Modeling Financial Return Dynamics via Decomposition

Stanislav Anatolyev
Nikolay Gospodinov

Modeling Financial Return Dynamics via Decomposition

Stanislav Anatolyev[*] Nikolay Gospodinov[†]
New Economic School Concordia University

First draft: January 2007
This version: September 2007

Abstract

While the predictability of excess stock returns is detected by traditional predictive regressions as statistically small, the direction-of-change and volatility of returns exhibit a substantially larger degree of dependence over time. We capitalize on this observation and decompose the returns into a product of sign and absolute value components whose joint distribution is obtained by combining a multiplicative error model for absolute values, a dynamic binary choice model for signs, and a copula for their interaction. Our decomposition model is able to incorporate important nonlinearities in excess return dynamics that cannot be captured in the standard predictive regression setup. The empirical analysis of US stock return data shows statistically and economically significant forecasting gains of the decomposition model over the conventional predictive regression.

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

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, 1988) and yields on short- and long-term Treasury and corporate bonds (Campbell, 1987) 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, 2007). 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 Ludvingson, 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: instead of trying to identify better predictors, we look for better ways of using these predictors. We accomplish this by modeling individual multiplicative components of excess stock returns and combining the components' information to recover the conditional expectation of the original variable of interest.

To fix ideas, 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 can be factored as

$$r_t = |r_t| \operatorname{sign}(r_t),$$

which is called "an intriguing decomposition" in Christoffersen and Diebold (2006). The conditional mean of r_t is then given by

$$E_{t-1}(r_t) = E_{t-1}(|r_t| \operatorname{sign}(r_t)),$$

where $E_{t-1}(.)$ denotes the expectation taken with respect to the available information up to time $t-1$. Our aim is to model the joint distribution of absolute values $|r_t|$ and signs $\operatorname{sign}(r_t)$ in order to pin down the conditional expectation $E_{t-1}(r_t)$. The approach we adopt to achieve this involves joint usage of a multiplicative error model for absolute values, a dynamic binary choice model for signs, and a copula for their interaction. We expect this detour to be successful for the following reasons.

First, the joint modeling of the multiplicative components is able to incorporate important hidden nonlinearities in excess return dynamics that cannot be captured in the standard predictive

1

regression setup. In fact, we argue that a conventional predictive regression lacks predictive power when the data are generated by our *decomposition model*. Second, the absolute values and signs exhibit a substantial degree of dependence over time while the predictability of returns seems to be statistically small as detected by conventional tools. Indeed, volatility (as measured by absolute values of returns) persistence and predictability has been extensively studied and documented in the literature (e.g., Andersen *et al.*, 2006). As far as signs are concerned, Christoffersen and Diebold (2006), Hong and Chung (2003) and Linton and Whang (2007) find convincing evidence of sign predictability of US stock returns for different data frequencies. Christoffersen and Diebold (2006) reconcile the standard finding of weak conditional mean predictability with possibly strong sign and volatility dependence.

Note that the joint predictive distribution of absolute values and signs provides a more general inference procedure than modeling directly the conditional expectation of returns as in the predictive regression literature. Studying the dependence between the sign and absolute value components over time is interesting in its own right and can be used for various other purposes. For example, the joint modeling would allow the researcher to explore trading strategies and evaluate their profitability (Satchell and Timmermann, 1996; Qi, 1999; Anatolyev and Gerko, 2005). In our empirical analysis of US stock return data we perform a similar portfolio allocation exercise, where an investment strategy requires information only about the predicted direction of returns. Another interesting aspect of the bivariate analysis is an important conclusion that in spite of a large unconditional correlation between the multiplicative components, they appear to be conditionally very weakly dependent.

The rest of the paper is organized as follows. Section 2 introduces our return decomposition, discusses the marginal density specifications and construction of the joint predictive density of sign and absolute value components, and demonstrates how mean predictions can be generated. Section 3 contains the empirical analysis of predictability of US excess returns using Campbell and Yogo's (2006) data set. The first two subsections describe the data and report the main findings from the commonly used linear predictive regression. Sections 3.3 and 3.4 present the results from the joint modeling and provides some in-sample and out-of-sample statistical comparisons with the benchmark predictive regression. Section 3.5 evaluates the performance of different models in the context of a portfolio allocation exercise, and Section 3.6 reports some simulation evidence about

the inability of the linear regression to detect predictability when the data are generated by the decomposition model. Section 4 concludes.

2 Methodological Framework

2.1 Decomposition and its motivation

The key identity that lies in the heart of our technique is the return decomposition

$$r_t = c + |r_t - c| \operatorname{sign}(r_t - c) = c + |r_t - c| \left(2\mathbb{I}\left[r_t > c\right] - 1\right), \tag{1}$$

where $\mathbb{I}\left[.\right]$ is the indicator function and c is an arbitrary constant. Our *decomposition model* will be based on the joint dynamic modeling of the two ingredients entering (1), the *absolute values* $|r_t - c|$ and *indicators* $\mathbb{I}\left[r_t > c\right]$ (or, equivalently, *signs* $\operatorname{sign}(r_t - c)$ related linearly to indicators). In case the interest lies in the mean prediction of returns, one can infer from (1) that

$$E_{t-1}\left(r_t\right) = c - E_{t-1}\left(|r_t - c|\right) + 2E_{t-1}\left(|r_t - c|\mathbb{I}\left[r_t > c\right]\right),$$

and the decomposition model can be used to generate optimal predictions of returns because it allows to deduce, among other things, the conditional mean of $|r_t - c|$ and conditional expected cross-product of $|r_t - c|$ and $\mathbb{I}\left[r_t > c\right]$ (for details, see subsection 2.4). In a different context, Rydberg and Shephard (2003) use a decomposition similar to (1) 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).

Recall that c is an arbitrary constant. Although our empirical analysis only considers the leading case $c = 0$, we develop the theory for arbitrary c for greater generality. The choice of c is dictated primarily by the application at hand. In the context of financial returns, Christoffersen and Diebold (2006) analyze the case when $c = 0$ while Hong and Chung (2003) and Linton and Whang (2007) 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 macroeconomic applications, in particular modeling GDP growth rates, c may be set to 0 if one is interested in recession/expansion analysis, or to 3%, for instance, if one is interested in modeling and forecasting a potential output gap. Likewise, it seems natural to set c to 2% if one considers modeling and forecasting inflation.

To provide further intuition and demonstrate the advantages of the decomposition model, consider an example in which we try to predict excess returns r_t with the lagged realized volatility RV_{t-1}. A linear predictive regression of r_t on RV_{t-1}, estimated on data from our empirical section, gives an in-sample $R^2 = 0.39\%$. Now suppose that we employ a simple version of the decomposition model where the same predictor is used linearly for absolute values, i.e. $E_{t-1}(|r_t|) = \alpha_{|r|} + \beta_{|r|} RV_{t-1}$, and for indicators in a linear probability model $\Pr_{t-1}(r_t > 0) = \alpha_{\mathbb{I}} + \beta_{\mathbb{I}} RV_{t-1}$. Assume for simplicity that the shocks in the two components are stochastically independent. Then, it is easy to see from identity (1) that $E_{t-1}(r_t) = \alpha_r + \beta_r RV_{t-1} + \gamma_r RV_{t-1}^2$ for certain constants α_r, β_r and γ_r. Running a linear predictive regression on both RV_{t-1} and RV_{t-1}^2 yields a much better fit with $R^2 = 0.72\%$. Even a linear predictive regression on RV_{t-1}^2 alone gives $R^2 = 0.69\%$, which indicates that RV_{t-1}^2 is a much better predictor than RV_{t-1}. This clearly suggests that the conventional predictive regression may miss important nonlinearities that are easily captured by the decomposition model.

Alternatively, suppose that the true model for indicators is trivial, i.e. $\Pr_{t-1}(r_t > 0) = \alpha_{\mathbb{I}} \neq \frac{1}{2}$, and the components are conditionally independent. Then, using again identity (1), it is straightforward to see that any parameterization of expected absolute values $E_{t-1}(|r_t|)$ leads to the same form of parameterization of the predictive regression $E_{t-1}(r_t)$. Augmenting the parameterization for indicators and accounting for the dependence between the multiplicative components then automatically delivers an improvement in the prediction of r_t by capturing hidden nonlinearities in its dynamics.

While the model setup used in the above example is fairly simplified (indeed, the regressor RV_t^2 is quite easy to find), the arguments that favor the decomposition model naturally extend to more complex settings. In particular, when the component models are quite involved and the components themselves are conditionally dependent, we find some simulation evidence that the standard linear regression framework has difficulties detecting any perceivable predictability as judged by the conventional criteria (see subsection 3.6). The driving force behind the predictive ability of the decomposition model is the predictability in the two components, documented in previous studies. Note also that, unlike the example above, the models for absolute values and indicators may in fact use different information variables.

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]

$$|r_t - c| = \psi_t \eta_t,$$

where $\psi_t \equiv E_{t-1}(|r_t - c|)$ and η_t is a positive multiplicative error with $E_{t-1}(\eta_t) = 1$ and conditional distribution $\mathcal{D}$. The conditional expectation ψ_t and conditional distribution $\mathcal{D}$ can be parameterized following the suggestions in the MEM and ACD literatures (Engle and Russell, 1998; Engle, 2002). A convenient dynamic specification for ψ_t is the logarithmic autoregressive conditional duration (LACD) model of Bauwens and Giot (2000) whose main advantage, especially when (weakly) exogenous predictors are present, 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 the 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. Let us denote the vector of shape parameters of $\mathcal{D}$ by ς.

The conditional expectation ψ_t is parameterized as

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

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 term $x'_{t-1} \delta_r$ accounts for the possibility that macroeconomic predictors such as valuation ratios and interest rates variables may have an effect on volatility dynamics proxied by $|r_t - c|$. In what follows, we refer to model (2) as the *volatility model*.

---

[1]The leading application of the MEM approach in the econometrics literature is that to durations between successive transactions in a high frequency financial market (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.

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

Now we turn our attention to the dynamic specification of the indicator $\mathbb{I}\left[r_t > c\right]$. The conditional distribution of $\mathbb{I}\left[r_t > c\right]$, given past information, is necessarily Bernoulli $\mathcal{B}\left(p_t\right)$ with probability mass function $f_{\mathbb{I}[r_t>c]}\left(v\right) = p_t^v\left(1 - p_t\right)^{1-v}$, $v \in \{0, 1\}$, where p_t denotes the conditional "success probability" $\Pr_{t-1}(r_t > c) = E_{t-1}\left(\mathbb{I}\left[r_t > c\right]\right)$.

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

$$\Pr{}_{t-1}(r_t > c) = 1 - F_\varepsilon\left(\frac{c - \mu_t}{\sigma_t}\right).$$

This expression suggests that time-varying volatility can generate sign predictability as long as $c - \mu_t \neq 0$. Furthermore, Christoffersen *et al.* (2006) derive a Gram–Charlier expansion of $F_\varepsilon(.)$ and show that $\Pr_{t-1}(r_t > c)$ 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 p_t using the dynamic logit model

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

with

$$\theta_t = \omega_s + \phi_s\mathbb{I}\left[r_{t-1} > c\right] + y'_{t-1}\delta_s, \tag{3}$$

where the set of predictors y_{t-1} includes macroeconomic variables (valuation ratios and interest rates) as well as realized measures such as realized variance (RV), bipower variation (BPV), realized third (RS) and fourth (RK) moments of returns as suggested above.[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 the presence of jumps (Barndorff-Nielsen and Shephard, 2004). In what follows, we refer to model (2) as the *direction model*.[4]

Of course, in other applications of the decomposition method, different specifications for ψ_t, $\mathcal{D}$ and p_t are possibly necessary, depending on the empirical context.

---

[3] We 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.

[4] de 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 $R_t \equiv (|r_t - c|, \mathbb{I}[r_t > c])'$ whose domain is $\mathbb{R}_+ \times \{0, 1\}$. Up to now we have dealt with the marginals[5] of the two components

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

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

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

If the two marginals were normal, a reasonable thing to do would be to postulate bivariate normality. If the two were exponential, a reasonable parameterization would be joint exponentiality. However, even though the literature documents a number of bivariate distributions with marginals from different families (e.g., Marshall and Olkin, 1985), it does not suggest a bivariate distribution whose marginals are Bernoulli and, say, exponential. Therefore, we use the copula theory to generate the joint distribution from the specified marginals. For introduction to copulas, see Nelson (1999) and Trivedi and Zimmer (2005), among others. Let $F_{R_t}(u, v)$ and $f_{R_t}(u, v)$ denote the joint CDF/CMF and joint density/mass of R_t, respectively. Then,

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

where $C(w_1, w_2)$ is a copula, a bivariate CDF on $[0, 1] \times [0, 1]$.

The unusual feature of the copula in our case is the continuity of one marginal and the discreteness of the other. 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:

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

---

[5]For 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.

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

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

where

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

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

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

Note that $\partial C(w_1, 1) / \partial w_1 = 1$ and $\partial C(w_1, 0) / \partial w_1 = 0$ due to the copula properties $C(w_1, 1) = w_1$ and $C(w_1, 0) = 0$ for all $w_1 \in [0, 1]$. Then the expression in the square brackets when evaluated at $v = 0$ is equal to

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

while when evaluated at $v = 1$ it is equal to

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

Now the conclusion easily follows. ∎

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 "deformed" Bernoulli mass of $\mathbb{I}[r_t > c]$. The "deformed" Bernoulli success probability parameter $\varrho_t \left(F^{\mathcal{D}}(u|\psi_t) \right)$ does not, in general, equal to the success probability parameter of the marginal distribution p_t (equality holds in the case of conditional independence between $|r_t - c|$ and $\mathbb{I}[r_t > c]$); 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 list three choices of copulas that will be used in the empirical section. The literature contains other examples (Trivedi and Zimmer, 2005). Let us denote the vector of copula parameters by α; usually α is one-dimensional and indexes dependence between the two marginals.

8

Frank copula. The Frank copula is

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

where $\alpha \in [-\infty, +\infty]$ and $\alpha < 0$ ($\alpha > 0$) implies negative (positive) dependence. 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 $\alpha \to 0$ implies independence between the marginals and $\varrho_t \to p_t$.

Clayton copula. The Clayton copula is

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

where $\alpha > 0$. 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 $\alpha \to +0$ implies independence between the marginals and $\varrho_t \to p_t$. Also note that this copula permits only positive dependence between the marginals, which should not be restrictive for our application.

Farlie–Gumbel–Morgenstern copula. The Farlie–Gumbel–Morgenstern (FGM) copula is

$$C(w_1, w_2) = w_1 w_2 \left(1 + \alpha \left(1 - w_1 \right) \left(1 - w_2 \right) \right),$$

where $\alpha \in [-1, +1]$ and $\alpha < 0$ ($\alpha > 0$) implies negative (positive) dependence. Note that this copula is useful only when the dependence between the marginals is modest, which again turns out not to be restrictive for our application. The joint density/mass is as (4) with

$$\varrho_t (z) = 1 - \left(1 - p_t \right) \left(1 + \alpha p_t \left(1 - 2z \right) \right).$$

Finally, $\alpha = 0$ implies independence between the marginals and $\varrho_t = p_t$.

Once all the three ingredients of the joint distribution of R_t, i.e. the volatility model, the direction model, and the copula, are specified, the vector $(\omega_r, \beta_r, \gamma_r, \rho_r, \delta'_r, \varsigma', \omega_s, \phi_s, \delta'_s, \varsigma', \alpha')'$ can

be estimated by maximum likelihood. From (4), the sample log-likelihood function to be maximized is given by

$$\sum_{t=1}^{T} \left\{ \mathbb{I}\left[r_t > c\right] \ln \varrho_t \left(F^{\mathcal{D}}(|r_t - c||\psi_t)\right) + (1 - \mathbb{I}\left[r_t > c\right]) \ln \left(1 - \varrho_t \left(F^{\mathcal{D}}(|r_t - c||\psi_t)\right)\right) \right\}$$

$$+ \sum_{t=1}^{T} \ln f^{\mathcal{D}}(|r_t - c||\psi_t).$$

2.4 Conditional mean prediction in decomposition model

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

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

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

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

where ψ_t is the conditional expectation of $|r_t - c|$, ξ_t is the conditional expected cross-product of $|r_t - c|$ and $\mathbb{I}\left[r_t > c\right]$, and $\widehat{\psi}_t$ and $\widehat{\xi}_t$ 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

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

so

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

and the returns can be predicted by

$$\widehat{r}_t = c + \left(2\widehat{p}_t - 1\right)\widehat{\psi}_t, \tag{6}$$

where $\widehat{p}_t$ denotes the predicted value of p_t. Note that one may ignore the dependence and use forecasts constructed as (6) even under conditional dependence between the components, but such forecasts will not be optimal. However, as it happens in our empirical illustration, if this conditional dependence is weak, the feasible forecasts (6) may well dominate the feasible optimal forecasts (5).

In the rest of this subsection, we discuss a technical subtlety of computing the conditional expected cross-product $\xi_t = E_{t-1}\left(|r_t - c|\mathbb{I}\left[r_t > c\right]\right)$ in the general case of conditional dependence.

The conditional distributions of $\mathbb{I}[r_t > c]$ given $|r_t - c|$ is

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

Then, the conditional expectation function of $\mathbb{I}[r_t > c]$ given $|r_t - c|$ is

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

and the expectation of the cross-product is given by

$$\xi_t = E_{t-1}\left(|r_t - c|\mathbb{I}[r_t > c]\right) = \int_0^{+\infty} u f^{\mathcal{D}}(u|\psi_t)\varrho_t\left(F^{\mathcal{D}}(u|\psi_t)\right) du. \tag{7}$$

In general, the integral (7) cannot be computed analytically (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

$$\xi_t = \int_0^1 Q^{\mathcal{D}}(z)\varrho_t(z)dz, \tag{8}$$

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 (8) 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 (8).

3 Empirical Analysis

3.1 Data

In our empirical study, we use Campbell and Yogo's (2006) data set that covers the period January 1952 – December 2002 at monthly frequency.[6] While monthly observations for the period 1927–2002 are also available, 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.

The excess stock returns and dividend-price ratio (dp) are constructed from the NYSE/AMEX value-weighted index and one-month T-bill rate from the Center for Research in Security Prices (CRSP) database. The earnings-price ratio (ep) is computed from S&P500 data and Moody's Aaa

---

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

corporate bond yield data are used to obtain the yield spread (irs). We also use the three-month T-bill rate ($ir3$) from CRSP as a predictor variable. The dividend-price and earnings-price ratios are in logs.

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

$$RV_t = \sum_{s=1}^{m} \widetilde{r}_{t,s}^2,$$

$$BPV_t = \frac{\pi}{2} \frac{m}{m-1} \sum_{s=1}^{m-1} |\widetilde{r}_{t,s}| \, |\widetilde{r}_{t,s+1}|,$$

$$RS_t = \sum_{s=1}^{m} \widetilde{r}_{t,s}^3,$$

$$RK_t = \sum_{s=1}^{m} \widetilde{r}_{t,s}^4.$$

3.2 Predictive regressions for excess returns

In this section, we present some empirical evidence on conditional mean predictability of excess stock returns from a linear predictive regression model estimated by OLS. In addition to the macroeconomic predictors that are commonly used in the literature, we follow Guo (2006) and include a proxy for stock market volatility (RV) as a predictor of future returns. We also attempted to match exactly the information variables that we use later in the decomposition model but the inclusion of the other realized measures generated large outliers in the predicted returns that deteriorated significantly the predictive ability of the linear model.

It is now well known that if the predictor variables are highly persistent, which is the case with the four macroeconomic predictors dp, ep, $ir3$ and irs, the coefficients in the predictive regression are biased (Stambaugh, 1999) and their limiting distribution is non-standard (Elliott and Stock, 1994) when the innovations of the predictor variable are correlated with returns. For example, Campbell and Yogo (2006) report that these correlations are -0.967 and -0.982 for dividend-price

and earnings-price ratios while the innovations of the three-month T-bill rate and the long-short interest rate spread are only weakly correlated with returns (correlation coefficients of -0.07). 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; Torous and Valkanov, 2000; 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 ratios is non-standard whereas the t-statistics for the interest rates variables and realized volatility can be roughly compared to the standard normal critical values due to their near-zero correlation with the returns innovations and low persistence, respectively. The results in the last two columns of Table 1 suggest some in-sample predictability with a value of the LR test statistic for joint significance of 27.8 and an R^2 of 4.45%. Even though the value of the R^2 coefficient is statistically small, Campbell and Thompson (2007) argue that it can still be economically meaningful when compared to the squared Sharpe ratio. Also, while some of the predictors (realized volatility, 3-month rate and earning-price ratio) do not appear statistically significant, they help to improve the out-of-sample predictability of the model as will be seen in the out-of-sample forecasting and the portfolio management exercises presented below.

3.3 Decomposition model for excess returns

Before we present the results from the decomposition model, we provide some details regarding the selected specification and estimation procedure. We postulate $\mathcal{D}$ to be Weibull with shape parameter $\varsigma > 0$ (the exponential distribution corresponds to the special case $\varsigma = 1$),

$$
\begin{aligned}
F^{\mathcal{D}}(u|\psi_t) &= 1 - \exp\left(-\left(\psi_t^{-1}\Gamma\left(1+\varsigma^{-1}\right)u\right)^\varsigma\right), \\
f^{\mathcal{D}}(u|\psi_t) &= \psi_t^{-\varsigma}\varsigma\Gamma\left(1+\varsigma^{-1}\right)^\varsigma u^{\varsigma-1}\exp\left(-\left(\psi_t^{-1}\Gamma\left(1+\varsigma^{-1}\right)u\right)^\varsigma\right),
\end{aligned}
$$

where $\Gamma\left(\cdot\right)$ is the gamma function. Then, the sample log-likelihood function is

$$
\sum_{t=1}^{T}\left\{\mathbb{I}\left[r_t > c\right]\ln\varrho_t\left(1-\exp\left(-\zeta_t\right)\right) + \left(1-\mathbb{I}\left[r_t > c\right]\right)\ln\left(1-\varrho_t\left(1-\exp\left(-\zeta_t\right)\right)\right)\right\}
$$

$$
+\sum_{t=1}^{T}\left\{\ln(\varsigma) - \ln|r_t - c| - \zeta_t + \ln\zeta_t\right\},
$$

where $\zeta_t = \left(\psi_t^{-1} | r_t - c | \Gamma \left(1 + \varsigma^{-1} \right) \right)^\varsigma$.

The results from the return decomposition model are reported for the case $c = 0$. Note that even though the results pertaining to the direction and volatility specifications are discussed separately, all estimates are obtained from maximizing the sample log-likelihood of the full decomposition model with Clayton copula.[7]

*** Table 2 ***

Table 2 presents the estimation results from the direction model. Several observations regarding the estimated dynamic logit specification 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 macroeconomic predictors are the same as in the linear predictive regression but the combined effect of the two realized volatility measures, RV and BPV, on the direction of the market is positive. The realized measures of the higher moments of returns do not appear to have a statistically significant effect on the direction of excess returns although they still turn out to be important in the out-of-sample exercise below.

*** Table 3 ***

Table 3 reports the results from the volatility model. The adequacy of the Weibull specification is tested using the excess dispersion and Pearson's goodness-of-fit tests. The excess dispersion test compares the residual variance to the estimated variance of a random variable distributed according to the normalized Weibull distribution:

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

where $\widehat{\sigma}_\eta^2 = \Gamma \left(1 + 2\widehat{\varsigma}^{-1} \right) / \Gamma \left(1 + \widehat{\varsigma}^{-1} \right)^2 - 1$, hats denote estimated values, and bars denote sample averages. Under the null of correct Weibull specification, ED is distributed as a standard normal random variable. 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 χ_{18}^2 and χ_{19}^2 because of the presence of an

---

[7]The results in Table 4 suggest that the Clayton copula leads to most precise estimates of the dependence between the components.

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 that is evident from our results is well documented in the literature. The nonlinear term $\rho_r \mathbb{I}[r_{t-j} > c]$ 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. The higher interest rates and earnings-price ratio appear to increase volatility while higher dividend-price ratio and yield spread tend to have the opposite effect although none of these effects is statistically significant.

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, Figures 1 and 2 plot the predicted probabilities from the direction model and the actual and predicted absolute returns from the volatility model. 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.

*** Table 4 ***

Now we consider the dependence between the two components – absolute values $|r_t - c|$ and indicators $\mathbb{I}[r_t > c]$. The dependence between these components is expected to be positive and big, and indeed, from the raw data, the estimated coefficient of unconditional correlation between them equals 0.768. Interestingly, though, after conditioning on the past, the two variables no longer exhibit any dependence. The results for the Frank, Clayton and FGM copulas are reported in Table 4 and show that the dependence parameter α is not significantly different from zero in any of the copula specifications. Insignificance aside, the point estimates are close to zero and imply near independence. The insignificance of the dependence parameter is compatible with the estimated conditional correlation between standardized residuals in the two submodels, $\psi_t^{-1}|r_t - c|$ and $p_t^{-1}\mathbb{I}[r_t > c]$, which is another indicator of dependence. These conditional correlations are

close to zero and are statistically insignificant. The result on conditional weak dependence, if any, between the components is quite surprising: once the absolute values and indicators are appropriately modeled conditionally on the past, the uncertainties left in both are statistically unrelated to each other. Furthermore, the fact of (near) independence is somewhat relieving because it facilitates the computation of the conditional mean of future returns: as discussed in section 2.4, under conditional independence (or even conditional uncorrelatedness) between the components there is no need to compute the most effort-consuming ingredient, the numerical integral (7). For illustration, however, we report later the results obtained when the conditional dependence is shut down, or equivalently, α is set to zero (ignoring dependence), and when no independence is presumed using the estimated value of α from the full model (exploiting dependence).

Table 4 also reports the values of mean log-likelihood and pseudo-R^2 goodness-of-fit measure. The log-likelihood values for the different copula specifications are of similar magnitude with a slight edge for the Clayton copula which holds also in terms of t-ratios of the dependence parameter. The LR test for joint significance of the predictor variables strongly rejects the null using the asymptotic χ^2 approximation with 16 degrees of freedom. The pseudo-R^2 goodness-of-fit measure is computed as the squared correlation coefficient between the actual and fitted excess returns from different copula specifications. A rough comparison with the R^2 from the predictive regression in Table 1 indicates an economically large improvement in the in-sample performance of the decomposition model over the linear predictive regression.

*** Figure 3 ***

Furthermore, an inspection of the fitted returns reveals some interesting differences across models. Figure 3 plots the in-sample predicted returns from our model and the predictive regression. We see that the decomposition 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 in the beginning of the sample and especially in the post-1990.

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, 2007) find that the commonly used predictive regressions would not help an

investor to profitably time the market. Campbell and Thompson (2007), 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 model proposed in this paper, predictive regression and unconditional mean (historical average) model. The forecasts are obtained from a rolling sample scheme with a fixed sample size $R = 360$. The results are reported using an out-of-sample coefficient of predictive performance OS (Campbell and Thompson, 2007) computed as

$$OS = 1 - \frac{\sum_{j=T-R+1}^{T} \partial \left(r_j - \widehat{r}_j \right)}{\sum_{j=T-R+1}^{T} \partial \left(r_j - \overline{r}_j \right)},$$

where $\partial \left(u \right) = u^2$ if it is based on squared errors and $\partial \left(u \right) = |u|$ if it is based on absolute errors, $\widehat{r}_j$ is the one-step forecast of r_j from the conditional (decomposition or predictive regression) model and $\overline{r}_j$ 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 model 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.

*** Figure 4 ***

Figure 4 plots the one-step ahead forecasts of returns from the predictive regression and the decomposition model with Clayton copula. As in the in-sample analysis, the predicted return series reveal substantial differences between the two models over time. The largest disagreement between the forecasts from the two models occurs in the 1990's when the linear regression completely misses the bull market by predicting predominantly negative returns while our model is able to capture the upward trend in the market and the increased volatility in the early 2000's.

*** Table 5 ***

Table 5 presents the results from the out-of-sample forecast evaluation. As in Goyal and Welch (2003, 2007) and Campbell and Thompson (2007), we find that the unconditional model based on the historical average performs better out-of-sample than the conditional linear model and the difference in the relative forecasting performance is close to 5%.

The results from the decomposition model estimated with the three copulas are reported separately for the cases of ignoring dependence and exploiting dependence. In all specifications, our

17

model dominates the unconditional mean forecast with forecast gains of $1.33 - 2.42\%$ for absolute errors and $1.80 - 2.64\%$ for squared errors. Although these forecast gains do not seem statistically large, Campbell and Thompson (2007) 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.

The results from the decomposition model when ignoring and exploiting dependence reveal little difference although the specification with $\alpha = 0$ appears to dominate in the case with absolute forecast errors and is outperformed by the full model in the case of squared losses. Interestingly, the Clayton copula does not show best out-of-sample performance among the three copulas, even though it fares best in-sample. Nonetheless, we will only report the findings using the Clayton copula in the decomposition model in all empirical experiments in the remainder of the paper; the other two choices of copulas deliver similar results.

It is well documented that the performance of the predictive regression deteriorates in the post-1990 period (Campbell and Yogo, 2006; Goyal and Welch, 2003; among others). To see if the decomposition model suffers from a similar forecast breakdown, we report separately the latest sample period January 1995 – December 2002. The OS statistics for this period are presented in the bottom part of Table 5. The forecasts from the linear model are highly inaccurate as the decreasing valuation ratios predict negative returns while the actual stock index continues to soar. In contrast, the forecast performance of the decomposition model tends to be rather stable over time even though it uses the same set of macroeconomic predictors.

To gain some intuition about the source of the forecasting improvements, we considered two nested versions of our model: one that contains only the own dynamics of the indicator variable and the absolute returns and a model that includes only macroeconomic predictors and realized measures without any autoregressive structure (the results are not reported to preserve space). 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. While the pure dynamic model is outperformed by the structural specification, it still dominates the linear predictive regression and the deterioration in its forecasting performance appears to be due to poor sign predictability that arises from the weak persistence in the indicator variable mentioned

above.

Test of predictive ability.   To determine the statistical significance of the differences in the out-of-sample performance of the decomposition model, predictive regression and historical average reported in Table 5, we adopt Giacomini and White's (2006) conditional predictive ability framework. Let L_{t+1}^i and L_{t+1}^j denote the loss functions (quadratic or absolute losses) of models i and j (for example, the predictive regression and the decomposition model) correspondingly, at time $t + 1$, and let $\triangle L_{t+1} = L_t^i - L_t^j$. Then, the null of equal predictive ability of two models can be expressed as $H_0 : E_t (\triangle L_{t+1}) = 0$ almost surely for all $t = R, ..., T - 1$.

For all $q \times 1$ vectors h_t that belong to the information set at time t, the null can be rewritten as $H_0 : E (h_t \triangle L_{t+1}) = 0$ and can be tested using the test statistic

$$W_{i,j} = \left(n^{-1/2} \sum_{t=R}^{T-1} h_t \triangle L_{t+1} \right)' \widehat{\Omega}_n^{-1} \left(n^{-1/2} \sum_{t=R}^{T-1} h_t \triangle L_{t+1} \right),$$

where $\widehat{\Omega}_n$ is a consistent estimator of $\lim_{n \to \infty} var \left(n^{-1/2} \sum_{t=R}^{T-1} h_t \triangle L_{t+1} \right)$ and $n = T - R - 1$. If R is assumed fixed as $n \to \infty$ and some weak regularity conditions are satisfied (Giacomini and White, 2006), $W_{i,j} \to^d \chi_q^2$ under the null of equal predictive ability. In our empirical application, $\widehat{\Omega}_n$ is a HAC estimator of Ω_n and $h_t = (1, \triangle L_t)'$. The relative performance of the models over time can be visualized by plotting the predicted loss differences $\{h_t'\widehat{\gamma}\}_{t=R}^{T-1}$, where $\widehat{\gamma}$ are the OLS estimates from a regression of $\triangle L_{t+1}$ on h_t (Giacomini and White, 2006). Finally, model i is preferred to model j if $I_{i,j} = n^{-1} \sum_{t=R}^{T-1} 1\{h_t'\widehat{\gamma} > 0\} < 0.5$. That is, a value of $I_{i,j}$ that is close to one indicates that model j dominates model i, while a value close to zero gives preference for model i over model j.

*** Table 6 ***

Table 6 presents the values of the $W_{i,j}$ test of equal conditional predictive ability of two models along with the corresponding p-values and the indicators $I_{i,j}$. The tests computed from the squared errors do not reveal any statistically significant differences across models although the indicator variable suggests that the decomposition model dominates both the historical average and predictive regression and the historical average in turn outperforms the linear model. The test based on the absolute errors, however, provides a convincing statistical evidence of superior predictive performance of the decomposition model and historical average over the predictive regression. The differences between the decomposition model and historical average are not statistically significant

although the indicator again suggests some out-of-sample superiority of the decomposition model. Consistent with the results in Table 5, exploiting dependence between the two components is a bit better in terms of squared forecast errors but a bit worse in terms of absolute losses.

*** Figure 5 ***

Figure 5 plots the relative performance of the predictive regression and decomposition model over time in terms of absolute forecast errors. Since all of the predicted absolute differences are positive, the decomposition model forecasts dominate uniformly the forecasts from the predictive regression for the entire out-of-sample period. The largest gains in terms of forecast accuracy appear to occur in the second part of the 1990's.

Mincer–Zarnowitz regressions. Another convenient approach to evaluating forecasts from competing models is the Mincer–Zarnowitz regression (Mincer and Zarnowitz, 1969). The Mincer–Zarnowitz regression has the form

$$r_t = a_0 + a_1 \widehat{r}_t + error,$$

for $t = R + 1, ..., T$, where r_t is the actual return and $\widehat{r}_t$ is the predicted return. Table 7 reports the estimates and R^2's from the Mincer–Zarnowitz regressions for the different models along with the Wald test of unbiasedness of the forecast $H_0 : a_0 = 0$, $a_1 = 1$.

*** Table 7 ***

The Mincer–Zarnowitz regression results in Table 7 reveal some interesting features of the forecasts from the competing models. Despite its relatively good performance in terms of symmetric forecast errors, the historical average forecasts prove to be severely biased. The forecasts from the predictive regressions also tend to be biased and the unbiasedness hypothesis is overwhelmingly rejected. None of the copula specifications reject the null of $a_0 = 0$ and $a_1 = 1$ and their forecasts, especially the forecasts from the decomposition model exploiting dependence, appear to possess very appealing properties.

3.5 Economic significance of return predictability: Profit-based evaluation

In order to assess the economic importance of our results, we use a profit rule for timing the market based on forecasts from different models. More specifically, we evaluate the model forecasts

in terms of the profits from a trading strategy for active portfolio allocation between stocks and bonds as in Breen *et al.* (1989), Pesaran and Timmermann (1995), Guo (2006), among others. The trading strategy consists of investing in stocks if the predicted excess return is positive or investing in bonds if the predicted excess return is negative. Note that these investment strategies require information only about the future direction (sign) of returns although the sign forecasts are obtained from the estimation of the full model. The initial investment is $100 and the value of the portfolio is recalculated and reinvested every period.

To make the profit exercise more realistic, we introduce proportional transaction costs of 0.25% of the portfolio value when the investor rebalances the portfolio between stock and bonds (Guo, 2006). The profits from this trading strategy are computed from actual stock return and risk-free rate after accounting for transaction costs and are compared to the benchmark buy-and-hold strategy.

*** Figure 6 ***

We first illustrate graphically the performance over time of the portfolios constructed from the decomposition model and predictive regression using in-sample predicted returns. The values of the portfolios from our model, linear regression and buy-and-hold strategy are plotted in Figure 6. The values of the portfolios at the end of the sample are $20,747$ for the buy-and-hold strategy, $52,154$ from the trading strategy based on the predictive regression and $80,430$ from the decomposition-based trading rule. The corresponding average annualized returns (standard deviations), after accounting for transaction costs, are 11.00% (14.44%), 13.39% (11.67%) and 14.41% (12.36%), respectively.

Now we turn our attention to the more realistic investment strategies based on out-of-sample predictions. The setup is the same as in the previous section when the model is estimated from a rolling sample of 360 observations and is used to produce 252 one-step ahead forecasts of excess returns. Table 8 reports some summary statistics of the different trading strategies such as average annualized return, standard deviation, Sharpe ratio and Jensen measure (alpha).

*** Table 8 ***

While the results in Table 8 are not as impressive as the in-sample exercise, they still provide strong evidence for the economic relevance of our approach. It is worth stressing that the out-of-

21

sample period that we examine (January 1982 – December 2002) coincides with arguably one of the greatest bull markets in history which explains the excellent performance of the buy-and-hold strategy (average annualized return of 12.55%). It is also interesting to note that the historical average forecasts give rise to a trading strategy that is equivalent to the buy-and-hold strategy since all forecasts are positive.

Despite the favorable setup for the buy-and-hold strategy, the trading strategy based on the decomposition model produces similar returns, 12.8% under independence and 11.53% with dependence, but accompanied with a large reduction in the portfolio standard deviation from 14.96% to 13.69% for the model under independence and to 12.75% for the full copula specification. As a result, the portfolio based on the independence specification has a Sharpe ratio of 0.485 (versus 0.428 for the market portfolio) and 1.37% risk-adjusted return measured by the Jensen alpha. In sharp contrast, the portfolio constructed from the linear predictive regression has a Sharpe ratio of 0.330 (average annualized return 9.96% and standard deviation 12.02%) and a negative Jensen alpha. As before, considering only the 1995–2002 period (results are not reported due to space limitations) leads to a significant deterioration of the statistics for the linear model whereas the performance of the decomposition model is practically unchanged.

3.6   Simulation experiment

In this section, we conduct a small simulation experiment that evaluates the performance of the linear predictive regression when the data are generated from the multiplicative components model analyzed in the paper. We do this for several reasons. First, it is interesting to see if this strategy can replicate the empirical findings of relatively strong predictability in the individual sign and volatility components of returns and the weak predictability of composite returns in a linear framework. This can also help us gain intuition about the importance of the nonlinearities implicit in the data generation process but not explicitly picked up by the linear predictive regression. Finally, it is instructive to investigate the effect of different degrees of dependence between the individual components on detecting predictability in the linear specification.

The simulation setup is the following. We generate 10,000 artificial samples from a DGP calibrated to the decomposition model with Clayton copula which is estimated in our empirical section, setting the predictor variables (we use only macroeconomic predictors) to their actual

values in the sample. For each artificial sample, we draw an IID series η_t distributed scaled Weibull and an IID series ν_t distributed standard uniform. The estimated volatility model is used to generate the paths of conditional means of absolute returns ψ_t, which is then transformed into a series of absolute returns by $|r_t| = \psi_t \eta_t$. The estimated direction model is used to obtain the process θ_t, which is subsequently transformed into a series of conditional success probabilities p_t. Next, we compute the series of ϱ_t implied by the Clayton copula and Weibull distribution conditional on the series of $|r_t|$, ψ_t and p_t, and generate a series of binary outcomes $\mathbb{I}[r_t > 0]$, each distributed Bernoulli with success probability ϱ_t, by setting $\mathbb{I}[r_t > 0] = \mathbb{I}[\nu_t < \varrho_t]$. Finally, we construct a sequence of simulated returns using $r_t = (2\mathbb{I}[r_t > 0] - 1)|r_t|$.

*** Figure 7 ***

Figure 7 depicts the actual and five arbitrary simulated paths of cumulative returns. We plot cumulative rather than raw returns in order to enhance the readability of the graph. Note that the simulated returns are almost twice as volatile as actual returns which appears to be due to the inclusion of a set of predictors of questionable statistical significance. Apart from that, the actual and simulated paths look quite similarly and the simulated paths do not exhibit unexpected (e.g., explosive) patterns.

*** Table 9 ***

Table 9 contains results from the linear predictive regression on simulated data generated using different values of the dependence parameter α. The upper panel corresponds to the value of the copula parameter α estimated from the data that implies weak conditional dependence between components, while the two lower panels correspond to tenfold and hundredfold values of such α implying strong and very strong conditional dependence. In all three cases the average unconditional component correlation is high and approximately matches the value 0.768 in the data, but the average conditional component correlation increases substantially as α increases. Two remarkable facts pertaining to the predictive regressions from Table 9 are worth stressing. The first is that the average t-statistics and R^2 in the upper panel are low with even smaller values than we find in the data. This indicates that the linear predictive framework has difficulties detecting the predictability in the components even for low degrees of dependence between the components. Moreover, and somewhat surprisingly, the average t-statistics and R^2 get even smaller

when the component dependence increases. This is perhaps due to the fact that the greater degree of dependence between the components increases the nonlinearities implicit in the multiplicative model and further obscures the relationship between the returns and the predictors in the linear framework. Overall, these results suggest that the linear approximation is unable to capture the predictive content of the multiplicative model.

4   Conclusion

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

Our empirical analysis of US excess stock returns for the period January 1952 – December 2002 delivers some interesting findings. In addition to the conventional statistical comparisons in- and out-of-sample, we carry out a portfolio allocation exercise that evaluates the models in terms of dollar profits. The in-sample results show that our model dominates the standard predictive regression and reveal some substantial differences in fitted returns from these methods over the sample period, especially in the late 1990s. The estimation results for the decomposition model tend to suggest that even though the sign and absolute value components exhibit substantial unconditional correlation, they have an almost zero conditional correlation which is reflected in a conditional near-independence in the copula specification.

In the out-of-sample analysis, we demonstrate that the forecasting improvements of the decomposition model over the linear predictive regression are statistically significant. While the historical average also appears to outperform the predictive regression out-of-sample as in Goyal and Welch (2003, 2006), the Mincer–Zarnowitz regressions show that the forecasts based on the unconditional mean are severely biased. In contrast, the forecasts from the decomposition model cannot reject the null of unbiasedness. Finally, the profit-based portfolio allocation exercise confirms the economic

usefulness of our model by producing risk-adjusted returns well in excess of the returns from the investment strategies based on the historical average (buy-and-hold) and the linear model.

5 Acknowledgments

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., Christoffersen, P.F., and Diebold, F.X. (2006), "Volatility and correlation forecasting," in G. Elliott, C.W.J. Granger, and Allan Timmermann (eds.), *Handbook of Economic Forecasting*, Amsterdam: North-Holland, 778-878.

[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] 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.

[7] Breen, W., Glosten, L.R., and R. Jagannathan (1989), "Economic significance of predictable variations in stock index returns," *Journal of Finance*, 44, 1177–1189.

[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. (1988), "Stock prices, earnings and expected dividends," *Journal of Finance*, 43, 661–676.

[11] Campbell, J.Y., and Thompson, S.B. (2007), "Predicting the equity premium out of sample: Can anything beat the historical average?" *Review of Financial Studies*, forthcoming.

[12] Campbell, J.Y., and Yogo, M. (2006), "Efficient tests of stock return predictability," *Journal of Financial Economics*, 81, 27–60.

[13] 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.

[14] Christoffersen, P.F., and Diebold, F.X. (2006), "Financial asset returns, direction-of-change forecasting, and volatility dynamics," *Management Science*, 52, 1273–1288.

[15] 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.

[16] Cochrane, J.H., 2005, *Asset Pricing*, second edition, Princeton University Press.

[17] de Jong, R.M., and Woutersen, T. (2005), "Dynamic time series binary choice," Manuscript, Ohio State University.

[18] 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.

[19] Engle, R.F. (2002), "New frontiers for ARCH models," *Journal of Applied Econometrics*, 17, 425–446.

[20] Engle, R.F., and Gallo, G.M. (2006), "A multiple indicators model for volatility using intra-daily data," *Journal of Econometrics*, 131, 3–27.

[21] Engle, R.F., and Russell, J.R. (1998), "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica* 66, 1127–1162.

[22] Fama, E., and French, F. (1988), "Dividend yields and expected stock returns," *Journal of Financial Economics*, 22, 3–26.

[23] Geweke, J. (1986), "Modeling the persistence of conditional variances: A comment," *Econometric Reviews*, 5, 57–61.

[24] Giacomini, R. and White, H. (2006), "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

[25] Goyal, A., and Welch, I. (2003), "The myth of predictability: Does the dividend yield forecast the equity premium?" *Management Science*, 49, 639–654.

[26] Goyal, A., and Welch, I. (2007), "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, forthcoming.

[27] Guo, H. (2006), "On the out-of-sample predictability of stock market returns," *Journal of Business*, 79, 645–670.

[28] Granger, C.W.J. (1998), "Comonotonicity," Lecture to the Econometrics and Financial Time Series Workshop, Cambridge University.

[29] Hong, Y., and Chung, J. (2003), "Are the directions of stock price changes predictable? Statistical theory and evidence," Manuscript, Cornell University.

[30] Judd, K. (1998), *Numerical Methods in Economics*, MIT Press.

[31] Kendall, M.G., and Stuart, A. (1973), *The Advanced Theory of Statistics.* London: Griffin.

[32] Lettau, M., and Ludvigson, S. (2001), "Consumption, aggregate wealth and expected stock returns," *Journal of Finance*, 56, 815–849.

[33] Linton, O., and Whang, Y.J. (2007), "The quantilogram: With an application to evaluating directional predictability," *Journal of Econometrics*, forthcoming.

[34] Marshall, A.W., and Olkin, I. (1985), "A family of bivariate distributions generated by the bivariate Bernoulli distribution," *Journal of the American Statistical Association*, 80, 332–338.

[35] Mincer, J. and V. Zarnowitz (1969), "The Evaluation of Economic Forecasts," in: J.Mincer (ed.) *Economic Forecasts and Expectations*, New York: Columbia University Press, 3–46.

[36] Nelson, R.B. (1999), *An Introduction to Copulas*, New York, NJ: Springer.

[37] Patton, A.J. (2006), "Modelling asymmetric exchange rate dependence", *International Economic Review*, 47, 527–556.

[38] Pesaran, M.H. and Timmermann, A. (1995), "Predictability of stock returns: Robustness and economic significance," *Journal of Finance*, 50, 1201–1228.

[39] Polk, C., Thompson, S.B., and Vuolteenaho, T. (2006), "New forecasts of the equity premium," *Journal of Financial Economics*, 81, 101–141.

[40] Qi, M. (1999), Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics,* 17, 419–429.

[41] Rydberg, T.H., and Shephard, N. (2003), "Dynamics of trade-by-trade price movements: Decomposition and models," *Journal of Financial Econometrics*, 1, 2–25.

[42] Satchell, S. and Timmermann, A. (1996), "Daily returns in international stocks markets: Predictability, nonlinearity and transaction costs," in: W.A. Barnett, A.P. Kirman and M. Salmon (eds.) *Nonlinear Dynamics and Economics,* Cambridge: Cambridge University Press, 369–391.

[43] Stambaugh, R.F. (1999), "Predictive regressions," *Journal of Financial Economics*, 54, 375–421.

[44] Torous, W., and Valkanov, R. (2000), "Boundaries of predictability: Noisy predictive regressions," unpublished manuscript, UCLA.

[45] Torous, W., Valkanov, R., and Yan, S. (2004), "On predicting stock returns with nearly integrated explanatory variables," *Journal of Business*, 77, 937–966.

[46] Trivedi, P.K., and Zimmer, D.M. (2005), "Copula modeling: An introduction for practitioners," *Foundations and Trends in Econometrics*, 1, 1–111.

Table 1. Estimation results from the predictive regression.

$t(dp)$	$t(ep)$	$t(ir3)$	$t(irs)$	$t(RV)$	LR	R^2
2.16	-1.42	-1.47	3.29	-0.99	27.8	4.45%

Notes: $t(z)$ denotes the t-statistic for the coefficient on variable z, and dp, ep, $ir3$, irs and RV stand for dividend-price ratio, earnings-price ratio, three-month T-bill rate, long-short yield spread and realized volatility. The t-statistics are computed using heteroskedasticity-robust standard errors. LR stands for a likelihood ratio test of joint significance of all predictors; its null distribution is χ_5^2 whose 5% critical value is 11.1.

Table 2. Estimation results from the direction model.

	ω_s	ϕ_s	$\delta_s(dp)$	$\delta_s(ep)$	$\delta_s(ir3)$	$\delta_s(irs)$	$\delta_s(RV)$	$\delta_s(BPV)$	$\delta_s(RS)$	$\delta_s(RK)$
coeff	3.418	0.190	2.526	-1.779	-8.739	15.35	5.200	-3.012	-0.324	-0.116
s.e.	1.096	0.172	1.137	1.108	3.891	6.93	2.445	2.440	0.449	0.074
t-stat	3.12	1.10	2.22	-1.61	-2.24	2.21	2.13	-1.23	-0.72	-1.56

Notes: $\delta_s(z)$ denotes the coefficient on variable z. See notes to Table 1 for the definition of variables, and additionally BPV, RS and RK stand for bipower variation, realized third moment and realized fourth moment. The estimates are obtained from the decomposition model with the Clayton copula. Shown are estimates together with robust standard errors and t-statistics of the coefficients in the logit equation $p_{t-1} = \exp(\theta_t) / (1 + \exp(\theta_t))$, 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
coeff	-0.504	0.808	0.035	-0.173	-0.077	0.065	0.348	-0.664	1.275	-0.08	19.96
s.e.	0.244	0.074	0.013	0.059	0.079	0.078	0.344	0.695	0.054		
t-stat	-2.07	10.9	2.69	-2.87	-0.98	0.83	1.01	-0.96	5.07		

Notes: $\delta_s(z)$ denotes the coefficient on variable z. See notes to Table 1 for the definition of variables. The estimates are obtained from the decomposition model with the Clayton copula. Shown are estimates together with robust standard errors and t-statistics of the coefficients in the MEM volatility equation $|r_t - c| = \psi_t \eta_t$, where ψ_t follows (2), and η_t is distributed as normalized Weibull with shape parameter ς. The t-statistic in the column for ς is computed for the restriction $\varsigma = 1$. The excess dispersion statistic ED is distributed as standard normal with a (right-tail) 5% critical value of 1.645 under the null of Weibull distribution. The Pearson test statistic PT compares the discretized empirical and Weibull distribution using 20 cells, and its null distribution is bounded between χ^2_{18} and χ^2_{19} whose 5% critical values are 28.87 and 30.14.

Table 4. Estimates and summary statistics from copula specifications.

	unconditional correlation	dependence parameter α coeff	s.e.	t-stat	conditional correlation	$LogL$	LR	pseudo-R^2
Frank copula	0.768	0.245	0.297	0.824	-0.026 (0.039)	1.8404	75.8	7.71%
Clayton copula	0.768	0.087	0.055	1.583	-0.027 (0.040)	1.8422	76.4	7.71%
FGM copula	0.768	0.123	0.149	0.825	-0.026 (0.039)	1.8405	75.9	7.71%

Notes: "Unconditional correlation" refers to the sample correlation coefficients between $|r_t - c|$ and $\mathbb{I}[r_t > c]$. "Conditional correlation" refers to the sample correlation coefficients between $\psi_t^{-1}|r_t - c|$ and $p_t^{-1}\mathbb{I}[r_t > c]$ estimated from the decomposition model, with robust standard errors in parentheses. $LogL$ denotes a sample mean loglikelihood value. LR stands for a likelihood ratio test of joint significance of all predictors; its null distribution is χ^2_{16} whose 5% critical value is 26.3. Pseudo-R^2 denotes squared correlation coefficients between excess returns and their in-sample predictions.

Table 5. Results of the out-of sample forecasting experiment.

	Linear model	Ignoring dependence Frank	Clayton	FGM	Exploiting dependence Frank	Clayton	FGM
1982:01-2002:12							
squared errors	-4.62	2.06	1.92	1.80	2.64	2.50	2.56
absolute errors	-4.81	2.42	2.21	2.21	1.54	1.33	1.40
1995:01-2002:12							
squared errors	-21.43	2.21	1.82	1.52	2.07	1.59	1.85
absolute errors	-15.84	0.88	0.43	0.36	-0.86	-1.34	-1.21

Notes: Shown are values of the OS statistic (in %). The rolling scheme uses a sample of fixed size $R = 360$. "Ignoring dependence" means that the decomposition model is estimated but predictions are constructed under the presumption of conditional independence between signs and absolute returns. "Exploiting dependence" means that the decomposition model is estimated and fully used in constructing predictions by (5), including numerical integration.

Table 6. Results of the test of predictive ability.

	Linear model	Ignoring dependence	Exploiting dependence
Squared errors			
Historical Average	1.519 (0.468) [0.016]	0.423 (0.809) [1.000]	0.846 (0.655) [0.865]
Linear Model		2.395 (0.302) [1.000]	3.294 (0.193) [1.000]
Ignoring dependence			0.129 (0.937) [0.761]
Absolute errors			
Historical Average	3.425 (0.180) [0.008]	1.106 (0.575) [0.789]	1.027 (0.599) [0.657]
Linear Model		6.751 (0.034) [1.000]	6.928 (0.031) [1.000]
Ignoring dependence			4.277 (0.118) [0.458]

Notes: The top entries in each cell are the values of test statistic $W_{i,j}$ based on the loss differences of models i and j in row i and column j, respectively, whose null distribution is χ_2^2; the corresponding p-values are in parentheses. The indicators $I_{i,j}$ are placed in square brackets.

Table 7. Results of the Mincer–Zarnowitz regression.

	Historical average	Linear model	Ignoring dependence	Exploiting dependence
$\widehat{a}_0$	0.046 (0.014)	0.005 (0.003)	0.000 (0.003)	0.003 (0.003)
$\widehat{a}_1$	−11.72 (3.96)	0.208 (0.223)	0.630 (0.228)	0.721 (0.268)
p-value	0.002	0.001	0.180	0.450
R^2	2.8%	0.4%	2.5%	2.4%

Notes: The Mincer–Zarnowitz regression is $r_t = a_0 + a_1 \widehat{r}_t + error$ for $t = R+1, ..., T$. Heteroskedasticity-robust standard errors are in parentheses. The last two rows report the p-value of the Wald test for $a_0 = 0$ and $a_1 = 1$ and the regression R^2.

Table 8. Summary statistics of different trading strategies for the out-of-sample period January

1982 – December 2002.

	Buy-and-hold	Linear model	Ignoring dependence	Exploiting dependence
average return	12.55%	9.96%	12.80%	11.53%
standard deviation	14.96%	12.02%	13.69%	12.75%
Sharpe ratio	0.428	0.330	0.485	0.426
Jensen measure		−0.27%	1.37%	0.88%

Table 9. Mean characterisics (with standard deviations) of predictive regressions run on simulated

samples.

	$t(dp)$	$t(ep)$	$t(ir3)$	$t(irs)$	R^2	UC	CC
$\alpha = 0.087$							
mean	1.42	−1.07	−1.14	1.90	1.99%	0.761	0.042
s.d.	1.00	1.00	1.03	1.02	1.11%	0.012	0.041
$\alpha = 0.869$							
mean	1.13	−0.84	−0.91	1.76	1.60%	0.760	0.303
s.d.	0.96	0.96	1.00	0.98	0.99%	0.013	0.038
$\alpha = 8.692$							
mean	0.16	−0.06	−0.12	0.93	0.71%	0.801	0.667
s.d.	0.91	0.91	0.95	0.94	0.55%	0.015	0.030

Notes: Shown are mean t-statistics, R^2, unconditional and conditional correlations, together with their standard deviations (s.d.), from predictive regressions run on 10,000 artificial samples. See notes to Table 1 for the meaning of $t(.)$ and definitions of variables. The DGP is calibrated to the decomposition model with Clayton copula. UC and CC denote unconditional and conditional, respectively, correlation of the two components of simulated returns.

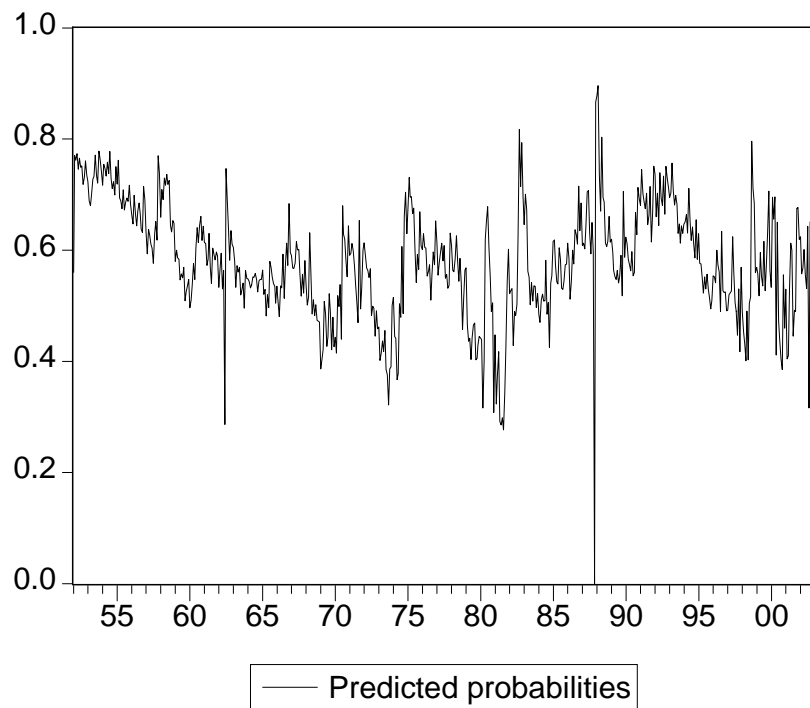Figure 1. Predicted probabilities from dynamic logit model.

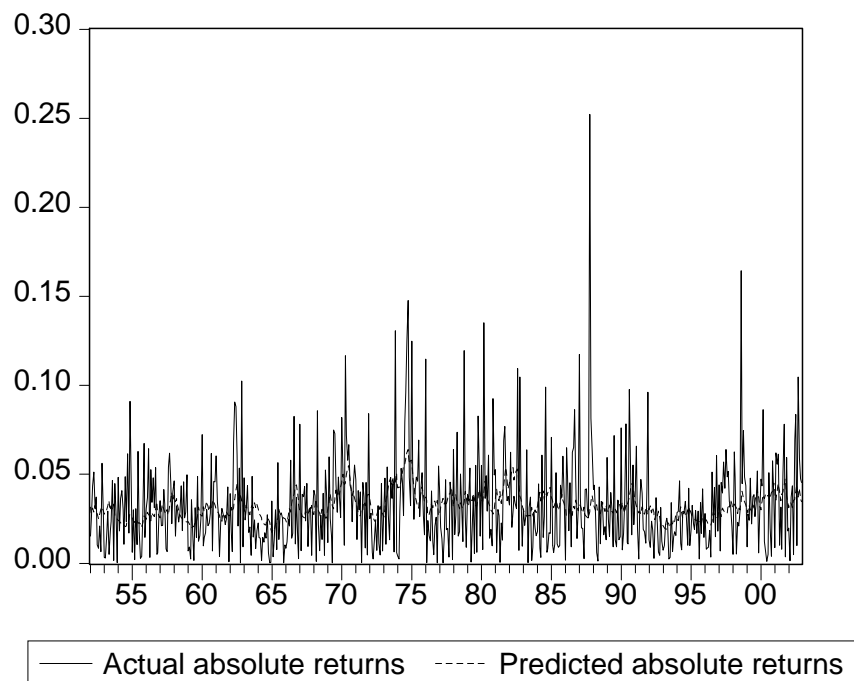Figure 2. Actual and predicted absolute returns from volatility model.

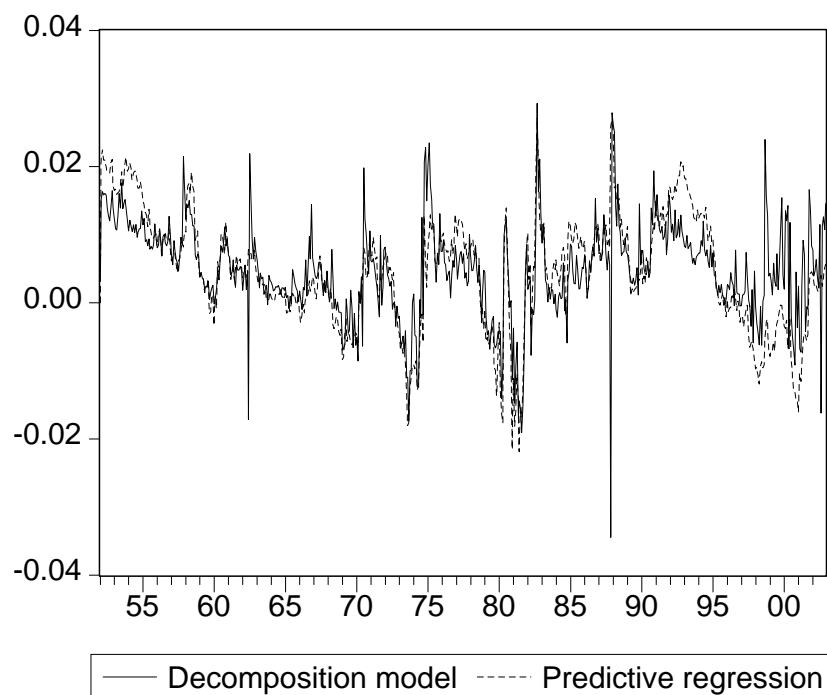Figure 3. Predicted (in-sample) returns from decomposition model and predictive regression.

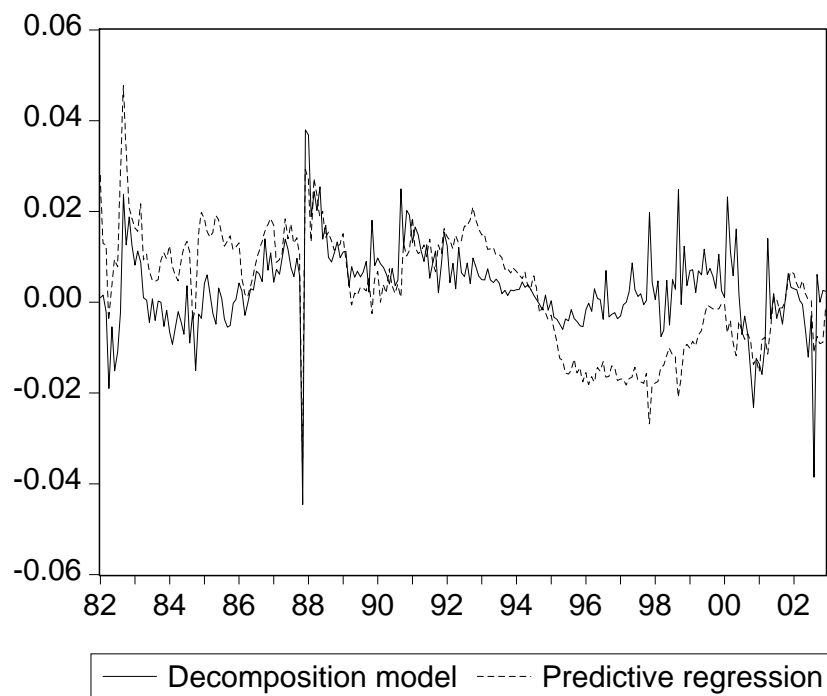Figure 4. Predicted (out-of-sample) returns from decomposition model and predictive regression.

Figure 5. Predicted absolute loss differences between predictive regression and decomposition model.
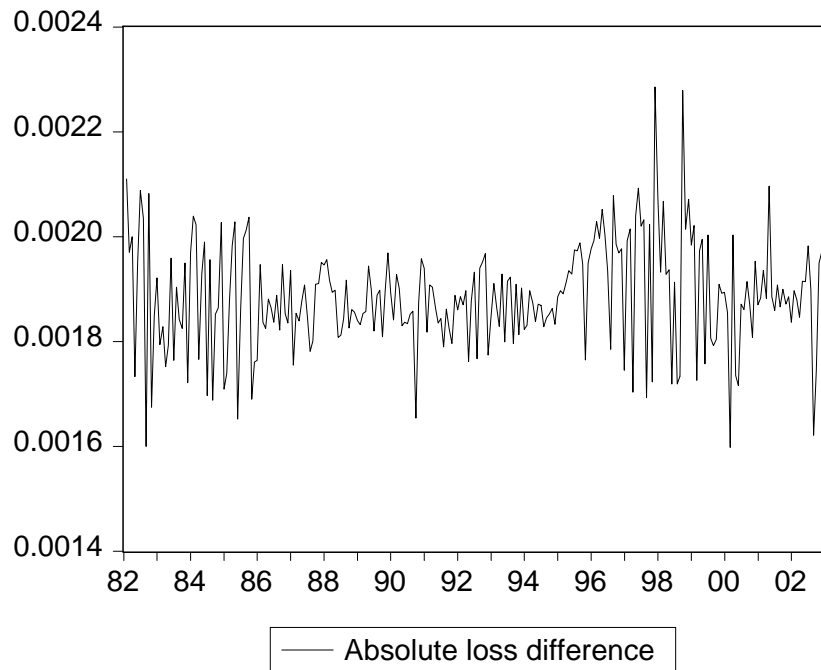

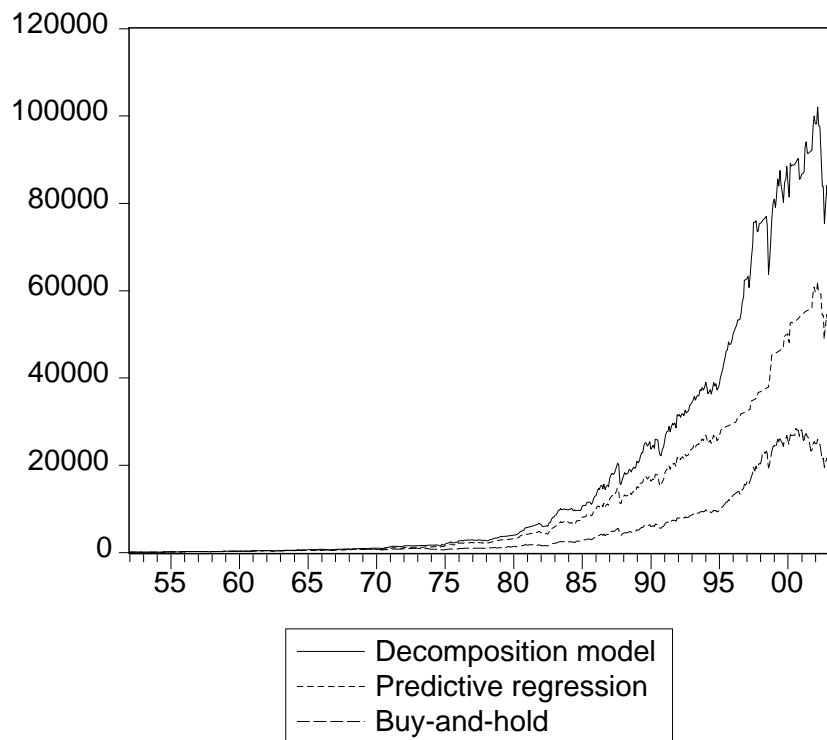Figure 6. Performance of portfolios constructed from different models.

Figure 7. Actual (bold) and five simulated paths of cumulative returns calibrated to estimated decomposition model.