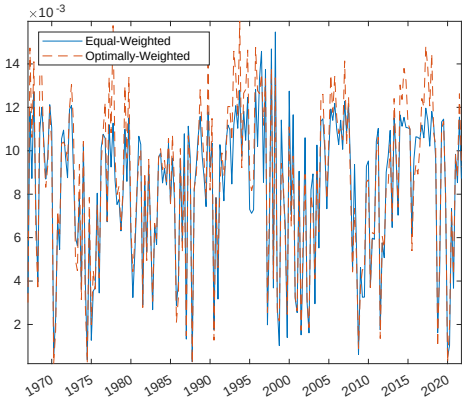
(a) **One quarter ahead**:
Equal-Weighted vs. Unconditional



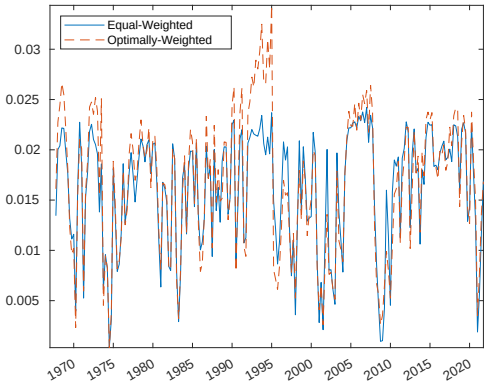
(b) **Four quarters ahead**:
Equal-Weighted vs. Unconditional



(c) **One quarter ahead**:
Equal-Weighted vs. Optimally-Weighted



(d) **Four quarters ahead**:
Equal-Weighted vs. Optimally-Weighted

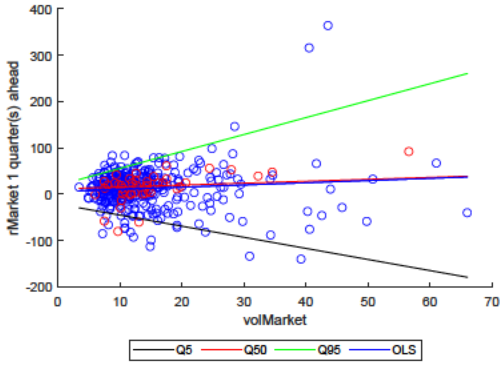


SA.3 Additional Figures

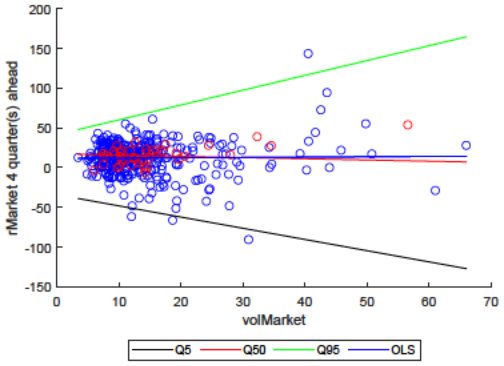
Figure SA-10: rMarket vs. volMarket

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 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.

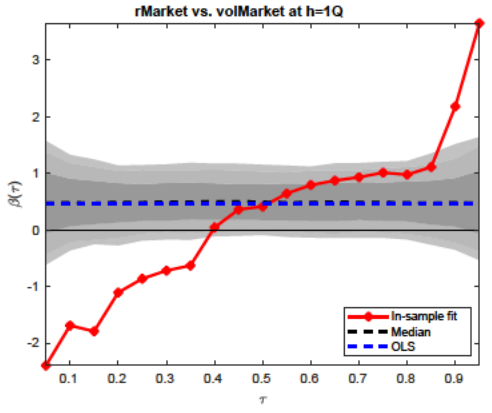
(a) **One quarter ahead**: Scatter



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



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



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

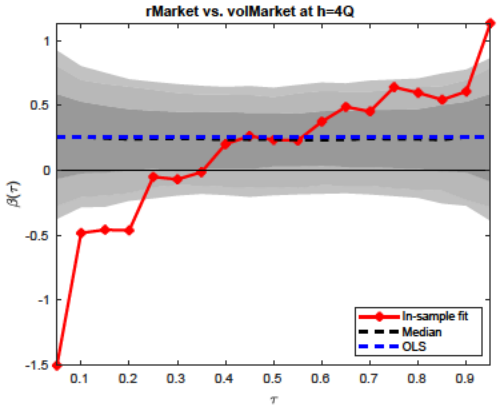
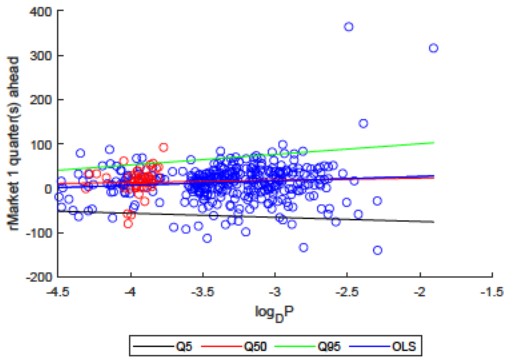


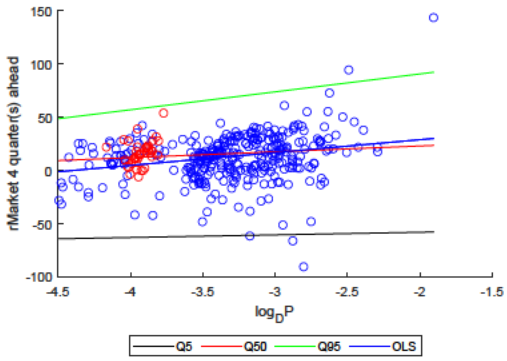
Figure SA-11: rMarket vs. logDP

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 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.

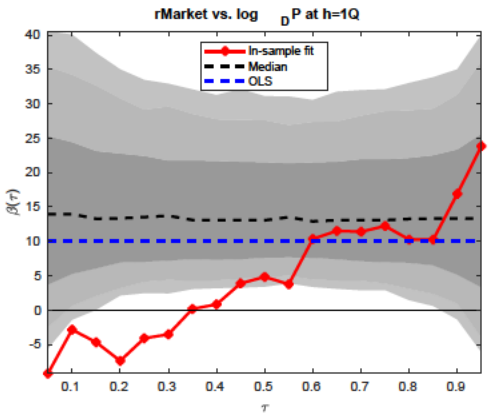
(a) **One quarter ahead**: Scatter



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



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



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

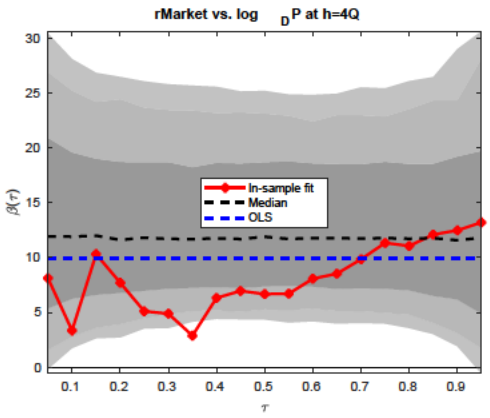
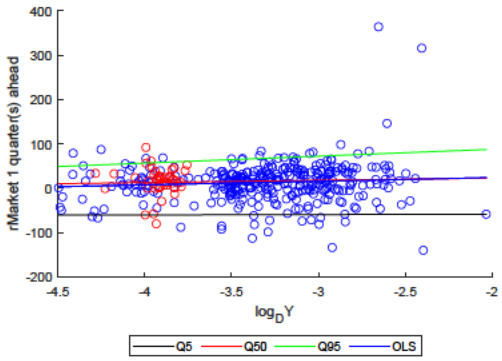


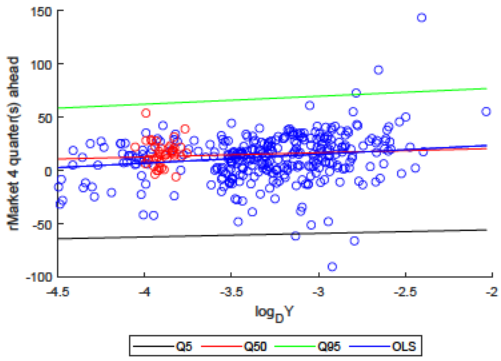
Figure SA-12: rMarket vs. logDY

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 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.

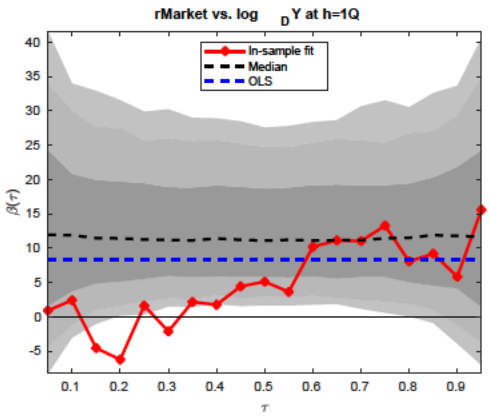
(a) **One quarter ahead**: Scatter



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



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



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

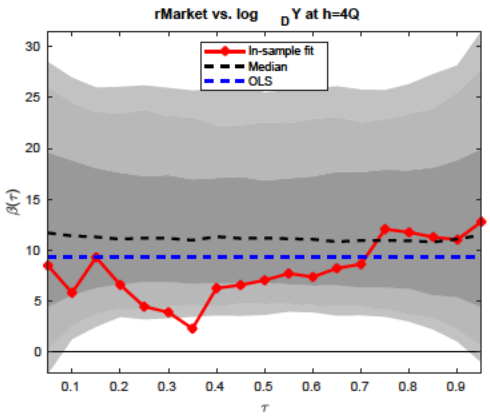
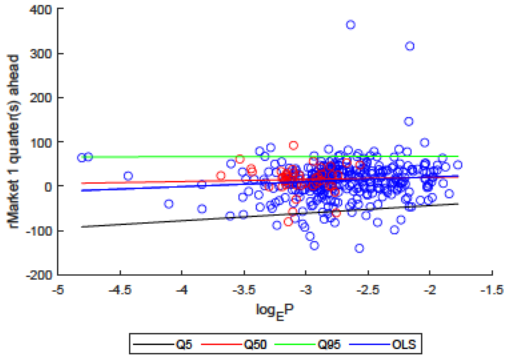


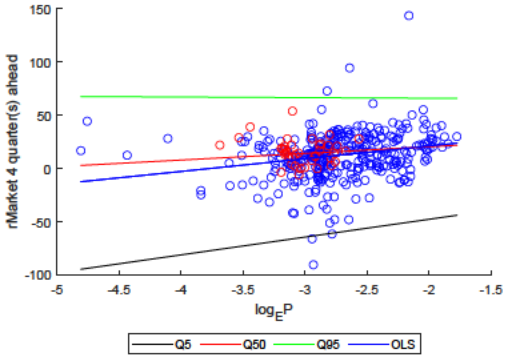
Figure SA-13: rMarket vs. logEP

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 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.

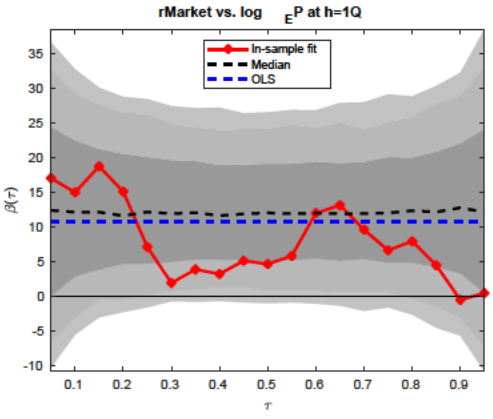
(a) **One quarter ahead**: Scatter



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



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



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

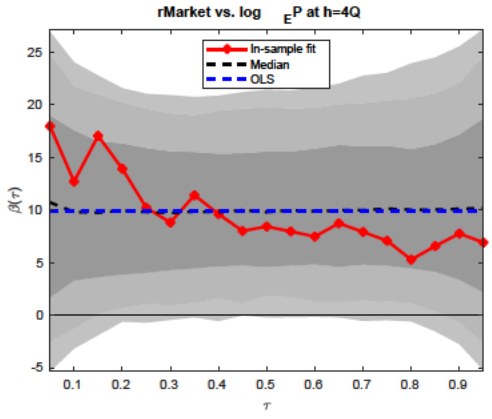
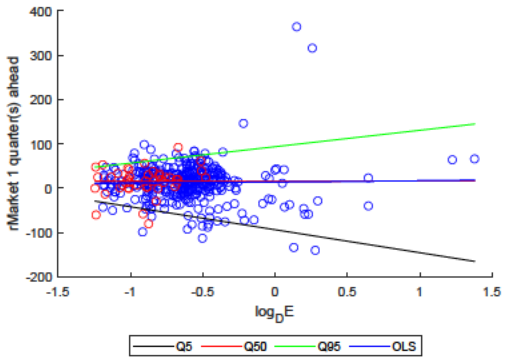


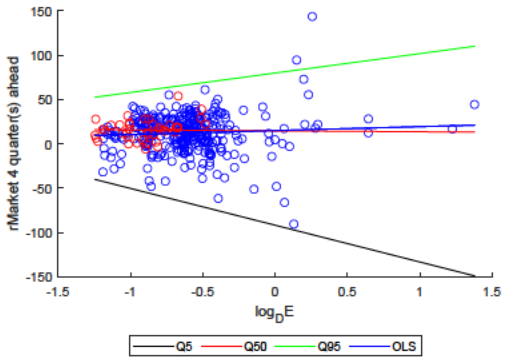
Figure SA-14: rMarket vs. logDE

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 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.

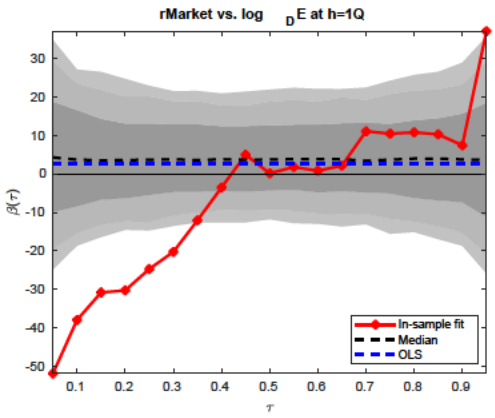
(a) **One quarter ahead**: Scatter



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



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



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

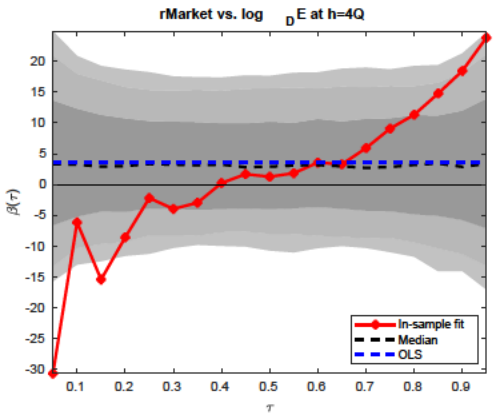
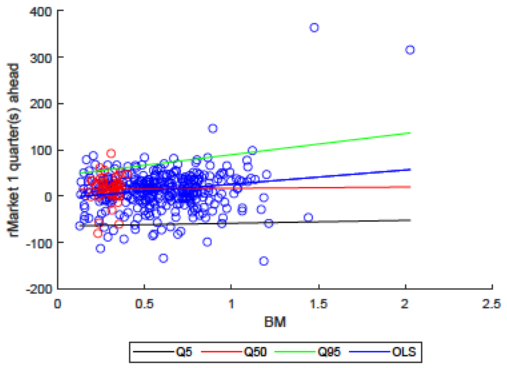


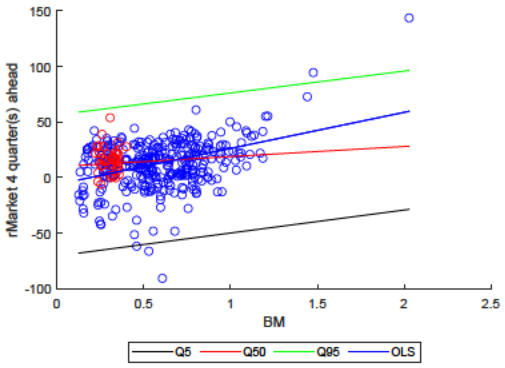
Figure SA-15: **rMarket vs. BM**

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 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.

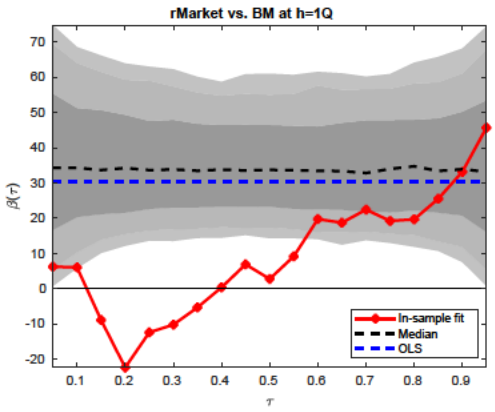
(a) **One quarter ahead**: Scatter



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



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



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

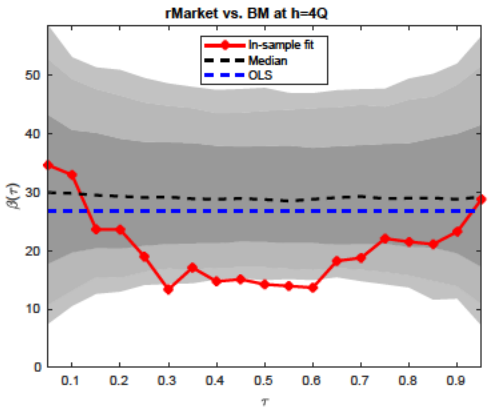
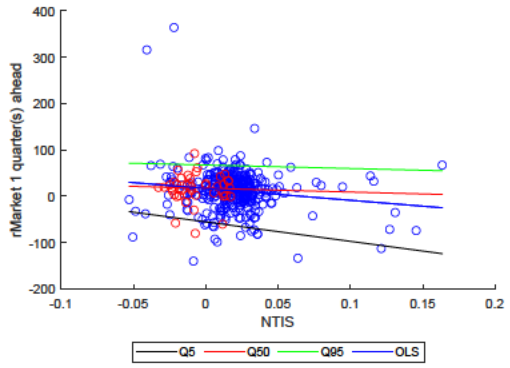


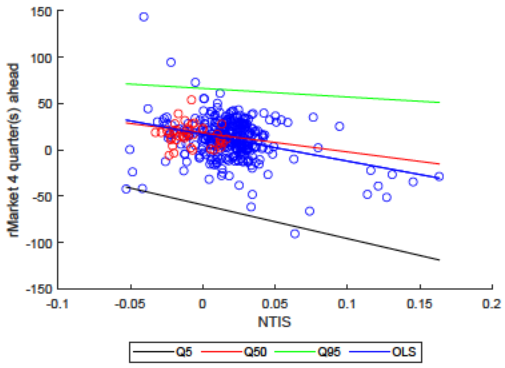
Figure SA-16: rMarket vs. NTIS

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 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.

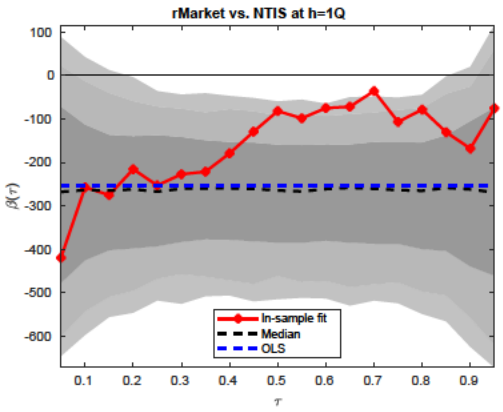
(a) **One quarter ahead**: Scatter



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



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



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

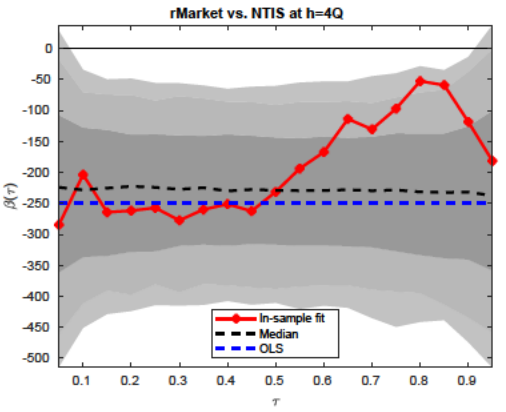
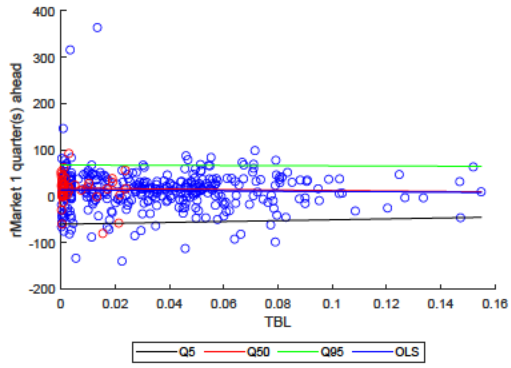


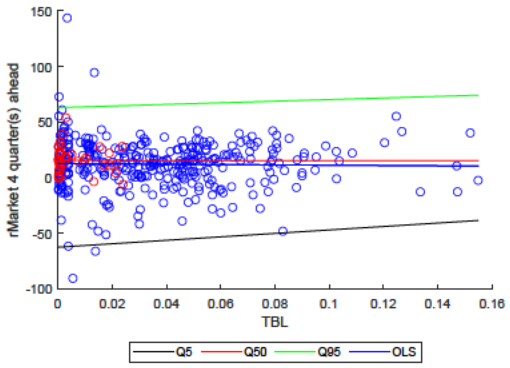
Figure SA-17: rMarket vs. TBL

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 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.

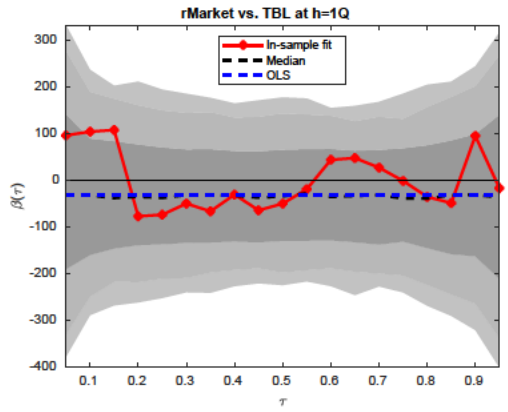
(a) **One quarter ahead**: Scatter



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



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



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

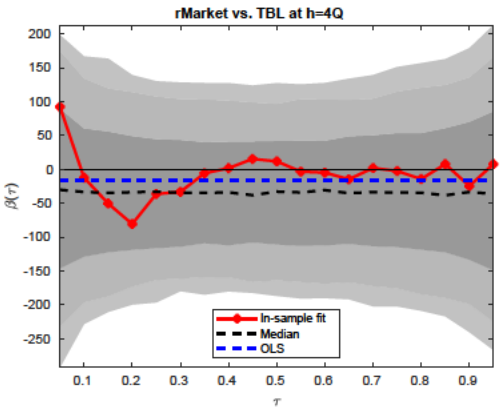
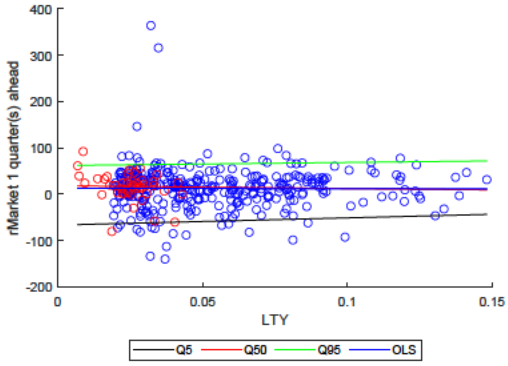


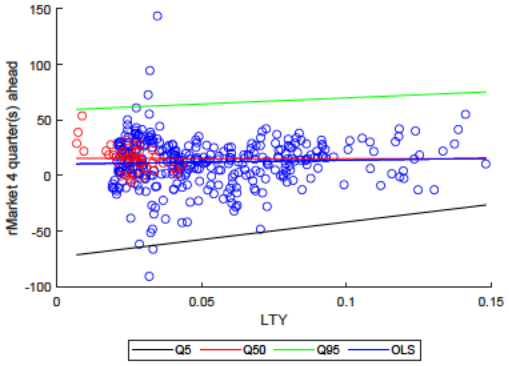
Figure SA-18: **rMarket vs. LTY**

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 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.

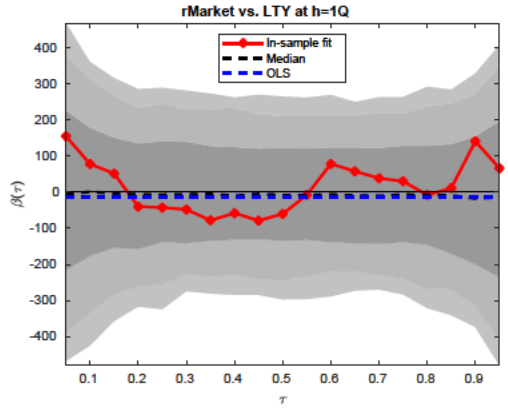
(a) **One quarter ahead**: Scatter



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



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



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

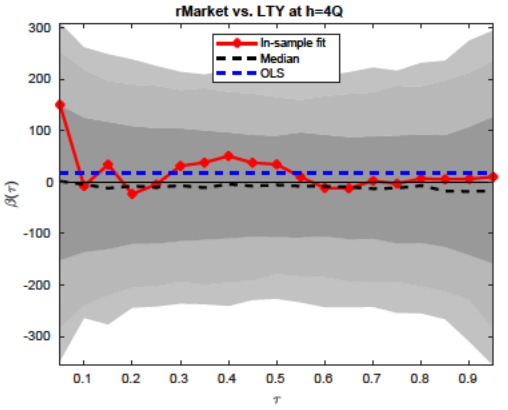
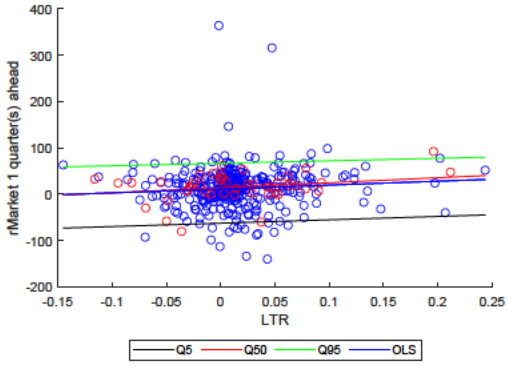


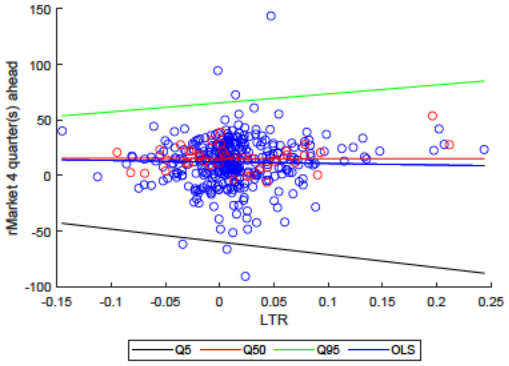
Figure SA-19: rMarket vs. LTR

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 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.

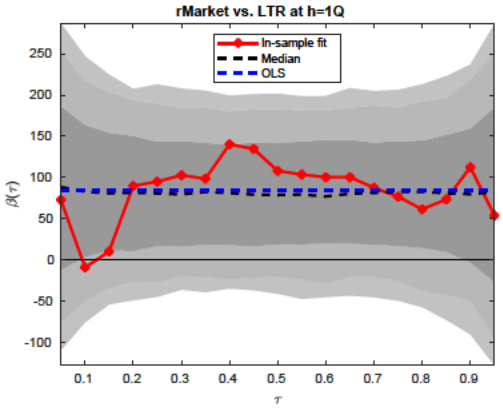
(a) **One quarter ahead**: Scatter



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



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



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

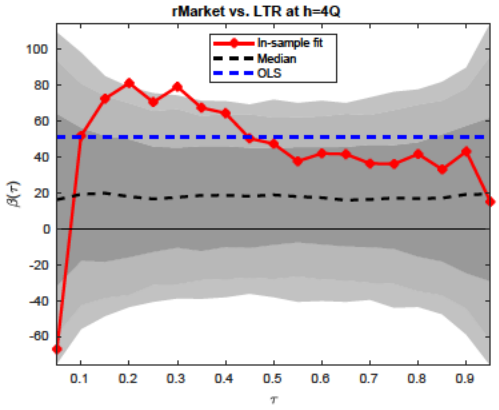
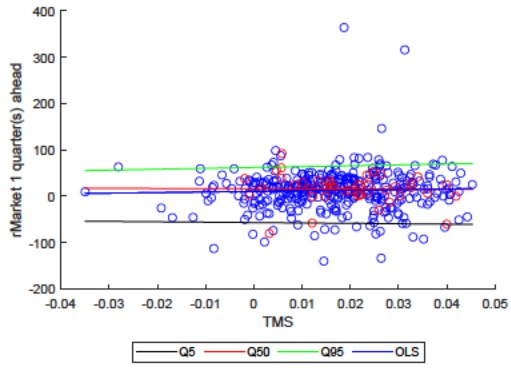


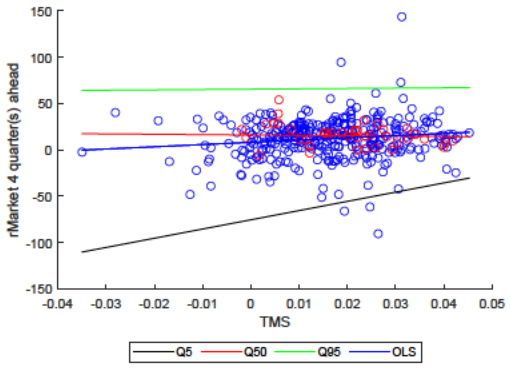
Figure SA-20: rMarket vs. TMS

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 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.

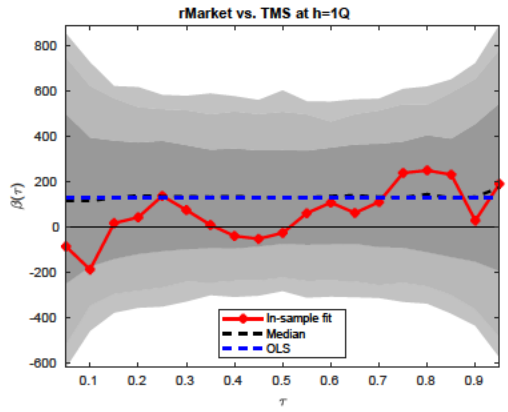
(a) **One quarter ahead**: Scatter



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



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



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

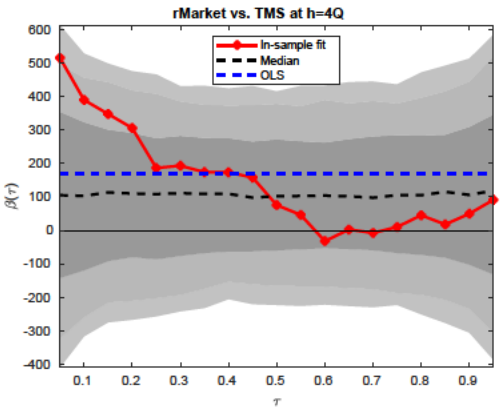
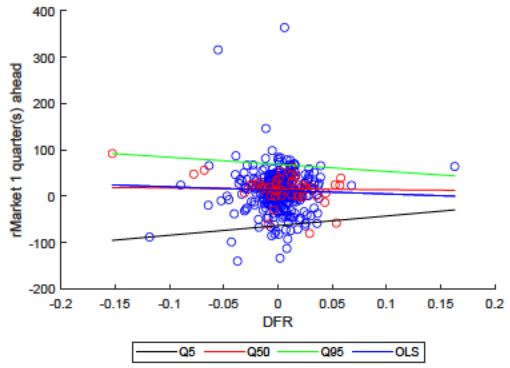


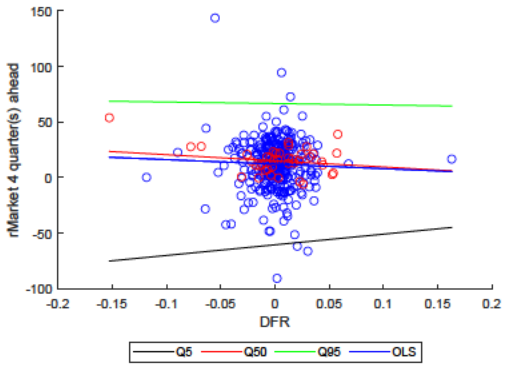
Figure SA-21: **rMarket vs. DFR**

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 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.

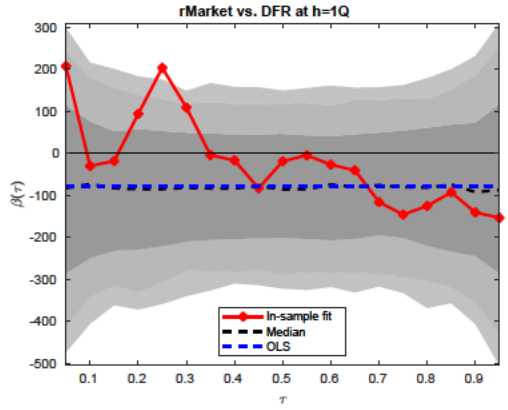
(a) **One quarter ahead**: Scatter



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



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



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

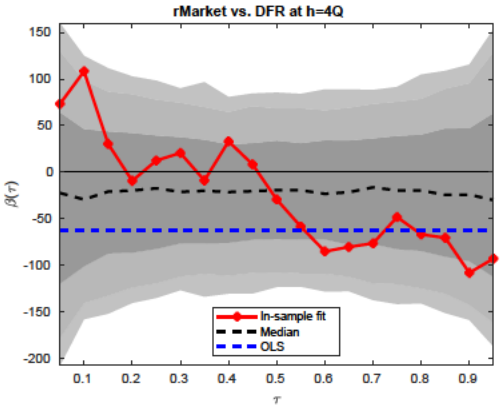
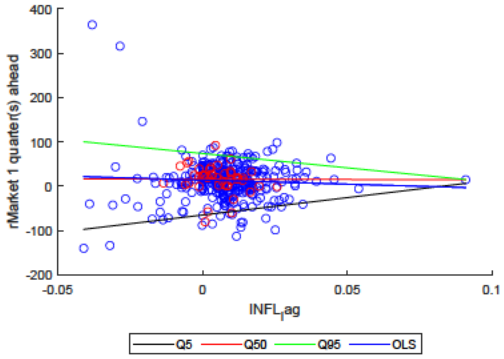


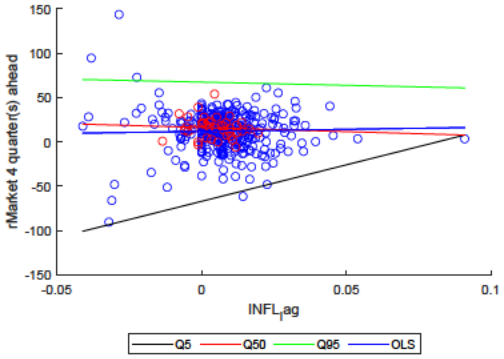
Figure SA-22: **rMarket vs. INFL**

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 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.

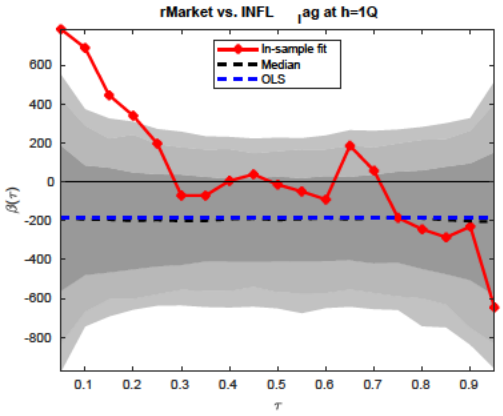
(a) **One quarter ahead**: Scatter



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



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



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

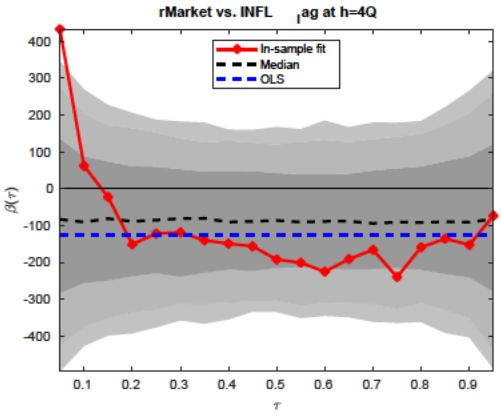
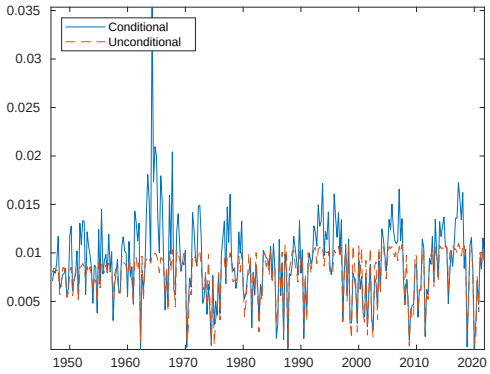


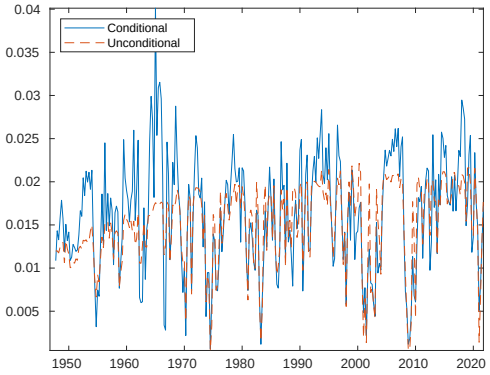
Figure SA-23: **Out-of-sample results: volMarket**

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.

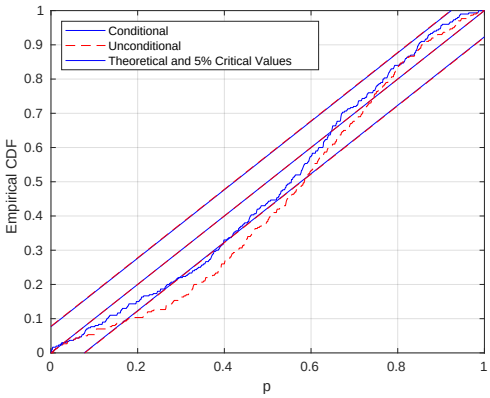
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

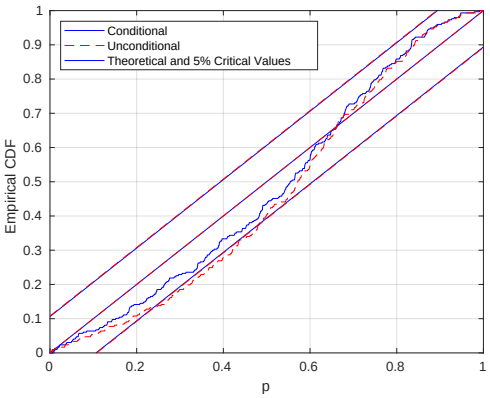
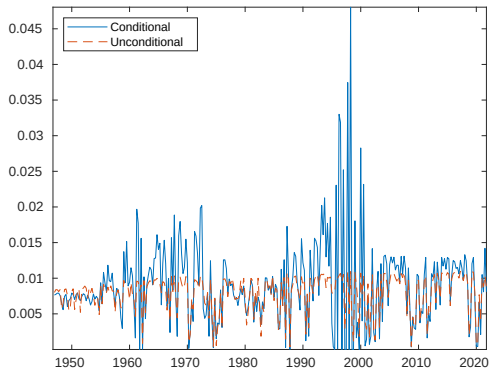


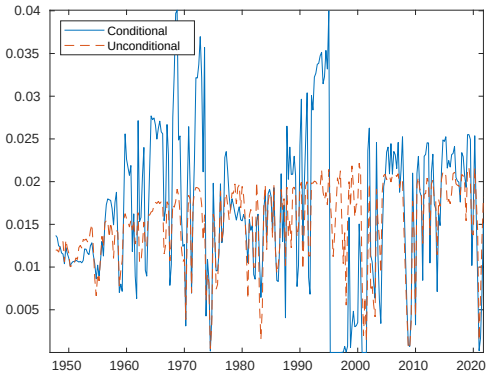
Figure SA-24: **Out-of-sample results: logDP**

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.

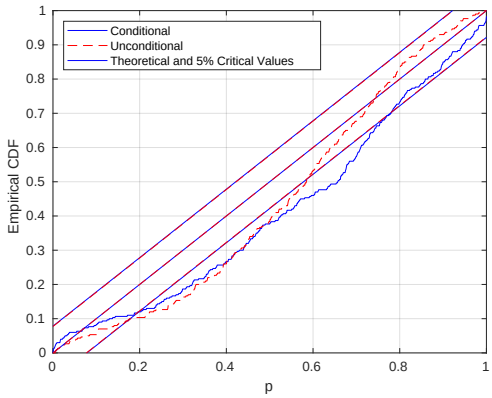
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

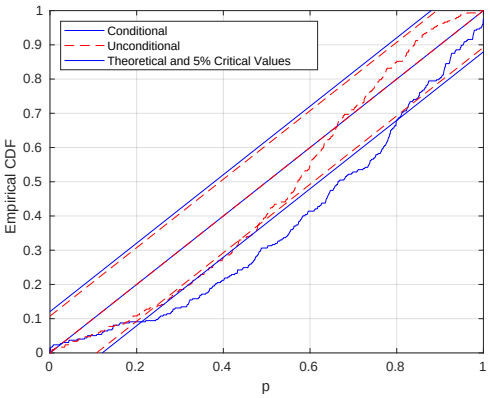
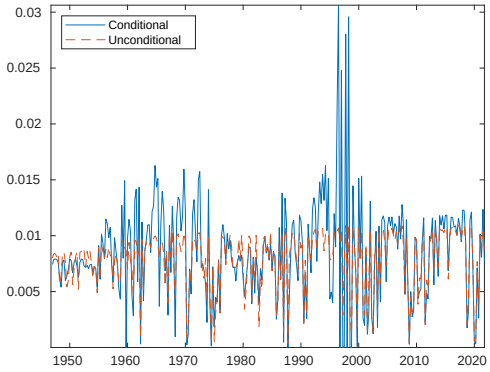


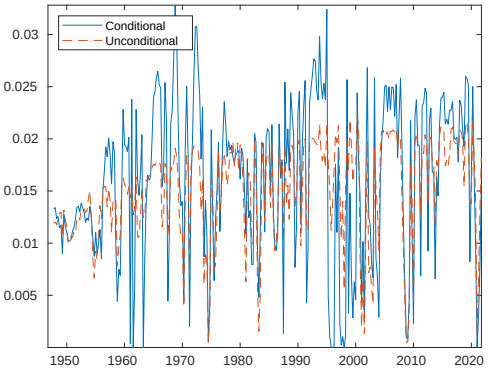
Figure SA-25: **Out-of-sample results: logDY**

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.

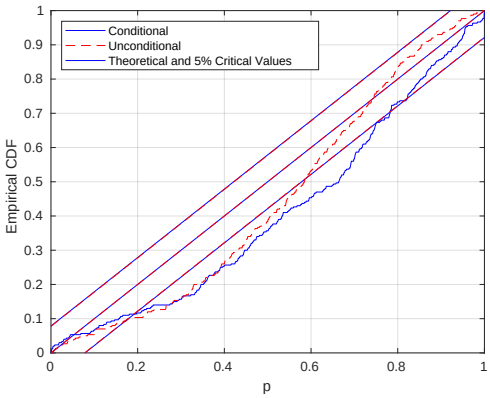
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

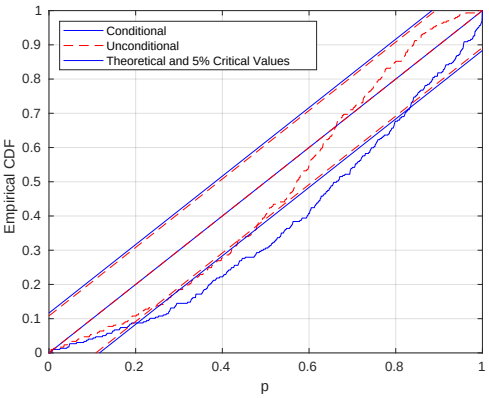
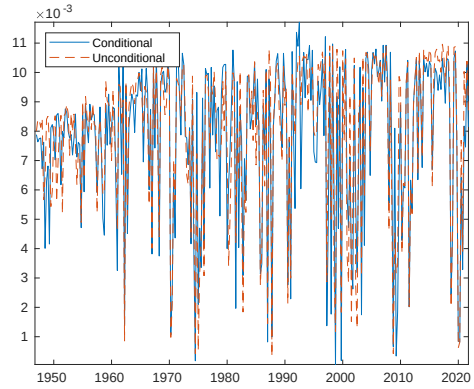


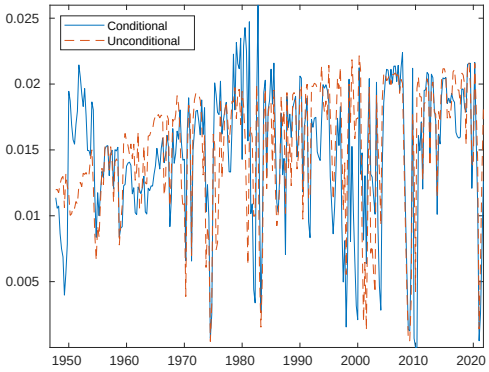
Figure SA-26: **Out-of-sample results: logEP**

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.

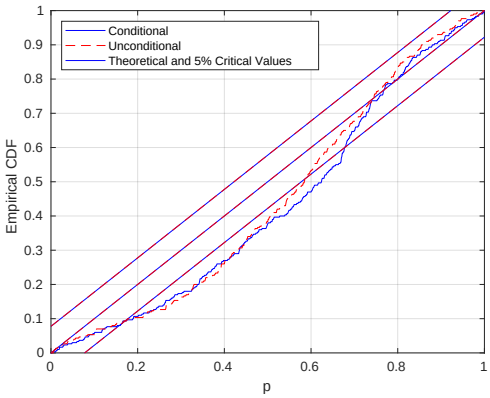
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

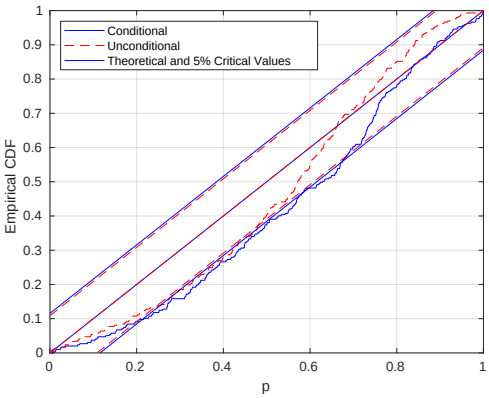
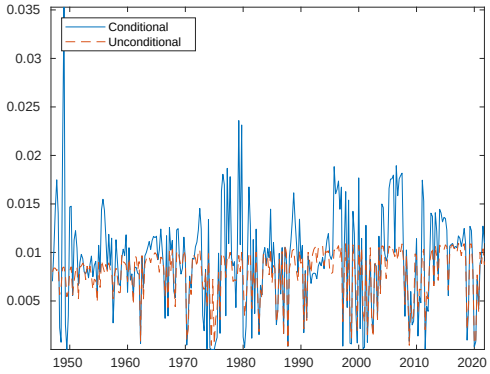


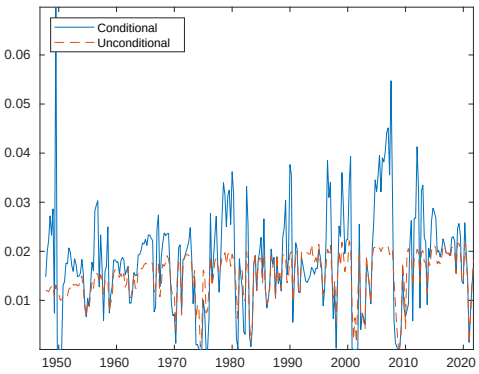
Figure SA-27: **Out-of-sample results: logDE**

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.

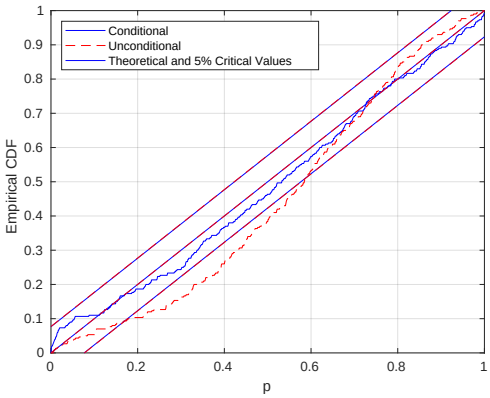
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

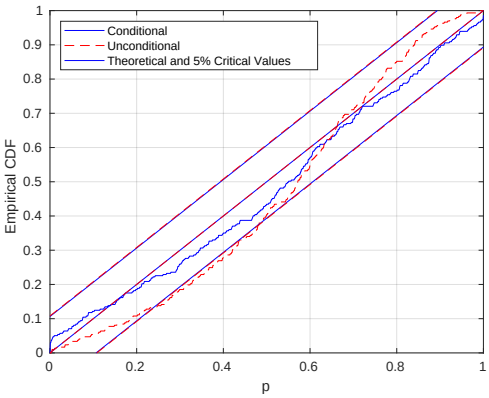
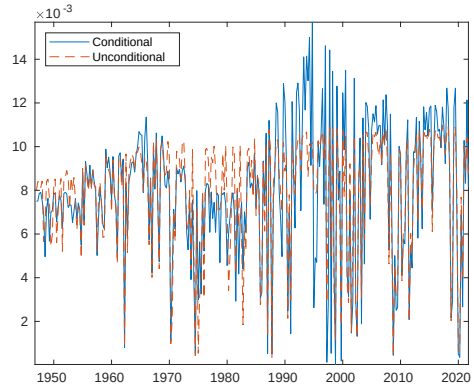


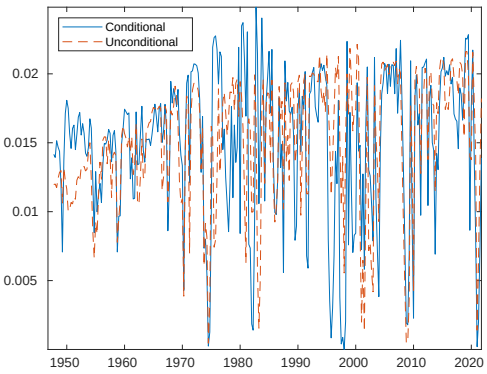
Figure SA-28: **Out-of-sample results: BM**

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.

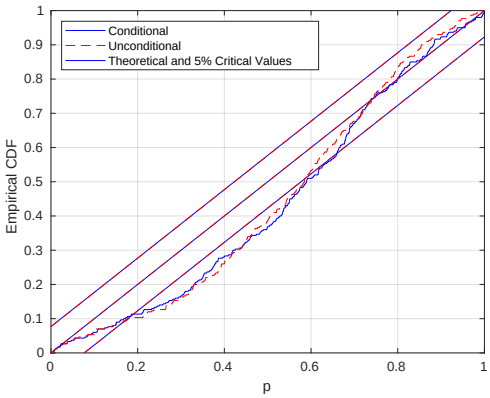
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

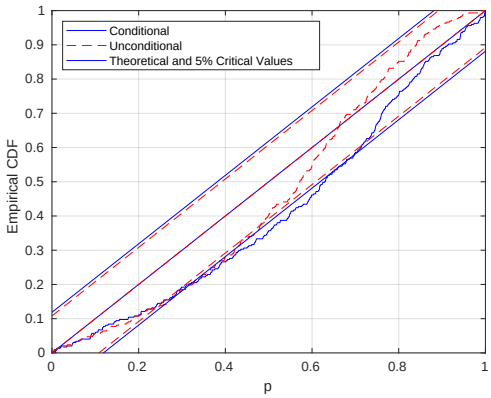
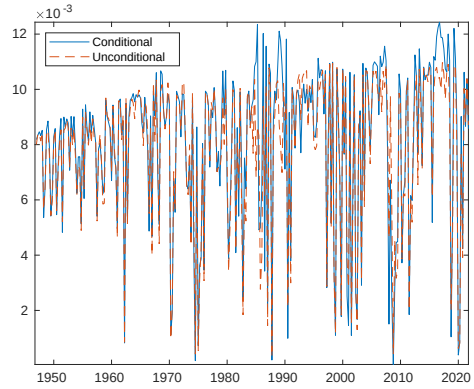


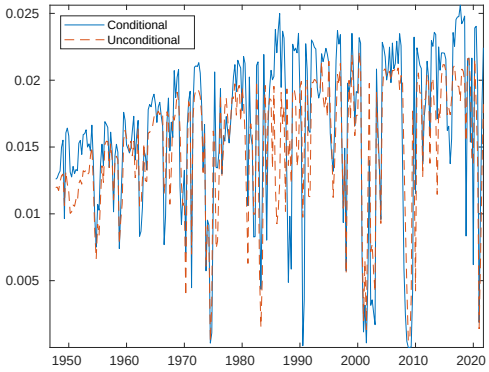
Figure SA-29: **Out-of-sample results: NTIS**

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.

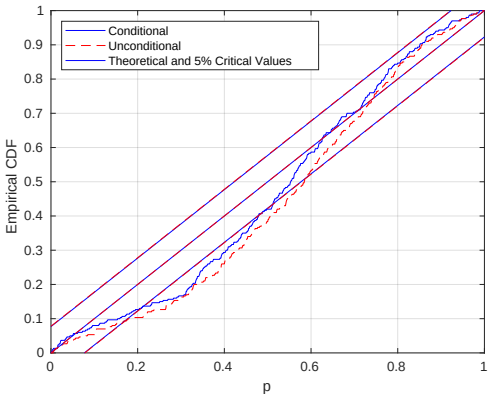
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

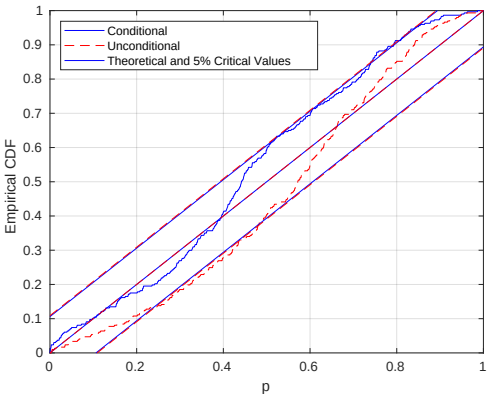
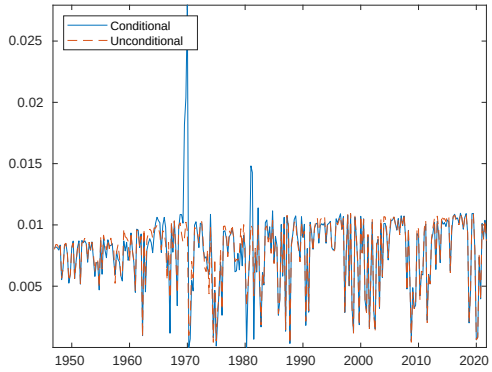


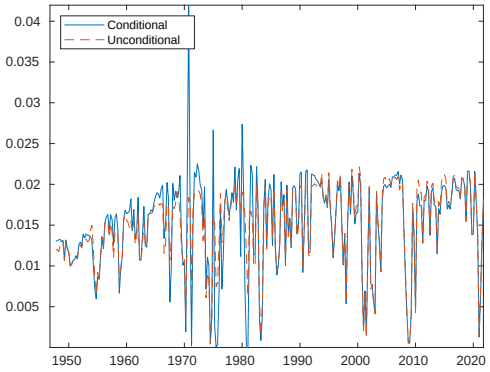
Figure SA-30: **Out-of-sample results: TBL**

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.

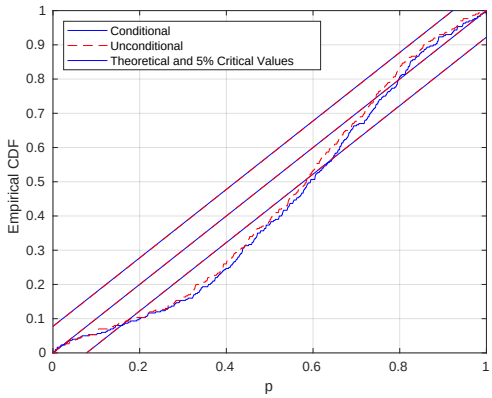
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

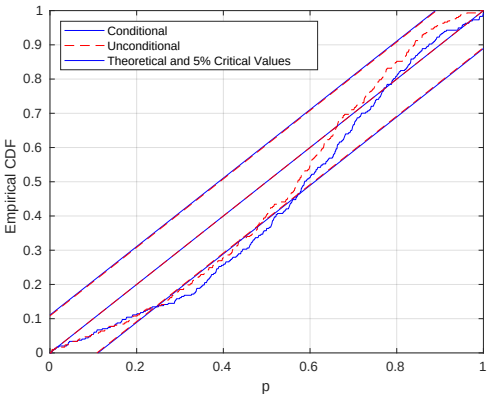
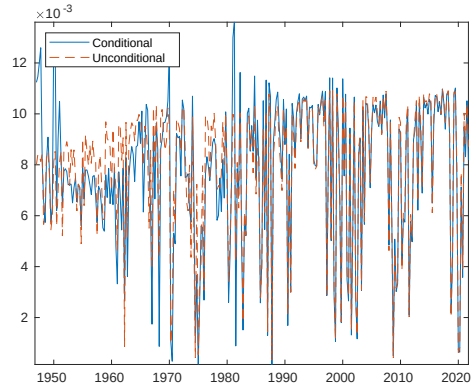


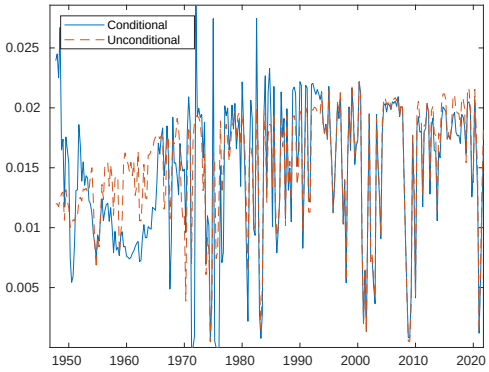
Figure SA-31: **Out-of-sample results: LTY**

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.

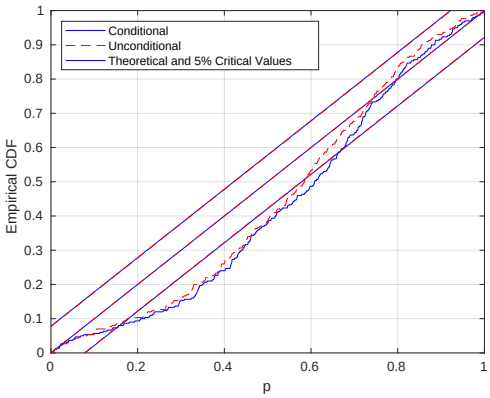
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

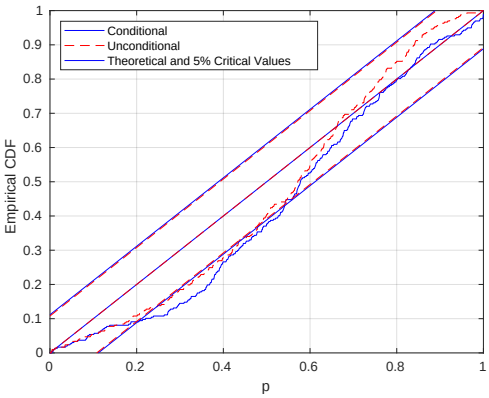
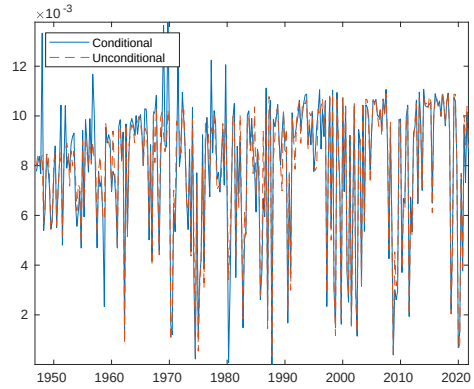


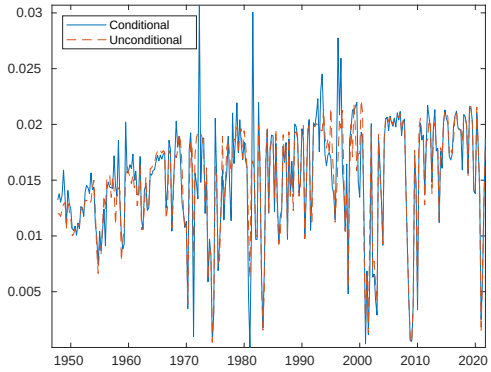
Figure SA-32: **Out-of-sample results: LTR**

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.

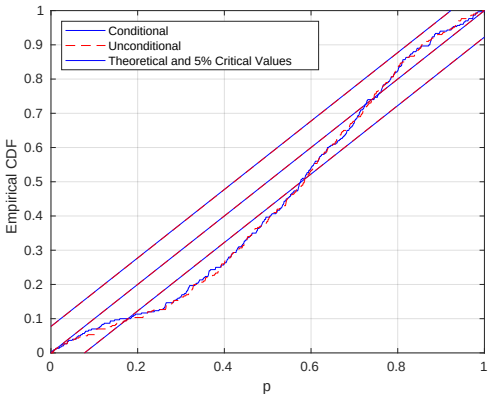
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

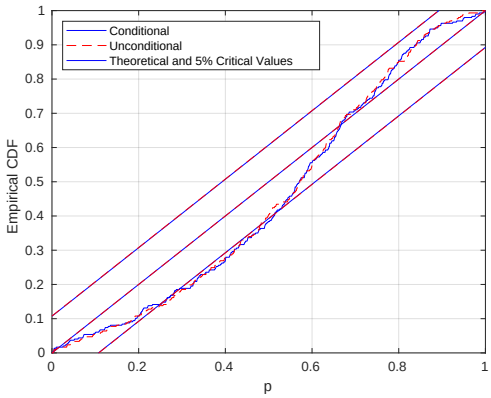
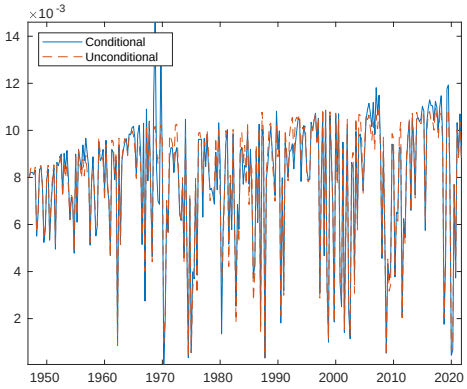


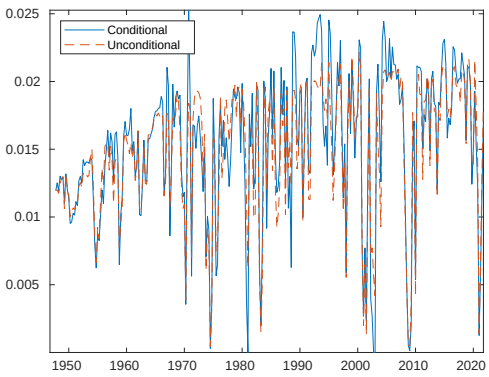
Figure SA-33: **Out-of-sample results: TMS**

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.

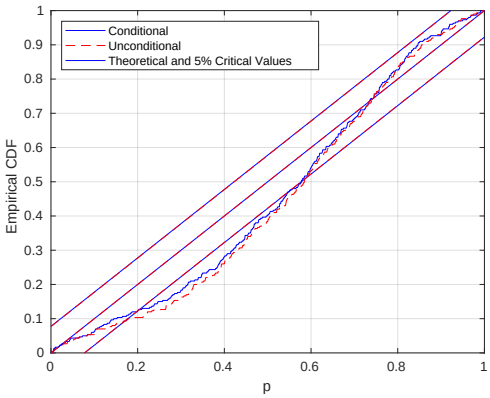
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

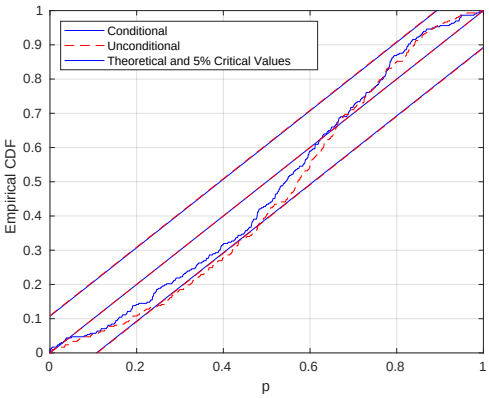
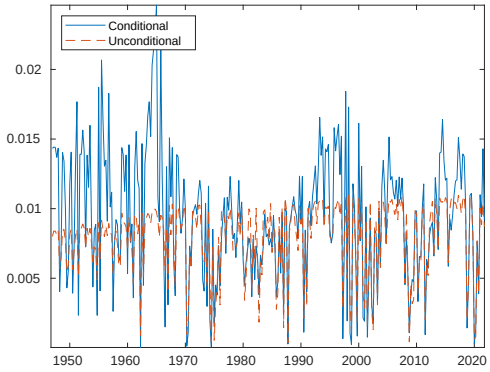


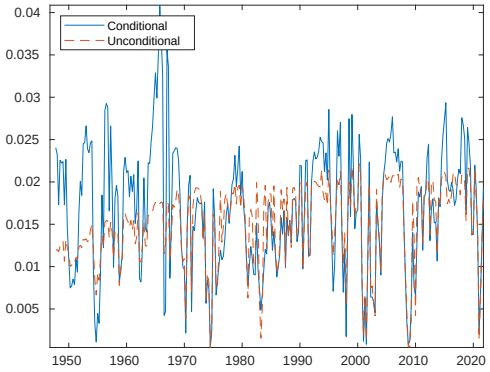
Figure SA-34: **Out-of-sample results: DFY**

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

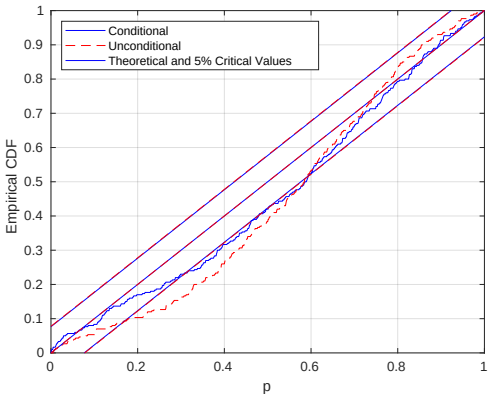
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

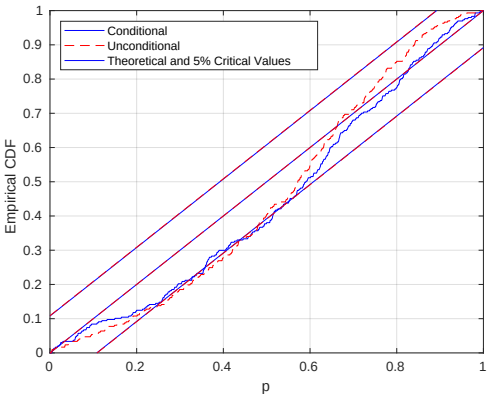
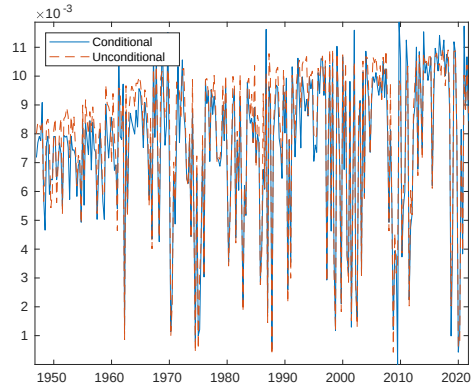


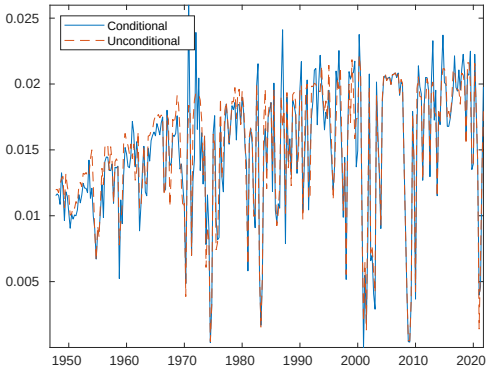
Figure SA-35: **Out-of-sample results: DFR**

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.

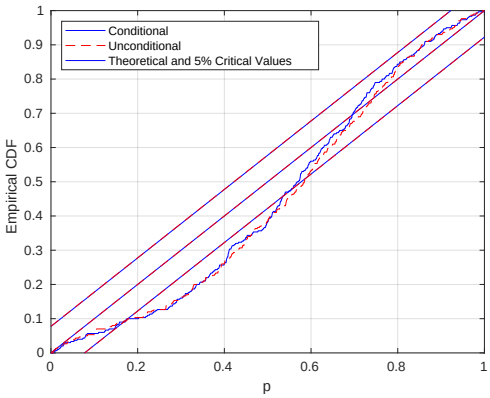
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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

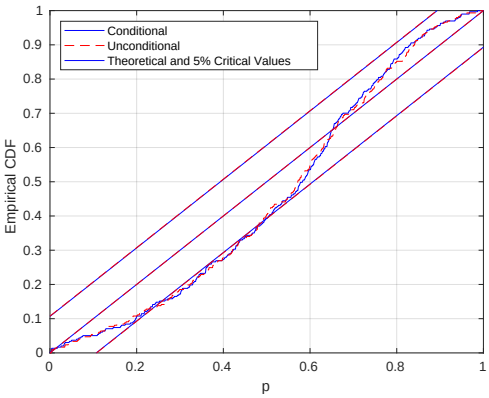
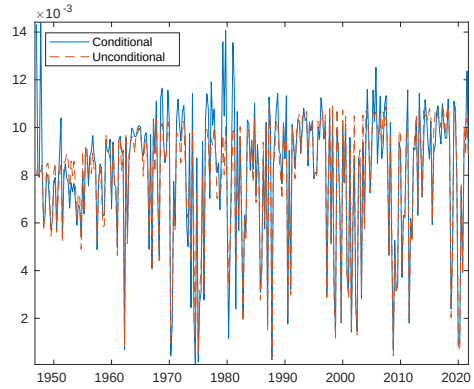


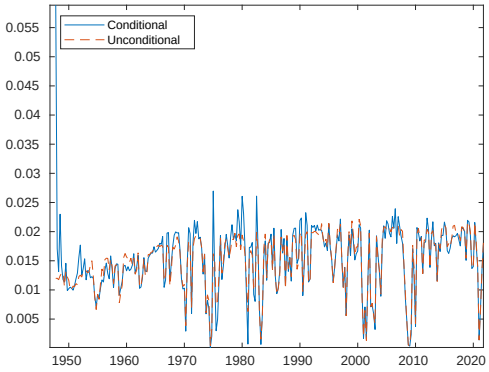
Figure SA-36: **Out-of-sample results: INFL**

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.

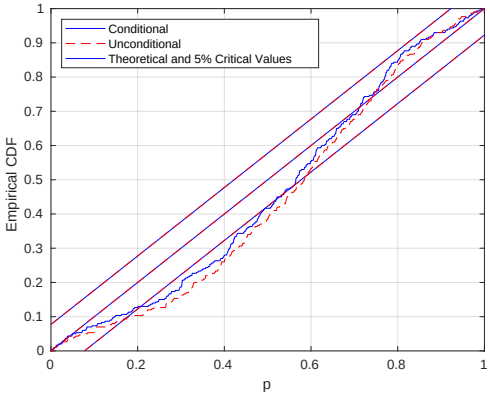
(a) **One quarter ahead**:
Out-of-sample scores



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



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



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