

**Centre for
Economic
and Financial
Research
at
New Economic
School**



January 2007

Modeling Financial Return Dynamics by Decomposition

Stanislav Anatolyev
Nikolay Gospodinov

Working Paper No 95

CEFIR / NES Working Paper series

Modeling Financial Return Dynamics by Decomposition

Stanislav Anatolyev*
New Economic School

Nikolay Gospodinov†
Concordia University

January 2007

Abstract

While the predictability of excess stock returns is statistically small, their sign and volatility exhibit a substantially larger degree of dependence over time. We capitalize on this observation and consider prediction of excess stock returns by decomposing the equity premium into a product of sign and absolute value components and carefully modeling the marginal predictive densities of the two parts. Then we construct the joint density of a positively valued (absolute returns) random variable and a discrete binary (sign) random variable by copula methods and discuss computation of the conditional mean predictor. Our empirical analysis of US stock return data shows among other interesting findings that despite the large unconditional correlation between the two multiplicative components they are conditionally very weakly dependent.

Key words: Stock returns predictability; Directional forecasting; Absolute returns; Joint predictive distribution; Copulas.

*Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru.

†Address: Nikolay Gospodinov, Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8. E-mail: gospodin@alcor.concordia.ca.

1 Introduction and Motivation

It is now widely believed that excess stock returns exhibit a certain degree of predictability over time (Cochrane, 2005). For instance, valuation (dividend-price and earnings-price) ratios (Fama and French, 1988; Campbell and Shiller, 1988a,b) and yields on short- and long-term Treasury and corporate bonds (Campbell, 1987; Hodrick, 1992) appear to possess statistically small but economically meaningful predictive power at short horizons that can be exploited for timing the market and active asset allocation (Campbell and Thompson, 2005).

Given the great practical importance of predictability of excess stock returns, there is a growing recent literature in search of new variables with incremental predictive power such as share of equity issues in total new equity and debt issues (Baker and Wurgler, 2000), consumption-wealth ratio (Lettau and Ludvigson, 2001), relative valuations of high- and low-beta stocks (Polk, Thompson and Vuolteenaho, 2006) etc. In this paper, we take an alternative approach to predicting excess returns by modeling individual multiplicative components of excess stock returns and combining the components' information using copula methods to recover the conditional expectation of the original variable of interest.

More specifically, suppose that we are interested in predicting excess stock returns based on past data and let r_t denote the excess return at period t . The return r_t can be decomposed (Christoffersen and Diebold, 2006) as

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

where $\text{sign}(\cdot)$ is the sign function and c is an arbitrary constant that may be chosen by the researcher (the leading case is $c = 0$). It is introduced for more generality because there may be more predictability with non-zero thresholds than with a zero threshold (Chung and Hong, 2006; Linton and Whang, 2005). Then, the conditional mean of r_t is given by

$$E_{t-1}[r_t] = c + E_{t-1}[|r_t - c| \text{sign}(r_t - c)],$$

where $E_{t-1}[\cdot]$ denotes the expectation taken with respect to the available information up to time $t - 1$.

Our aim is to model the joint distribution of $|r_t - c|$ and $\text{sign}(r_t - c)$ in order to obtain the conditional expectation $E_{t-1}[r_t]$. We expect this detour to be successful for the following reasons.

First, the models for $|r_t - c|$ and $\text{sign}(r_t - c)$ are likely to use different information variables than the covariates in a typical predictive regression of stock returns. Volatility persistence and predictability has been extensively studied and documented in the literature (see Andersen *et al.*, 2006, for an up-to-date comprehensive review on model-based and model-free volatility forecasting). Furthermore, Christoffersen and Diebold (2006) and Christoffersen *et al.* (2006) argue that time variability in the conditional second and higher-order moments of stock returns can generate sign predictability even in the absence of conditional mean dependence.

Second, there is evidence on sign predictability and volatility predictability of much larger degree than mean predictability. Christoffersen and Diebold (2006), Hong and Chung (2003) and Linton and Whang (2005) find convincing evidence of sign predictability of US stock returns for different threshold values and data frequencies. Christoffersen and Diebold (2006) reconcile the standard finding of weak conditional mean predictability with possibly strong sign and volatility dependence.

Third, the joint predictive density of $|r_t - c|$ and $\text{sign}(r_t - c)$ provides a more general inference procedure than modeling directly the conditional expectation of r_t as in the predictive regression literature. For example, the joint modeling would allow the researcher to explore trading strategies and evaluate their profitability (Gençay, 1998; Qi, 1999; Anatolyev and Gerko, 2005). Also, the joint modeling of the multiplicative components can incorporate some important nonlinearities in excess return dynamics that cannot be captured in the standard predictive regression setup.

Finally, we would like to stress that studying the dependence between the sign and absolute value components over time and the bivariate modeling that we propose are also interesting in their own right and can be used for various other purposes. For example, in our empirical analysis of US stock return data we reach a conclusion that despite the large unconditional correlation between the two multiplicative components, they are conditionally very weakly dependent or even independent.

The rest of the paper is organized as follows. Section 2 introduces our return decomposition and discusses the marginal density specifications and the construction of the joint predictive density of sign and volatility components. Section 3.2 summarizes some of the main findings in the literature of predictability of excess returns using Campbell and Yogo's (2006) data set. Sections 3.3 and 3.4 present the results from the joint modeling and provides some in-sample and out-of-sample comparisons with the benchmark predictive regression. Section 4 concludes.

2 Methodological Framework

2.1 Decomposition

Consider the return decomposition

$$r_t = c + |r_t - c| \text{sign}(r_t - c). \quad (1)$$

Christoffersen and Diebold (2006) analyze the case when $c = 0$ while Hong and Chung (2003) and Linton and Whang (2005) use threshold values for c that are multiples of the standard deviation of r_t or quantiles of the marginal distribution of r_t . The non-zero thresholds may reflect the presence of transaction costs and capture possible different dynamics of small, large positive and large negative returns (Chung and Hong, 2006). In a different context, Rydberg and Shephard (2003) use a similar decomposition to model the dynamics of the trade-by-trade price movements. The potential usefulness of decomposition (1) is also stressed in Granger (1998) and Anatolyev and Gerko (2005).

Suppose for now that the dynamics of r_t is a function only of its past values and the object of interest is the predictive density $f(r_t|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1} = \sigma(r_s : s \leq t-1)$ is an increasing sequence of sigma-fields ($\dots\mathcal{F}_{t-3} \subset \mathcal{F}_{t-2} \subset \mathcal{F}_{t-1}$) generated by the history of the series to date $t-1$. Later we will expand the sigma-field to include other information variables x_{t-1}, x_{t-2}, \dots , such as past values of dividend yield, interest rates etc. Part of our motivation to model the dynamics of the sign and volatility components on the right-hand side of (1), instead of directly modeling returns, is that r_t , $\text{sign}(r_t - c)$ and $|r_t - c|$ may use different parts of the information set. In order to emphasize the dependence of return components on different parts of the sigma-field, it may be instructive to decompose the natural filtration \mathcal{F}_{t-1} (Rydberg and Shephard, 2003) into $\mathcal{F}_{t-1}^{\text{sign}(r_t - c)}$ and $\mathcal{F}_{t-1}^{|r_t - c|}$, where $\mathcal{F}_{t-1} = \mathcal{F}_{t-1}^{\text{sign}(r_t - c), |r_t - c|}$. Note also that decomposition (1) allows for a time-varying conditional mean which is consistent with the empirical findings in the recent literature on predictability of excess stock returns.

2.2 Marginal distributions

Consider first the model specification for absolute returns. Since $|r_t - c|$ is a positively valued variable, the dynamics of absolute returns is specified using the multiplicative error modeling

(MEM) framework of Engle (2002)¹

$$|r_t - c| = \exp(\psi_t)\eta_t,$$

where $\exp(\psi_t) \equiv E_{t-1}[|r_t - c|]$ (or equivalently $\psi_t \equiv \ln E_{t-1}[|r_t - c|]$), and η_t is a positive multiplicative error with $E_{t-1}[\eta_t] = 1$ and conditional distribution \mathcal{D} . The dynamic specification will be placed on ψ_t in the spirit of the logarithmic autoregressive conditional duration (LACD) model of Bauwens and Giot (2000). The main advantage of the logarithmic specification is that no parameter restrictions are needed to enforce positivity of $E_{t-1}[|r_t - c|]$. Possible candidates for \mathcal{D} include exponential, Weibull, Burr and Generalized Gamma distributions, and potentially free parameters of \mathcal{D} may be parameterized as functions of the past. In the empirical section, we use the constant parameter Weibull distribution as it turns out that its flexibility is sufficient to provide adequate description of the conditional density of absolute excess returns.

The conditional mean function is parameterized as

$$\psi_t = \omega_r + \beta_r \psi_{t-1} + \gamma_r \ln |r_{t-1} - c| + \rho_r \mathbb{I}[r_{t-j} > c] + x'_{t-1} \delta_r. \quad (2)$$

If only the first three terms on the right-hand side of (2) are included, the structure of the model is analogous to the LACD model of Bauwens and Giot (2000) and log GARCH model of Geweke (1986) where the persistence of the process is measured by the parameter $|\gamma_r + \beta_r|$. We also allow for regime-specific mean volatility depending on whether $r_{t-j} > c$ or $r_{t-j} \leq c$.² Finally, the inclusion of $x'_{t-1} \delta_r$ accounts for the possibility that (functions of) other variables in the information set at time $t - 1$ may have an effect on volatility dynamics proxied by $|r_t - c|$. In what follows, we refer to model (2) as volatility model. The vector of unknown parameters $(\varsigma, \omega_r, \beta_r, \gamma_r, \rho_r, \delta'_r)'$ can be estimated by maximizing the log-likelihood function.

Now we turn our attention to the dynamic specification of the variable $\mathbb{I}[r_t > c]$, where $\mathbb{I}[\cdot]$ is the indicator function, which is linearly related to $\text{sign}(r_t - c)$. The conditional distribution of $\mathbb{I}[r_t > c]$, given past information, is necessarily Bernoulli $\mathcal{B}(p_t)$ with the density $f_{\mathbb{I}[r_t > c]}(v) = p_t^v (1 - p_t)^{1-v}$, where $p_t = E_{t-1}[\mathbb{I}[r_t > c]]$ denotes the conditional “probability of success” $\Pr_{t-1}(\mathbb{I}[r_t > c] = 1)$.

¹The leading application of the MEM approach in the econometrics literature is that to durations between successive transactions in a high frequency financial market (see, for example, Engle and Russel, 1998). There are other occasional applications of the MEM approach. Engle (2002) illustrates the MEM methodology using exchange rate realized volatilities. Chou (2005) models a high/low range of asset prices in the MEM framework. Engle and Gallo (2006) analyze the dynamics of three volatility indexes using a multivariate version of the MEM.

²We also interacted ψ_{t-1} and $\ln |r_{t-1} - c|$ terms with $\mathbb{I}[r_{t-j} > c]$ but the estimated coefficients on these variables were statistically insignificant.

If the data are generated by $r_t = \mu_t + \sigma_t \varepsilon_t$, where $\mu_t = E_{t-1}(r_t)$, $\sigma_t^2 = \text{Var}_{t-1}(r_t)$ and ε_t is a homoskedastic martingale difference with unit variance and distribution function $F_\varepsilon(\cdot)$, Christoffersen and Diebold (2006) show that

$$E_{t-1} [\mathbb{I} [r_t > c]] = 1 - F_\varepsilon \left(\frac{c - \mu_t}{\sigma_t} \right).$$

This expression implies that time-varying volatility will generate sign predictability so long as $c - \mu_t \neq 0$. Furthermore, Christoffersen *et al.* (2006) derive a Gram-Charlier expansion of $F_\varepsilon(\cdot)$ and show that $E_{t-1} [\mathbb{I} [r_t > c]]$ depend on the third and fourth conditional cumulants of the standardized errors ε_t . As a result, sign predictability would arise from time variability in second and higher-order moments.

We use these insights and parameterize $E_{t-1} [\mathbb{I} [r_t > c]]$ using the dynamic logit model

$$E_{t-1} [\mathbb{I} [r_t > c]] = \frac{\exp(\theta_t)}{1 + \exp(\theta_t)}$$

with

$$\theta_t = \omega_s + \phi_s \mathbb{I} [r_{t-1} > c] + x'_{t-1} \delta_s. \quad (3)$$

Model (3) on the right side includes the lagged values of the indicator as well as some other variables in x_{t-1} . These may be macroeconomic predictors such as interest rates and various valuation ratios. We include in particular $RV_{t-1}, BPV_{t-1}, RS_{t-1}, RK_{t-1}$, where RV , BPV , RS and RK denote the realized variance, bipower variation, realized third and fourth moments of returns constructed from daily data.³ We include both RV and BPV as proxies for the unobserved volatility process since the former is an estimator of integrated variance plus a jump component while the latter is unaffected by presence of jumps (Barndorff-Nielsen and Shephard, 2004, 2006). The unknown parameters in (3) can be estimated by maximum likelihood.⁴

³We experimented with some flexible nonlinear specifications of θ_t in order to capture the possible interaction between volatility and higher-order moments (Christoffersen *et al.*, 2006) but the nonlinear terms did not deliver incremental predictive power and are omitted from the final specification.

⁴de Jong and Woutersen (2005) provide conditions for the consistency and asymptotic normality of the parameters estimates in dynamic binary choice models.

2.3 Joint distribution using copulas

This section discusses the construction of the bivariate conditional distribution of $R_t \equiv (|r_t - c|, \mathbb{I}[r_t > c])'$. Up to now we have dealt with the marginals⁵

$$\begin{pmatrix} |r_t - c| \\ \mathbb{I}[r_t > c] \end{pmatrix} \sim \begin{pmatrix} \mathcal{D}(\psi_t) \\ \mathcal{B}(p_t) \end{pmatrix},$$

with marginal PDF/PMFs

$$\begin{pmatrix} f_{|r_t - c|}(u) \\ f_{\mathbb{I}[r_t > c]}(v) \end{pmatrix} = \begin{pmatrix} f^{\mathcal{D}}(u|\psi_t) \\ p_t^v (1 - p_t)^{1-v} \end{pmatrix},$$

and marginal CDF/CMFs

$$\begin{pmatrix} F_{|r_t - c|}(u) \\ F_{\mathbb{I}[r_t > c]}(v) \end{pmatrix} = \begin{pmatrix} F^{\mathcal{D}}(u|\psi_t) \\ 1 - p_t(1 - v) \end{pmatrix}.$$

Let $f_{R_t}(u, v)$ denote the joint density of $|r_t - c|$ and $\mathbb{I}[r_t > c]$. We will use the copula theory (for introduction to copulas, see Nelson, 1999) to generate the joint distribution from the specified marginals

$$F_{R_t}(u, v) = C(F_{|r_t - c|}(u), F_{\mathbb{I}[r_t > c]}(v)),$$

where $C(w_1, w_2)$ is a copula. The unusual feature of the copula in our case is the continuity of one marginal and discreteness of the other marginal. The typical case in bivariate modeling are two continuous marginals (for example, Patton, 2006) and much more rarely two discrete marginals (Cameron *et al.*, 2004).

Because the first component is continuously distributed while the second component is a discrete binary random variable, the joint density/mass function can be obtained as a partial derivative with respect to the continuous entry and a finite difference with respect to the binary entry:

$$f_{R_t}(u, v) = \frac{\partial F_{R_t}(u, v)}{\partial w_1} - \frac{\partial F_{R_t}(u, v - 1)}{\partial w_1}.$$

Theorem. The joint density/mass function $f_{R_t}(u, v)$ can be represented as

$$f_{R_t}(u, v) = f^{\mathcal{D}}(u|\psi_t) \varrho_t(F^{\mathcal{D}}(u|\psi_t))^v (1 - \varrho_t(F^{\mathcal{D}}(u|\psi_t)))^{1-v}, \quad (4)$$

where

$$\varrho_t(z) = \frac{\partial C(z, 1)}{\partial w_1} - \frac{\partial C(z, 1 - p_t)}{\partial w_1}.$$

⁵For brevity we use the terms “marginal distribution”, “joint distribution” and the like, although a more correct terminology would be “conditional marginal distribution”, “conditional joint distribution”, etc., where the qualifier “conditional” refers to conditioning on the past.

Proof. Differentiation of $F_{R_t}(u, v)$ yields

$$f_{R_t}(u, v) = f_{|r_t - c|}(u) \left[\frac{\partial C(F^{\mathcal{D}}(u|\psi_t), F_{1[r_t > c]}(v))}{\partial w_1} - \frac{\partial C(F^{\mathcal{D}}(u|\psi_t), F_{1[r_t > c]}(v - 1))}{\partial w_1} \right].$$

Denote by $\Delta_t(v)$ the expression in the square brackets. In particular, when $v = 0$,

$$\Delta_t(0) = \frac{\partial C(F^{\mathcal{D}}(u|\psi_t), 1 - p_t)}{\partial w_1} - \frac{\partial C(F^{\mathcal{D}}(u|\psi_t), 0)}{\partial w_1}$$

and when $v = 1$,

$$\Delta_t(1) = \frac{\partial C(F^{\mathcal{D}}(u|\psi_t), 1)}{\partial w_1} - \frac{\partial C(F^{\mathcal{D}}(u|\psi_t), 1 - p_t)}{\partial w_1}.$$

We will now show that these two components sum up to one: $\Delta_t(0) + \Delta_t(1) = 1$. Indeed,

$$\begin{aligned} \Delta_t(0) + \Delta_t(1) &= \left[\frac{\partial C(w_1, 1)}{\partial w_1} - \frac{\partial C(w_1, 0)}{\partial w_1} \right]_{w_1 = F^{\mathcal{D}}(u|\psi_t)} \\ &= \left[\int_0^1 \frac{\partial^2 C(w_1, w_2)}{\partial w_1 \partial w_2} dw_2 \right]_{w_1 = F^{\mathcal{D}}(u|\psi_t)} \\ &= [c(w_1)]_{w_1 = F^{\mathcal{D}}(u|\psi_t)}, \end{aligned}$$

where $c(w_1)$ is the marginal density of the copula with respect to the first argument. Because $F^{\mathcal{D}}(u|\psi_t)$ is distributed uniformly on $[0, 1]$, we get $c(F^{\mathcal{D}}(u|\psi_t)) = 1$.

It trivially holds that $\Delta_t(v) = \Delta_t(1)^v \Delta_t(0)^{1-v}$, $\Delta_t(1) = \varrho_t(F^{\mathcal{D}}(u|\psi_t))$. From the result just proved it follows that $\Delta_t(0) = 1 - \varrho_t(F^{\mathcal{D}}(u|\psi_t))$. Combining all results yields the final conclusion.

■

The representation (4) for the joint density/mass function has the form of a product of the marginal density of $|r_t - c|$ and the Bernoulli density of $\mathbb{I}[r_t > c]$. The “success probability” parameter $\varrho_t(F^{\mathcal{D}}(u|\psi_t))$ does not, in general, equal to p_t (equality holds in the case of independence between $|r_t - c|$ and $\mathbb{I}[r_t > c]$), the “success probability” parameter of the marginal distribution; it depends not only on p_t , but also on $F^{\mathcal{D}}(u|\psi_t)$, inducing dependence between the marginals of $|r_t - c|$ and $\mathbb{I}[r_t > c]$. Interestingly, the form of representation (4) does not depend on the marginal distribution of $|r_t - c|$, although the joint density/mass function itself does.

Below we present two examples of copulas that will be used in the empirical section.

Frank copula. The Frank copula is

$$C(w_1, w_2) = -\frac{1}{\alpha} \log \left(1 + \frac{(e^{-\alpha w_1} - 1)(e^{-\alpha w_2} - 1)}{e^{-\alpha} - 1} \right),$$

where $\alpha \in [-\infty, +\infty]$ indexes dependence between entries. The joint density/mass function is given in (4) with

$$\varrho_t(z) = \frac{1}{1 - \frac{1 - e^{-\alpha(1-p_t)}}{1 - e^{\alpha p_t}} e^{\alpha(1-z)}}.$$

Note that $\alpha \rightarrow 0$ implies independence between the marginals and $\varrho_t \rightarrow p_t$.

Clayton copula. The Clayton copula is

$$C(w_1, w_2) = (w_1^\alpha + w_2^\alpha - 1)^{\frac{1}{\alpha}},$$

where $\alpha < 0$ indexes dependence between entries. The joint density/mass is as (4) with

$$\varrho_t(z) = 1 - \left(1 + \frac{(1 - p_t)^\alpha - 1}{z^\alpha}\right)^{\frac{1}{\alpha} - 1}.$$

Note that $\alpha \rightarrow 0$ implies independence between the marginals and $\varrho_t \rightarrow p_t$.

2.4 Conditional mean prediction

In many cases, the interest lies in the mean prediction of returns that can be expressed as

$$\begin{aligned} E_{t-1}[r_t] &= c + E_{t-1}[|r_t - c| \text{sign}(r_t - c)] \\ &= c + E_{t-1}[|r_t - c| (2\mathbb{I}[r_t > c] - 1)] \\ &= c + 2E_{t-1}[|r_t - c| \mathbb{I}[r_t > c]] - E_{t-1}[|r_t - c|]. \end{aligned}$$

Hence, the prediction of returns at time t is given by

$$\hat{r}_t = c + 2\hat{\xi}_t - \hat{\psi}_t, \tag{5}$$

where ψ_t is the conditional mean of $|r_t - c|$, ξ_t is the conditional expected cross-product of $|r_t - c|$ and $\mathbb{I}[r_t > c]$, and $\hat{\psi}_t$ and $\hat{\xi}_t$ are feasible analogs of ψ_t and ξ_t .

If $|r_t - c|$ and $\mathbb{I}(r_t > c)$ happen to be conditionally independent, then

$$\xi_t = E_{t-1}[|r_t - c|] E_{t-1}[\mathbb{I}(r_t > c)] = \psi_t p_t,$$

so

$$E_{t-1}[r_t] = c + (2p_t - 1) \psi_t,$$

and the returns can be predicted by

$$\hat{r}_t = c + (2\hat{p}_t - 1) \hat{\psi}_t,$$

where \widehat{p}_t denotes the predicted value of p_t .

In the general case of conditional dependence, we can compute the conditional distributions

$$f_{1[r_t > c] | |r_t - c|}(v|u) = \frac{f_{R_t}(u, v)}{f_{|r_t - c|}(v)} = \varrho_t(F^{\mathcal{D}}(u|\psi_t))^v (1 - \varrho_t(F^{\mathcal{D}}(u|\psi_t)))^{1-v}.$$

The conditional expectation function (of $\mathbb{I}[r_t > c]$ given $|r_t - c|$) is

$$E_{t-1}[\mathbb{I}[r_t > c] | |r_t - c|] = \varrho_t(F^{\mathcal{D}}(u|\psi_t)),$$

and the expectation of the cross-product is given by

$$\xi_t = E_{t-1}[|r_t - c|\mathbb{I}[r_t > c]] = \int_0^{+\infty} u f^{\mathcal{D}}(u|\psi_t) \varrho_t(F^{\mathcal{D}}(u|\psi_t)) du. \quad (6)$$

In general, the integral (6) cannot be computed analytically in most cases (even in the simple case when $f^{\mathcal{D}}(u|\psi_t)$ is exponential), but can be easily evaluated numerically, keeping in mind that the domain of integration is infinite. Note that the change of variables $z = F^{\mathcal{D}}(u|\psi_t)$ yields

$$\xi_t = \int_0^1 Q^{\mathcal{D}}(z) \varrho_t(z) dz, \quad (7)$$

where $Q^{\mathcal{D}}(z)$ is a quantile function of the distribution \mathcal{D} . Hence, the returns can be predicted by (5), where $\widehat{\xi}_t$ is obtained by numerically evaluating integral (7) with a fitted quantile function and fitted function $\varrho_t(z)$. In the empirical section, we apply the Gauss–Chebyshev quadrature formulas (Judd, 1998, section 7.2) to evaluate (7).

3 Empirical Analysis

3.1 Data

In our empirical study, we use Campbell and Yogo’s (2006) data set that covers the period 1927–2002 at annual, quarterly and monthly frequency.⁶ The excess stock returns are constructed from NYSE/AMEX value-weighted index data and T-bill rates from the Center for Research in Security Prices (CRSP) database. The risk-free return at monthly, quarterly and annual frequency is proxied by one-month rate, three-month rate and the return from rolling over three-month rate every quarter, respectively. The CRSP data are combined with S&P500 data and Moody’s Aaa corporate bond yield data to obtain the predictor variables dividend-price ratio, earnings-price ratio and yield spread. The dividend-price and earnings-price ratios are included in the regression models in logs.

⁶We would like to thank Moto Yogo for making the data available on his website.

The realized measures of second and higher-order moments of stock returns are constructed from daily data on NYSE/AMEX value-weighted index from CRSP. Let m be the number of daily observations per period (month, quarter, year) and $r_{t,j}$ denote the demeaned daily log stock return for day j in period t . Then, the realized variance RV_t (Andersen and Bollerslev, 1998; Andersen *et al.*, 2001), bipower variation BPV_t (Barndorff-Nielsen and Shephard, 2004, 2006), realized third moment RS_t and realized fourth moment RK_t for period t are computed as

$$\begin{aligned} RV_t &= \sum_{s=1}^m r_{t,s}^2 \\ BPV_t &= \frac{\pi}{2} \frac{m}{m-1} \sum_{s=1}^{m-1} |r_{t,s}| |r_{t,s+1}| \\ RS_t &= \sum_{s=1}^m r_{t,s}^3 \\ RK_t &= \sum_{s=1}^m r_{t,s}^4. \end{aligned}$$

Finally, following Campbell and Yogo (2006), we consider the subsample 1952–2002 for which the data, especially the interest rate variables after the Federal Reserve-Treasury Accord in 1951, are more reliable. This also roughly corresponds to the period that is most extensively studied in the empirical studies on predictability of stock returns.

3.2 Predictive regressions for excess returns

In this section, we present some empirical evidence on conditional mean predictability of excess stock returns. It is now well known that if the predictor variables are highly persistent, which is the case with all four predictors that we consider, the coefficients in the predictive regression is biased (Stambaugh, 1999) and their limiting distribution is non-standard (Elliott and Stock, 1994) if the innovations of the predictor variable are correlated with returns. For example, Campbell and Yogo (2006) report that this correlation is in the range $[-0.69, -0.99]$ for dividend-price and earnings-price ratios at different frequencies while the innovations of the three-month T-bill rate and the long-short interest rate spread are only weakly correlated with returns (correlation coefficient of -0.07 for monthly observations). A number of recent papers propose inference procedures that take these data characteristics into account when evaluating the predictive power of the different regressors (Campbell and Yogo, 2006; Cavanagh *et al.*, 1994; Jansson and Moreira, 2006; Lewellen, 2004; Torous, Valkanov and Yan, 2004; among others).

*** Table 1 ***

Table 1 reports some regression statistics when all the predictors are included in the regression. As argued above, the distribution theory for the t -statistics of the dividend-price and earnings-price ratio is non-standard whereas the t -statistics for the interest rates variables can be roughly compared to the standard normal critical values. The results in Table 1 suggest some predictability at monthly and quarterly frequency for the subsample 1952–2002. Even though the values of the R^2 coefficients are statistically small, Campbell and Thompson (2005) argue that they can still be economically meaningful when compared to the squared Sharpe ratio.

In summary, using the predictor variables in Campbell and Yogo’s (2006) dataset, we find some evidence of in-sample predictability of excess returns at monthly and quarterly frequency in the 1952–2002 period.

3.3 Decomposition-based model for excess returns

The sample log-likelihood function to be maximized is given by

$$\begin{aligned} & \sum_{t=1}^T \mathbb{I}[r_t > c] \ln \varrho_t (1 - \exp(-\xi_t)) + (1 - \mathbb{I}[r_t > c]) \ln (1 - \varrho_t (1 - \exp(-\xi_t))) \\ & + \sum_{t=1}^T \ln(\varsigma) - \ln |r_t - c| - \xi_t + \ln \xi_t, \end{aligned}$$

where $\xi_t = (|r_t - c| \Gamma(1 + \varsigma^{-1}) / \exp(\psi_t))^\varsigma$ and $\Gamma(\cdot)$ is the gamma function. This likelihood is based on Weibull-distributed absolute returns with shape parameter $\varsigma > 0$ (the exponential distribution can be obtained as special case when $\varsigma = 1$). The results from the return decomposition (1) are for the case $c = 0$. The estimates reported below for the logit and volatility specifications are obtained using the Frank copula.

*** Table 2 ***

Table 2 presents the estimation results from the dynamic logit model for the indicator variable. Several observations regarding the estimated logit model are in order. First, the persistence in the indicator variable over time is relatively weak once we control for other factors such as macroeconomic predictors and realized high-order moments of returns. The estimated signs of the interest rate variables are consistent across the different frequencies. The coefficient on the 3-month T-bill indicates that the stock prices are likely to fall if the interest rate goes up. Also, the combined effect

of the two realized volatility measures, RV and BPV , on the direction of the market is negative with the exception of the subsample 1952–2002. The realized measures of the higher moments of returns appear to have a statistically significant effect on the direction of excess returns at quarterly frequency but not at the other frequencies.

*** Table 3 ***

Table 3 reports the results from the volatility model (2). The adequacy of the Weibull specification is tested using the excess dispersion and Pearson’s goodness-of-fit tests. The excess dispersion test is adapted to the use of Weibull distribution

$$ED = \sqrt{T} \frac{(\hat{\eta}_t - 1)^2 - \hat{\sigma}_\eta^2}{\sqrt{((\hat{\eta}_t - 1)^2 - \hat{\sigma}_\eta^2)^2}},$$

where $\hat{\sigma}_\eta^2 = \Gamma(1 + 2\hat{\varsigma}^{-1}) / \Gamma(1 + \hat{\varsigma}^{-1})^2 - 1$ is an estimate of the variance of Weibull-distributed η_t , hats denote estimated values, and bars denote sample averages. The Pearson goodness-of-fit test (e.g., Kendall and Stuart, 1973, chapter 30) compares the multinomial distribution induced by standardized residuals and that implied by the normalized Weibull density. We set the number of equiprobable classes to 20, so the null distribution of the Pearson statistic is bounded between $\chi^2(18)$ and $\chi^2(19)$ because of the presence of an additional shape parameter (Kendall and Stuart, 1973, sect. 30.11–30.19), under the null of correct distributional specification.

The high persistence in absolute returns is evident from the results at quarterly and monthly frequency. The nonlinear term also suggests that positive returns correspond to low-volatility periods and negative returns tend to occur in high volatility periods where the difference in the average volatility of the two regimes is statistically significant at monthly frequency. For the 1927–2002 sample, higher interest rates and earnings-price ratio appear to reduce volatility while higher dividend-price ratio tends to have the opposite effect. Interestingly, the effect of these predictors is reversed for the most recent subsample 1952–2002.

Table 3 also shows the statistically significant departure of ς from 1 implying exponentiality of the density. On the other hand, further generalization of the density is not required because neither the excess dispersion nor Pearson tests reject the null of Weibull density.

*** Figures 1 and 2 ***

In order to visualize the outcome of our estimation procedure, we plot the predicted probabili-

ties form the dynamic logit model and the actual and predicted absolute returns from the volatility model in Figures 1 and 2. The predicted probabilities inherit the high persistence of volatility dynamics and are clearly inversely related to volatility movements: negative predicted returns tend to be associated with periods of high volatility and positive returns are predicted when volatility is low. The predicted absolute returns appear to follow closely the dynamics of stock return volatility.

Now we turn to modeling the joint distribution of sign and absolute returns using the Frank and Clayton copulas, where dependence is imposed through the copula parameter α . The result for the two copulas are reported in Table 4.

*** Table 4 ***

Table 4 shows that α is not significantly different from zero with numerical values close to zero that imply near-independence at all frequencies. The signs of parameter estimates are also consistent across the data frequencies.

The insignificance of the dependence parameter is compatible with the figures on the unconditional correlation of the raw variables $|r_t - c|$ and $\mathbb{I}(r_t > c)$, and the conditional correlation of the standardized variables $\psi_t^{-1}|r_t - c|$ and $p_t^{-1}\mathbb{I}[r_t > c]$, reported in Table 5.

*** Table 5 ***

We see from Table 5 that the conditional correlations are close to zero, while the unconditional ones are large. The bigger value for annual data corresponds to higher copula estimates of $|\alpha|$.

*** Table 6 ***

Finally, we computed the squared correlation coefficient between the actual and fitted excess returns from our model at different frequencies and report this pseudo- R^2 goodness-of-fit measure in Table 6. A rough comparison with the R^2 from the predictive regression in Table 1 indicates a similar in-sample performance of the two models. However, an inspection of the fitted returns of the two models reveals some interesting differences.

*** Figures 3 and 4 ***

Figures 3 and 4 plot the monthly fitted returns from our model and the predictive regression over the whole sample 1927–2002 and the subsample period 1952–2002. We see that our model is able to predict large volatility movements which is not the case for the predictive regression model.

Moreover, there are substantial differences in the predicted returns around the Great Depression period and towards the end of the sample.

3.4 Out-of-sample forecasting results

While there is some consensus in the finance literature on a certain degree of in-sample predictability of excess returns (Cochrane, 2005), the evidence on out-of-sample predictability is mixed. Goyal and Welch (2003, 2006) find that the commonly used predictive regressions would not help an investor to profitably time the market. Campbell and Thompson (2005), however, show that the out-of-sample predictive performance of the models is improved after imposing restrictions on the sign of the estimated coefficients and the equity premium forecast.

In our out-of-sample experiments, we compare the one-step ahead forecasting performance of the decomposition-based model proposed in this paper, conditional predictive regression and unconditional mean model. The forecast are computed from monthly observations for the subsample 1952–2002 using a rolling sample scheme with fixed sample size $R = 360$. The results are reported using an out-of-sample coefficient of predictive performance OS computed as

$$OS = 1 - \frac{\sum_{j=T-R+1}^T \partial(r_j - \hat{r}_j)}{\sum_{j=T-R+1}^T \partial(r_j - \bar{r}_j)},$$

where $\partial(u) = u^2$ if it is based on squared errors and $\partial(u) = |u|$ if it is based on absolute errors, \hat{r}_j is the one-step forecast of r_j from the conditional predictive model and \bar{r}_j denotes the unconditional mean of r_j computed from the last R observations in the rolling scheme. If the value of OS is equal to zero, the conditional models and the unconditional mean predict equally well the next period excess return; if $OS < 0$, the unconditional mean performs better; and if $OS > 0$, the conditional model dominates.

*** Table 7 ***

Table 7 presents the results from the out-of-sample forecast evaluation. We refer in Table 7 to the forecasting model that does not impose any restrictions on the forecasts as unrestricted forecast. As in Campbell and Thompson (2005), we consider a forecasting procedure that rules out negative values for the equity premium by setting the forecast to zero if the next period predicted excess return is negative (restricted forecast in Table 7).

The upper left corner of Table 7 presents the results from the predictive regression of excess

returns. As in Goyal and Welch (2003, 2006) and Campbell and Thompson (2005), we find that the unconditional model based on the historical mean performs better out-of-sample than the model that uses conditional information and, in some cases, the difference in the relative forecasting performance is up to 5%. When the non-negativity restriction on the equity premium forecast is imposed, the unconditional forecast no longer dominates the conditional prediction. In summary, using the predictor variables in Campbell and Yogo's (2006) dataset, we find some evidence of improved out-of-sample predictability after imposing a non-negativity constraint on the equity premium forecast.

The results from our model described in Section 2 are reported in the second panel of Table 7. We present separately the results under conditional independence and conditional dependence. In all cases, our model dominates the unconditional mean forecast with forecast gains of 1.1-2.7%. Although these forecast gains do not seem statistically large, Campbell and Thompson (2005) argue that a 1% increase in the out-of-sample statistic OS implies economically large increases in portfolio returns. This forecasting superiority over the unconditional mean forecast is even further reinforced by the fact that our model is overly parameterized compared to the benchmark model. Restricting the equity premium forecast to positive values and imposing conditional independence do not seem to have a large effect on the forecasts.

To gain some intuition about the source of the forecasting improvements, we consider two nested versions of our model: one that contains only the own dynamics of the indicator variable and the absolute returns (pure dynamic model) and a model that includes only macroeconomic predictors and realized measures without any autoregressive structure (pure structural model). Interestingly, the forecasting gains of the full model appear to have been generated by the information contained in the predictors and not in the dynamic behavior of the sign and volatility components.⁷ In fact, the forecasts from the pure structural model dominate the forecasts from the full model with up to 3% improvements over the unconditional mean forecasts.

⁷The poor forecasting performance of the pure dynamic model appears to be due to poor probability predictability that arises from the weak persistence in the indicator variable mentioned above.

4 Conclusion

This paper proposes a new method for analyzing the dynamics of excess returns by modeling the joint distribution of their sign and volatility multiplicative components using copulas. Our framework attempts to capitalize on the stronger degree of directional and volatility predictability and judiciously exploit possible nonlinearities and different information sets in dynamics of the two parts. Furthermore, the paper develops copula modeling with one discrete and one continuous marginal, which is new to the literature, and discusses computation of the conditional mean predictor under conditional dependence of the two components

Our empirical analysis of US excess stock returns at monthly, quarterly and annual frequencies delivers some interesting findings. While the in-sample fits of our model and the standard predictive regression are of similar magnitude, there are some substantial differences in fitted returns from these methods during the Great Depression period and late 1990s. Furthermore, the predicted probabilities and absolute returns from the fitted marginals of the two multiplicative components of returns can be used for directional and volatility forecasting. Our estimation results reveal that these two components exhibit substantial unconditional correlation but an almost zero conditional correlation that is reflected in a conditional near-independence in the copula specification. We also report some economically significant forecasting gains from our procedure out of sample that appear to arise from efficiently incorporating the information contained in the macroeconomic predictors and realized volatility and higher-order moments of returns.

5 Acknowledgments

The second author would like to thank FQRSC, IFM2 and SSHRC for financial support.

References

- [1] Anatolyev, S., and Gerko, A. (2005), “A trading approach to testing for predictability”, *Journal of Business and Economic Statistics*, 23, 455–461.
- [2] Andersen, T.G., and Bollerslev, T. (1998), “Answering the skeptics: Yes, standard volatility models do provide accurate forecasts,” *International Economic Review*, 39, 885-905.
- [3] Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (2001), “The distribution of realized exchange rate volatility,” *Journal of the American Statistical Association*, 96, 42-55.
- [4] Baker, M., and Wurgler, J. (2000), “The equity share in new issues and aggregate stock returns,” *Journal of Finance*, 55, 2219-2257.
- [5] Barndorff-Nielsen, O.E., and Shephard, N. (2004), “Power and bipower variation with stochastic volatility and jumps,” (with discussion) *Journal of Financial Econometrics*, 2, 1-48
- [6] Barndorff-Nielsen, O.E., and Shephard, N. (2006), “Econometrics of testing for jumps in financial economics using bipower variation,” *Journal of Financial Econometrics*, 4, 1-30.
- [7] Bauwens, L., and Giot, P. (2000), “The logarithmic ACD model: an application to the bid/ask quote process of three NYSE stocks,” *Annales d'Economie et de Statistique*, 60, 117–149.
- [8] Cameron, A.C, Li, T., Trivedi, P.K., and Zimmer, D.M. (2004), “Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts,” *Econometrics Journal*, 7, 566–584.
- [9] Campbell, J.Y. (1987), “Stock returns and the term structure,” *Journal of Financial Economics*, 18, 373-399.
- [10] Campbell, J.Y., and Shiller, R.J. (1988a), “Stock prices, earnings and expected dividends,” *Journal of Finance*, 43, 661-676.
- [11] Campbell, J.Y., and Shiller, R.J. (1988b), “The dividend-price ratio and expectations of future dividends and discount factors,” *Review of Financial Studies*, 1, 195-228.
- [12] Campbell, J.Y., and Thompson, S.B. (2005), “Predicting the equity premium out of sample: Can anything beat the historical average?” NBER Working Paper No. 11468.

- [13] Campbell, J.Y., and Yogo, M. (2006), “Efficient tests of stock return predictability,” *Journal of Financial Economics*, 81, 27-60.
- [14] Cavanagh, C.L., Elliott, G., and Stock, J.H. (1995), “Inference in models with nearly nonstationary regressors”, *Econometric Theory*, 11, 1131-1147.
- [15] 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.
- [16] Christoffersen, P.F., and Diebold, F.X. (2006), “Financial asset returns, direction-of-change forecasting, and volatility dynamics,” *Management Science*, 52, 1273–1288.
- [17] Christoffersen, P.F., Diebold, F.X., Mariano, R.S., Tay, A.S., and Tse, Y.K. (2006), “Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: International evidence,” Working Paper No. 06-016, Penn Institute for Economic Research.
- [18] Chung, J., and Hong, Y. (2006), “Model-free evaluation of directional predictability in foreign exchange markets,” Manuscript, Cornell University.
- [19] Cochrane, J.H., 2005, *Asset Pricing*, second edition, Princeton University Press.
- [20] de Jong, R.M., and Woutersen, T. (2005), “Dynamic time series binary choice,” Manuscript, Ohio State University.
- [21] Elliot, G., and Stock, J.H. (1994), “Inference in time series regression when the order of integration of a regressor is unknown,” *Econometric Theory*, 10, 672-700.
- [22] Engle, R.F. (2002), “New frontiers for ARCH models,” *Journal of Applied Econometrics*, 17, 425–446.
- [23] Engle, R.F., and Gallo, G.M. (2006), “A multiple indicators model for volatility using intradaily data,” *Journal of Econometrics*, 131, 3–27.
- [24] Engle, R.F., and Russell, J.R. (1998), “Autoregressive conditional duration: a new model for irregularly spaced transaction data,” *Econometrica* 66, 1127–1162.
- [25] Fama, E., and French, K. (1988), “Dividend yields and expected stock returns,” *Journal of Financial Economics*, 22, 3-26.

- [26] Gençay, R. (1998), “Optimization of Technical Trading Strategies and the Profitability in Security Markets,” *Economics Letters*, 59, 249–254.
- [27] Geweke, J. (1986), “Modeling the persistence of conditional variances: A comment,” *Econometric Reviews*, 5, 57–61.
- [28] Goyal, A., and Welch, I. (2003), “The myth of predictability: Does the dividend yield forecast the equity premium?” *Management Science*, 49, 639-654.
- [29] Goyal, A., and Welch, I. (2006), “A comprehensive look at the empirical performance of equity premium prediction,” *Review of Financial Studies*, forthcoming.
- [30] Granger, C.W.J. (1998), “Comonotonicity,” Lecture to the Econometrics and Financial Time Series Workshop, Cambridge University.
- [31] Hodrick, R.J. (1992), “Dividend yields and expected stock returns: Alternative procedures for inference and measurement,” *Review of Financial Studies*, 5, 357-386.
- [32] Hong, Y., and Chung, J. (2003), “Are the directions of stock price changes predictable? Statistical theory and evidence,” Manuscript, Cornell University.
- [33] Jansson, M., and Moreira, M.J. (2006), “Optimal inference in regression models with nearly integrated regressors,” *Econometrica*, 74, 681-715.
- [34] Judd, K. (1998), *Numerical Methods in Economics*, MIT Press.
- [35] Kendall, M.G., and A. Stuart (1973): *The Advanced Theory of Statistics*. London: Griffin.
- [36] Lettau, M., and Ludvigson, S. (2001), “Consumption, aggregate wealth and expected stock returns,” *Journal of Finance*, 56, 815-849.
- [37] Lewellen, J. (2004), “Predicting returns with financial ratios,” *Journal of Financial Economics*, 74, 209-235.
- [38] Linton, O., and Whang, Y.J. (2005), “A quantilogram approach to evaluating directional predictability,” Manuscript, London School of Economics.
- [39] Nelson, R. B. (1999), *An Introduction to Copulas*, New York, NJ: Springer.

- [40] Patton, A.J. (2006), “Modelling asymmetric exchange rate dependence”, *International Economic Review*, 47, 527–556.
- [41] Polk, C., Thompson, S.B., and Vuolteenaho, T. (2006), “New forecasts of the equity premium,” *Journal of Financial Economics*, 81, 101-141.
- [42] Qi, M. (1999), Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics*, 17, 419–429.
- [43] Rydberg, T.H., and Shephard, N. (2003), “Dynamics of trade-by-trade price movements: Decomposition and models,” *Journal of Financial Econometrics*, 1, 2-25.
- [44] Stambaugh, R.F. (1999), “Predictive regressions,” *Journal of Financial Economics*, 54, 375-421.
- [45] Torous, W., Valkanov, R., and Yan, S. (2004), “On predicting stock returns with nearly integrated explanatory variables,” *Journal of Business*, 77, 937-966.

Table 1. Estimation results from predictive regression of excess returns.

	$t(dp)$	$t(ep)$	$t(ir3)$	$t(irs)$	R^2
	Whole sample 1927–2002				
Annual	0.14	0.76	−0.85	0.53	0.1036
Quarterly	−0.92	1.71	−1.20	−0.05	0.0346
Monthly	−1.02	1.64	−1.24	0.34	0.0124
	Subsample 1952–2002				
Annual	1.44	−0.73	−1.39	1.46	0.2108
Quarterly	2.01	−1.22	−1.66	2.33	0.0819
Monthly	2.19	−1.44	−1.71	2.91	0.0413

Notes: $t(z)$ denotes the t -statistic for the coefficient on variable z , and dp , ep , $ir3$ and irs stand for dividend-price ratio, earnings-price ratio, three-month T-bill rate and long-short yield spread. The t -statistics are computed using Newey–West HAC standard errors with automatic bandwidth.

Table 2. Estimation results from the dynamic logit model for return indicators

	ω_s	ϕ_s	$\delta_s(ir3)$	$\delta_s(irs)$	$\delta_s(RV)$	$\delta_s(BPV)$	$\delta_s(RS)$	$\delta_s(RK)$
Annual data 1927–2002								
coeff	1.37	−0.68	−8.34	23.8	2.21	−2.77	−0.02	0.01
s.e.	1.19	0.78	8.46	22.3	1.69	2.24	0.09	0.01
t-stat	1.15	−0.87	−0.99	1.07	1.31	−1.23	−0.22	0.59
Quarterly data 1927–2002								
coeff	0.74	0.12	−5.74	8.19	−1.45	0.95	−0.29	0.02
s.e.	0.40	0.26	4.22	10.9	1.42	1.43	0.13	0.01
t-stat	1.82	0.47	−1.36	0.75	−1.02	0.66	−2.17	2.08
Monthly data 1927–2002								
coeff	0.43	0.19	−5.56	6.18	−0.69	0.31	0.00	0.00
s.e.	0.21	0.14	2.35	5.74	0.70	0.66	0.04	0.00
t-stat	2.03	1.33	−2.37	1.08	−0.97	0.47	0.00	0.21
Monthly data 1952–2002								
coeff	0.30	0.18	−8.09	9.92	4.70	−3.49	−0.30	−0.09
s.e.	0.27	0.17	3.28	6.41	2.50	2.42	0.41	0.07
t-stat	1.11	1.04	−2.46	1.55	1.88	−1.44	−0.73	−1.25

Notes: The model is estimated on the basis of the logit equation $E_{t-1} [\mathbb{I}[r_t > c]] = \exp(\theta_t) / (1 + \exp(\theta_t))$, where θ_t is determined by (3).

Table 3. Estimation results from the volatility model.

	ω_r	β_r	γ_r	ρ_r	$\delta_r(dp)$	$\delta_r(ep)$	$\delta_r(ir3)$	$\delta_r(irs)$	ς	ED	PT
	Annual data 1927–2002										
coeff	4.97	−0.21	0.13	−0.44	−12.31	57.39	−5.50	−9.42	1.66	−0.39	11.89
s.e.	1.30	0.19	0.07	0.14	64.76	59.12	2.37	6.33	0.17		
t-stat	3.81	−1.12	1.97	−3.09	−0.19	0.97	−2.32	−1.49	9.85		
	Quarterly data 1927–2002										
coeff	0.09	0.89	0.09	−0.14	9.66	−14.51	−0.05	−0.38	1.26	0.09	13.63
s.e.	0.21	0.09	0.03	0.08	6.35	6.82	0.38	0.95	0.06		
t-stat	0.42	9.38	2.67	−1.82	1.52	−2.13	−0.14	−0.41	21.3		
	Monthly data 1927–2002										
coeff	0.24	0.89	0.06	−0.13	3.35	−1.88	−0.19	−0.28	1.21	−0.01	29.45
s.e.	0.07	0.03	0.01	0.03	2.24	2.39	0.13	0.37	0.04		
t-stat	3.39	36.2	5.96	−4.34	1.49	−0.78	−1.48	−0.77	31.7		
	Monthly data 1952–2002										
coeff	−0.52	0.81	0.03	−0.17	−8.62	7.14	0.36	−0.69	1.27	−0.08	20.61
s.e.	0.26	0.08	0.01	0.06	7.97	7.90	0.35	0.70	0.05		
t-stat	−2.03	10.4	2.65	−2.85	−1.08	0.90	1.01	−0.98	23.6		

Notes: The model is estimated on the basis of the MEM equation $|r_t - c| = \exp(\psi_t)\eta_t$, where ψ_t follows (2), and η_t is distributed as normalized Weibull with shape parameter ς . ED and PT denote the excess dispersion and Pearson tests. The Pearson test compares the discretized empirical and Weibull distribution using 20 cells. Under the correct distributional specification, the distribution of Pearson statistic is bounded between $\chi^2(18)$ and $\chi^2(19)$ whose 95% critical values are 28.87 and 30.14, and the excess dispersion statistic is distributed as standard normal with a (right-sided) 5% critical value of 1.645.

Table 4. Estimates of the dependence parameter.

Frank copula			Clayton copula		
coeff	s.e.	t-stat	coeff	s.e.	t-stat
Annual data 1927–2002					
0.976	1.208	0.809	−0.144	0.216	−0.668
Quarterly data 1927–2002					
0.213	0.479	0.445	−0.112	0.087	−1.281
Monthly data 1927–2002					
0.070	0.245	0.285	−0.030	0.043	−0.693
Monthly data 1952–2002					
0.285	0.296	0.964	−0.089	0.055	−1.625

Notes: Shown are estimates of dependence parameter α in copula specifications.

Table 5. Unconditional and conditional correlations.

	Unconditional	Conditional
Annual data 1927–2002	0.801	−0.133
Quarterly data 1927–2002	0.699	−0.029
Monthly data 1927–2002	0.704	−0.043
Monthly data 1952–2002	0.768	−0.011

Notes: “Unconditional correlation” refers to the sample correlation coefficients between $|r_t - c|$ and $1[r_t > c]$. “Conditional correlation” refers to the sample correlation coefficients between $\psi_t^{-1}|r_t - c|$ and $p_t^{-1}1[r_t > c]$ estimated from the decomposition-based model with the Frank copula.

Table 6. In-sample goodness-of-fit measures.

	pseudo- R^2
Annual data 1927–2002	0.1663
Quarterly data 1927–2002	0.0151
Monthly data 1927–2002	0.0034
Monthly data 1952–2002	0.0418

Notes: pseudo- R^2 denotes squared correlation coefficients between excess returns and their in-sample predictions.

Table 7. Out-of sample *OS* statistic (in %) for the monthly subsample 1952–2002.

	No decomposition	Ignoring dependence		Using dependence	
Copula		Frank	Clayton	Frank	Clayton
	Linear model				
Unrestricted forecast	$\begin{pmatrix} -4.91 \\ -4.18 \\ 1.37 \\ 0.62 \end{pmatrix}$				
Restricted forecast					
	Full parametric model				
Unrestricted forecast		$\begin{pmatrix} 1.49 \\ 2.71 \\ 1.08 \\ 2.13 \end{pmatrix}$	$\begin{pmatrix} 1.26 \\ 2.48 \\ 0.86 \\ 1.90 \end{pmatrix}$	$\begin{pmatrix} 1.42 \\ 1.64 \\ 1.76 \\ 1.27 \end{pmatrix}$	$\begin{pmatrix} 1.26 \\ 1.46 \\ 1.57 \\ 1.08 \end{pmatrix}$
Restricted forecast					
	Pure dynamic model				
Unrestricted forecast		$\begin{pmatrix} 0.03 \\ -0.28 \\ 0.35 \\ -0.09 \end{pmatrix}$	$\begin{pmatrix} 0.03 \\ -0.27 \\ 0.34 \\ -0.09 \end{pmatrix}$	$\begin{pmatrix} -1.55 \\ -1.95 \\ -0.16 \\ -0.92 \end{pmatrix}$	$\begin{pmatrix} -1.42 \\ -1.88 \\ -0.07 \\ -0.87 \end{pmatrix}$
Restricted forecast					
	Pure structural model				
Unrestricted forecast		$\begin{pmatrix} 2.34 \\ 3.01 \\ 1.51 \\ 2.40 \end{pmatrix}$	$\begin{pmatrix} 2.25 \\ 2.96 \\ 1.41 \\ 2.35 \end{pmatrix}$	$\begin{pmatrix} 1.68 \\ 1.77 \\ 1.76 \\ 1.37 \end{pmatrix}$	$\begin{pmatrix} 1.85 \\ 1.94 \\ 1.76 \\ 1.46 \end{pmatrix}$
Restricted forecast					

Notes: The rolling scheme uses a sample of fixed size $R = 360$. Restricted forecast sets the forecast of equity premium to zero if it is negative. The top figure in a cell is *OS* based of squared prediction errors, the lower figure shows *OS* based on absolute values of prediction errors. “Linear model” and “No decomposition” refer to the linear predictive regression of excess returns as in Table 1. “Full parametric model” refers to the decomposition-based model that includes all predictors corresponding to Tables 2 and 3. “Pure dynamic model” refers to a smaller decomposition-based model where only a constant, ψ_{t-1} , $\ln|r_{t-1} - c|$ and $\mathbb{I}(r_{t-j} > c)$ are included in equation (2) for absolute returns and only a constant and $\mathbb{I}(r_{t-1} > c)$ are included in equation (3) for indicators. “Pure structural model” refers to a decomposition-based model where instead only a constant, dp , ep , $ir3$ and irs are included in equation (2) for absolute returns and only a constant, $ir3$, irs , RV , BPV , RS and RK are included in equation (3) for indicators. “Ignoring dependence” means that the decomposition-based model is estimated but predictions are constructed under the presumption of conditional independence between signs and absolute returns. “Using dependence” means that the decomposition-based model is estimated and fully used in constructing predictions by (5), including numerical integration.

Figure 1. Predicted probabilities from the dynamic logit model, monthly data 1927–2002.

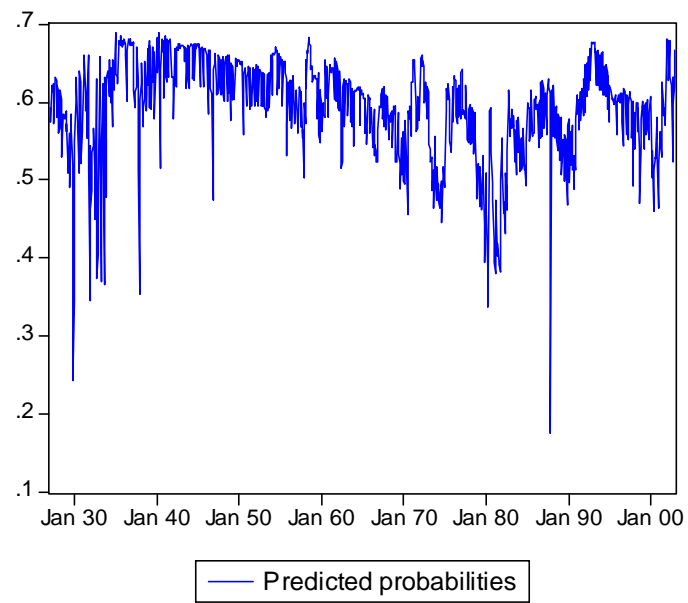


Figure 2. Actual and predicted absolute returns from the volatility model, monthly data 1927–2002.

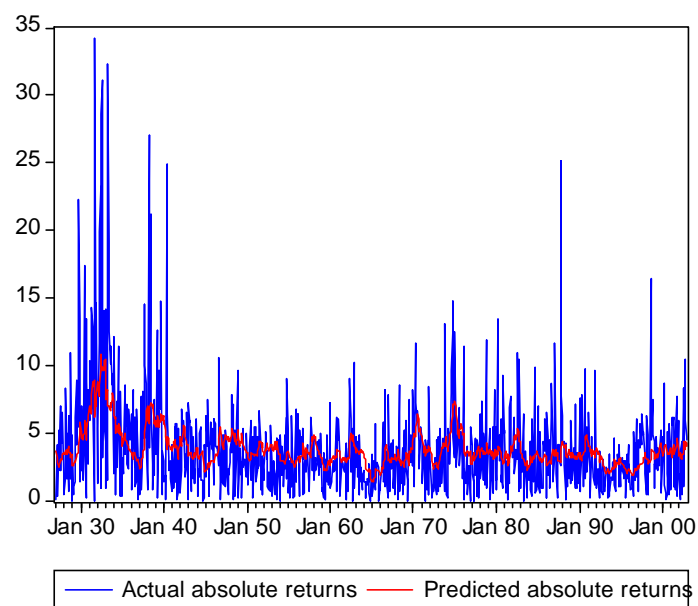


Figure 3. Predicted returns from the predictive regression and from decomposition-based model, monthly data 1927–2002.

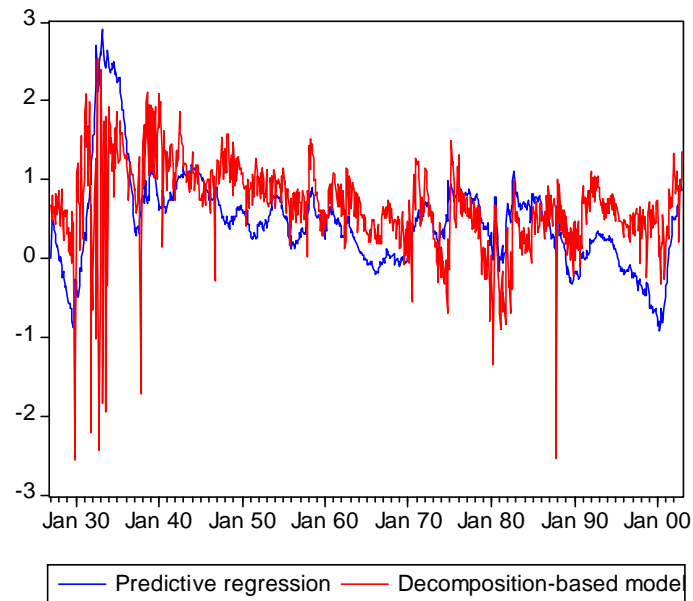


Figure 4. Predicted returns from the predictive regression and from decomposition-based model, monthly data 1952–2002.

