Modeling The Time-Varying Skewness via Decomposition For Out-of-Sample Forecast

Liu, Xiaochun

Department of Economics, Emory University

30 August 2011

Modeling The Time-Varying Skewness via Decomposition For Out-of-Sample Forecast*

Xiaochun Liu†

August 28, 2012

Abstract

This paper models time-varying skewness for financial return dynamics. We decompose financial returns into the product of the absolute returns and signs, so-called the intriguing decomposition. The joint distribution between the decomposed components is modeled through a copula function with marginals. Allowing the copula dependence parameter time-varying, we estimate the dynamic nonlinear dependence between absolute returns and signs, which governs time-varying skewness for out-of-sample forecast of financial returns.

The empirical results in this paper show that the proposed models with dynamic dependence obtain better gains of out-of-sample forecast, and suggest the robust strategy for a risk-averse investor in response to the market timing. This paper also explores the sources of the forecasting performance via a recently developed econometric pin-down approach. Beyond the pure statistical sense, we find that the forecasts of time-varying skewness trace closely to NBER-dated business-cycle phases.

Keywords: Time-varying skewness, Dynamic nonlinear dependence, Copulas, Out-of-sample forecast, Sources of forecasting performance

JEL: G17 G12 C53 C58

*This paper was previously titled as 'Modeling Robust Financial Return Dynamics: The Timing and Sources of Forecasting Performance and Links to Real Economy'.

†Department of Economics, Emory University, USA. Email: xiaochun.liu@emory.edu

1 Introduction

An intriguing decomposition framework decomposes a financial return into the product of its absolute return and its sign, e.g., Anatolyev and Gospodinov (2010), Christoffersen and Diebold (2006), Christoffersen et. al. (2007), Hong and Chung (2003), Linton and Whang (2007), Thomakos and Wang (2010), Chung and Hong (2007), among others. Let r_t denote the excess return at time t, the intriguing decomposition is given by.

$$r_t = |r_t| sign(r_t) \tag{1.1}$$

This simple return decomposition presents the potential nonlinear dependence between the return higher moments such as volatility, skewness, downside risk, and the direction-of-change, etc. For instance, the negative signs of financial returns materialize the market downturn generally associated with higher return volatility and intriguing the asymmetric effects among investors. As such, the intuition behind the intriguing decomposition is that the nonlinear dependence between absolute returns and signs contain valuable information for return forecasting, through a direct connection between return volatility forecastability and sign forecastability. This connection has been revealed in the recent works. Christoffersen and Diebold (2006) show that volatility dependence produces sign dependence to help forecast the market timing, which is not likely to be detected via traditional methods due to the special nonlinear nature of sign dependence and the dynamics between absolute returns and signs. Anatolyev and Gospodinov (2010) show promising out-of-sample forecasting results through the modeling of the constant nonlinear dependence between absolute returns and signs.

As in Christoffersen and Diebold (2006), the nonlinear "common feature" raises the questions to understand and model the nonlinearities between absolute returns and signs in any possible forms. The conditionally non-Gaussian environments, for example, with time-varying conditional skewness remain untapped for the link between volatility dependence and sign dependence. In particular, this paper focuses on the higher moments' dynamics of financial returns. We pretest the constancy of nonlinear dependence structure (see details in Section 2) and show strong evidence for dynamic dependence and time-varying asymmetry across different quantiles/tails. The pretesting results imply that time-varying higher moments should be appropriately modeled and estimated. The test results are also in line with Jondeau and Rockinger (2006, 2007, 2012) and Patton (2004) who show the importance of time variability in higher moments for asset allocation.

This paper models the time-varying skewness for the out-of-sample forecast of financial returns. We propose a class of novel dynamic dependence structures within the intriguing decomposition framework: (i) to forecast out-of-sample returns, (ii) to constructively explore both the timing and sources of forecast performance via an econometric pin-down approach, (iii) and to economically link the forecasting performance to real economy for economic justification. The joint distribution with dynamic dependence between absolute returns and signs is modeled as a copula function with two marginals: a multiplicative error model for the marginal of absolute returns and a dynamic logistic linked model for the marginal of signs. The proposed evolution equations for copula functions, depending on the previous values of the dependence parameters and the historical data, allow modeling the financial return dynamics in tail dependence and varying in asymmetry over time.

Further, the forecast performance is econometrically pin-downed to three source components, namely instability in models' forecast ability, difference in predictive content and the over-fitting issue. This econometric pin-down approach based on Rossi and Sekhposyan (2011) attempts to understand the reasons why a competing model perform better or worse than benchmarks. In addition beyond a pure statistic sense, it is also important to see if the forecast performance is driven by the fluctuation of real economy. To our best knowledge, the latter two investigations have barely been studied in literature.

We apply the proposed models to the forecast of U.S. financial returns. The empirical results in this paper show that the proposed dynamic decomposition models consistently perform better out-of-sample than benchmarks including historical average, linear predictive regression, and constant decomposition model. The largest difference in forecast gains has been observed as 15.88% of the proposed dynamic decomposition models over the constant decomposition models.

More interestingly, the out-of-sample performance exhibits specific market timing forecastability, which is remarkably useful for a risk-averse investor. For instance, the dynamic decomposition models assuming a conditional dependence between absolute returns and signs have much better out-of-sample performance during the studied turmoil times, whereas they seem not performing equally well in the tranquil periods; by contrast, we obtain the opposite performance for the dynamic models assuming a conditional independence. Intuitively, the better forecasting performance of the dynamic conditional dependence models reflects the significantly increasing correlations between the volatility (absolute returns) and asymmetric effects (signs) in a turmoil period. In contrast, the comovement between the volatility and asymmetric effects is low in a normal time. These empirical results seem to suggest the robust strategy that an investor effectively benefits from applying the different dynamic decomposition models in different economic states.

We employ a range of recently proposed approaches to test the empirical results, including a direct comparison based on conditional predictive ability test of Giacomini and White (2006) and an indirect comparison for out-of-sample density forecasts based on Diks et al. (2010). As a result, the tests further consolidate the empirical findings in this paper.

This paper also emphasizes in a good understanding for the reasons why the forecasting models perform differently. We apply the econometric pin-down approach, recently proposed by Rossi and Sekhposyan (2011), to extract the uncorrelated sources of forecasting performance. The identified sources of a model's superior forecasting performance is capable of providing valuable information for improvement, for instance, the abilities to model the market timing associated with time-varying skewness.

We gain better understanding of the forecasting performance sources in a statistical sense; nevertheless, it tells very little about the economic sense of forecasting performance. To provide insights on the economic sources of forecasting performance, we analyze whether the forecast performance can be significantly explained by real economy. A few studies in the literature have pursued in this respect, e.g., Rapach et al. (2010), Ghysels et al. (2011), among others. Our results show that the forecasts of time-varying return higher moments exhibit well-defined patterns to trace closely to NBER-dated business-cycle phases, and precisely predict turning points for those peaks and troughs. A further investigation shows the significant relations with macroeconomic variables, i.e., term spread, term premium, interest rate, housing price, inflation and industrial production growth.

The rest of this paper is structured as follows. Section 2 pretests the U.S. financial returns for the existence of dependence and the constancy of copulas between absolute returns and signs. Section 3 presents our dynamic decomposition models. Section 4 provides forecasting methods. Section 5 describes data. Section 6 reports and evaluates the empirical results of out-of-sample forecasts. Section 7 applies the econometric pin-down approach to understand the sources of forecasting performance. Section 8 links the out-of-sample forecasts to real economy. Section 9 concludes this paper.

2 Pretests Based on Empirical Copula Process

We pretest the U.S. stock return data for the following concerns: (1) whether the empirical evidence exists for the dependence between the absolute returns and signs? (2) whether the empirical evidence supports the time-varying dependence in the different parts of the distribution? (3) which copula function is chosen by the copula goodness-of-fit tests?

Panel A in Table 1 reports the test results for the existence of dependence based on Genest and Rémillard (2004, 2006, 2007). Parametric bootstrap is used to obtain critical values. The tests, based on the empirical copula processes, have asymptotic distributions of Cramér-von Mises. The null hypotheses of both independence and serial independence are rejected at 5% and 1% significance levels, respectively. The results present a statistically substantial degree of dependence between absolute returns and signs. In Panel B of Table 1, the copula goodness-of-fit tests, based on Genest, Remillard and Beaudoin (2009), Genest and Remillard (2008), Kojadinovic, Yan and Holmes (2011), Kojadinovic and Yan (2010), and Kojadinovic and Yan (2011), are in favor of Clayton copula, while reject Gumbel copula and other elliptical copulas. The favor in Clayton copula, a distribution with a lower tail dependence, indicates the significant asymmetric effects in the U.S. return data.

[Table 1 about here]

Panel C in Table 1 reports the copula constancy tests based on Busetti and Harvey (2011) and Harvey (2010). Their test approaches are flexible and useful in the way that tests associated with different quantiles may point to changes in the different parts of a copula distribution. τ is the chosen quantile level. For $\tau = 50\%$, the test results reject both the constant lower and the constant upper tail dependencies, while the constancy of overall copula is not rejected and there is no evidence showing a jointly changing in the lower-upper tails. Further, the test results show the significantly changing in the lower tail dependence at the quantile level, $\tau = 25\%$, whereas the upper tail counterpart remains constant at the quantile level, $\tau = 75\%$. Evidently, changing in asymmetry is statistically significant across all tested quantile levels. As a result, the evidence of changing in the lower tail dependence and varying in asymmetry over time is substantially strong for the joint behavior of absolute returns and signs.

Overall, the preliminary evidences empirically show both the existence of significant dependence, and time-varying tail dependence and asymmetry between the U.S. absolute returns and signs. Consequently, to forecast the financial returns, we pay our attention to modeling dynamic tail behaviors and time-varying asymmetry by adopting a Clayton copula.

3 Methodological Framework

The intriguing decomposition in eq(1.1) is generalized as:

$$r_t = c + sign(r_t - c)\,|r_t - c|$$

where c is a user-determined threshold for generality. Nonzero thresholds may be considered for e.g., transaction costs, and targeting rates, etc. Christoffersen and Diebold (2006) and Anatolyev and Gospodinov (2010) consider the leading case, $c = 0$. The return decomposition can be rewritten with the leading case as:

$$r_t = 2\left|r_t\right| \mathbb{I}(r_t > 0) - \left|r_t\right| \tag{3.1}$$

where $\mathbb{I}(\cdot)$ is an indicator function. The conditional expected return is the interest of a dynamic model given by

$$E_{t-1} r_t = 2 E_{t-1} \left|r_t\right| \mathbb{I}(r_t > 0) - E_{t-1} \left|r_t\right| \tag{3.2}$$

where $E_{t-1}(\cdot)$ denotes the expectation operator conditional on the information, I_{t-1}, available up to time $t-1$. For simplicity, I_{t-1} is suppressed as the subscript $t-1$.

The key to predict returns is to modeling the joint distribution of absolute values, $|r_t|$ and indicators, $\mathbb{I}(r_t > 0)$. This paper models the joint distribution through a copula function due to the mixed marginals of a continuous distribution for the absolute returns and a discrete binary distribution for signs. The dependence parameter of the copula function governs the nonlinear relation between absolute returns and signs. In this paper, we allow the copula dependence parameter time varying for the dynamic skewness.

Copula models are widely used in financial economics and risk management. Copulas, which are applied to financial time series data, are most of time treated as constant over time. However, it has become a stylized fact that correlations between financial data are not constant through time. The dynamic structure of the dependence between the data can be modeled by allowing either the copula function or the dependence parameter to be time varying. Patton (2006) proposes a parametric model to describe the evolution of the Gaussian copula parameter, while Jondeau and Rockinger (2006) estimate a time-varying skewed Student-t copula GARCH model for conditional dependencies between international stock markets.

Next we briefly introduce Anatolyev and Gospodinov's constant decomposition model; then, our dynamic dependence models are proposed.

3.1 Constant Decomposition Model

Let $u_t = |r_t|$ and $v_t = \mathbb{I}(r_t > 0)$. The joint density/mass function $f_{r_t}(u, v)$ between absolute returns and signs is derived in Anatolyev and Gospodinov (2010) as:

$$f_{r_t}(u_t, v_t) = f(u_t|\psi_t)\varrho_t[F(u_t|\psi_t)]^v \{1 - \varrho_t[F(u_t|\psi_t))]\}^{1-v} \tag{3.3}$$

where $\varrho_t(z) = 1 - \partial C(z, 1-p_t, \alpha)/\partial w_{1t}$ with $w_{1t} = F(u_t|\cdot)$ and $\psi_t = E_{t-1}|r_t|$. $F(u_t|\cdot)$ and $f(u_t|\cdot)$ are the marginal CDF and PDF of the absolute returns which will be specified in later sections. $p_t = E_{t-1}\mathbb{I}(r_t > 0)$ are the conditional expectations of signs. α, as the dependence parameter of a copula function, $C(\cdot)$, is constant.

Remark 3.1. The decomposition model estimates the dependence between absolute values and signs, and then predict the conditional expectation of returns via eq(3.2). It might be helpful to understand how the sample moments of returns are affected by the changes in α. We conduct a small simulation experiment.[1] The simulation results in Figure 1 show that the dependence parameter of Clayton copula is negatively related to the second and third moments of returns. For instance, when increasing in α, the return skewness is more negative, while volatility decreases; this indicates that increasing the dependence between absolute returns and signs raises the asymmetric effects and tail dependence, but reduces the risk of returns.

[Figure 1 about here]

[1]We take 200 points for $\alpha \in (0, 5]$. At each point, we randomly draw 10,000 pairs of data from Clayton copula, and then transform the data to absolute values and indicators through their marginals to generate simulated returns. We take the estimated scale parameter, $k = 1.275$, from Anatolyev and Gospodinov (2010) for the Weibull marginal distribution of absolute values. The conditional expectation of absolute values is estimated from the sample data and then resampled with replacement. We use the simulated returns to compute the sample volatility and skewness against 200 points of α.

Remark 3.2. In Figure 1, the solid lines are the levels of sample moments estimated from the U.S. return data, and the dot lines are the estimated dependence parameter, $\alpha = 0.087$, from the constant decomposition model of Anatolyev and Gospodinov (2010). As seen, the sample volatility and skewness request the corresponding dependence around 2.5 and 0.9, respectively, much higher than the estimated constant dependence parameter. The dependence underestimation consequently leads to the underestimated return skewness.

3.2 Dynamic Decomposition Models

The emphasis of this paper is to modeling the time-varying skewness associated with the dynamic tail behaviors and varying in asymmetry within the intriguing decomposition. As the joint distribution for absolute returns and signs, Clayton copula is given by:

$$C(w_{1t}, w_{2t}) = (w_{1t}^{-\alpha_t} + w_{2t}^{-\alpha_t} - 1)^{-1/\alpha_t} \tag{3.4}$$

and hence, $\varrho_t(z)$ in eq(3.3) is derived as

$$\varrho_t(z_t) = 1 - \left(1 + \frac{(1-p_t)^{-\alpha_t} - 1}{z_t^{-\alpha_t}}\right)^{-1/\alpha_t - 1} \tag{3.5}$$

where $\alpha_t > 0$ is varying over time as specified in eq(3.6)-(3.13) and $w_{2t} = 1 - p_t$. α_t, depending on its past values and historical data, governs the time-varying skewness.

3.2.1 Dynamic Tail Dependence

Patton (2006) proposes the observation driven copula models for which the time-varying dependence parameter of a copula is a parametric function of transformations of the lagged data and an autoregressive term. Patton (2006) assumes that the functional form of the copula remains fixed over the sample period, whereas the parameters vary according to some evolution equation. In general, the model for the evolution of a dependence parameter (or tail dependence) of a copula is specified as:

$$\alpha_t = \Lambda\left(\omega + \beta\Lambda^{-1}(\alpha_{t-1}) + \gamma\frac{1}{m}\sum_{i=1}^{m}|w_{1,t-i} - w_{2,t-i}| + \delta\mathbf{z}_{t-1}\right) \tag{3.6}$$

where $\Lambda(x) = exp(x)$ is the transformation for Clayton copula. $\mathbf{z}_t$ is a vector of exogenous variables. $w_{1t} = F(u_t|\psi_t)$ and $w_{2t} = 1 - p_t$ are the marginal CDFs of absolute returns and signs from eq(3.15) and (3.17). Patton choose $m = 10$. Note that the expectation of this distance measure is inversely related to the concordance ordering of copulas. In the empirical section, we refer this specification as Patton.

3.2.2 Exponentially Weighted (ExpWeight) Dynamic Dependence

As pointed out in Patton (2006), the difficulty in specifying how the parameters evolve over time lies in defining the forcing variables of the evolution equation. The eq(3.6) suggested by Patton (2006) is a martingale process depending on the past, m, path.

Based on eq(3.6), we propose a new evolution equation using the exponentially-weighted dynamic dependence as:

$$\alpha_t = \Lambda\left(\omega + \beta\Lambda^{-1}(\alpha_{t-1}) + \gamma\frac{1-\lambda}{1-\lambda^{t-1}}\sum_{i=1}^{t-1}\lambda^{i-1}|w_{1,t-i} - w_{2,t-i}| + \delta\mathbf{z}_{t-1}\right) \tag{3.7}$$

where $0 < \lambda < 1$ and $\frac{1-\lambda}{1-\lambda^{t-1}}\sum_{i=1}^{t-1}\lambda^{j-1} = 1$. The exponentially weighted dynamic dependence assigns more weights to the recent observations closer to time t, whereas it gives less weights to the past observations far away from time t. The intuition of this evolution equation is from the general fact that the latest events and observations have larger influence on current and near future dependence evolution than the far past.

3.2.3 GARCH-type (Gtype) Dynamic Dependence

Jondeau and Rockinger (2003) present various possible specifications for the dynamics of Skewed Student-t distribution parameters. Similar to the specification in their model (3), we take a linear structure of the type of heteroskedesticity autoregressive form with the exponential transformation for Clayton copula as:

$$\alpha_t = \Lambda\left(\omega + \beta\Lambda^{-1}(\alpha_{t-1}) + \gamma\left|w_{1,t-1} - w_{2,t-1}\right|\right) \tag{3.8}$$

with that the stability of the dynamics is assumed, for example, $|\beta| < 1$.

3.2.4 Integrated GARCH-type (IGtype) Dynamic Dependence

The integrated-GARCH type dynamics is a variant of eq(3.8) as follows, by restricting $\omega = 0$ and $\beta + \gamma = 1$:

$$\alpha_t = \beta\alpha_{t-1} + \gamma\left|w_{1,t-1} - w_{2,t-1}\right| \tag{3.9}$$

where $0 \leq \beta, \gamma < 1$.

Alternatively, eq(3.9) can be written in the recursive form of eq(3.7) as a simple exponential smoothing:

$$\alpha_t = \lambda\alpha_{t-1} + (1 - \lambda)\left|w_{1,t-1} - w_{2,t-1}\right|$$

3.2.5 One-Sided Asymmetric (OSA) Dynamic Dependence

The dynamic dependence parameter is modeled for the nonlinear relation between absolute returns and signs; it can be described that α_t measures the interdependence of absolute returns and signs as the uncentered product, $\vartheta_t = |r_t|\,\mathbb{I}(r_t > 0)$. Hence, we specify a new dynamic dependence as:

$$\alpha_t = \omega + \beta\alpha_{t-1} + \gamma\vartheta_{t-1} \tag{3.10}$$

The uncentered product is the second term in eq(1.1). Despite ϑ_t reveals the information for modeling α_t, the expectation of α_t, is not equal to the expectation of ϑ_t. For stationary, following the proof in Zakoian (1994) and Nelsen (1990, 1991), we have the conditions, $\omega > 0$, $0 \leq \beta < 1$, and $\gamma \geq 0$.

In the empirical section, we estimate eq(3.10) via a logarithm transformation for easy convergence purpose as

$$log\alpha_t = \omega + \beta log\alpha_{t-1} + \gamma\vartheta_{t-1} + \delta\mathbf{z}_{t-1}$$

with $|\beta| < 1$.

3.2.6 Two-Sided Asymmetric (TSA)Dynamic Dependence

An extension to eq(3.10) is a two-sided asymmetric tail dependence as:

$$log\alpha_t = \omega + \beta log\alpha_{t-1} + \gamma\vartheta^+_{t-1} + \lambda\vartheta^-_{t-1} + \delta\mathbf{z}_{t-1} \tag{3.11}$$

where $|\beta| < 1$. $\vartheta^+_t = |r_t|\,\mathbb{I}(r_t > 0)$, $\vartheta^-_t = |r_t|\,\mathbb{I}(r_t < 0)$. Such a specification has been also suggested by Glosten et al. (1993) and Zakoian (1994). Jondeau and Rockinger (2003) use the similar specification for a generalized skew-t distribution with time-varying parameters. Eq(3.11) allows modeling the dependence in terms of time-varying asymmetry.

3.2.7 TVC Dynamic Dependence

The time-varying dependence parameter, α_t, is related to Kendall's τ via $\tau_t = \frac{\alpha_t}{\alpha_t+2}$, which implies that $0 < \tau_t < 1$, due to $\alpha_t > 0$. The time variation in α_t can be modeled as $\alpha_t = \frac{2\tau_t}{1-\tau_t}$ with τ_t itself governed by the TVC-type equation of Tse and Tsui (2002) as:

$$\tau_t = \omega + \beta\tau_{t-1} + \gamma\tilde{\tau}_{t-1} \tag{3.12}$$

where $\omega > 0$, $0 \leq \beta, \gamma < 1$ and $\beta + \gamma < 1$. $\tilde{\tau}_{t-1}$ is the non-negative estimates transformed from sample Kendall's $\hat{\tau}$ between periods $t - m$ and $t - 1$; i.e., for each of the $m(m-1)/2$ possible pairs, an estimate of Kendall's $\hat{\tau}$ between times $t - m$ and $t - 1$ is first computed as

$$\hat{\tau}_{t-1} = \frac{2}{m(m-1)} \sum_{t-m \le t_1 < t_2 \le t-1} sign\left[(w_{1,t_1} - w_{1,t_2}) \, (w_{2,t_1} - w_{2,t_2}) \right]$$

and then, since the domain of sample Kendall's $\hat{\tau}$ is $[-1,\ 1]$, we take the following transformation to ensure $\tilde{\tau}_{t-1} \in [0,1]$ by keeping the original movement of sample Kendall's $\hat{\tau}$ intact:

$$\tilde{\tau}_{t-1} = \frac{exp(\hat{\tau}_{t-1})}{1 + exp(\hat{\tau}_{t-1})}$$

The eq(3.12) implicitly assumes a martingale process to capture the variation of sample Kendall's $\hat{\tau}_t$. The similar TVC approach is also specified in Jondeau and Rockinger (2006) for a time-varying Student-t copula.

In this paper for the estimation of TVC dependence, we apply a variance-targeting-like method suggested by Engle and Sheppard (2005) and the chapter 11 of Engle (2009). In particular, we model the Kendall's τ as

$$\tau_t = (1 - \beta - \gamma)\bar{\psi} + \beta\tau_{t-1} + \gamma\tilde{\tau}_{t-1}$$

where $0 \le \beta, \gamma < 1$, $\beta + \gamma < 1$. $\bar{\psi} = \sum_{i=t-R}^{t-1} \tilde{\tau}_i / R$ with R being the length of the estimation window specified in later sections.

3.2.8 Integrated-TVC (ITVC) Dynamic Dependence

The integrated-TVC is simply a special case of TVC specification by restricting $\omega = 0$ and $\beta + \gamma = 1$, such that

$$\tau_t = \beta\tau_{t-1} + \gamma\tilde{\tau}_{t-1} \tag{3.13}$$

where $0 \le \beta, \gamma < 1$.

3.3 Marginals

The ingredients for estimating the decomposition models also require for the specifications of marginals for absolute returns and signs. Since $|r_t|$ is a positively valued variable, the dynamics of absolute returns is specified using the multiplicative error model (MEM) framework:

$$|r_t| = \psi_t \eta_t \tag{3.14}$$

where $\psi_t = E_{t-1}|r_t|$ and η_t is a positive multiplicative error with $E_{t-1}(\eta_t) = 1$, and its conditional marginal distribution $\mathcal{D}$. Such multiplicative error model (MEM) has recently devoted to modeling volatility measured based on ultra-high frequency data, volumes, number of trades, and durations, etc.[2] MEM is particularly suited to model non-negative time series. Brownlees et al. (2011) provide a review of MEM.

The dynamics for ψ_t is specified in a logarithmic conditional autoregressive model as:

$$ln\psi_t = \omega_v + \beta_v ln\psi_{t-1} + \gamma_v ln|r_{t-1}| + \rho_v \mathbb{I}(r_{t-1} > 0) + x'_{t-1}\delta_v \tag{3.15}$$

where x_{t-1} are economic predictors and observables at time $t-1$. The main advantage of eq(3.15) is that no parameter restrictions are needed to enforce positivity of $E_{t-1}|r_t|$, especially when (weakly) exogenous predictors are present. The persistence of the process is governed by the parameter $|\beta_v + \gamma_v|$. Eq(3.15)is referred to as the volatility model.

The possible choices for $\mathcal{D}$ include exponential, Weibull, Burr, and Generalized Gamma distributions, etc. In empirical section following Anatolyev and Gospodinov (20010), we use the Weibull distribution with shape parameter, $k > 0$ and scale parameter, $\varsigma > 0$. The advantage to apply the Weibull distribution is that the Weibull parameters are reduced to one parameter distribution by using the condition of $E_{t-1}(\eta_t) = 1$. The cumulative probability function of a Weibull distribution of a random variable X is given by

$$F(x;\ k, \varsigma) = 1 - e^{-\left(\frac{x}{\varsigma}\right)^k}$$

with the probability density function, $f(x,\ k, \varsigma)$,

[2]see, e.g., Engle et al. (2011), Chou(2005), Engle and Gallo(2006), Engle and Russel (1998), Anatolyev (2009), among others

$$f(x;\ k,\ \varsigma) = \begin{cases} \frac{k}{\varsigma}\left(\frac{x}{\varsigma}\right)^{k-1} e^{-\left(\frac{x}{\varsigma}\right)^{k}} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3.16}$$

The mean of a Weibull random variable can be expressed as $E(X) = \varsigma\Gamma\left(1 + 1/k\right)$. $\Gamma(\cdot)$ is a gamma function. In our case, $E_{t-1}(\eta_t) = 1$, such that $\varsigma = \Gamma^{-1}\left(1 + 1/k\right)$. Hence, the Weibull distribution has been reduced to one parameter function. By setting $x = \eta_t = |r_t|/\psi_t$ with the specification of ψ_t in eq(3.15), we have the marginal PDF and CDF functions for eq(3.3).

Specifying the sign dynamics, we take a Bernoulli marginal of the form:

$$B(p_t) = p_t^v\left(1 - p_t\right)^{1-v}$$

with p_t is specified as a dynamic logit-linked model:

$$p_t = \frac{exp(\theta_t)}{1 + exp(\theta_t)}$$

with

$$\theta_t = \omega_d + \phi_d \mathbb{I}(r_{t-1} > 0) + y_{t-1}^{'}\delta_d \tag{3.17}$$

where eq(3.17) is referred to as the direction model. y_{t-1} includes the set of predictors such as macroeconomic variables as well as realized return higher moments.

3.4 Likelihood Function

With the marginals of eq(3.15) and (3.17), the copula function eq(3.4) and (3.5), and the dynamic dependence structures of eq(3.6)-(3.13), we have the sample log-likelihood function as:

$$\begin{aligned} L(\Phi) = & \sum_{t=1}^{T}\{\mathbb{I}(r_t > 0) ln\varrho_t[F(u_t|\psi_t))] \\ & +[1 - \mathbb{I}(r_t > 0)]ln(1 - \varrho_t[F(u_t|\psi_t))])\} \\ & +\sum_{t=1}^{T} lnf(u_t|\psi_t) \end{aligned}$$

With all the specified ingredients, the set of parameters to be estimated by maximum likelihood estimation is $\Phi = (\omega_v,\ \beta_v,\ \gamma_v,\ \rho_v,\ \delta_v,\ k,\ \omega_d,\ \phi_d,\ \delta_d,\ \Theta)$, where Θ contains the dynamic dependence parameters, $(\omega,\ \beta,\ \gamma,\ \lambda)$ from eq(3.6)-(3.13). Note that in this paper, all parameters in the set, Φ, are obtained from maximizing the sample log-likelihood of the full decomposition models.

4 Forecasting Methods

4.1 Conditional Mean Forecasts

Based on the eq(3.2), the one-period ahead forecast of financial returns at time $t+1$ conditional on the past information is given by

$$\hat{r}_{t+1} = 2\hat{\xi}_{t+1|t} - \hat{\psi}_{t+1|t} \tag{4.1}$$

where $\hat{\psi}_{t+1|t}$ is the conditional expectation forecast of $|r_{t+1}|$ using eq(3.15). $\hat{\xi}_{t+1|t} = E_t\,\widehat{|r_{t+1}|\,\mathbb{I}(r_{t+1}} > 0)$ is the conditional expectation forecast for the cross product of absolute returns and indicators. Anatolyev and Gospodinov (2010) find that the expectation of the cross product happens to be weakly conditional dependence. Thus, one might ignore the dependence by assuming the conditional independence between absolute returns and indicators as

$$\hat{r}_{t+1} = (2\hat{p}_{t+1|t} - 1)\hat{\psi}_{t+1|t} \tag{4.2}$$

where $\hat{p}_{t+1|t}$ is the conditional forecast from eq(3.17). Note that eq(4.2) will not be optimal under conditional dependence between the components.

For the general case of conditional dependence, the conditional expected cross-product is given by

$$\begin{aligned}\hat{\xi}_{t+1|t} &= E_t\,\widehat{|r_{t+1}|\,\mathbb{I}(r_{t+1}} > 0) \\ &= \int_0^{+\infty} uf\left(u|\hat{\psi}_{t+1|t}\right)\hat{\varrho}_{t+1}\left(F\left(u|\hat{\psi}_{t+1|t}\right)\right)du \end{aligned} \tag{4.3}$$

which depends on the estimated parameters, $\hat{\Phi}$, at time t. Eq(4.3) is integrated by using the approach proposed in Anatolyev and Gospodinov (2010).

4.2 Return Higher Moments Forecasts

Using the model estimators, we can also forecast the time-varying higher moments, similar to eq(4.3). Particularly, we consider the out-of-sample forecasts for downside risk and skewness used in the section 8.

Let $Q_t(z)$ denote the quantile function of the marginal distribution for absolute returns. The conditional skewness at time $t+1$ is forecasted as

$$\widehat{skew_t(r_{t+1})} = \frac{\widehat{E_t\left(r_{t+1}^3\right)} - 3\widehat{E_t(r_{t+1})}\widehat{var_t(r_{t+1})} - \left[\widehat{E_t(r_{t+1})}\right]^3}{\left[\widehat{var_t(r_{t+1})}\right]^{\frac{3}{2}}} \tag{4.4}$$

where

$$\widehat{var_t(r_{t+1})} = \widehat{E_t(r_{t+1}^2)} - \left[\widehat{E_t(r_{t+1})}\right]^2$$

with $\widehat{E_t(r_{t+1})}$ is estimated by eq(4.1) and $\widehat{E_t(r_{t+1}^2)} = \int_0^1 \left[\hat{Q}_{t+1|t}(z)\right]^2 dz$, and,

$$\widehat{E(r_t^3)} = 2E_t\left[|r_{t+1}|^3 \widehat{\mathbb{I}(r_{t+1}} > 0)\right] - E_t\left(\widehat{|r_{t+1}|^3}\right)$$

with

$$E_t|r_{t+1}\widehat{|^3\mathbb{I}(r_{t+1}} > 0) = \int_0^1 \left[\hat{Q}_{t+1|t}(z)\right]^3 \hat{\varrho}_{t+1|t}(z)\,dz \tag{4.5}$$

$$\widehat{E_t|r_{t+1}}|^3 = \int_0^1 \left[\hat{Q}_{t+1|t}(z)\right]^3 dz \tag{4.6}$$

The conditional downside risk is also referred to semi-variance as

$$semivariance = \frac{1}{T}\sum_{t=1}^{T}\left[(r_t-\mu)^2\mathbb{I}(r_t<\mu)\right] \to E\left[(r_t-\mu)^2\mathbb{I}(r_t<\mu)\right]$$

where μ is the mean or average of returns. In our case, we set $\mu = 0$. Thus, the conditional downside risk (semi-variance) forecast ($\widehat{drisk}_{t+1}$) is given by

$$\widehat{drisk_t(r_{t+1})} \quad = \quad \widehat{E_t(r_{t+1}^2)} - E_t\left[r_{t+1}^2 \widehat{\mathbb{I}(r_{t+1}} > 0)\right] \tag{4.7}$$

with $E_t\left[r_{t+1}^2 \widehat{\mathbb{I}(r_{t+1}} > 0)\right] = \int_0^1 \left[\hat{Q}_{t+1|t}(z)\right]^2 \hat{\varrho}_{t+1|t}(z)\, dz$. See *Appendix A* for the proofs.

Note that each solution of these higher moments has a corresponding conditional independence counterpart. However, in this paper, we only estimate higher moments for conditional dependence cases. In the empirical section, we apply both Monte Carlo simulation method and the Gauss-Chebyshev quadrature formulas (Judd 1998, section 7.2) to evaluate integrals. Since these methods have shown similar results, we only report the results obtained from the Gauss-Chebyshev quadrature.

5 Data

The monthly U.S. return data from January 1952 to December 2002 is available from Anatolyev and Gospodinov's website. We extend their dataset to cover the period of the recent financial crisis till December 2010. For the extended period, the value-weighted excess return is taken from the Center for Research in Security Prices (CRSP); the earnings-price ratio (ep) and dividend-price ratio (dp) data in logs are constructed using the dataset provided by Shiller (2005); the three-month T-bill rate ($ir3$), and Moody's Aaa and Baa corporate bond yield data, are from Federal Reserve Bank of St. Louis. We construct the yield spread (irs) by the difference between Moody's Aaa and Baa.

In this paper, the realized higher moments, i.e., realized variance (RV), bipower variation (BPV), realized third (RS) and fourth (RK) moments, are also extended till December 2010 and constructed from daily data on the NYSE/AMEX value-weighted index from CRSP. As the proxies for the unobserved volatility process, the realized variance is an estimator of integrated variance plus a jump component while the latter is unaffected by the presence of jumps. For the detail formula, see e.g., Barndorff-Nielsen and Shephard (2004, 2006, 2007), and Anatolyev and Gospodinov (2010), among others.

We use predictors $x_t = (dp_t,\ ep_t,\ ir3_t,\ irs_t)$ in eq(3.15) for the marginal dynamics of absolute returns, and $y_t = (dp_t,\ ep_t,\ ir3_t,\ irs_t,\ RV_t,\ BPV_t,\ RS_t,\ RK_t)$ in eq(3.17) for the dynamics of the Bernoulli marginal. We employ a rolling window with the length of the past 360 months for out-of-sample forecast. It results in the out-of-sample period from January 1982 to December 2010, which covers the forecasting period in Anatolyev and Gospodinov (2010), and extends to the recent financial crisis of 2007-2009.

We briefly introduce the name convention used in the empirical section. This paper estimates both conditional independence (CI, eq(4.2)) and conditional dependence (CD, eq(4.1)) scenarios. Constant decomposition models are denoted as CDM, and DDM for dynamic decomposition models. Hence, we represent a constant decomposition model from the conditional independence estimation as $CDM - CI$, and $DDM - CI$ for a dynamic decomposition model from the conditional independence estimation. For instance, the dynamic decomposition model of OSA (eq(3.10)), if it were estimated from the conditional dependence, would be denoted as $OSA - CD$.

6 Empirical Results

Table 2 presents the in-sample estimation results. The subsample estimation from January 1952 to December 2002 is made comparable to the results of Anatolyev and Gospodinov (2010). The parameter estimations in Panel A are statistically significant at conventional confidence levels. The Wald tests in Panel B show that the estimators of the dynamic decomposition models are jointly significant. As seen in Panel B, the dynamic decomposition models obtain higher MLE log-likelihood values than the constant decomposition models, and the likelihood ratio tests reject the null hypothesis at 1% significance level.

[Table 2 about here]

The persistence of a dependence structure is measured by the parameter β. Among the dynamic models, OSA and TSA have the highest persistence, whereas Patton shows less persistence in dependence structure. The smoothing weighting parameter, λ, is significant at 1% level, which provides the evidence for the weighting of the time series observations. The integrated models of IGtype and ITVC have the persistence of around 0.75 which is lower than 0.94, the value calibrated for integrated GARCH model by J.P. Morgan RiskMetrics.

The in-sample evidences support the hypothesis of time-varying nonlinear dependence between absolute returns and signs. The results are also consistent with the pretesting results of rejecting the constancy of copulas.

6.1 Out-of-Sample Forecast

The out-of-sample coefficient of predictive performance (OS), suggested by Campbell and Thompson (2008), is used for forecasting evaluation as

$$OS = 1 - \frac{\sum_{t=R+1}^{T} L(r_t - \hat{r}_t)}{\sum_{t=R+1}^{T} L(r_t - \bar{r}_t)}$$

where $L(s) = s^2$ if it is based on squared errors, and $L(s) = |s|$ if it is based on absolute errors. We use a rolling sample window of length, $R = 360$ for out-of-sample forecast, while T is the total observations in the sample period, such that the one-step ahead out-of-sample forecast period is $H \in [R+1, T]$. $\bar{r}_t$ is the unconditional mean of r_t computed from the last R observations in the rolling scheme, which is referred to as the historical average. $\hat{r}_t$ is a forecast at time t from individual models.

The OS statistics measures the reduction in forecasting errors relative to the historical average forecast. Thus, if $OS > 0$, the $\hat{r}_t$ forecast outperforms the historical average forecast; if $OS < 0$, the historical average forecast performs better.

Figure 2 plots the return forecasts from the linear predictive regression, constant and dynamic decomposition models. We have the following observations in order. First, similar to Anatolyev and Gospodinov (2010), Figure

2 show that in the late of 1990s the linear predictive model underestimates the bull market by predicting predominantly negative returns, while both constant and dynamic models capture well the upward trend in the market and the increased volatility in the early 2000s.

[Figure 2 about here]

Second, the difference of the forecasts between the constant and dynamic models exists over time. The largest disagreements occur during: (1) the time from late 1990s (around the late of 1997 when Asia financial crisis started) to the early 2000s (the period of IT bubble bust); and (2) the recent financial crisis time.

Based on the observations above, the out-of-sample period is divided into sub-samples: (1) a relatively tranquil period: 1982:01 - 1997:12; (2) the first relatively turmoil period: 1998:01 - 2002:12; (3) the second turmoil period: 2007:07 - 2010:12; (4) the comparison period with Anatolyev and Gospodinov: 1982:01 - 2002:12.

Table 3 presents the out-of-sample forecast results. The average performance of the dynamic decomposition models are also reported. The forecasts from the conditional independence are estimated by ignoring the conditional dependence structure between absolute returns and signs through eq(4.2), while the conditional dependence forecasts are computed via eq(4.1) through numerical integrations.

[Table 3 about here]

In line with the previous studies, e.g., Rapach et al. (2010), Welch and Goyal(2008), and Campbell and Thompson (2008), etc., the out-of-sample forecast of historical average in Table 3 performs better than the conditional linear model and the difference in forecast gain is around 5%. The decomposition models dominate both the historical average and conditional linear models in terms of OS statistics. Observably, the dynamic decomposition models achieve the most outstanding performance.

Of interest, we turn next to the detail analysis for comparisons between the constant and dynamic decomposition models, and between conditional independence and conditional dependence.

6.1.1 Tranquil Period vs. Turmoil Time

As shown in Table 3, the proposed dynamic decomposition models consistently perform better out-of-sample than the constant models both individually and on average across sample periods. For instance, $CDM-CI$ has the OS statistics of 0.18% in the full sample period, while the dynamic decomposition models have 1.41% on average with the highest individual gain, 2.53%, by $TSA-CI$ model. The best out-of-sample performance in the full sample period is 3.59% and 3.57% from $IGtype-CD$ and $OSA-CD$, respectively, while the best for the comparison subperiod is 4.62% from $IGtype-CI$.[3]

Interestingly, among the out-of-sample forecasts, the largest difference in gains is 8.63% during the recent financial crisis time between CDM ($CDM-CI$ with -3.57%) and DDM ($TSA-CI$ with 5.06%); moreover, for the conditional dependence case, the largest gain differences between CDM and DDM are 4.64% ($CDM-CI$ with 5.18% and $Patton-CD$ with 9.82%), and 5.59% ($CDM-CI$ with 6.72% and $IGtype-CD$ with 12.31%) in the two turmoil periods. However, the difference for the tranquil time, around 1% or less, is much less than those in the turmoil periods.

In summary, we see that the difference of forecasting gains between the constant and dynamic decomposition models is considerably large during the turmoil times, at least for our sample periods, whereas the difference in the tranquil periods is still large in the individual level, but relatively small on average. As a result, these results imply that an investor will benefit out-

[3]Note that the OS statistics in Table 3 of the constant model for the subperiod of 1982:01-2002:12 is similar to, but higher than the results of Anatolyev and Gospodinov (2010). The reasons are the following: (i) we drop the first $m = 10$ observations for all constant and dynamic models, as required by eq(3.6), to make them comparable and to some extent reduce the initial value effects. We investigate the dataset and find the first ten observations have large volatility, which might have influence on the estimation results; (ii) The extreme events and outliers of returns and their realized higher moments produce some extreme values for the probability of Bernoulli marginal distribution, such as p_t might reach the extreme values by $p_t = 0, \, or \, 1$. These extreme values, most likely occurring in crisis times such as the Black Friday in 1987, will make the sample log-likelihood function exploded or collapsed. We consult this issue by empirically restricting $0.002 \leq p_t \leq 0.998$. For those values out of this range, we set $p_t = 0.002$ if $p_t < 0.002$ and $p_t = 0.998$ if $p_t > 0.998$. The restrictions should be consistent with the argument of Campbell and Thompson (2008) that imposing theoretically and economically motivated restrictions on individual models can improve their out-of-sample performance.

of-sample forecasting from the dynamic dependence models in general, and even more during a turmoil time.

6.1.2 Conditional Independence vs. Conditional Dependence

In this paper, the forecasting performance comparison between the conditional independence and dependence is also great of interest. First, in the full sample period, the forecast gains from the conditional dependence have better performance than those from the conditional independence. In particular, $CDM - CD$ has higher forecasting gain than $CDM - CI$ by 2% or so, while $DDM - CDs$ have 2.62% forecasting gain on average, larger than $DDM - CIs$ with 1.41%. All the individual $DDM - CDs$ perform better than their counterparts from $DDM - CIs$. For instance, $OSA - CD$ gains 2% more than from $OSA - CI$.

Second, we observe the much larger difference in forecasting gains between conditional dependence and conditional independence in the two turmoil times. The difference in the first turmoil period is 4.06% for $CDM-CD$ over $CDM - CI$, and 7.18% for $DDM - CD$ over $DDM - CI$ on average, and the difference for the recent financial crisis is 10.29% for $CDM - CD$ over $CDM - CI$, and 7.25% for $DDM - CD$ over $DDM - CI$ on average.

Conversely, the conditional independence in tranquil periods performs better out-of-sample forecast than the conditional dependence. During 1982:01 - 1997:12, $DDM - CI$ perform better out-of-sample than $DDM - CD$ by 3.26%. Most of the individual $DDM-CIs$ outperform $DDM-CDs$ in tranquil periods. Additionally, the comparison period also shows the conditional independence slightly better on average.

As a result, we see that the conditional dependence provides better overall out-of-sample forecast than the conditional independence, especially during the turmoil times; however, the conditional independence does a better job in the tranquil period relative to the conditional dependence.

Intuitively, the better forecasting performance of the conditional dependence reflects that the volatility (absolute returns) is higher and asymmetric effects (signs) are larger in a turmoil period, and they are more correlated and comoved in a bad time. On the other hand, markets are more stable in a normal time such that volatility is low and the skewness is less negative, and their correlation is less significant in a tranquil time. As such, the empirical evidence in this paper supports that the proposed dynamic decomposition models well capture these stylized facts and nonlinear dependence in different market timings, by estimating conditional independence in a normal time and conditional dependence during a turmoil period.

In summary, the robust strategy for an investor is to adopt the dynamic decomposition models by exploiting conditional dependence for a turmoil time. Suggestively, the better choice for a tranquil period is to take the dynamic decomposition models by assuming conditional independence. Hence, the implication is useful for an investor to determine the model choices based on the current state of an economy or the timing of financial markets.

6.1.3 Time-Varying Dependence

The concern has also been raised by *Remark 3.2* as shown in Figure 1: whether the dynamic dependence specifications reduce the gaps between the dependence requested by the sample return moments and estimated by dynamic decomposition models. Figure 3 plots the out-of-sample dependence parameters. The parameters estimated from the constant model are relatively stable around its mean (0.0816) with the minimum (0.035) and maximum (0.179). By contrast, the dynamic decomposition models have the dependence parameters more volatile than the constant model over time. For instance, The exponentially weighted dynamic models have the estimated dependence parameters with the minimum (0.00004) and maximum (1.55). Its minimum value close to zero implies a conditional independence, while its maximum value tends to close the gap.

[Figure 3 about here]

In addition, the estimated dependence degree is remarkably higher for the turmoil times. *IGtype* model has the highest estimated mean dependence around 0.275 which is 3.4 times higher than that of the constant model; more importantly, IGtype model has accomplished better out-of-sample forecast than most of others. Compared to the constant model, the dynamic models improve the out-of-sample forecast performance by having the estimated dependence structure closer to those requested by sample return moments.

6.2 Tests of Predictive Ability

To evaluate the statistical significance of the forecasting results, we adopt Giacomini and White (2006) test of conditional predictive ability. Table 4 presents the test results. The rejection of the null of equal predictive ability is denoted by *, **, *** at 10%, 5%, and 1% significance levels. Relative performance summarizes the relative out-of-sample performance by computing the proportion of times over the whole forecasting period that the foregoing decision rule chooses the model given in column heading over the model given in row heading. The entries of decision rule, scaled by 10^4, is the mean values of fitted loss function difference. Following the two-step procedure of Giacomini and White (2006), in the case of rejection, the decision rule chooses the model in column heading if its entry is positive, and vice versa. For technical details, see Giacomini and White (2006). Panel A and Panel B report the test results from the squared and absolute errors, respectively. The test results for the full and comparison forecast periods are reported.[4]

[Table 4 about here]

For the constant models in the comparison sample period, Table 4 has the similar results to Anatolyev and Gospodinov (2010) that the constant decomposition models do not show statistically significant differences from historical average and linear models, although the relative performance suggests that the constant models dominate both historical average and linear models.

However, the dynamic decomposition models tell different stories that the statistically significant differences from historical average and linear models

[4]Other subsample periods are too short to conduct Giacomini and White's test.

have been observed in many dynamic model estimations. For instance, in the comparison period based on absolute errors, most $DDM-CIs$ are significantly different from historical average, and most of $DDM-CDs$ are significantly different from linear model. Furthermore, the tests based on the full sample period provide more convincing evidence for the superior predictive performance based on both squared and absolute errors. Table 4 also shows that the dynamic decomposition models are statistically different from constant decomposition model in both sample periods and different measures of loss functions.

More importantly, in these significant cases, the values of decision rule are positive, which imply the decisions of choosing dynamic decomposition models given in column headings. In addition, all the dynamic decomposition models have dominated benchmarks in terms of relative forecasting performance. For instance, the dynamic decomposition models have on average approximately 90% of the times over the whole forecasting period performing better than benchmarks.

In general, the plot of the predicted loss differences is useful for assessing the relative performance of the competing methods at different times. Figure 4 plots the predicted loss differences over time: Panel A uses historical average as benchmark and Panel B uses $CDM-CI$ as benchmark. A positive value at time t indicates the competing models perform better than the benchmarks.

[Figure 4 about here]

Panel A is able to distinguish the under- vs. out-performance periods of the models. Both $CDM-CD$ and $DDM-CD$ have positive loss differences in most of time periods above the zero line. Specifically, during the financial crisis, the decomposition models are improved dramatically by exploiting the dependence between return absolute values and signs, which turns the loss difference to be positive, compared to the conditional independence cases with negative loss difference. The evidence is even more consolidated when comparing dynamic models to constant models in Panel B. The most impressive improvement is during the two turmoil times shown as the positive loss difference over the period of time.

6.3 Copula Specifications: Out-of-Sample Density Forecast Comparison

It has been recently developed to assess relative predictive accuracy by comparing out-of-sample density forecast through the Kullback-Leibler information criterion (KLIC). Amisano and Giacomini (2007) provide an interesting interpretation of the KLIC-based comparison in terms of scoring rules, which are loss functions depending on the density forecast and the actually observed data. Diks et al. (2010) compare predictive accuracy of various copula specifications from an out-of-sample forecasting perspective by using out-of-sample log-likelihood scores obtained from copula density forecasts. Their test method is valid under general conditions on the competing copulas including the density forecasts from time-varying copula parameters.

In this paper, we apply Diks et al. (2010) approach to evaluate the predictive accuracy of density forecasts among our copula decomposition specifications, particularly eq(3.6)-(3.13). The KLIC of density forecast obtained from a copula-based model can be decomposed as $KLIC_{t+1} = \sum_{j=1}^{d} log\hat{f}_{j,t}(Y_{j,t+1}) + log\hat{c}_t(\hat{U}_{t+1})$, where $\hat{c}_t$ is the conditional copula density associated with the density forecast, and $\hat{f}_{j,t}$ is the forecast of marginal distribution, j. In our case, $d = 2$ for absolute returns and signs. Since in this paper the conditional marginals are identically specified, the marginal log-likelihood, $log\hat{f}_t(Y_{t+1})$, are canceled out, so that the test for the null hypothesis of equal predictive accuracy is $H_0 : E(\triangle KLIC) = E(log\hat{c}_{A,t}(\hat{U}_{t+1})) - E(log\hat{c}_{B,t}(\hat{U}_{t+1})) = 0$, where the term $log\hat{c}_{i,t}(\hat{U}_{t+1})$ is a score assigned to the copula model i based on the obtained probability integral transform (PIT) $\hat{U}_{t+1}$. The test statistic of equal KLIC scores, based on a heteroskedasticity and auto-covariance consistent (HAC) estimator, asymptotically follows a standard normal distribution by applying Theorem 4 of Giacomini and White (2006). For details, see Diks et al. (2010).

Table 5 presents the test results of out-of-sample density forecasts for the decomposition models. We use the constant decomposition model as benchmark, c_A, and the dynamic models as c_B. The values in Table 5 is the mean of $\triangle KLIC$. The value signs indicate which copula specifications perform better. In particular, a negative value indicates the better predictive accuracy from the copula model given in column heading, as a higher average copula score is preferred.

[Table 5 about here]

The test results in Table 5 show that the differences between the copula models are statistically significant. In most of the significant cases, the differences between KLIC scores are negative. For instance, *IGtype*, OSA and TSA have the best predictive accuracy of density forecasts. All these models statistically significantly outperform the constant decomposition model due to their negative $\triangle KLIC$ signs across various sample periods. Table 5 also exhibits that ExpWeight and Gtype dynamic models have negative signs but insignificant. As a result, the test results for out-of-sample density forecast accuracy are consistent with the tests of Giacomini and White (2006) that both show the forecasting improvement from modeling time-varying skewness by allowing the dynamic copula specifications .

7 The Sources of Forecasting Performance

The existing literature on out-of-sample forecasting has devoted to test whether one model performs equally well or superior to the other models in a statistical sense, e.g., Diebold and Mariano (1995), West (1996), White (2000) and Clark and West (2006), among others. Elliott and Timmermann (2008) provided a review. However, the comparison in a pseudo out-of-sample forecasting environment informs the researcher only about which model forecasts best, and do not shed light on why that is the case. In contrast, the recent paper, Rossi and Sekhposyan (2011), proposes a new econometric pin-down method to decompose the existing measures of forecasting performance into the uncorrelated components, namely predictive content, over-fitting issue, and unstable forecasting ability, in order to exploiting the sources of forecasting performance.

This paper particularly emphasize in the source component of unstable forecast ability. The unstable forecasting ability might be caused by changes in the parameters of the models, as well as by unmodeled changes in the stochastic processes generating the variables. The instability of models' forecasting ability measures the presence of time variation in the models' performance relative to their average performance in the presence of no time variation in the expected relative forecasting performance.

Table 6 presents the test results for the unstable forecast ability with the forecasting period: 1982:01-2010:12.[5] The models in the first column as benchmarks are compared to the models in the first row. The entries in table 6 are the test statistics for the null hypotheses of no difference in time variation of forecasting ability, denoted as $\Gamma_p^{(A)}$ in Rossi and Sekhposyan (2011). Panel A and B report the test results from the loss functions of squared errors and absolute errors, respectively.

[Table 6 about here]

The test results show that the dynamic decomposition models accounting for time-varying skewness have significantly different forecasting ability over time. The null hypothesis of no difference in the time variation of forecasting ability are rejected for all pairwise model comparisons based on both squared and absolute error loss functions. For instance, the dynamic decomposition models could outperform constant decomposition models within a certain time period, but possibly underperform for the other subperiod. Prominently, these test results substantially support our empirical findings in previous sections that the out-of-sample forecast of our dynamic decomposition models generates much bigger gains than the constant models; nonetheless, the gains are much smaller in a tranquil period than during a turmoil time. Also, an investor benefits largely from the forecast models based on conditional dependence within a turmoil time. This timing specific nature is highly in line with the test results for unstable forecasting ability. Tests for the other two components, not reported here, also provide the empirical evidences for

[5]Other subsample periods are too short to conduct the test of Rossi and Sekhposyan (2011).

understanding the different forecasting performances between models. The significant difference in predictive content is due to the dynamic function form in dependence parameters for modeling time-varying tail dependence and asymmetry. By contrast, no strong evidence show the difference in over-fitting issue between the models.

8 Links to Real Economy

The further forecasting improvement could also lie in whether the forecasts are significantly and economically connected to real economy and justified by economic theories. Campbell and Thompson (2008) argue that the empirical models can yield useful out-of-sample forecasts if one restricts their parameters in economically justified ways. Recently, the finance literature has been exploiting the implications and evidence of economic theories in out-of-sample forecasting. Engle and Rangel (2008) and Engle et al. (2008) find that imputing economic fundamentals into stock volatility models pays off in terms of both long and short horizon forecasting. Ghysels et al. (2011) show a strong relationship between the conditional skewness and macroeconomic variables. Rapach et al. (2010) link the combination forecasts to the real economy for an economic rationale of the out-of-sample gains.

Particularly, we investigate whether the forecasts of time-varying skewness and downside risks can be explained by a set of predetermined state variables and well trace to economic fluctuations. The monthly economic state variables, including NBER-dated business-cycle, GDP growth rate, inflation, short-term interest rate, term premium and term spread, were taken from Federal Reserve Bank of St. Louis and the housing price from Federal Housing Finance Agency. The difference between corporate bond Aaa and Baa yields represents term premium as a proxy of credit risk, while term spread is computed as the difference between 10-year and 3-month T-bill rates as an expectation of term structure for future economic condition. Using the out-of-sample rolling window, we also compute the volatility for short-term interest rate, GDP growth, inflation and housing price as the proxy of macroeconomic uncertainty.

8.1 Trace to NBER-dated Business Cycle Phases

In the spirit of Rapach et al. (2010), we trace the time-varying skewness and down risk forecasts to NBER-dated business cycle phases, to see if there is any evidence for the connections to macroeconomic fluctuation. Figure 5 depicts the out-of-sample forecasts along with vertical lines indicating NBER-dated business-cycle phases of recessions (Rec) and expansions (Expan). Peaks (P) and troughs (T) in Figure 5 are used to track turning points of these phases. Over the 1982:01-2010:12 out-of-sample period, there are four recessions: (1) 1982:01-1982:11; (2) 1990:08-1991:03; (3) 2001:04-2001:11; (4) 2008:01-2009:06.

[Figure 5 about here]

Figure 5 shows well-defined patterns for these forecasts around the peaks and troughs. We have the following observations in order:

(i) The forecasts of return downside risk have sharp increases during these recession periods and quick drops at the end of each recession period to become turning points of recovering from bad times. These forecasts also show low values closely near the peaks preceding these recessions, which together with the followed sharp increases form turning points of entering bad times. The forecasts of downside risk during expansion times are low with downtrend.

(ii) The forecast of return skewness for the recession of early 1990s is modest, while the largest changes in return skewness forecast occur in the recent financial crisis. This makes sense, since the former recession is much milder than the recent financial crisis. The skewness forecasts have well reflected the turning points of in and out of bad times. On the contrary, there is not much fluctuations in the skewness forecasts of historical average, which has only a sudden and permanent decreasing in 1987, corresponding to the stock market collapse.

Overall, Figure 5 demonstrates that the forecasts of time-varying return higher moments closely track NBER-dated business-cycle phases. The historical average forecast appears too smooth; that is, from an economic perspective, the historical average forecast ignores business-cycle fluctuations and thus fails to incorporate meaningful macroeconomic information. Opposite to the historical average forecast, the decomposition model is capable of including relevant macroeconomic information. In what follows, we further investigate the relation significance between the forecasts and macroeconomic variables.

8.2 Relation with Macroeconomic Variables

To investigate the connections of these forecasts to macroeconomic variables, following Ghysels et al. (2011), we run the regressions of the forecasting values at time $t+1$ on state variables at time t:

$$\hat{\upsilon}_{t+1} = \phi_0 + \phi_1 x_t + e_{t+1}$$

where $\hat{\nu}_{t+1} = \left\{ \widehat{drisk}_{t+1},\ \widehat{skew}_{t+1} \right\}$ are forecast values at time $t+1$ of time-varying downside risk and skewness from (4.4) and(4.7); and, $x_t = \{GG_t,\ Infl_t,\ IR_t,\ HP_t,\ TP_t,\ TS_t,\}$ contains the set of macroeconomic variables at time t, namely GDP growth, inflation, short-term interest rate, housing price, term premium and term spread, respectively. e_{t+1} is error terms. The significance of the coefficient, ϕ_1 is of interest for a relationship testing purpose. Table 7 reports the regression results of the coefficient, ϕ_1 along with its significance at 1%, 5%, and 10% levels denoted by ***, **, and *.

[Table 7 about here]

8.2.1 Term Premium and Term Spread

The correlations between the forecasts and term spread are statistically significant. In particular, term spread is positively correlated with downside risk and negatively with skewness. Such significant correlations indicate that the forecasts from the decomposition models effectively incorporate macroeconomic information (e.g., the expectation of future economic conditions), and their forecast performance can be explained (justified) to some extent by economic conditions. The forecasts are also substantially correlated with credit risk(term premium). This result indicates that credit risk is more likely a type of tail risk or extreme events which can be better modeled by the tail dependence.

8.2.2 Interest Rate

All the decomposition models have shown significant negative relations of the short-term interest rate with downside risk, whereas the significant positive correlations with return skewness forecast can be seen only in ExpWeight and IGtype decomposition models. The negative correlations of the interest rate with downside risk forecasts make sense in that an increase in interest rate will generally have a negative effect on stock markets.

The volatility of short-term interest rate is not correlated with downside risk and skewness forecasts. These results show that the interest rate uncertainty does not explain well the forecast performance of decomposition models; and, there is less evidence that the asymmetric effects of returns (skewness) are affected by short-term interest rate.

8.2.3 GDP Growth

We see that the skewness forecasts are not significantly correlated with GDP growth rate; however, GDP growth rate can significantly explain the forecast performance of decomposition models, having negative correlations with downside risk. The negative correlations support the hypothesis that a good economic condition generates lower level of downside risk.

In addition, the macroeconomic uncertainty measured by the volatility of GDP growth is correlated positively with downside risk, and negatively with its skewness forecast. Both the positive and negative correlations consolidate the risk-aversion and asymmetric hypothesis that an investor requires higher premium for bearing macroeconomic risk. As a result, these results indicate that the overall economic condition and investment environment are the important factors to the forecast performance of decomposition models. The return asymmetric effects and extreme event effects are in a large scale related to the current economic state.

8.2.4 Inflation

There is no evidence to show inflation correlated with the forecasts. However, we do find strong evidence in the inflation fluctuation, which is significantly related positively with downside risk and negatively with skewness. This result seems to show that an investor more care about bearing inflation risk, but ignoring the level of inflation. Regardless of the inflation level high or low, as long as it remains stable, an investor can stay with it quite well without requiring much risk premium. In other words, even in the case of a low inflation level, once the inflation risk increases, a risk-aversion investor will require a significant level of risk premium for bearing such inflation uncertainty, and even more in recession times due to the asymmetric effects (the increased degree of skewness).

8.2.5 Housing Price

Since the credit crunch and defaults in mortgage markets have triggered the recent financial crisis, housing price has drawn much more researcher's attentions nowadays. Table 7 shows that no forecasts are significantly correlated with housing price. However, the negative economic connection is strongly established between the housing price fluctuation and the forecasts of return downside risk. This new evidence shows that the expectation for a "bad" thing occurring in the future stock performance will be risen substantially by the housing price uncertainty. If an investor is less clear about the housing price risk, he or she will be more anxiety to expect a worse future in stock performance.

9 Conclusion

This paper proposes the intriguing decomposition framework for modeling the time-varying return higher moments in out-of-sample forecast. The empirical results show that the proposed dynamic decomposition models, by allowing the nonlinear dependence between absolute returns and signs time-varying, consistently perform better out-of-sample than benchmark models.

More interestingly, the time-varying nonlinear dependence shows some specific capability of timing markets, which provides remarkably useful implications for a risk-averse investor. It suggests the best strategies for an investor who can employ the proposed dynamic decomposition model with conditional dependence within a turmoil time, and in contrast with conditional independence in a tranquil period. This timing specific nature is also supported by the source test of unstable forecasting ability based on Rossi and Sekhposyan (2011). To further gain the economic sense of forecasting performance, we find that the forecasts of time-varying downside risk and skewness exhibit well-defined patterns to trace closely to NBER-dated business-cycle phases. They precisely forecast turning points for those peaks and troughs.

References

[1] Amisano, G., and R. Giacomini (2007) Comparing Density Forecasts via Weighted Likelihood Ratio Tests. Journal of Business & Economic Statistics 25(2): 177-190

[2] Anatolyev, S. (2009) Dynamic Modeling Under Linear-Exponential Loss. Economic Modelling 26: 82–89

[3] Anatolyev, S., and N. Gospodinov (2010) Modeling Financial Return Dynamics via Decomposition. Journal of Business and Economic Statistics 28(2): 232-245

[4] Barndorff-Nielsen, O. E., and N. Shephard (2004) Power and Bipower Variation With Stochastic Volatility and Jumps (with discussion). Journal of Financial Econometrics 2: 1–48.

[5] Barndorff-Nielsen, O. E., and N. Shephard (2006) Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. Journal of Financial Econometrics 4(1): 1-30

[6] Barndorff-Nielsen, O. E., and N. Shephard (2007). Variation, jumps and high frequency data in financial econometrics. In: Blundell, R., Torsten, P. Newey, W. K., eds. Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress. Econometric Society Monograph, Cambridge University Press.

[7] Brownlees, C.T., F. Cipollini, and G.M. Gallo (2011) Multiplicative Error Models. Working Paper

[8] Busetti, F., and A. Harvey (2011) When is a Copula Constant? A Test for Changing Relationships. Journal of Financial Econometrics 9(1): 106-131

[9] Campbell, J. Y., and S. B. Thompson (2008) Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies 21: 1509–1531

[10] Chou, R. Y. (2005) Forecasting Financial Volatilities With Extreme Values: The Conditional Autoregressive Range (CARR) Model. Journal of Money, Credit and Banking 37: 561–582.

[11] Christoffersen, P. F., and F. X. Diebold (2006) Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics. Management Science 52(8): 1273–1288

[12] Christoffersen, P. F., F. X. Diebold, R. S. Mariano, A. S. Tay, and Y. K. Tse (2007) Direction-of-Change Forecasts Based on Conditional Variance, Skewness and Kurtosis Dynamics: International Evidence. Journal of Financial Forecasting 1(2): 1–22

[13] Chung, J., and Y. Hong (2007) Model-Free Evaluation of Directional Predictability in Foreign Exchange Markets. Journal of Applied Econometrics 22: 855–889

[14] Clark, T.E., and K.D. West (2006) Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. Journal of Econometrics 135 (1–2): 155–186.

[15] Diks, C., V. Panchenko and D. van Dijk (2010) Out-of-sample comparison of copula specifications in multivariate density forecasts. Journal of Economic Dynamics & Control 34: 1596-1609

[16] Diebold, F.X., and R.S. Mariano (1995) Comparing predictive accuracy. Journal of Business and Economic Statistics 13 (3), 253–263.

[17] Elliott, G., and A. Timmermann (2008) Economic forecasting. Journal of Economic Literature 46 (1), 3–56.

[18] Engle, R. F., and K. Sheppard (2005) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working paper, University of California, San Diego.

[19] Engle, R.F. (2009) Anticipating Correlations: A New Paradigm for Risk Management. Princeton University Press, Princeton and Oxford

[20] Engle, R. F., G.M. Gallo, and M. Velucchi (2011) Volatility Spillovers in East Asian Financial Markets: A MEM-based Approach. Review of Economics and Statistics, forthcoming.

[21] Engle, R.F., and J.G. Rangel (2008) The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes. The Review of Financial Studies 21(3): 1187-1222

[22] Engle, R. F., E. Ghysels, and B. Sohn (2008) On the Economic Sources of Stock Market Volatility. Working Paper

[23] Engle, R. F., and G.M. Gallo (2006) A Multiple Indicators Model for Volatility Using Intra-Daily Data. Journal of Econometrics 131: 3–27

[24] Engle, R. F., and J.R. Russell (1998) Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica, 66: 1127–1162

[25] Genest, C., and B. Rémillard (2004) Tests of independence and randomness based on the empirical copula process. Test 13: 335–369

[26] Genest, C., and B. Rémillard (2006) Local efficiency of a Cramer-von Mises test of independence. Journal of Multivariate Analysis 97(1): 274–294.

[27] Genest, C., and B. Rémillard (2007) Asymptotic local efficiency of Cramér-von Mises tests for multivariate independence. The Annals of Statistics 35(1): 166–191

[28] Genest, C., and B. Remillard (2008) Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models.

Annales de l'Institut Henri Poincare-Probabilites et Statistiques 44(6): 1096-1127.

[29] Genest, C. , B. Remillard and D. Beaudoin (2009) Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics 44: 199-214.

[30] Giacomini, R., and H. White (2006) Tests of Conditional Predictive Ability. Econometrica 74: 1545–1578

[31] Glosten, R.T., R. Jagannathan, and D. Runkle (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48(5): 1779-1801

[32] Ghysels, E., A. Plazzi, and R. Valkanov (2011) Conditional Skewness of Stock Market Returns in Developed and Emerging Markets and its Economic Fundamentals. Working Paper

[33] Harvey, A. (2010) Tracking a changing copula. Journal of Empirical Finance 17: 485-500

[34] Hong, Y., and J. Chung (2003) Are the Directions of Stock Price Changes Predictable? Statistical Theory and Evidence. Cornell University, Working Paper

[35] Judd, K. (1998) Numerical Methods in Economics. Cambridge, MA: MIT Press.

[36] Jondeau, E., and M. Rockinger (2006) The copula-garch model of conditional dependencies: An international stock market application. Journal of International Money and Finance 25: 827-853

[37] Jondeau, E., and M. Rockinger (2003) Conditional volatility, skewness, and kurtosis: existence, persistence, and comovement. Journal of Economic Dynamics & Control 27: 1699-1737

[38] Kojadinovic, I., J. Yan and M. Holmes (2011). Fast large-sample goodness-of-fit tests for copulas. Statistica Sinica. 21: 841-871

[39] Kojadinovic, I. and J. Yan (2011) A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. Statistics and Computing 21(1): 17-30

[40] Kojadinovic, I. and J. Yan (2010) Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. Journal of Statistical Software 34(9): 1-20.

[41] Linton, O., and Y.J. Whang (2007) The Quantilogram: With an Application to Evaluating Directional Predictability. Journal of Econometrics 141: 250–282

[42] Nelson, D. B. (1990) Stationarity and Persistence in the GARCH(1,1) Model. Econometric Theory 6(3): 318-334

[43] Nelson, D. B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59(2): 347-370

[44] Patton, A. (2006) Modelling asymmetric exchange rate dependence. International Economic Review 47: 527-556

[45] Rapach, D.E., J.K. Strauss, and G. Zhou (2010) Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. The Review of Financial Studies 223(2):821-862

[46] Rossi, B., and T. Sekhposyan (2011) Understanding model's forecasting performance. Journal of Econometrics 164(1): 158-172

[47] Shiller, R. J. (2005) Irrational Exuberance. 2nd edition, Princeton University Press

[48] Thomakos, D. D., and T. Wang (2010) 'Optimal' probabilistic and directional predictions of financial returns. Journal of Empirical Finance 17: 102-119

[49] Tse, Y.K., and A.K.C. Tsui (2002) A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations. Journal of Business & Economic Statistics 20(3): 351-362

[50] Welch, I., and A. Goyal (2008) A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21(4): 1455–1508.

[51] West, K.D. (1996) Asymptotic inference about predictive ability. Econometrica 64 (5), 1067–1084.

[52] White, H. (2000) A Reality Check for Data Snooping. Econometrica 68: 1097-1136

[53] Zakoian, J. (1994) Threshold heteroskedastic models. Journal of Economic Dynamics and Control 18: 931-955

Table 1: Pretesting the independence, constancy and choice of Copulas. All tests, based on the empirical copula processes, have asymptotic distributions of Cramér-von Mises with different degrees of freedom. The independence and serial independence tests in Panel A are as proposed by Genest and Rémillard (2004, 2006, 2007). The goodness-of-fit tests in Panel B are based on Genest and Remillard (2008), Genest, Remillard and Beaudoin (2009), Kojadinovic, Yan and Holmes (2011), Kojadinovic and Yan (2010 & 2011). The constancy tests in Panel C are based on Busetti and Harvey (2011), and Harvey (2010). τ denotes quantiles.

Panel A: Independence Tests	Indep. Test	Serial Indep. Test
p value	0.041	<0.001

Panel B: Goodness-of-fit Tests	Normal Copula	Student-t Copula	Frank Copula	FGM Copula	Gumbel Copula	Clayton Copula
p value	0.040	0.043	0.040	0.035	<0.001	0.154

Panel C: Constancy Tests	τ	p value	τ	p value	τ	p value
Overall Constant Copula			0.5	0.378		
Joint Constancy of Lower-Upper Tails	0.25	0.384				
Constant Lower Tail Dependence	0.25	0.031	0.5	0.019		
Constant Upper Tail Dependence			0.5	0.006	0.75	0.076
Constant Asymmetry	0.25	0.006	0.5	<0.001	0.75	<0.001

Table 2: In-Sample Estimations. The monthly sample periods estimated are Jan. 1952-Dec. 2010 with 708 observations, and Jan. 1952-Dec. 2002 with 612 observations. Panel A is the estimated parameters and their statistics. Panel B is robustness tests. Log-likelihood values are computed as the mean of the sample log-likelihood values. Standard errors are reported in square brackets. *, **, *** represent the significant levels of 10%, 5%, and 1%, respectively. p-values are reported for likelihood ratio tests and Wald joint significance tests.

	Constant Decomposition Model	Dynamic Decomposition Models							
		Patton	Exp-Weighted	IGtype	Gtype	OSA	TSA	TVC	ITVC
Full Sample: 1/1952-12/2010									
Panel A									
ω		-2.940***	-2.222*		-1.249	-0.221*	-0.314		
		[0.895]	[1.198]		[0.948]	[0.133]	[0.324]		
β		0.218	0.512	0.767***	0.489**	0.932***	0.901***	0.645*	0.732**
		[0.455]	[0.395]	[0.138]	[0.219]	[0.062]	[0.125]	[0.358]	[0.361]
ϒ		4.815***	3.909***		0.313	4.865***	3.785**	0.087	
		[0.637]	[0.571]		[0.545]	[1.120]	[1.712]	[0.111]	
λ			0.702***				1.848		
			[0.286]				[3.027]		
α	0.112**								
	[0.052]								
Panel B									
Log-Likelihood Values	1.830	1.831	1.822	1.827	1.833	1.831	1.833	1.831	1.834
Likelihood Ratio Test		<0.001	<0.001	<0.001	0.004	0.040	0.005	0.002	0.036
Wald Joint Significance Test		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.041
Subsample: 1/1952-12/2002									
Panel A									
ω		-2.664**	-2.236*		-1.366	-0.235	-0.247		
		[1.255]	[1.325]		[1.439]	[0.153]	[0.200]		
β		0.436	0.492	0.768***	0.436	0.933***	0.927***	0.608	0.753
		[0.381]	[0.415]	[0.148]	[0.480]	[0.063]	[0.090]	[0.791]	[0.492]
ϒ		4.749*	3.802		0.092	5.301**	4.031***	0.046	
		[2.814]	[2.398]		[0.825]	[2.310]	[1.590]	[0.153]	
λ			0.677**				1.016		
			[0.292]				[1.890]		
α	0.093*								
	[0.055]								
Panel B									
Log-Likelihood Values	1.855	1.856	1.857	1.850	1.855	1.856	1.855	1.855	1.855
Likelihood Ratio Test		<0.001	<0.001	<0.001	0.003	0.041	0.006	0.001	0.030
Wald Joint Significance Test		0.002	0.012	<0.001	0.011	<0.001	<0.001	0.065	0.120

Table 3: Out-of-Sample Forecast Results. Values in this table are the OS statistic in percentage. Conditional Independence is ignoring the dependence by assuming conditionally independence between absolute returns and signs; its forecasts are obtained by eq(4.2). In the conditional dependence, the forecasts are computed by eq(4.1). The average performance of dynamic dependence models are computed.

	Linear	Conditional Independence										Conditional Dependence									
	Model	Constant	Dynamic Dependence Models									Constant	Dynamic Dependence Models								
		Model	Patton	ExpWeight	Igtype	Gtype	OSA	TSA	TVC	ITVC	Average	Model	Patton	ExpWeight	Igtype	Gtype	OSA	TSA	TVC	ITVC	Average
1982:02-2010:12																					
Squared errors	-3.87	0.18	1.33	0.88	1.69	0.87	1.57	2.53	1.76	0.62	1.41	2.16	3.03	1.58	3.59	0.99	3.57	1.91	2.98	3.32	2.62
Absolute errors	-3.54	0.70	1.67	1.08	1.59	1.46	1.67	1.74	1.45	0.53	1.40	2.23	2.81	1.76	2.79	2.16	3.12	2.51	2.67	2.94	2.59
1982:01-2002:12																					
Squared errors	-4.61	3.23	3.80	3.89	4.62	3.18	3.34	3.46	3.32	3.62	3.66	2.79	3.18	3.11	3.44	2.56	2.88	2.92	2.81	3.22	3.01
Absolute errors	-4.81	2.45	3.05	3.05	3.39	3.38	2.76	2.49	2.78	2.44	2.92	2.30	2.76	2.81	2.83	3.77	2.53	2.40	2.72	3.50	2.92
1982:01-1997:12																					
Squared errors	-4.42	4.13	4.22	5.26	5.05	5.26	4.56	4.75	3.68	4.34	4.64	0.79	0.89	1.14	1.66	1.76	1.25	1.66	0.95	1.73	1.38
Absolute errors	-4.95	3.61	4.06	4.61	4.20	5.32	4.07	3.76	3.69	3.62	4.17	1.87	1.83	2.36	2.31	3.73	1.79	1.93	2.23	2.98	2.39
1998:01-2002:12																					
Squared errors	-5.08	1.12	2.82	0.70	3.62	-1.72	0.49	0.44	2.48	1.94	1.35	5.18	9.82	8.52	9.36	8.58	8.77	5.34	8.25	9.62	8.53
Absolute errors	-4.44	-0.50	0.49	-0.92	1.33	-1.56	-0.58	-0.74	0.44	-0.57	-0.26	3.42	5.15	3.97	4.16	3.88	4.43	3.59	3.98	4.82	4.25
2007:07-2010:12																					
Squared errors	0.67	-3.57	-0.41	0.91	1.79	-1.36	1.57	5.06	3.52	-2.58	1.06	6.72	9.30	7.30	12.31	-0.01	11.21	6.49	10.84	9.02	8.31
Absolute errors	2.42	2.06	4.18	5.15	6.15	1.72	5.73	6.78	5.43	1.42	4.57	6.55	8.40	6.74	10.99	1.78	11.38	8.37	9.09	7.22	8.00

Table 4: Tests of conditional predictive ability based on Giacomini and White (2006). [a]: The relative performance indicates the percentage of times the model in column j dominates the model in row i over the whole forecasting period. [b]: decision rule is the two-step procedure for adaptively selecting a forecasting method at time t. The entries in decision rule are the loss differences, $E(\Delta L_{m,t+\tau}|\mathcal{F}_T)$. Step 1: Test the null hypothesis of equal conditional predictive ability between the model in column j and the model in column i. The rejection at 10%, 5% and 1% are denoted by *, **, ***, respectively. Step 2: in case of rejection, the decision rule is: use the model in column j if $E(\Delta L_{m,t+\tau}|\mathcal{F}_T) > c$ and use the model in row i if $E(\Delta L_{m,t+\tau}|\mathcal{F}_T) < c$. c is a user-specified threshold. In our case, we use $c = 0$. Test details are referred to Giacomini and White (2006).

				Conditional Independence								Conditional Dependence							
	Linear Model	CDM-CI	CDM-CD	Patton	ExpWeight	IGtype	Gtype	OSA	TSA	TVC	ITVC	Patton	ExpWeight	IGtype	Gtype	OSA	TSA	TVC	ITVC
Panel A: Squared Errors																			
1982:02-2010:12																			
Historical Average																			
[a]: relative performance	0.081	0.418	0.885	0.767	0.697	0.723	0.637	0.749	0.807	0.778	0.599	0.954	0.948	0.914	0.741	0.902	0.752	0.983	0.994
[b]: decision rule	-0.762	-0.047	0.371	0.179	0.090	0.247**	0.089***	0.224	0.414**	0.262	0.038*	0.541	0.258	0.653	0.144	0.649	0.324**	0.531	0.599
Linear Model																			
[a]: relative performance		0.856	0.804	0.862	0.816	0.939	1.000	0.925	0.934	0.890	0.890	0.821	0.726	0.816	0.830	0.841	0.856	0.810	0.824
[b]: decision rule		0.715*	1.133***	0.941**	0.852**	1.010	0.852	0.986	1.176	1.024	0.801	1.303***	1.020***	1.415***	0.906***	1.410***	1.086***	1.293***	1.361***
CDM-CI																			
[a]: relative performance			0.784	0.885	0.879	0.942	0.882	0.954	0.899	0.954	0.879	0.833	0.700	0.824	0.986	0.824	0.804	0.813	0.850
[b]: decision rule			0.417***	0.226***	0.137	0.295	0.136	0.270*	0.460***	0.309	0.086	0.588***	0.304***	0.700***	0.191	0.695***	0.371***	0.577***	0.646***
CDM-CD																			
[a]: relative performance				0.314	0.242	0.343	0.285	0.311	0.548	0.352	0.265	0.937	0.104	0.827	0.130	0.965	0.268	0.879	0.850
[b]: decision rule				-0.19***	-0.28**	-0.12***	-0.28***	-0.14***	0.043**	-0.10***	-0.33***	0.171	-0.113	0.282***	-0.22**	0.278	-0.04**	0.160***	0.228***
1982:01-2002:12																			
Historical Average																			
[a]: relative performance	0.000	0.797	0.932	0.916	0.988	0.908	0.972	0.988	0.841	0.884	0.928	1.000	0.857	0.984	0.960	0.924	0.821	0.980	0.976
[b]: decision rule	-0.846	0.521	0.201	0.629	0.645	0.781	0.513	0.540	0.566	0.540	0.594	0.365	0.324	0.443	0.410	0.357	0.216	0.284	0.466
Linear Model																			
[a]: relative performance		0.980	0.841	0.988	1.000	1.000	0.964	0.996	1.000	0.980	1.000	0.857	0.817	0.833	0.857	0.873	0.880	0.837	0.857
[b]: decision rule		1.367	1.046***	1.475	1.491	1.627	1.358	1.386	1.412	1.386	1.440	1.210**	1.169***	1.289***	1.255***	1.202**	1.061*	1.129***	1.311***
CDM-CI																			
[a]: relative performance			0.378	0.896	0.968	0.841	0.287	0.857	0.964	0.912	0.801	0.462	0.474	0.514	0.486	0.454	0.355	0.430	0.542
[b]: decision rule			-0.32***	0.108	0.123	0.259***	-0.009	0.019	0.045	0.018	0.073**	-0.15***	-0.19***	-0.07***	-0.11***	-0.16***	-0.30***	-0.23***	-0.05***
CDM-CD																			
[a]: relative performance				0.677	0.681	0.705	0.677	0.673	0.673	0.637	0.653	0.992	0.968	1.000	0.892	0.952	0.964	0.992	0.964
[b]: decision rule				0.428***	0.444***	0.580***	0.311***	0.339***	0.365***	0.339***	0.393***	0.164	0.123	0.243	0.209	0.156	0.015	0.083	0.265
Panel B: Absolute Errors																			
1982:02-2010:12																			
Historical Average																			
[a]: relative performance	0.003	0.516	0.746	0.963	0.764	0.850	0.960	0.899	0.772	0.827	0.352	0.761	0.671	0.738	0.839	0.769	0.697	0.841	0.914
[b]: decision rule	-12.943	0.291	5.746	3.514	1.538	3.225	2.810	3.483	3.772	2.802	-0.286	7.697**	4.205*	7.641**	5.542	8.727**	6.692***	7.211	8.134
Linear Model																			
[a]: relative performance		0.813	0.893	0.847	0.850	0.882	0.971	0.902	0.916	0.856	0.853	0.908	0.873	0.928	0.914	0.911	0.963	0.890	0.879
[b]: decision rule		13.23**	18.68**	16.45***	14.48*	16.16*	15.753	16.42*	16.714	15.74**	12.65**	20.63**	17.14**	20.58**	18.48**	21.66***	19.635	20.15***	21.07***
CDM-CI																			
[a]: relative performance			0.793	0.847	0.720	0.827	0.876	0.994	0.798	0.968	0.271	0.844	0.784	0.957	0.919	0.816	0.824	0.807	0.914
[b]: decision rule			5.455**	3.223***	1.247***	2.934***	2.520	3.193	3.481***	2.512	-0.57*	7.406*	3.915	7.351	5.252	8.436**	6.401*	6.920*	7.844
CDM-CD																			
[a]: relative performance				0.352	0.098	0.323	0.271	0.219	0.297	0.314	0.205	0.879	0.308	0.617	0.354	0.980	0.712	0.804	0.862
[b]: decision rule				-2.23**	-4.207	-2.52*	-2.936	-2.262	-1.974	-2.94**	-6.03**	1.952	-1.54***	1.896***	-0.204	2.982	0.946***	1.465***	2.388**
1982:01-2002:12																			
Historical Average																			
[a]: relative performance	0.020	0.665	0.892	0.717	0.845	0.801	0.781	0.737	0.677	0.673	0.705	0.944	0.996	0.984	0.948	0.876	0.876	0.928	1.000
[b]: decision rule	-17.215	5.234**	5.281	7.192**	7.147	8.260*	8.252**	6.192	5.381**	6.291***	5.175*	6.808	6.954	7.034	10.085	6.090	5.616	6.655	9.198
Linear Model																			
[a]: relative performance		0.936	0.944	0.952	1.000	0.996	1.000	0.996	0.984	0.936	0.980	0.952	0.920	0.936	0.964	0.936	0.964	0.944	0.924
[b]: decision rule		22.449	22.496	24.408	24.362	25.475	25.468	23.407	22.597	23.507	22.391	24.023	24.16*	24.24*	27.300	23.305	22.831	23.870	26.41*
CDM-CI																			
[a]: relative performance			0.546	0.884	0.976	0.801	0.944	0.880	0.622	0.900	0.426	0.594	0.598	0.590	0.717	0.574	0.582	0.598	0.701
[b]: decision rule			0.046***	1.958*	1.913	3.025***	3.018	0.957	0.147	1.057	-0.059	1.573***	1.719***	1.799***	4.850***	0.855***	0.382	1.420***	3.963**
CDM-CD																			
[a]: relative performance				0.554	0.558	0.590	0.633	0.526	0.470	0.498	0.474	1.000	0.964	0.857	0.972	0.896	0.992	0.932	1.000
[b]: decision rule				1.911***	1.866**	2.979**	2.971*	0.911	0.100**	1.010***	-0.10***	1.527	1.673	1.753	4.804	0.809	0.335	1.374	3.917

Table 5: Comparison of copula specifications in out-of-sample density forecasts. The values, scaled by 100, are reported the averages of the difference in density forecasts between the benchmark model and competing models. The benchmark model is the constant decomposition model. The null hypothesis is the equality of the density forecasts between the benchmark model and a competing model. *, **, *** represent the significant levels of 10%, 5%, and 1%, respectively. In the case of rejection, a negative value provides the evidence of the superior density forecast ability of a competing model over the benchmark model. The tests in this table are based on Diks et al. (2010).

	Patton	ExpWeight	IGtype	Gtype	OSA	TSA	TVC	ITVC
1982:02-2010:12	0.236***	0.052	-0.909***	-0.521	-0.243***	-0.516***	0.175	0.052
1982:01-2002:12	0.484	-1.020	-1.272***	-0.204	-0.277	-0.423***	0.181***	0.247
1982:01-1997:12	0.551	-1.441***	-1.435***	-0.364	-0.439***	-0.445***	0.255	0.312**
1998:01-2002:12	-0.025	0.235**	-0.806***	0.245	0.214	-0.348***	-0.106***	-0.017
2007:07-2010:12	-0.229***	2.664***	-0.189	2.039***	0.376	-0.636	-0.612	-0.604

Table 6: Sources of Forecasting Performance. The entries are test statistics for the null hypotheses of no difference in time variation of forecasting ability. The historical average, linear model, and constant models in the first column as benchmarks are compared to competing models in the first row. *, **, *** denote the significance levels at 10%, 5%, and 1%. Panel A and B use squared and absolute error as a loss function, respectively.

				Conditional Independence								Conditional Dependence							
	Linear Model	CDM-CI	CDM-CD	Patton	ExpWeight	IGtype	Gtype	OSA	TSA	TVC	ITVC	Patton	ExpWeight	IGtype	Gtype	OSA	TSA	TVC	ITVC
Panel A: Squared Errors																			
Historical Average	10.54	12.17***	12.44***	13.38***	20.42***	18.02***	11.16	14.04***	20.52***	14.45***	11.54	13.06***	19.43***	17.22***	10.11	14.19***	21.94***	14.77***	11.85
Linear Model		13.14***	13.21***	13.27***	19.42***	17.39***	16.11***	14.15***	14.70***	14.07***	13.77***	13.08***	18.66***	16.73***	16.11***	14.40***	16.22***	14.39***	13.98***
CDM-CI			9.102	19.45***	28.51***	25.56***	27.80***	23.50***	27.01***	35.86***	15.41***	14.68***	26.60***	22.35***	25.83***	20.83***	25.03***	35.39***	11.17
CDM-CD				21.70***	27.44***	23.83***	29.43***	23.59***	28.87***	34.34***	20.58***	18.30***	26.37***	23.17***	28.32***	23.34***	27.65***	36.70***	16.94***
Panel B: Absolute Errors																			
Historical Average	12.39***	14.96***	14.69***	14.84***	21.78***	18.90***	15.64***	13.75***	17.39***	14.59***	14.10***	15.65***	21.89***	19.40***	15.66***	15.11***	19.69***	15.80***	14.05***
Linear Model		19.31***	19.29***	17.73***	17.98***	16.81***	18.16***	17.53***	17.39***	19.35***	17.92***	17.44***	18.22***	16.84***	18.22***	17.24***	17.18***	19.15***	18.14***
CDM-CI			13.72***	15.96***	24.63***	20.82***	22.64***	21.85***	26.65***	23.46***	8.978	15.11***	24.79***	21.30***	20.06***	23.53***	25.05***	25.82***	8.376
CDM-CD				15.91***	22.74***	17.71***	24.13***	21.00***	28.09***	23.68***	11.41	16.42***	23.45***	20.76***	22.74***	24.21***	26.89***	25.15***	7.707

Table 7: Links to real economy. TP is term premium. TS is term spread. IR is short-term interest rate, which is 3-month T-bill in this context. GG is GDP growth computed as the changes in industrial production. Infl is inflation. HP is housing price changes. *, **, *** denote the significance levels at 10%, 5%, and 1%.

	Macro Variable in Level						Macro Variable in Volatility			
	TP	TS	IR	GG	infl	HP	IR	GG	infl	HP
Panel A: Downside Risk										
Historical Average	0.009***	-9.30	-0.00***	-0.00	-0.00**	-0.00	-0.00***	-0.03**	0.036	-0.00***
Constant Model	-0.06***	0.008**	-0.00**	-0.02***	0.014	0.0006	0.001	0.446***	2.392***	-0.0014***
Patton	-0.06***	0.009**	-0.00**	-0.02***	0.019	0.0006	0.001	0.488***	2.677***	-0.0012**
ExpWeight	-0.06***	0.012***	-0.00	-0.02***	0.016	0.0007	0.003	0.448***	2.371***	-0.0010**
IGtype	-0.06***	0.012***	-0.00**	-0.02***	0.019	0.0006	0.004**	0.543***	2.719***	-0.0010**
Gtype	-0.05***	0.004	-0.00	-0.02***	0.020	0.0003	0.001	0.375***	1.932***	-0.0012***
OSA	-0.06***	0.009**	-0.00**	-0.02***	0.017	0.0007	0.002	0.468***	2.460***	-0.0013***
TSA	-0.05***	0.007**	-0.00**	-0.02***	0.014	0.0002	0.001	0.444***	2.313***	-0.0013***
TVC	-0.06***	0.010**	-0.00**	-0.02***	0.014	0.0003	0.002	0.496***	2.653***	-0.0013***
ITVC	-0.06***	0.009**	-0.00**	-0.02***	0.016	0.0003	0.001	0.474***	2.575***	-0.0013***
Panel B: Skewness										
Historical Average	-7.54***	5.259***	6.033***	6.289***	11.57**	0.150	4.736***	7.878	-638.***	1.410***
Constant Model	4.789***	-2.77***	0.330	0.361	3.795	0.089	-0.68	-32.3**	-187.***	-0.12
Patton	3.857**	-1.26	-0.27	0.996	4.399	0.152	-0.23	7.189	34.12	-0.08
ExpWeight	2.285	-0.84	-1.81***	3.110	3.348	0.099	0.204	78.75***	524.7***	-0.22
IGtype	0.859	-0.13	-1.28***	1.564	3.509	0.144	-0.10	42.68**	275.6***	-0.15
Gtype	4.149**	-1.27**	0.465	1.543	5.111	0.156	-0.42	-27.4**	-190.***	-0.04
OSA	2.454	-1.94**	-0.12	-0.17	1.333	0.093	-0.58	-4.24	-49.2	-0.17**
TSA	3.492**	-2.29***	0.072	0.597	3.397	0.125	-0.65	-8.61	-85.4	-0.14
TVC	4.964***	-2.76***	0.223	0.124	4.714	0.100	-0.77	-32.5**	-189.***	-0.13
ITVC	4.774**	-2.44***	0.310	0.734	4.767	0.101	-0.70	-34.7**	-223.***	-0.13

Figure 1: Scatter-plots of simulated return moments against changes in dependence parameters, α, of Clayton copula. The solid lines represent sample return moments of U.S. returns. The dotted lines are the estimated dependence structure from constant decomposition model.

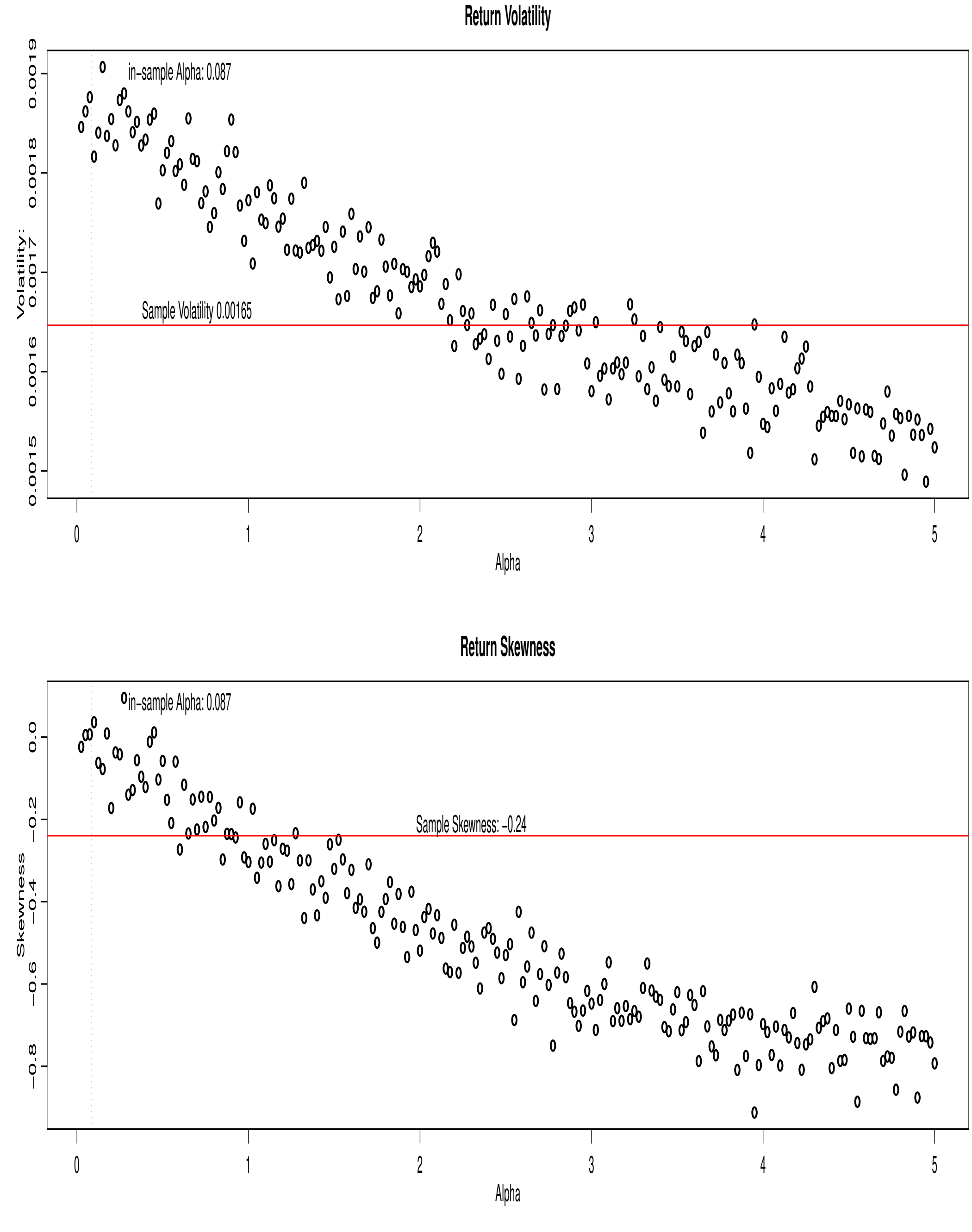


Figure 2: Plot of Out-of-Sample Forecasts from Constant decomposition model (solid line), linear model (dash-dotted line), and dynamic decomposition model (dash line). The out-of-sample forecast period is 1982:01-2010:12.

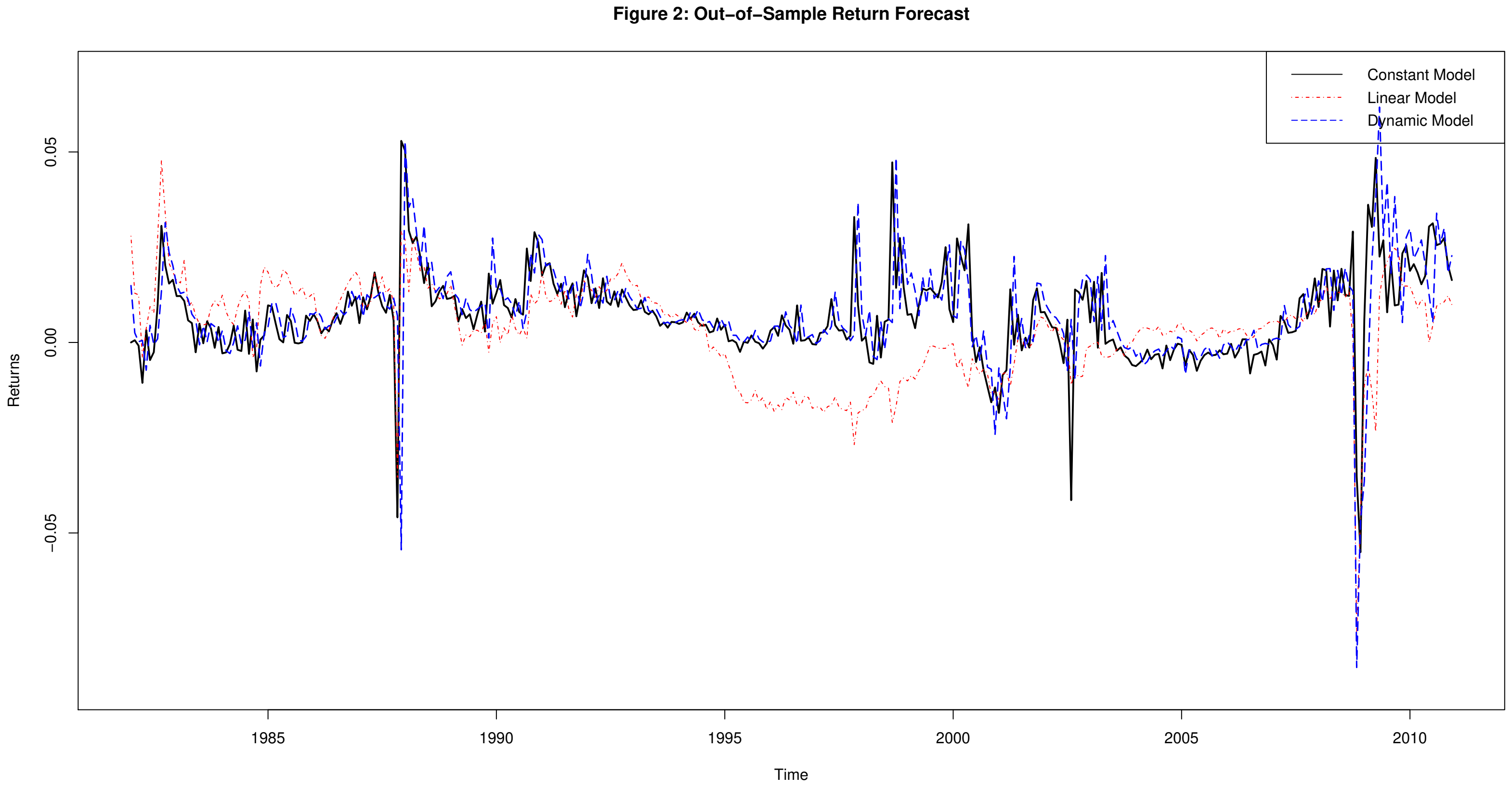


Figure 3: Out-of-sample dependence structure estimation from the constant decomposition model (solid line) and the dynamic decomposition models (non-solid lines).

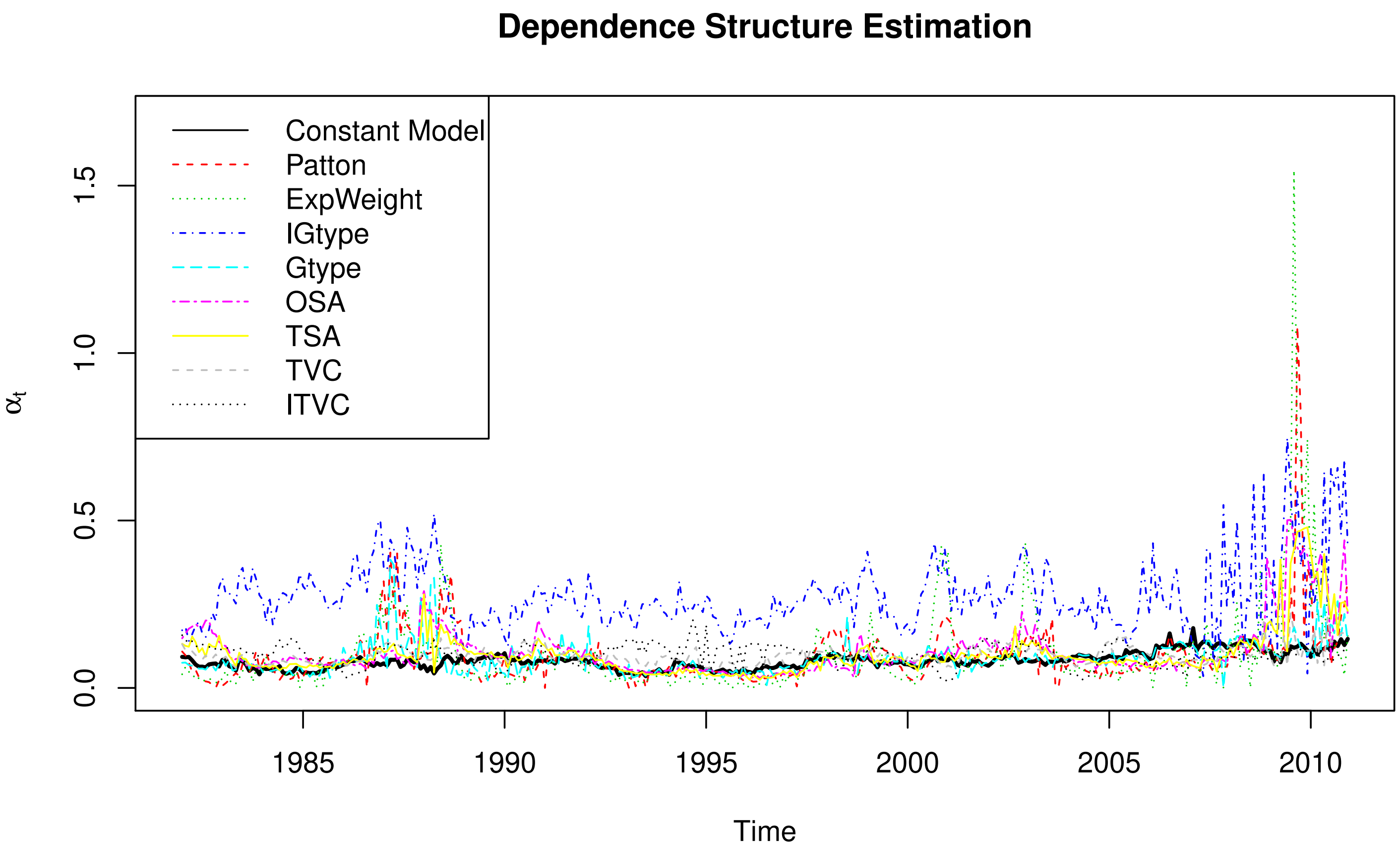
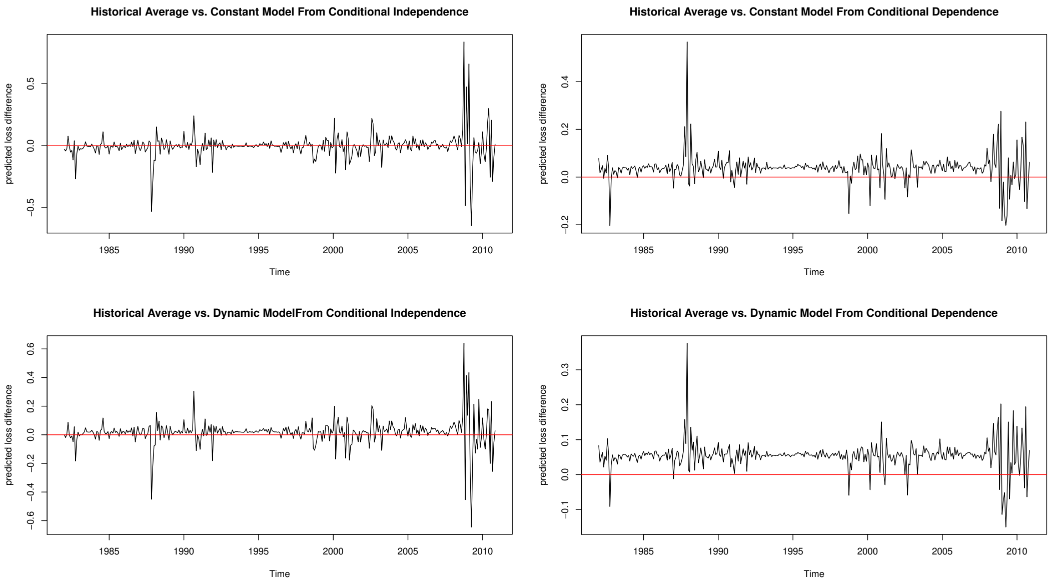


Figure 4: Plots of predicted loss differences over the entire time path. The horizontal line denotes a zero level. Positive values represent the time periods that a competing model performs better out-of-sample than benchmarks.

Panel A: Historical average as benchmark



Panel B: Constant decomposition model as benchmark

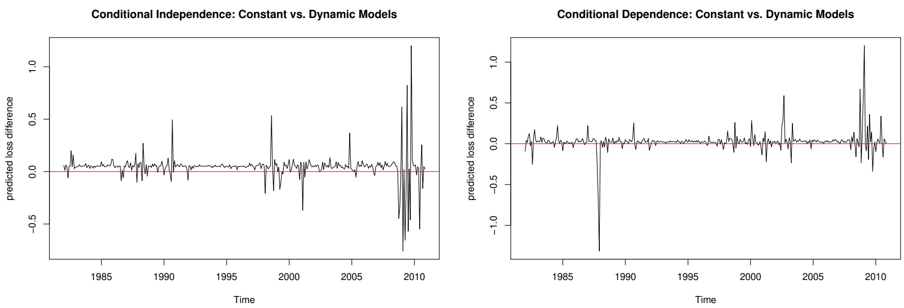
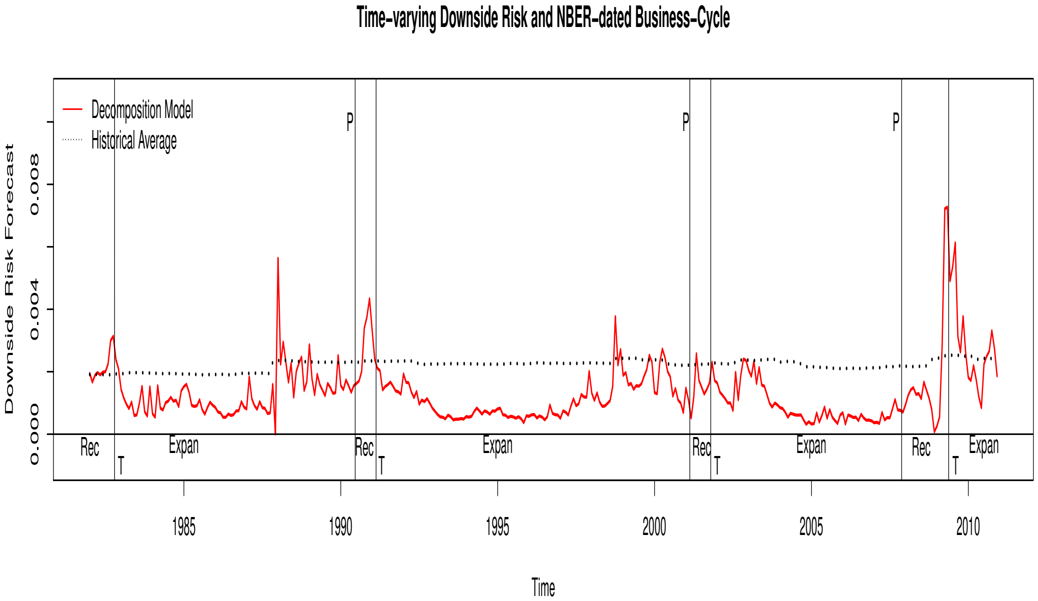
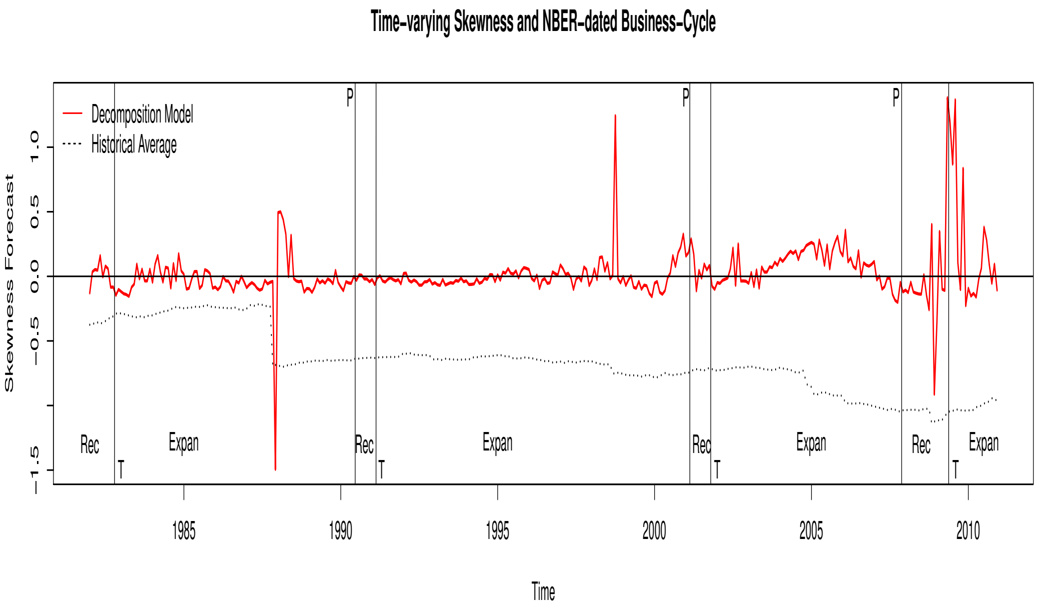


Figure 5: Plots of out-of-sample forecasts for time-varying downside risk and skewness along with vertical lines indicating NBER-dated business-cycle phases. "Rec" and "Expan" denote recession and expansion periods, and "P" and "T" represent a peak and a trough, respectively.





A Higher Moments' Forecasting

We integrate the following integral

$$
\begin{aligned}
\xi_t &= E_{t-1}\left(|r_t - c|\,\mathbb{I}[r_t > c]\right) \\
&= \int_0^{+\infty} u f\left(u|\psi\right) \varrho_t\left(F\left(u|\psi\right)\right) du
\end{aligned} \tag{A.1}
$$

We transform the infinity domain to be $[0,1]$ by letting $z = F\left(u|\psi\right)$. Hence, we have $u = F^{-1}(z|\psi)$ with $0 \le z \le 1$ since $F(\cdot)$ is a CDF function. Now, we rewrite (A.1) as:

$$
\xi_t = \int_0^1 Q(z) \varrho_t\left(z\right) dz \tag{A.2}
$$

where $Q(z) = F^{-1}(z|\psi) f\left(F^{-1}(z|\psi)|\psi\right) \frac{\partial F^{-1}(z|\psi)}{\partial z}$.

In our case, $z = F(u|\psi) = 1 - e^{-\left(\frac{u}{\lambda}\right)^k}$ with $k, \lambda > 0$ is a Weibull CDF distribution, and $u = F^{-1}(z|\psi) = \lambda\left[-ln\left(1-z\right)\right]^{\frac{1}{k}}$. Further, the PDF for a Weibull distribution is:

$$
f(u;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{u}{\lambda}\right)^{k-1} e^{-\left(\frac{u}{\lambda}\right)^k} & u \ge 0 \\ 0 & u < 0 \end{cases}
$$

such that

$$
\begin{aligned}
f\left(F^{-1}(z;\lambda,k);\ \lambda,k\right) &= f\left\{\lambda\left[-ln\left(1-z\right)\right]^{\frac{1}{k}}\right\} \\
&= \frac{k}{\lambda}\left(\frac{\lambda\left[-ln\left(1-z\right)\right]^{\frac{1}{k}}}{\lambda}\right)^{k-1} e^{-\left(\frac{\lambda\left[-ln\left(1-z\right)\right]^{\frac{1}{k}}}{\lambda}\right)^k} \\
&= \frac{k(1-z)}{\lambda}\left[-ln\left(1-z\right)\right]^{1-\frac{1}{k}}
\end{aligned} \tag{A.3}
$$

and

$$\begin{aligned} \frac{\partial F^{-1}(z;\lambda,k)}{\partial z} &= \frac{\partial\left\{\lambda\left[-ln\left(1-z\right)\right]^{\frac{1}{k}}\right\}}{\partial z} \\ &= \frac{\lambda}{k(1-z)}\left[-ln\left(1-z\right)\right]^{\frac{1}{k}-1} \end{aligned} \tag{A.4}$$

Combine (A.3) and (A.4) to obtain

$$Q(z) \;=\; \lambda\left[-ln\left(1-z\right)\right]^{\frac{1}{k}} \tag{A.5}$$

Therefore, $Q(z) = u = F^{-1}(z|\psi)$ is the quantile function of the Weibull distribution for z.

The solution for $E(r_t)^2$ is derived as:

$$\begin{aligned} E(r_t)^2 &= \int_0^{+\infty} u^2 f(u|\psi_t)du \\ &= \int_0^1 \left[F^{-1}(z|\psi_t)\right]^2 f\left(F^{-1}(z|\psi_t)|\psi_t\right) dF^{-1}(z|\psi_t) \\ &= \int_0^1 \left[F^{-1}(z|\psi_t)\right]^2 f\left(F^{-1}(z|\psi_t)|\psi_t\right) \frac{\partial F^{-1}(z;\lambda,k)}{\partial z}dz \\ &= \int_0^1 \left[F^{-1}(z|\psi_t)\right]^2 dz \end{aligned}$$

Note that $u = F^{-1}(z|\psi_t) = Q(z)$, and based eq(A.3) and A.4, $f\left(F^{-1}(z|\psi_t)|\psi_t\right)\frac{\partial F^{-1}(z;\lambda,k)}{\partial z} = 1$. Thus, we have

$$E(r_t)^2 = \int_0^1 \left[Q(z)\right]^2 dz \tag{A.6}$$

for $Q(z) = \lambda\left[-ln\left(1-z\right)\right]^{\frac{1}{k}}$ as in eq(A.5). Following the proof of eq(A.6), we can obtain the solutions of $E\left|r_t\right|^3$ as in (4.6).

The solution for $E(r_t^2\mathbb{I}(r_t > 0))$ is derived as

$$
\begin{aligned}
E(r_t^2 \mathbb{I}(r_t > 0)) &= \int_0^{+\infty} u^2 f(u|\psi_t) \varrho_t(F(u|\psi_t)) du \\
&= \int_0^1 \left[F^{-1}(z|\psi_t)\right]^2 f\left(F^{^{-1}}(z|\psi_t)|\psi_t\right) \varrho_t(z) \frac{\partial F^{^{-1}}(z;\lambda,k)}{\partial z} dz \\
&= \int_0^1 \left[F^{-1}(z|\psi_t)\right]^2 \varrho_t(z) dz \\
&= \int_0^1 [Q(z)]^2 \varrho_t(z) dz
\end{aligned}
$$

Similarly, we can prove the solution of $E(|r_t|^3 \mathbb{I}(r_t > 0))$ as in eq(4.5).