Multivariate Return Decomposition: Theory and Implications*

Stanislav Anatolyev†
CERGE-EI and New Economic School

Nikolay Gospodinov‡
Federal Reserve Bank of Atlanta

May 2017

Abstract

In this paper, we propose a model based on multivariate decomposition of multiplicative – absolute values and signs – components of asset returns. In the m-variate case, the marginals for the m absolute values and the binary marginals for the m directions are linked through a $2m$-dimensional copula. The approach is detailed in the case of a bivariate decomposition. We outline the construction of the likelihood function and the computation of different conditional measures. The finite-sample properties of the maximum likelihood estimator are assessed by simulation. An application to predicting bond returns illustrates the usefulness of the proposed method.

Keywords: multivariate decomposition, multiplicative components, volatility and direction models, copula, dependence.

JEL classification codes: C13, C32, C51, G12.

*We are grateful to the Editor and two anonymous referees for useful suggestions that significantly improved the paper. We have benefited from discussions with Richard Luger and Ángel León. Our thanks also go to seminar and conference audiences at the Universidad de Alicante, Universidad Adolfo Ibáñez, the 2016 Econometric Society European meeting at the Université de Genève, and the 3rd Annual Conference of International Association for Applied Econometrics at the University of Milano-Bicocca. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

†Address: Stanislav Anatolyev, CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic; e-mail stanislav.anatolyev@cerge-ei.cz.

‡Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St. N.E., Atlanta, Georgia, 30309, USA; e-mail: nikolay.gospodinov@atl.frb.org.

1 Introduction

Any variable can be decomposed, by identity, into multiplicative absolute value and sign components. One widely documented finding in empirical work is that while the two multiplicative components exhibit a substantial degree of predictability, the variable itself is often linearly unpredictable. Anatolyev and Gospodinov (2010) capitalize on this observation and propose, in a univariate setting, a model of joint dynamics of the components that is able to exploit implicit nonlinearities, predictability in the marginals, dependence of the components etc. This analytical setup also helps to construct the whole conditional predictive density (and various conditional measures), uncover the sources of possible prediction failures of linear conditional mean models, etc.

Given the rich information content and wide applicability of this approach, it is desirable to extend it to a multivariate framework with m variables. For expositional purposes, we focus the analysis on asset returns but other stationary, weakly dependent processes can also be accommodated. In this paper, we propose a multivariate extension of the decomposition model. We link the continuous marginals for the m absolute values and the binary marginals for the m signs via a $2m$-dimensional copula generated by the inversion method. A leading choice is the normal copula which is prompted by its flexibility and computability and whose parsimonious parameterization is readily interpretable; however, asymmetric copulas (for instance, skewed normal) are also possible. We show how the likelihood function is constructed from the data, how various conditional measures of interest (such as conditional mean, variance, covariance and correlation, skewness and co-skewness, and so on) can be computed, and how the parameter estimates behave in finite samples. We further detail the model for a special case of bivariate processes, i.e. when $m = 2$, which keeps the model parsimonious and avoids the curse of dimensionality, with higher values of m implying a much higher risk of overparameterization.[1] We also conduct a small simulation experiment to get a feel for the quality of estimation in realistic samples. Finally, we provide an empirical application to two bond returns of different maturity.

It should be stressed that our approach is multi-purpose and trades off flexibility in modeling the marginals (for volatility and direction) for analytical tractability of the joint density of the $2m$ components. More flexible functional and distributional forms could be allowed provided that this preserves the analytical convenience and internal consistency of the model. Instead, and this is the approach adopted in this paper, one could further improve the specification of

[1] In addition, the case $m > 2$ brings in a need to consider a multivariate framework for binary directions-of-change which does not often happens in financial econometrics, with models like bivariate probit just beginning to gain attention recently (e.g., Nyberg, 2014); a rare exception is Anatolyev (2010).

the marginals by incorporating (functions of) additional predictors.

The article is organized as follows. In Section 2 we discuss the construction of the joint density and likelihood function as well as the computation of conditional measures. The numerical properties of the proposed maximum likelihood estimator are evaluated in a Monte Carlo experiment reported in Section 3. The usefulness of the method in a multivariate context is illustrated by studying the predictability of intermediate-term and long-term government bond returns. The empirical results from various different models, including the bivariate decomposition model, are presented in Section 4. Section 5 concludes. The Appendix contains proofs, derivations and auxiliary technical details.

2 Decomposition Approach

The decomposition approach is based on modeling the joint distribution of multiplicative components of asset returns – their absolute values and signs, or, equivalently, directions. In a univariate case, a positive marginal for the absolute values and a binary marginal for the signs are linked by a copula, all three ingredients being conditional on the history of returns. In the m-variate case, the ingredients of the decomposition model are an m-variate positive 'marginal' for m absolute values, an m-variate binary 'marginal' for m directions, and a $2m$-dimensional copula that links all components.

The marginals for absolute values have positive support;[2] each of these m marginals may be Weibull, Generalized Gamma, Burr, or another positive distribution, with the dynamics following some form of a multiplicative error model (MEM) (Engle, 2002). Each of m binary marginals is, of course, Bernoulli; candidates for a convenient direction submodel are the probit and logit models. The choice of the copula is vast. Anatolyev and Gospodinov (2010) in their application used the Clayton, Frank, and Farley-Gumbel-Morgenstern copulas; Liu and Luger (2015) in addition used the rotated Clayton copula, etc. In the multivariate setting, we suggest using copulas generated by the inversion method (Trivedi and Zimmer, 2005). This class, in addition to providing analytical convenience and integrity of the decomposition model, yields other attractive features. First, such copulas are fairly flexible in a multivariate context: their covariance matrix is parameterized by $m(2m-1)$ parameters, which in the bivariate case $m = 2$ equals 6. These parameters are easily interpretable as degrees of dependence among different components, which may not be the case with other copula choices. Second, the submodel for absolute values in this case is a multivariate MEM model (as in Cipollini, Engle and Gallo, 2017), and the submodel for directions is a multivariate binary choice model (as, for example,

[2] Strictly speaking, the support should be non-negative, but we assume a continuous distribution of returns which makes the difference inconsequential.

the multivariate probit in Ashford and Sowden, 1970). A special case of this class of copulas is the normal copula which can facilitate computations of various conditional distributions involved in the likelihood. The general theory developed below, however, is applicable to other choices of copulas from this class; we also work out an asymmetric case of the skewed normal copula.

2.1 Univariate decomposition

To illustrate the main idea of our approach, we first present the univariate decomposition method of Anatolyev and Gospodinov (2010). To show a more clear-cut result, we intentionally set the copula to be normal, the volatility marginal to be Weibull, and the direction marginal to be probit. Our multivariate extension in the next subsection is applicable to more general marginals and copulas.

Let r_t, $t = 1, ..., T$, be a time series of asset returns. It can be decomposed into two multiplicative components as

$$r_t = |r_t| \operatorname{sign}(r_t) = |r_t|(2I_t - 1),$$

where $I_t = \mathbb{I}\{r_t > 0\}$, and $\mathbb{I}\{\cdot\}$ denotes the indicator function. The univariate decomposition method of Anatolyev and Gospodinov (2010) is based on joint dynamic modeling of the two multiplicative components – 'volatility' $|r_t|$ and 'direction' I_t, which is a linear transformation of sign $\operatorname{sign}(r_t)$.

Let $\varphi_t = E(|r_t||\mathcal{F}_{t-1})$ be the conditional expectation of a conditionally Weibull distributed absolute value $|r_t|$ with a shape parameter ς, denoted as $|r_t||\mathcal{F}_{t-1} \sim \mathcal{W}(\varphi_t, \varsigma)$. Let $p_t = \Pr\{r_t > 0|\mathcal{F}_{t-1}\} = \Phi(\theta_t)$ be the 'success' (i.e. the market going up) probability of the Bernoulli distributed direction I_t denoted as $I_t|\mathcal{F}_{t-1} \sim \mathcal{B}(p_t)$. The joint distribution of the two multiplicative components can be expressed as

$$(|r_t|, I_t)\,|\mathcal{F}_{t-1} \sim C(\mathcal{W}(\varphi_t, \varsigma), \mathcal{B}(p_t), \varrho),$$

where $C(w, y) = \Phi_2(\Phi^{-1}(w), \Phi^{-1}(y), \varrho)$ is a bivariate normal copula with correlation parameter ϱ. The processes φ_t and θ_t can be specified as functions of the variables in $\mathcal{F}_{t-1}$ adding to the dimension of the parameter vector.

Let us temporarily suppress the time indexing. Denote by $f_v(u)$ and $F_v(u)$ the PDF and CDF of the volatility component, and by p the success probability of the direction component.

Proposition 1. *The joint density/mass of the pair* $(|r|, I)$ *is equal to*

$$f(u, v) = f_v(u)\, f_d^C(u, v),$$

where $f_d^C(u,v)$ is the Bernoulli PMF with 'distorted' probability

$$p^C(u) = \Phi\left(\frac{\Phi^{-1}(p) - \varrho\Phi^{-1}(F_v(u))}{\sqrt{1-\varrho^2}}\right).$$

In our case, $f_v(u)$ and $F_v(u)$ are those of the Weibull distribution, and $p = \Phi(\theta)$ is probit success probability. Restoring time indexing, the joint log-likelihood is

$$\ell_r = \sum_{t=1}^{T} \log f^{\mathcal{W}(\varphi_t,\varsigma)}(|r_t|) + \sum_{t=1}^{T} I_t \log p_t^C + (1-I_t)\log(1-p_t^C),$$

where the series of 'distorted' probabilities is, for $t = 1, ..., T$,

$$p_t^C = \Phi\left(\frac{\theta_t + \varrho\Phi^{-1}(F^{\mathcal{W}(\varphi_t,\varsigma)}(|r_t|))}{\sqrt{1-\varrho^2}}\right).$$

2.2 Multivariate decomposition

Now let $r_{1,t}$, $r_{2,t}$, ..., $r_{m,t}$ be m time series of returns. Each of them can be decomposed as

$$r_{i,t} = |r_{i,t}|(2I_{i,t} - 1),$$

$i = 1, ..., m$, where $I_{i,t} = \mathbb{I}\{r_{i,t} > 0\}$. Each absolute value $|r_{i,t}|$ is distributed as $\mathcal{D}(\varphi_{i,t}, \varsigma_i)$, where $\mathcal{D}$ is some distribution with a positive support, $\varphi_{i,t} = E(|r_{i,t}||\mathcal{F}_{t-1})$ is conditional mean, and ς_i is (possibly multidimensional) shape parameter. Each direction $I_{i,t}$ is distributed as Bernoulli $\mathcal{B}(p_{i,t})$, where $p_{i,t} = \Pr\{r_{i,t} > 0|\mathcal{F}_{t-1}\}$ is conditional success probability. The information set $\mathcal{F}_{t-1}$ now embeds individual information sets $\mathcal{F}_{i,t-1}$ and possibly information beyond the history of the returns. Together, there are $2m$ components that are linked through $2m$-variate copula C. Arranging the arguments by m absolute values first and then by m indicators, we consider copulas generated by the inversion method (Trivedi and Zimmer, 2005, section 3.1):

$$C(w_1, ..., w_m, y_1, ..., y_m) = \Psi_{2m}(\Psi^{-1}(w_1), ..., \Psi^{-1}(w_m), \Psi^{-1}(y_1), ..., \Psi^{-1}(y_m), R, \nu),$$

where Ψ_{2m} is CDF of some $2m$-variate distribution (with corresponding PDF ψ_{2m}), Ψ is its one-dimensional marginal CDF (with corresponding PDF ψ), R is a $2m \times 2m$ copula correlation matrix, and ν denotes an additional (possibly multidimensional) shape parameter, if any. This class contains, in particular, the popular in financial econometrics normal copula, in which case $\Psi_{2m} = \Phi_{2m}$, $\Psi = \Phi$, and ν is nil. Another possibility is Student's t copula, in which case $\Psi_{2m} = T_{2m,\nu}$, $\Psi = T_\nu$, and ν is degrees of freedom. In the empirical application, we use the normal copula, as well as the asymmetric copula based on the multivariate skewed normal distribution (Azzalini and Dalla Valle, 1996), in which case ν contains asymmetry parameters.[3]

[3] See Appendix A.2 for details when $m = 2$.

For notational convenience, let us again temporarily suppress time indexing and parameter dependence. Denote the marginal CDFs of volatility components by $F_1(u_1), ..., F_m(u_m)$ and their marginal PDFs by $f_1(u_1), ..., f_m(u_m)$. Furthermore, denote the success probabilities of the direction components by $p_1, ..., p_m$, and the corresponding CMFs by $G_1(v_1), ..., G_m(v_m)$. The following Proposition gives an expression for the $2m$-variate joint density/mass function.

Proposition 2. *The joint density/mass of the* $2m$*-tuple* $(|r_1|, ..., |r_m|, I_1, ..., I_m)$ *is equal to*

$$f(u_1, ..., u_m, v_1, ..., v_m) = f_u(u_1, ..., u_m) f^C(v_1, ..., v_m, u_1, ..., u_m),$$

$(u_1, ..., u_m) \in [0, \infty)^m$, $(v_1, ..., v_m) \in \{0, 1\}^m$, where

$$f_u(u_1, ..., u_m) = f_1(u_1)...f_m(u_m)c(F_1(u_1), ..., F_2(u_m))$$

is the m*-variate PDF of the volatility submodel linked by the* m*-variate copula* $C(w_1, ..., w_m) = \Psi_m(\Psi^{-1}(w_1), ..., \Psi^{-1}(w_m))$ *with density*

$$c(w_1, ..., w_m) = \frac{\psi_m(\Psi^{-1}(w_1), ..., \Psi^{-1}(w_m))}{\psi\left(\Psi^{-1}(w_1)\right)...\psi\left(\Psi^{-1}(w_m)\right)},$$

and

$$f^C(v_1, ..., v_m, u_1, ..., u_m) = \sum_{\ell_1 \in \{0,1\}} ... \sum_{\ell_m \in \{0,1\}} (-1)^{\ell_1 + ... + \ell_m} \pi_{\ell_1 ... \ell_m}(v_1, ..., v_m, u_1, ..., u_m),$$

where

$$\pi_{\ell_1 ... \ell_m}(v_1, ..., v_m, u_1, ..., u_m) =$$

$$\Psi_m(\Psi^{-1}(G_1(v_1 - \ell_1)), ..., \Psi^{-1}(G_m(v_m - \ell_m))|\Psi^{-1}(F_1(u_1)), ..., \Psi^{-1}(F_m(u_m)))$$

is the 'distorted' PMF of the m*-variate PDF direction submodel.*

Note that the volatility submodel is the copula-based m-variate MEM (though with different marginals) from Cipollini, Engle and Gallo (2017). The direction submodel is represented by the probability mass function for an m-variate binary vector. It is straightforward to write out the joint log-likelihood function (see below for the case $m = 2$) to be maximized over the parameter collection that includes parameters of m marginal volatility models, m marginal direction models, and copula parameters. For the former two subcollections, it is natural to use as starting values the parameter estimates from the stand-alone volatility and stand-alone direction models, respectively, and the copula parameters can be initialized at the values implying independence and normality.

2.3 Bivariate decomposition

We now specialize the results to the case $m = 2$. The copula correlation matrix then has the following structure:

$$R = \begin{bmatrix} 1 & \varrho_v & \varrho_1 & \varrho_{vd} \\ \varrho_v & 1 & \varrho_{dv} & \varrho_2 \\ \varrho_1 & \varrho_{dv} & 1 & \varrho_d \\ \varrho_{vd} & \varrho_2 & \varrho_d & 1 \end{bmatrix} \equiv \begin{bmatrix} R_v & R_{vd} \\ R_{dv} & R_d \end{bmatrix},$$

where each block R_v, R_d, R_{vd} and R_{dv} is 2×2. The coefficient ϱ_v is responsible for the dependence between the two volatilities in the volatility submodel, the coefficient ϱ_d – for the dependence between the two directions in the direction submodel, the coefficient ϱ_1 – for the dependence between the volatility and direction in the decomposition submodel for the first asset, etc. The following Corollary is an $m = 2$ refinement of Proposition 2.

Corollary 3. *The joint density/mass of the quartuple* $(|r_{1,t}|, |r_{2,t}|, I_{1,t}, I_{2,t})$ *is equal to*

$$f(u_1, u_2, v_1, v_2) = f_v(u_1, u_2) f_d^C(u_1, u_2, v_1, v_2),$$

$(u_1, u_2) \in [0, \infty)^2$, $(v_1, v_2) \in \{0, 1\}^2$, *where*

$$f_v(u_1, u_2) = f_1(u_1) f_2(u_2) c(F_1(u_1), F_2(u_2))$$

is the bivariate PDF of the volatility submodel linked by the bivariate copula $C(w_1, w_2) = \Psi_2(\Psi^{-1}(w_1), \Psi^{-1}(w_2))$ with density

$$c(w_1, w_2) = \frac{\psi_2(\Psi^{-1}(w_1), \Psi^{-1}(w_2))}{\psi\left(\Psi^{-1}(w_1)\right) \psi\left(\Psi^{-1}(w_2)\right)},$$

and

$$f_d^C(u_1, u_2, v_1, v_2) = p_{11}^C(u_1, u_2)^{v_1 v_2} p_{01}^C(u_1, u_2)^{(1-v_1)v_2} p_{10}^C(u_1, u_2)^{v_1(1-v_2)} p_{00}^C(u_1, u_2)^{(1-v_1)(1-v_2)}$$

is the bivariate Bernoulli PMF of the direction submodel with 'distorted' probabilities

$$\begin{aligned} p_{11}^C(u_1, u_2) &= 1 - \pi_1(u_1, u_2) - \pi_2(u_1, u_2) + \pi_{12}(u_1, u_2), \\ p_{01}^C(u_1, u_2) &= \pi_1(u_1, u_2) - \pi_{12}(u_1, u_2), \\ p_{10}^C(u_1, u_2) &= \pi_2(u_1, u_2) - \pi_{12}(u_1, u_2), \\ p_{00}^C(u_1, u_2) &= \pi_{12}(u_1, u_2), \end{aligned}$$

where[4]

$$\begin{aligned} \pi_1(u_1, u_2) &= \Psi_2(\Psi^{-1}(1 - p_1), \Psi^{-1}(1) | \Psi^{-1}(w_1), \Psi^{-1}(w_2))\big|_{w_1 = F_1(u_1), w_2 = F_2(u_2)}, \\ \pi_2(u_1, u_2) &= \Psi_2(\Psi^{-1}(1), \Psi^{-1}(1 - p_2) | \Psi^{-1}(w_1), \Psi^{-1}(w_2))\big|_{w_1 = F_1(u_1), w_2 = F_2(u_2)}, \\ \pi_{12}(u_1, u_2) &= \Psi_2(\Psi^{-1}(1 - p_1), \Psi^{-1}(1 - p_2) | \Psi^{-1}(w_1), \Psi^{-1}(w_2))\big|_{w_1 = F_1(u_1), w_2 = F_2(u_2)}. \end{aligned}$$

[4] The following expressions can be simplified using that $\Psi_2(y_1, \Psi^{-1}(1)) = \Psi(y_1)$ and $\Psi_2(\Psi^{-1}(1), y_2) = \Psi(y_2)$. However, we prefer not to do it for the sake of generality of further computations of conditional CDFs.

It is straightforward to deduce that when the copula is normal and there is no link between the direction and volatility submodels, then

$$p_{ij}^C = \Phi_{(-1)^{|i-j|}\varrho_d}((-1)^{i+1}\theta_1, (-1)^{j+1}\theta_2), \quad i,j \in \{0,1\},$$

where $\Phi_\varrho(\cdot,\cdot)$ denotes a standard bivariate normal CDF with correlation coefficient ϱ, reducing to the bivariate probit model (Ashford and Sowden, 1970).[5]

The probabilities $\pi.(u_1, u_2)$ can be computed using that the conditional CDF is

$$\Psi_2(v_1, v_2|u_1, u_2) = \int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\psi_4(u_1, u_2, x_3, x_4)}{\psi_2(u_1, u_2)} dx_3 dx_4.$$

Restoring time indexing, the joint log-likelihood equals

$$\begin{aligned}\ell_r \;=\; & \sum_{t=1}^{T}\sum_{i=1,2} \log f^{\mathcal{D}(\varphi_{i,t})}(|r_{i,t}|) + \sum_{t=1}^{T} c(F^{\mathcal{D}(\varphi_{1,t})}(|r_{1,t}|), F^{\mathcal{D}(\varphi_{2,t})}(|r_{2,t}|)) + \\ & \sum_{t=1}^{T} I_{1,t}I_{2,t}\log p_{11,t}^C + (1-I_{1,t})I_{2,t}\log p_{01,t}^C + I_{1,t}(1-I_{2,t})\log p_{10,t}^C + (1-I_{1,t})(1-I_{2,t})\log p_{00,t}^C,\end{aligned}$$

where $p_{ij,t}^C = p_{ij}^C(|r_{1,t}|, |r_{2,t}|)$, $i,j \in \{0,1\}$, $t = 1, ..., T$, is a collection of the series of 'distorted' probabilities. We rely on standard results in the literature on maximum likelihood estimation to obtain asymptotic normality of parameter estimates and compute their asymptotic variance.

2.4 Computation of conditional measures

The decomposition model is a fully specified parametric model, and hence allows computation of various conditional measures such as conditional mean values, conditional variances, covariances and correlations, and so on. In this subsection we give technical details how one can compute conditional expectations of various functions of r_1, ..., r_m.

Suppose one is interested in the conditional expectation of $g(r_1, ..., r_m)$ for some function $g(\cdot, ..., \cdot)$. The predictor for a general function of returns is, temporarily omitting conditioning on $\mathcal{F}_{t-1}$ and time indexes,

$$\begin{aligned}E\left[g(r_1,...,r_m)\right] \;=\; & \sum_{v_1\in\{0,1\}} ... \sum_{v_m\in\{0,1\}} \int_{u_1=0}^{+\infty} ... \int_{u_m=0}^{+\infty} g(u_1(2v_1-1), ..., u_m(2v_m-1)) \\ & f_d^C(v_1, .., v_m) f_v(u_1, ..., u_m) du_1 ... du_m.\end{aligned}$$

The summation over all m-tuples of $(v_1, .., v_m) \in \{0,1\}^m$ implies 2^m terms, and integration over $(u_1, ..., u_m) \in [0,\infty)^m$ can be implemented numerically, at least for smaller m.

[5] For example, $p_{00}^C = \Phi_{\varrho_d}(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)) = \Phi_{\varrho_d}(-\theta_{1,t}, -\theta_{2,t})$, or $p_{01}^C = 1 - \Phi(\theta_{1,t}) - \Phi_{\varrho_d}(-\theta_{1,t}, -\theta_{2,t}) = \Phi(-\theta_{1,t}) - (\Phi(-\theta_{1,t}) - \Phi_{-\varrho_d}(-\theta_{1,t}, \theta_{2,t})) = \Phi_{-\varrho_d}(-\theta_{1,t}, \theta_{2,t})$, etc.

Consider again the case $m = 2$. The predictor for a general function of returns is

$$\begin{aligned}
E[g(r_1, r_2)] &= \sum_{v_1 \in \{0,1\}} \sum_{v_2 \in \{0,1\}} \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} g(u_1(2v_1 - 1), u_2(2v_2 - 1)) f_d^C(v_1, v_2) f_v(u_1, u_2) du_1 du_2 \\
&= \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} \left[\begin{array}{c} g(-u_1, -u_2) p_{00}^C(u_1, u_2) + g(u_1, -u_2) p_{10}^C(u_1, u_2) \\ +g(-u_1, u_2) p_{01}^C(u_1, u_2) + g(u_1, u_2) p_{11}^C(u_1, u_2) \end{array} \right] f_v(u_1, u_2) du_1 du_2,
\end{aligned}$$

where $f_v(u_1, u_2)$ and $p_{ij}^C(u_1, u_2)$, $i, j \in \{0, 1\}$ are defined in Corollary 3. If the function $g(\cdot, \cdot)$ is defined over absolute values only, then, denoting $g(r_1, r_2) = h(|r_1|, |r_2|)$,

$$E[h(|r_1|, |r_2|)] = \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} h(u_1, u_2) f_v(u_1, u_2) du_1 du_2.$$

If g is a function of only one of returns, r_1 say, the expression simplifies:

$$\begin{aligned}
E[g(r_1)] &= \sum_{v_1 \in \{0,1\}} \sum_{v_2 \in \{0,1\}} \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} g(u_1(2v_1 - 1)) f_d^C(v_1, v_2) f_v(u_1, u_2) du_1 du_2 \\
&= \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} [g(-u_1)\pi_1(u_1, u_2) + g(u_1)(1 - \pi_1(u_1, u_2))] f_v(u_1, u_2) du_1 du_2.
\end{aligned}$$

Alternatively and more simply, one can proceed as in Anatolyev and Gospodinov (2010):

$$E[g(r_1)] = \int_{u_1=0}^{+\infty} \left(g(-u_1) p_{1,t}^C(u_1) + g(u_1)\left(1 - p_{1,t}^C(u_1)\right)\right) f_1(u_1) du_1.$$

As a consequence, the conditional means can be computed as

$$E_{t-1}[r_{1,t}] = \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} u_1 (1 - 2\pi_{1,t}(u_1, u_2)) f_{v,t}(u_1, u_2) du_1 du_2,$$

or alternatively, as

$$E_{t-1}[r_{1,t}] = 2E_{t-1}[|r_{1,t}| I_{1,t}] - E_{t-1}[|r_{1,t}|] = 2\xi_{1,t} - \varphi_{1,t},$$

where $\xi_{1,t} = \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} u_1 (1 - \pi_{1,t}(u_1, u_2)) f_{v,t}(u_1, u_2) du_1 du_2$. The expressions for $E_{t-1}[r_{2,t}]$ are obtained similarly. The conditional means can be used, among other things, for constructing the pseudo-R^2 measure.

The conditional variances are simply

$$\text{var}_{t-1}(r_{1,t}) = E_{t-1}\left[|r_{1,t}|^2\right] - E_{t-1}[r_{1,t}]^2,$$

where $E_{t-1}\left[|r_{1,t}|^2\right] = \int_{u_1=0}^{+\infty} u_1^2 \int_{u_2=0}^{+\infty} f_{v,t}(u_1, u_2) du_1 du_2$, and similarly for $\text{var}_{t-1}(r_{2,t})$. The conditional correlations are

$$\text{corr}_{t-1}(r_{1,t}, r_{2,t}) = \frac{E_{t-1}[r_{1,t} r_{2,t}] - E_{t-1}[r_{1,t}] E_{t-1}[r_{2,t}]}{\sqrt{\text{var}_{t-1}(r_{1,t}) \text{var}_{t-1}(r_{2,t})}},$$

where $E_{t-1}[r_{1,t} r_{2,t}] = \int_{u_1=0}^{+\infty} \int_{u_2=0}^{+\infty} u_1 u_2 (1 - 2\pi_{1,t}(u_1, u_2) - 2\pi_{2,t}(u_1, u_2) + 4\pi_{12,t}(u_1, u_2)) f_{v,t}(u_1, u_2) du_1 du_2$.

In our simulations and empirical work, we compute the two-dimensional integrals involved in these formulas via a product Gauss–Chebychev quadrature with 100 Chebychev quadrature nodes on $[0, 1]$; see Judd (1998, p. 270).

3 Simulation evidence

In this section, we investigate the finite-sample properties of the maximum likelihood estimator of the bivariate decomposition model. The simulation design is an 'autoregressive' (not containing extraneous predictors) version of the model used in the empirical section, with similar parameter values.

The volatility equations are specified as

$$\ln \varphi_{i,t} = \omega_{vi} + \beta_{vi} \ln \varphi_{i,t-1} + \alpha_{vij} \ln |r_{j,t-1}| + \gamma_{vij} I_{j,t-1}$$

for $i, j = 1, 2$. The parameter values are $\omega_{vi} = 0$, $\beta_{vi} = 0.8$, $\alpha_{vij} = 0.1$ for $i = j$ and $\alpha_{vij} = 0.05$ for $i \neq j$, and $\gamma_{vij} = -0.3$ for $i = j$ and $\gamma_{vij} = 0.2$ for $i \neq j$. That is, the persistence in volatility is high, and its reaction to news about own components is higher than that to news about the other variable's components. The Weibull distribution shape parameters are $\varsigma_1 = \varsigma_2 = 1.2$. We also try non-symmetric shapes with higher curvature: $\varsigma_1 = 1.5$, $\varsigma_2 = 2.5$.

The direction equations are specified as

$$\theta_{i,t} = \omega_{di} + \phi_{dij} I_{j,t-1}$$

for $i, j = 1, 2$. The parameter values are $\omega_{di} = 0.3$, $\phi_{dij} = 0.3$ for $i = j$ and $\phi_{dij} = -0.1$ for $i \neq j$. That is, the reaction of direction to own past directions is higher in absolute value than that the other variable's directions, and opposite in sign.

The elements of the dependence matrix R are set at $\varrho_v = \varrho_d = 0.6$ and $\varrho_1 = \varrho_2 = \varrho_{vd} = \varrho_{dv} = 0.2$. That is, the namesake components are moderately correlated across variables; the opposite components are weakly correlated both within the same variable and across variables.

Table 1 presents the mean and standard deviation of the estimates across 1,000 replications for sample sizes $n = 500$ and $n = 2000$. To assess the accuracy of the asymptotic standard errors and asymptotic normality of the estimates, we also report the empirical size of the individual t-tests at the 5% significance level. Overall, the estimates appear unbiased and well identified. The empirical size of tests for the true value is very close to the nominal level of the test. The results also suggest that the volatility equation coefficients are estimated on average twice as precisely as the direction equation coefficients, the other variable's news impact coefficients beating the record. Another interesting thing is that the degree of volatility dependence is estimated twice as precisely as all other dependence coefficients. Higher curvature of the shape of the absolute value marginals in fact brings a better precision of some parameter estimates, not significantly altering the empirical rejection rates.

To evaluate the accuracy of the predictions from the bivariate decomposition model, we compute various predictive performance measures, and compare them to their analogs from a linear VAR(1) model fitted to the simulated data. One measure is the pseudo-R^2 for returns which is a measure that favors linear projections over complex nonlinear structures. The other two measures, for which the decomposition model is supposed to show advantage, are the pseudo-R^2 for absolute values and the fraction of successful sign predictions.[6,7]

Figures 1–3 plot histograms for these three measures from the linear and decomposition models for $n = 500$. It turns out that for this DGP the figures of R^2 for returns are small, indicating that the noise by far exceeds the signal, and one should not expect big R^2 values in a corresponding application. The decomposition model, being the true DGP, produces a much higher R^2 for returns than the linear model, and the corresponding distribution dominates that for the linear model. Even sharper is the contrast for the values of R^2 for absolute returns which are several-fold higher for the decomposition model than those for the linear model. The sign predictability is also markedly more successful for the decomposition model, though by a narrower margin. This is also natural as directions are intrinsically less persistent and hence less predictable than the absolute values.

The discrepancy between these figures in actual applications is not expected to be that sharp as in these simulations, for two reasons. First, the actual data generation mechanism may be more complex and its type is likely to fall 'in between' linear and nonlinear. Second, these quantities, especially pseudo-R^2, are very noisy measures of time-series fit, at least for such sample sizes. We will return to this issue in the empirical section of the paper.

4 Bond return predictability

4.1 Motivation and data description

Predictability of bond returns has been the focus of renewed research interest in the recent literature. Cochrane and Piazzesi (2005) provide strong evidence of predictability of excess bond returns by a linear combination of forward rates. Some subsequent studies show that the first few principal components from a large panel of US economic and financial time series (Ludvigson and Ng, 2009), survey inflation expectations (Chernov and Mueller, 2012) and a

[6] The pseudo-R^2 for returns and pseudo-R^2 for absolute values are computed as

$$\text{pseudo-}R^2_r = \frac{\left(\sum_t \left(\hat{r}_{t|t-1} - \bar{\hat{r}}_{t|t-1}\right)\left(\hat{r}_t - \overline{r}_t\right)\right)^2}{\sum_t \left(\hat{r}_{t|t-1} - \bar{\hat{r}}_{t|t-1}\right)^2 \sum_t \left(\hat{r}_t - \overline{r}_t\right)^2}, \quad \text{pseudo-}R^2_{|r|} = \frac{\left(\sum_t (|\hat{r}_{t|t-1}| - \overline{|\hat{r}_{t|t-1}|})(|\hat{r}_t| - \overline{|r_t|})\right)^2}{\sum_t (|\hat{r}_{t|t-1}| - \overline{|\hat{r}_{t|t-1}|})^2 \sum_t (|\hat{r}_t| - \overline{|r_t|})^2},$$

[7] The proportions of successful sign predictions are computed as

$$S = \frac{1}{T}\sum_t \left(I_t^+ \mathbb{I}_{\{r_t>0\}} + \left(1 - I_t^+\right)\mathbb{I}_{\{r_t \le 0\}}\right),$$

where $I_t^+ = \mathbb{I}_{\{\hat{r}_{t|t-1}>0\}}$ for the linear model and $I_t^+ = \mathbb{I}_{\{p_t^C>0.5\}}$ for the decomposition model.

cyclical component of past inflation (Cieslak and Povala, 2015) also tend to be strong predictors of future bond returns. The statistical magnitude of the bond return predictability and the robustness of these findings are summarized in Duffee (2013). This predictive evidence may at first appear to be at odds with the results that these additional factors are ineffective in explaining the term structure of bond yields where the level, slope and the curvature of the yield curve explain in excess of 99.5% of the cross-sectional variation of bond yields. However, Duffee (2011) argues that this evidence can be reconciled if these are hidden factors; i.e., they do not affect the cross-section of yields but help to predict the future dynamics of bond returns. In other words, the dimension of the state vector that determines current yields is smaller than the dimension of the state vector that determines the expected bond yields and returns (Duffee, 2013).

Following Duffee (2013), this could be best illustrated using the conditional expectation version of the main relationship linking long-term and short-term yields:

$$y_t^{(n)} = \frac{1}{n} E_t \left(\sum_{h=0}^{n-1} y_{t+h}^{(1)} \right) + \frac{1}{n} E_t \left(\sum_{h=0}^{n-1} r_{t+h,t+h+1}^{(n)} - y_{t+h}^{(1)} \right),$$

where $y_t^{(n)} = -\frac{1}{n}\ln(P_t^{(n)})$ is the yield on an n-period zero-coupon bond with price $P_t^{(n)}$ at time t and $r_{t,t+1}^{(n)} - y_t^{(1)} = \ln(P_{t+1}^{(n-1)}/P_t^{(n)}) - y_t^{(1)}$ is the excess bond return between time t and $t+1$. It is then plausible to envision a situation when a hidden factor has a non-zero equal but opposite effect on both expectational terms of the right-hand side while passing undetected through the cross-section of yields at time t. For more rigorous discussion of this, see Duffee (2011). Joslin, Priebsch and Singleton (2014) propose a new framework for estimating dynamic term structure model with such hidden (unspanned) macro factors.

In this application, we explore possible nonlinearities in the predictive relationship between long-term (with average maturity of 20 years) and medium-term (with average maturity of 5 years) bond returns and two popular predictors: the Cochrane-Piazzesi factor (cp_t) and S&P500 stock returns (sp_t). The Cochrane-Piazzesi factor is constructed from the following monthly predictive regression (Cochrane and Piazzesi, 2005; Duffee, 2013):

$$\overline{rx}_{t,t+1} = F_t'\gamma + \varepsilon_{t+1},$$

where $\overline{rx}_{t,t+1} = \frac{1}{4}\sum_{n=2}^{5} r_{t,t+1}^{(n)} - y_t^{(1)}$ are excess returns for a portfolio of bonds with 2-, 3-, 4- and 5-year maturities, $F_t = [1, y_t^{(1)}, f_t^{(2)}, ..., f_t^{(5)}, f_{t-1}^{(2)}, ..., f_{t-1}^{(5)}, ..., f_{t-11}^{(2)}, ..., f_{t-11}^{(5)}]'$ and $f_t^{(i)} = \ln(P_t^{(i-1)}) - \ln(P_t^{(i)})$ for $i = 2, ..., 5$ is the forward rate at time t for loans between time $t+i-1$ and $t+i$. The Cochrane-Piazzesi factor is then computed as $cp_t = F_t'\hat{\gamma}$, where $\hat{\gamma}$ is the OLS estimate from the above regression.

The yield data, used for constructing the Cochrane-Piazzesi factor, is obtained from the U.S. Treasury yield curve of Gürkaynak, Sack and Wright (2007), maintained by the Federal Reserve Board.[8] The data for bond returns and S&P500 returns is from Ibbotson SBBI 2014 yearbook. We use returns on long-term government bonds (with an approximate maturity of 20 years) and intermediate-term government bonds (with an approximate maturity of 5 years), which are denoted by r_1 (LT) and r_2 (IT), respectively. The data are monthly observations covering the period January 1953 – December 2013.

The predictive regressions for bond returns are typically linear in the predictors and are estimated separately for each maturity. The decomposition method allows for possible nonlinearities while the multivariate version of the method exploits possible dependencies between long- and intermediate-term bond returns and their components.

4.2 Dynamic specifications and empirical results

In all specifications, we employ only one lag of the explanatory variables; extension to a higher-lag versions are obvious, though they may lead to poorer precision of asymptotic inference because of possible overparameterization. When the data are abundant and allow this, the lag orders can be selected via information criteria such as BIC.

As a benchmark, we use the linear univariate and bivariate models. The univariate version of our benchmark linear models is

$$r_{i,t} = \omega_{\ell i} + \alpha_{\ell i} r_{i,t-1} + \xi_{\ell i} sp_{t-1} + \zeta_{\ell i} cp_{t-1}$$

for $i = 1, 2$, and the multivariate version is

$$r_{i,t} = \omega_{\ell i} + \alpha_{\ell ii} r_{i,t-1} + \alpha_{\ell ij} r_{j,t-1} + \xi_{\ell i} sp_{t-1} + \zeta_{\ell i} cp_{t-1}$$

for $i, j = 1, 2$. The estimation is performed via univariate and bivariate Gaussian QML, respectively. The estimation results for the univariate and bivariate model are presented in Table 2. For both (univariate and bivariate) versions, the external predictors have strong predictive power: past stock returns have a negative effect on bond returns, in line with the 'great rotation' hypothesis between stocks and bonds, and the Cochrane-Piazzesi factor tends to increase future bond returns. In all cases, long-term bonds appear to react more strongly to changes in the predictors. There is also a strong cross-effect of lagged IT bond returns on LT bond returns but not vice versa. As expected, the residuals of the two equations are highly positively correlated with a correlation coefficient of $\rho = 0.82$.

[8] Available at http://www.federalreserve.gov/Pubs/feds/2006/200628/200628abs.html.

Along with the bivariate decomposition model, we also estimate the univariate decomposition models for both variables separately. In addition, we estimate bivariate stand-alone models for absolute values only (bivariate MEM) and directions only (bivariate probit). Finally, the univariate decomposition model combines the univariate volatility and directions submodels. The bivariate decomposition model combines the bivariate volatility and directions submodels, or, from the other perspective, it combines the two univariate decomposition models.

The specifications for the latent processes $\varphi_{i,t}$ and $\theta_{i,t}$ are the same in these models as long as they model the same number of variables. The conditional mean in the bivariate model for absolute returns ('volatility submodel') is specified as

$$\ln \varphi_{i,t} = \omega_{vi} + \beta_{vi} \ln \varphi_{i,t-1} + \alpha_{vi} \ln |r_{i,t-1}| + \gamma_{vi} I_{i,t-1} + \xi_v sp_{t-1} + \zeta_v cp_{t-1}$$

for $i = 1, 2$. The individual log-likelihood for the univariate volatility submodel for variable i is given by

$$\ell_{vi} = \sum_{t=1}^{T} \log f^{\mathcal{W}(\varphi_{i,t},\varsigma_i)}(|r_{i,t}|)$$

for $i = 1, 2$. This is maximized separately for each individual volatility model (or jointly, which is equivalent). The conditional means in the bivariate model for absolute returns are specified as

$$\ln \varphi_{i,t} = \omega_{vi} + \beta_{vi} \ln \varphi_{i,t-1} + \alpha_{vii} \ln |r_{i,t-1}| + \gamma_{vii} I_{i,t-1} + \alpha_{vij} \ln |r_{j,t-1}| + \gamma_{vij} I_{j,t-1} + \xi_{vi} sp_{t-1} + \zeta_{vi} cp_{t-1}$$

for $i = 1, 2$. The joint log-likelihood for the bivariate volatility submodel is

$$\ell_v = \ell_{v1} + \ell_{v2} + \sum_{t=1}^{T} \log c(F^{\mathcal{W}(\varphi_{1,t},\varsigma_1)}(|r_{1,t}|), F^{\mathcal{W}(\varphi_{2,t},\varsigma_2)}(|r_{2,t}|), \varrho_v).$$

An additional parameter involved is the degree of conditional dependence ϱ_v between the two absolute values. The construction of the excess dispersion test that tests for adequacy of Weibull marginals is described in Anatolyev and Gospodinov (2010). The estimation results for the univariate and bivariate (standalone) volatility submodels are reported in the middle part of Table 3, and those for the univariate (bivariate) decomposition models are presented in the left (right) panel of Table 3. The results indicate that both volatility processes are persistent. For LT, past positive (negative) returns cause lower (higher) current volatility. As in the linear model, the external predictors have a significant effect on volatility but both of these effects are now negative. Another difference with the linear models of the conditional mean is that the cross-effects (of absolute returns and direction) are now from LT to IT. The two volatility processes are moderately strongly dependent with $\varrho_v = 0.6$.

The latent variables that determine the conditional success probabilities in a univariate probit model for directions ('direction submodel') is given by

$$\theta_{i,t} = \omega_{di} + \phi_{di} I_{i,t-1} + \xi_d sp_{t-1} + \zeta_d cp_{t-1}$$

for $i = 1, 2$. The individual log-likelihood for the univariate direction submodel for variable i are given by

$$\ell_{di} = \sum_{t=1}^{T} I_{i,t} \log p_{1i,t} + (1 - I_{i,t}) \log p_{0i,t}.$$

The latent variables that determine the conditional success probabilities in a univariate probit model for directions is given by

$$\theta_{i,t} = \omega_{di} + \phi_{dii} I_{i,t-1} + \phi_{dij} I_{j,t-1} + \xi_{di} sp_{t-1} + \zeta_{di} cp_{t-1}$$

for $i = 1, 2$. The joint log-likelihood for the bivariate direction submodel is

$$\ell_d = \sum_{t=1}^{T} I_{1,t} I_{2,t} \log p_{11,t} + I_{1,t}(1 - I_{2,t}) \log p_{10,t} + (1 - I_{1,t}) I_{2,t} \log p_{01,t} + (1 - I_{1,t})(1 - I_{2,t}) \log p_{00,t}.$$

The degree of conditional dependence between the two directions is characterized by additional parameter ϱ_d. The estimation results for the stand-alone direction submodels, as well as the univariate and bivariate decomposition models, are reported in Table 4. The direction model for LT exhibits positive persistence. Furthermore, the past direction of LT returns affects positively the direction of IT returns. The lagged stock returns and the Cochrane-Piazzesi factor again have a significant (negative and positive, respectively) effect on the direction of bond returns. The dependence of the directional components is strong with $\varrho_d = 0.84$.

The estimates of the dependence matrix R for the decomposition model are collected in Table 5. The parameters for the univariate decomposition models are the degree of component dependence ϱ_i for variable $i = 1, 2$. For the bivariate decomposition model, the two new elements are ϱ_{dv} and ϱ_{vd}. The former indicates dependence between direction of LT and volatility of IT. It is moderately large and highly significant. The latter indicates dependence between direction of IT and volatility of LT. It is close to zero and statistically insignificant. The positive and highly significant dependence between volatility of shorter-term bonds and the direction of longer-term bond returns is interesting. It suggests that the volatility of the short-term bond returns appears to be a priced risk that may not be necessarily revealed in a linear specification.

4.3 Model comparison and prediction

Our model comparison includes univariate and bivariate linear models, bivariate stand-alone (direction and volatility) models and univariate and bivariate decomposition models. Table 6 reports the values of the log-likelihood and Bayesian information criterion (BIC) for different models. The linear model is dominated by the other models. The bivariate decomposition model performs the best despite the large number of estimated parameters. This is also confirmed by conducting of a likelihood ratio test in the bivariate decomposition model with restrictions imposed by the nested standalone and univariate decomposition models. In both cases, the restrictions are strongly rejected.[9] Also, it is interesting to note that the pair of standalone bivariate volatility and direction models outperforms the pair of univariate decomposition models. This finding can be attributed to the fact that the two series are strongly dependent 'in dynamics' and much less 'in multiplicative components'.

We construct return predictions from the decomposition model as described in subsection 2.4. The actual and predicted returns from the decomposition model are plotted in Figure 4. Table 7 provides information on the quality of the model predictions measured by the pseudo-R^2 for returns and for absolute values, as well as fractions of correct sign predictions. In terms of return prediction, the decomposition model tends to generate better predictions than the linear model, with the bivariate version offering noticeable improvements only for the long-term bond returns. In terms of absolute return prediction, the predictions from the decomposition model substantially dominate those from the linear model. In terms of sign prediction, again, the decomposition model dominates, although by a narrower margin. We would like to stress that a valuable feature of a fully specified non-linear model for the components, such as the decomposition model, is its ability to predict any function of these components, in contrast to the linear model which is intrinsically tied to the objective of return mean prediction.

As indicated above, a fully-specified model, such as the bivariate decomposition model, can be used to derive the dynamics of any moments and co-moments of the predictive distribution of returns. Figure 5 plots the conditional variances of r_1 and r_2 as well as their conditional correlation. The conditional variances are characterized by sharp rises in the early 70s, early 80s and during the recent financial crisis. While the conditional correlation is large and stable, it also exhibits sharp movements over the business cycle.

4.4 Robustness analysis

We estimate the bivariate decomposition model with some deviations from baseline specifications to verify the robustness of the results obtained. The first analysis concerns the marginal distributions of the absolute value components. Recall that they are specified as conditionally Weibull, $|r_t||\mathcal{F}_{t-1} \sim \mathcal{W}(\varphi_t, \varsigma)$, and that the shape parameter ς is estimated to be statistically

[9] The values of the LR statistic are $2 \cdot 732 \cdot (5.8085 - 5.7676) \approx 60$ and $2 \cdot 732 \cdot (5.8085 - 5.3232) \approx 710$ which are far larger than conventional critical values of the χ^2 distribuition with 4 and 10 degrees of freedom, respectively.

significantly larger than unity (the unity would imply conditional exponentiality) for both types of bonds; at the same time, the excess dispersion test does not reveal deviations from conditional Weibullianity. However, we also fit a more general Burr distribution (e.g., Grammig and Maurer, 2000) having Weibull as a limiting case. The PDF and CDF are, respectively,

$$\begin{aligned} f(\eta;\varsigma,\varrho) &= \frac{\varsigma}{\chi^{\varsigma}}\eta^{\varsigma-1}\left(1+\varrho\left(\frac{\eta}{\chi}\right)^{\varsigma}\right)^{-1-\varrho^{-1}}, \\ F(\eta;\varsigma,\varrho) &= 1-\left(1+\varrho\left(\frac{\eta}{\chi}\right)^{\varsigma}\right)^{-\varrho^{-1}}, \end{aligned}$$

where

$$\chi=\frac{\Gamma\left(1+\varrho^{-1}\right)\varrho^{1+\varsigma^{-1}}}{\Gamma\left(1+\varsigma^{-1}\right)\Gamma\left(\varrho^{-1}-\varsigma^{-1}\right)},\quad \varsigma>\varrho>0.$$

When $\varrho\to 0$, the Burr distribution reduces to Weibull. When the bivariate decomposition model is estimated with Burr absolute value marginals, the estimated parameter ϱ^{-1} tends to take very large values, with the mean log-likelihood not reaching the Weibull-implied magnitude of 5.8085. This confirms that conditional Weibullianity is flexible enough to describe the conditional marginal distributions of the absolute value components.

Next, we deviate from the conditional probit model for the indicators and fit the logit model

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

instead. Note that the probit function is conformable with the normal copula specification while a use of logit in this context appears clumsy. When the bivariate decomposition model is estimated with logit indicator marginals, the estimated parameters in the evolution of θ_t stay similar; with the mean loglikelihood infinitesimally going up from 5.80851 to 5.80866. This increase is statistically insignificant according to the Voung (1989) test; the Voung test statistic equals 0.62 which corresponds to the p-value of 0.54.

We also fit the dependence structure of the bivariate decomposition model with a non-symmetric copula. We replace the normal copula by the skewed normal copula implied by the so-called skewed normal density (Azzalini, 1985; Azzalini and Dalla Valle, 1996). The univariate skewed normal density (Azzalini, 1985) is

$$\psi(x)=2\phi(x)\,\Phi\left(\frac{\delta x}{\sqrt{1-\delta^2}}\right),$$

where the parameter δ indexes asymmetry; the density reduces to $\phi(x)$ when $\delta=0$. The multivariate version of the skewed normal density and how to construct a corresponding copula density are described in Appendix A.2. When the bivariate decomposition model is estimated with the skewed normal copula, the estimated asymmetry parameters of both absolute value and direction components, δ_v and δ_d, are very close to zero, with the mean loglikelihood staying practically the same at the magnitude of 5.80851 implied by the normal copula.

Finally, we verify optimality of using one lag of explanatory variables in evolution specifications for $\ln \varphi_{i,t}$ and $\theta_{i,t}$. In order to avoid severe parameterization, we separately append those with second lags of both endogenous and exogenous explanatory variables $\ln |r_{i,t-2}|$, $I_{i,t-2}$, $\ln |r_{j,t-2}|$, $I_{j,t-2}$, sp_{t-2}, cp_{t-2} in the case of $\ln \varphi_{i,t}$ (leading to 12 additional parameters) and $I_{i,t-2}$, $I_{j,t-2}$, sp_{t-2}, cp_{t-2} in the case of $\theta_{i,t}$ (leading to 8 additional parameters). In both cases, an increase in the likelihood does not justify higher model complexity: the BIC criterion increases from -8279.39 to -8228.12 in the former case and to -8236.00 in the latter case.

5 Conclusions

This paper is concerned with the development of a multivariate version of the multiplicative decomposition approach of Anatolyev and Gospodinov (2010). A particular attention is paid to the parsimony, tractability and interpretability of this multivariate extension. The marginals for the m absolute values and the binary marginals for the m directions are linked through a $2m$-dimensional flexible copula which is parameterized by $m(2m-1)$ correlation parameters and possibly some shape parameters. The computation of various conditional measures of interest is also discussed. We show how this approach allows one to uncover some important dependencies that remain hidden in the usual analysis of multivariate models.

References

[1] Anatolyev, S., 2009, Multi-market direction-of-change modeling using dependence ratios, *Studies in Nonlinear Dynamics & Econometrics*, 13(1), article 5.

[2] Anatolyev, S., and N. Gospodinov, 2010, Modeling financial return dynamics via decomposition, *Journal of Business and Economic Statistics* 28, 232–245.

[3] Ashford, J. R., and R. R. Sowden, 1970, Multivariate probit analysis, *Biometrics* 26, 535–546.

[4] Azzalini, A., 1985, A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178.

[5] Azzalini, A. and A. Dalla Valle, 1996, The multivariate skew-normal distribution. *Biometrika* 83, 715–726.

[6] Chernov, M., and P. Mueller, 2012, The term structure of inflation expectations, *Journal of Financial Economics* 106, 367–394.

[7] Cieslak, A., and P. Povala, 2015, Expected returns in Treasury bonds, *Review of Financial Studies* 28, 2859–2901.

[8] Cipollini, F., R. F. Engle, and G.M. Gallo, 2017, Copula-based vMEM specifications versus alternatives: the case of trading activity, *Econometrics* 5(2), 16.

[9] Cochrane, J. H., and M. Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.

[10] Duffee, G. R., 2011, Information in (and not in) the term structure, *Review of Financial Studies* 24, 2895–2934.

[11] Duffee, G. R., 2013, Forecasting interest rates, Chapter 7 in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol.2A, Elsevier: Amsterdam, 385–426.

[12] Engle, R. F., 2002, New Frontiers for ARCH Models, *Journal of Applied Econometrics* 17, 425–446.

[13] Grammig, J., and K.-O. Maurer, 2000, Non-monotonic hazard functions and the autoregressive conditional duration model, *Econometrics Journal* 3, 16–38.

[14] Gürkaynak, R. S., B. Sack, and J. H. Wright, 2007, The U.S. Treasury yield curve: 1961 to present, *Journal of Monetary Economics* 54, 2291–2304.

[15] Joslin, S., M. Priebsch, and K. J. Singleton, 2014, Risk premiums in dynamic term structure models with unspanned macro risks, *Journal of Finance* 69, 1197–1233.

[16] Judd, K., 1998, *Numerical Methods in Economics*, MIT Press.

[17] Liu, X. and R. Luger, 2015, Unfolded GARCH models, *Journal of Economic Dynamics and Control* 58, 186–217.

[18] Ludvigson, S. C., and S. Ng, 2009, Macro factors in bond risk premia, *Review of Financial Studies* 22, 5027–5067.

[19] Nyberg, H., 2014, A bivariate autoregressive probit model: Business cycle linkages and transmission of recession probabilities. *Macroeconomic Dynamics*, 18, 838–862.

[20] Trivedi, P. K., and Zimmer, D. M., 2005, Copula modeling: an introduction for practitioners, *Foundations and Trends in Econometrics* 1, 1–111.

[21] Vuong, Q. H., 1989, Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica* 57, 307–333.

Appendix

A.1 Proofs

Proof of Proposition 1. We will suppress the time index throughout. Anatolyev and Gospodinov (2010) derive the multivariate structure of the density as in Proposition 1. What is left is to compute the distorted success probability in the case of Gaussian copula. Because

$$\begin{aligned}\frac{\partial \Phi_2(x_1,x_2)}{\partial x_1} &= \frac{\partial}{\partial x_1}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\phi_2(t_1,t_2)dt_1dt_2 \\ &= \int_{-\infty}^{x_2}\left(\frac{\partial}{\partial x_1}\int_{-\infty}^{x_1}\phi_2(t_1,t_2)dt_1\right)dt_2 \\ &= \int_{-\infty}^{x_2}\phi_2(x_1,t_2)dt_2 \\ &= \phi(x_1)\int_{-\infty}^{x_2}\phi(t_2|x_1)dt_2 \\ &= \phi(x_1)\,\Phi(x_2|x_1)\end{aligned}$$

and hence

$$\begin{aligned}\frac{\partial C(w,y)}{\partial w} &= \frac{\partial \Phi_2(\Phi^{-1}(w),\Phi^{-1}(y),\varrho)}{\partial w} = \left.\frac{\partial \Phi_2(x_1,x_2)}{\partial x_1}\right|_{x_1=\Phi^{-1}(w),x_2=\Phi^{-1}(y)}\frac{\partial \Phi^{-1}(w)}{\partial w} \\ &= \left.\phi(x_1)\,\Phi(x_2|x_1,\varrho)\right|_{x_1=\Phi^{-1}(w),x_2=\Phi^{-1}(y)}\left.\frac{1}{\phi(x_1)}\right|_{x_1=\Phi^{-1}(w)} = \Phi\left(\Phi^{-1}(y)|\Phi^{-1}(w),\varrho\right) \\ &= \Phi\left(\frac{\Phi^{-1}(y)-\varrho\Phi^{-1}(w)}{\sqrt{1-\varrho^2}}\right)\end{aligned}$$

we have, from Anatolyev and Gospodinov (2010), the distorted success probability is

$$p^C = 1-\left.\frac{\partial C(w,y)}{\partial w}\right|_{w=F(u),y=1-\Phi(\theta)} = \Phi\left(\frac{\theta+\varrho\Phi^{-1}(F(u))}{\sqrt{1-\varrho^2}}\right).$$

□

Lemma A1. For any $2m$-variate copula $C(w_1,...,w_m,y_1,...,y_m)$ from the class considered, the following holds:

$$\frac{\partial^m C(w_1,...,w_m,y_1,...,y_m)}{\partial w_1...\partial w_m} = c(w_1,...,w_m)\Psi_m(\Psi^{-1}(y_1),...,\Psi^{-1}(y_m)|\Psi^{-1}(w_1),...,\Psi^{-1}(w_m)),$$

where $c(w_1,...,w_m)$ is the density of the m-variate copula $C(w_1,...,w_m)$ implied by the $2m$-variate copula $C(w_1,...,w_m,y_1,...,y_m)$:

$$c(w_1,...,w_m) = \frac{\psi_m(\Psi^{-1}(w_1),...,\Psi^{-1}(w_m))}{\psi\left(\Psi^{-1}(w_1)\right)...\psi\left(\Psi^{-1}(w_m)\right)}.$$

Proof of Lemma A1. Note the following property:

$$
\begin{aligned}
&\frac{\partial^m \Psi_{2m}(x_1,...,x_m,x_{m+1},...,x_{2m})}{\partial x_1 ... \partial x_m} \\
&= \frac{\partial}{\partial x_1}...\frac{\partial}{\partial x_m}\int_{-\infty}^{x_1}...\int_{-\infty}^{x_m}\int_{-\infty}^{x_{m+1}}...\int_{-\infty}^{x_{2m}}\psi_{2m}(t_1,...,t_m,t_{m+1},...,t_{2m})dt_1...dt_m dt_{m+1}...dt_{2m} \\
&= \int_{-\infty}^{x_{m+1}}...\int_{-\infty}^{x_{2m}}\left(\frac{\partial}{\partial x_1}...\frac{\partial}{\partial x_m}\int_{-\infty}^{x_1}...\int_{-\infty}^{x_m}\psi_{2m}(t_1,...,t_m,t_{m+1},...,t_{2m})dt_1...dt_m\right)dt_{m+1}...dt_{2m} \\
&= \int_{-\infty}^{x_{m+1}}...\int_{-\infty}^{x_{2m}}\psi_{2m}(x_1,...,x_m,t_{m+1},...,t_{2m})dt_{m+1}...dt_{2m} \\
&= \int_{-\infty}^{x_{m+1}}...\int_{-\infty}^{x_{2m}}\psi_{2m}(t_{m+1},...,t_{2m}|x_1,...,x_m)\psi_m(x_1,...,x_m)dt_{m+1}...dt_{2m} \\
&= \psi_m(x_1,...,x_m)\int_{-\infty}^{x_{m+1}}...\int_{-\infty}^{x_{2m}}\psi_{2m}(t_{m+1},...,t_{2m}|x_1,...,x_m)dt_{m+1}...dt_{2m} \\
&= \psi_m(x_1,...,x_m)\Psi_m(x_{m+1},...,x_{2m}|x_1,...,x_m).
\end{aligned}
$$

This leads to

$$
\begin{aligned}
&\frac{\partial^m C(w_1,...,w_m,y_1,...,y_m)}{\partial w_1...\partial w_m} \\
&= \frac{\partial^m \Psi_{2m}(\Psi^{-1}(w_1),...,\Psi^{-1}(w_m),\Psi^{-1}(y_1),...,\Psi^{-1}(y_m))}{\partial w_1...\partial w_m} \\
&= \left.\frac{\partial^m \Psi_{2m}(x_1,...,x_m,x_{m+1},...,x_{2m})}{\partial x_1...\partial x_m}\right|_{x_1=\Psi^{-1}(w_1),...,x_m=\Psi^{-1}(w_m),x_{m+1}=\Psi^{-1}(y_1),...,x_{2m}=\Psi^{-1}(y_m)} \\
&\quad \cdot\frac{\partial\Psi^{-1}(w_1)}{\partial w_1}...\frac{\partial\Psi^{-1}(w_m)}{\partial w_m} \\
&= \psi_m(x_1,...,x_m)\Psi_m(x_{m+1},...,x_{2m}|x_1,...,x_m)|_{x_1=\Psi^{-1}(w_1),...,x_m=\Psi^{-1}(w_m),x_{m+1}=\Psi^{-1}(y_1),...,x_{2m}=\Psi^{-1}(y_m)} \\
&\quad \cdot\left.\frac{1}{\psi(x_1)}\right|_{x_1=\Psi^{-1}(w_1)}...\left.\frac{1}{\psi(x_{k_u})}\right|_{x_m=\Psi^{-1}(w_m)} \\
&= \frac{\psi_m(\Psi^{-1}(w_1),...,\Psi^{-1}(w_m))}{\psi(\Psi^{-1}(w_1))...\psi(\Psi^{-1}(w_m))}\Psi_m(\Psi^{-1}(y_1),...,\Psi^{-1}(y_m)|\Psi^{-1}(w_1),...,\Psi^{-1}(w_m)) \\
&= c(w_1,...,w_m)\Psi_m(\Psi^{-1}(y_1),...,\Psi^{-1}(y_m)|\Psi^{-1}(w_1),...,\Psi^{-1}(w_m)),
\end{aligned}
$$

as asserted. □

Proof of Proposition 2. We will suppress the time index throughout. Denote the marginal (Bernoulli) CDFs of direction components by $G_1(v_1),...,G_m(v_m)$ and their marginal success probabilities by $p_1,...,p_m$. The joint CDF of the $2m$-tuple $(u_1,...,u_m,v_1,...,v_m)$ is

$$F(u_1,...,u_m,v_1,...,v_m) = C(F_1(u_1),...,F_m(u_m),G_1(v_1),...,G_m(v_m)).$$

The joint PDF/PMF is derived by taking the second derivative with respect to the m continuous components and second-order difference with respect to the m discrete components. Then, using the properties of finite differences,

$$\begin{aligned} f(u_1, ..., u_m, v_1, ..., v_m) &= \sum_{\ell_1 \in \{0,1\}} ... \sum_{\ell_m \in \{0,1\}} (-1)^{\ell_1 + ... + \ell_m} \times \\ & \quad \frac{\partial^m C}{\partial u_1 ... \partial u_m} (F_1(u_1), ..., F_m(u_m), G_1(v_1 - \ell_1), ..., G_m(v_m - \ell_m)) \\ &= f_1(u_1)...f_m(u_m) f_{\partial^m}(u_1, ..., u_{k_u}, v_1, ..., v_m), \end{aligned}$$

where the last term $f_{\partial^m}(u_1, ..., u_{k_u}, v_1, ..., v_m)$ equals

$$\begin{aligned} & \sum_{\ell_1 \in \{0,1\}} ... \sum_{\ell_m \in \{0,1\}} (-1)^{\ell_1 + ... + \ell_m} \times \\ & \quad \frac{\partial^m C}{\partial w_1 ... \partial w_m} (w_1, ..., w_m, G_1(v_1 - \ell_1), ..., G_m(v_m - \ell_m)) \Big|_{w_1 = F_1(u_1), ..., w_m = F_m(u_m)} \\ = \; & c(F_1(u_1), ..., F_m(u_m)) \sum_{\ell_1 \in \{0,1\}} ... \sum_{\ell_m \in \{0,1\}} (-1)^{\ell_1 + ... + \ell_m} \pi_{\ell_1 ... \ell_m}(v_1, ..., v_m, u_1, ..., u_m), \end{aligned}$$

using Lemma A1. Collecting the pieces,

$$f(u_1, ..., u_m, v_1, ..., v_m) = f_u(u_1, ..., u_m) f^C(v_1, ..., v_m, u_1, ..., u_m),$$

where

$$f_u(u_1, ..., u_m) = f_1(u_1)...f_m(u_m) c(F_1(u_1), ..., F_2(u_m))$$

is the joint density of all 'volatilities' modelled through the m-variate copula $C(w_1, ..., w_m)$, and $f^C(v_1, ..., v_m)$ is distorted PMF of the m-variate Bernoulli random variable representing all 'directions':

$$f^C(v_1, ..., v_m, u_1, ..., u_m) = \sum_{\ell_1 \in \{0,1\}} ... \sum_{\ell_m \in \{0,1\}} (-1)^{\ell_1 + ... + \ell_m} \pi_{\ell_1 ... \ell_m}(v_1, ..., v_m, u_1, ..., u_m),$$

where each $v_l \in \{0, 1\}$, $l = 1, ..., m$. □

Proof of Corollary 3. When $m = 2$,

$$\begin{aligned} f^C(v_1, v_2, u_1, u_2) &= \Psi_2(\Psi^{-1}(G_1(v_1)), \Psi^{-1}(G_2(v_2)) | \Psi^{-1}(F_1(u_1)), \Psi^{-1}(F_2(u_2))) \\ & \quad - \Psi_2(\Psi^{-1}(G_1(v_1 - 1)), \Psi^{-1}(G_2(v_2)) | \Psi^{-1}(F_1(u_1)), \Psi^{-1}(F_2(u_2))) \\ & \quad - \Psi_2(\Psi^{-1}(G_1(v_1)), \Psi^{-1}(G_2(v_2 - 1)) | \Psi^{-1}(F_1(u_1)), \Psi^{-1}(F_2(u_2))) \\ & \quad + \Psi_2(\Psi^{-1}(G_1(v_1 - 1)), \Psi^{-1}(G_2(v_2 - 1)) | \Psi^{-1}(F_1(u_1)), \Psi^{-1}(F_2(u_2))). \end{aligned}$$

Considering the four points of support for $(v_1, v_2) \in \{0,1\}^2$ and using that $\Psi_2(y,0) = \Psi_2(0,y) = 0$ for any y and $\Psi_2(\Psi^{-1}(1), \Psi^{-1}(1)) = 1$, we get

$$f^C(v_1, v_2, u_1, u_2) = p_{11}^C(u_1,u_2)^{v_1 v_2} p_{01}^C(u_1,u_2)^{(1-v_1)v_2} p_{10}^C(u_1,u_2)^{v_1(1-v_2)} p_{00}^C(u_1,u_2)^{(1-v_1)(1-v_2)},$$

where $p_{ij}^C(u_1, u_2)$, $i, j \in \{0,1\}$ are as in the statement. □

A.2 Skewed normal copula

Let the parameter vector δ index asymmetries of the components. The $(2m)$-variate skewed normal density (Azzalini and Dalla Valle, 1996) is

$$\psi_{2m}(x) = 2\phi_{2m}(x, \Omega)\,\Phi\left(\alpha' x\right),$$

where $\Omega = \Delta\left(R + \lambda\lambda'\right)\Delta$,

$$\alpha = \frac{\Delta^{-1} R^{-1} \lambda}{\sqrt{1 + \lambda' R^{-1} \lambda}},$$

and $\Delta = \text{diag}\left\{\left(1-\delta_i^2\right)^{1/2}\right\}_{i=1}^{2m}$, $\lambda = \left\|\delta_i\left(1-\delta_i^2\right)^{-1/2}\right\|_{i=1}^{2m}$. When $m = 2$, the conditional CDF is

$$\Psi(v|u) = \int_{-\infty}^{v} \frac{\phi_2(u, x, \omega_v)}{\phi(u)} \frac{\Phi(\alpha_{v,1} u + \alpha_{v,2} x)}{\Phi(\delta_u\left(1-\delta_u^2\right)^{-1/2} u)} dx.$$

The corresponding bivariate copula density is

$$\begin{aligned} c_{SN}(w_1, w_2) &= \frac{1}{2} \frac{\phi_2(x_1, x_2, \omega_v)\,\Phi(\alpha_{v,1} x_1 + \alpha_{v,2} x_2)}{\phi(x_1)\,\phi(x_2)\,\Phi(\lambda_1 x_1)\,\Phi(\lambda_2 x_2)}\bigg|_{x_1 = \Psi^{-1}(w_1), x_2 = \Psi^{-1}(w_2)} \\ &= \frac{\phi_2\left(\Psi^{-1}(w_1), \Psi^{-1}(w_2), \omega_v\right)}{\phi\left(\Psi^{-1}(w_1)\right)\phi\left(\Psi^{-1}(w_2)\right)} \frac{\Phi\left(\alpha_{v,1}\Psi^{-1}(w_1) + \alpha_{v,2}\Psi^{-1}(w_2)\right)}{2\Phi\left(\lambda_1 \Psi^{-1}(w_1)\right)\Phi\left(\lambda_2 \Psi^{-1}(w_2)\right)}. \end{aligned}$$

Note that the first term is related to the normal copula density, and the second term is a correction for asymmetry. Finally, the conditional CDF needed to compute $\pi.(u_1, u_2)$, is

$$\Psi_2(v_1, v_2|u_1, u_2) = \int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\phi_4(u_1, u_2, x_1, x_2, \Omega)}{\phi_2(u_1, u_2, \omega_v)} \frac{\Phi(\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 x_1 + \alpha_4 x_2)}{\Phi(\alpha_{v,1} u_1 + \alpha_{v,2} u_2)} dx_1 dx_2.$$

Table 1. Simulation results for the bivariate decomposition model.

		symm, $n = 500$			symm, $n = 2000$			asymm, $n = 2000$		
parameter	true value	mean	stdev	t,%	mean	stdev	t,%	mean	stdev	t,%
ω_{v1}	0	−0.018	0.065	4.9	−0.004	0.029	4.8	−0.003	0.024	4.6
β_{v1}	0.8	0.800	0.032	6.8	0.800	0.015	6.4	0.800	0.015	6.2
α_{v11}	0.1	0.095	0.024	7.8	0.099	0.011	5.1	0.099	0.011	5.6
γ_{v11}	−0.3	−0.302	0.054	7.4	−0.301	0.027	6.2	−0.301	0.021	4.2
α_{v12}	0.05	0.049	0.017	7.5	0.050	0.008	4.7	0.050	0.009	4.8
γ_{v12}	0.2	0.203	0.052	7.2	0.200	0.025	5.5	0.200	0.020	4.8
ς_1	1.2	1.208	0.041	4.2	1.202	0.021	5.3			
	1.5							1.502	0.026	5.8
ω_{v2}	0	−0.017	0.064	5.1	−0.003	0.028	5.4	−0.003	0.014	6.8
β_{v2}	0.8	0.801	0.031	8.1	0.801	0.014	5.3	0.800	0.013	7.0
α_{v21}	0.05	0.050	0.017	6.9	0.050	0.008	5.6	0.050	0.004	5.6
γ_{v21}	0.2	0.202	0.053	6.3	0.201	0.025	5.6	0.200	0.012	4.2
α_{v22}	0.1	0.094	0.024	8.4	0.099	0.011	5.7	0.100	0.011	7.4
γ_{v22}	−0.3	−0.303	0.054	6.5	−0.301	0.025	4.7	−0.301	0.012	4.2
ς_2	1.2	1.207	0.041	4.0	1.203	0.020	5.1			
	2.5							1.503	0.026	6.0
ω_{d1}	0.3	0.308	0.117	5.7	0.302	0.056	4.7	0.301	0.056	4.4
ϕ_{d11}	0.3	0.294	0.133	5.0	0.299	0.063	3.5	0.298	0.063	3.0
ϕ_{d12}	−0.1	−0.106	0.136	5.2	−0.103	0.067	5.7	−0.99	0.068	6.2
ω_{d2}	0.3	0.309	0.118	6.3	0.302	0.055	5.3	0.303	0.056	5.8
ϕ_{d21}	−0.1	−0.103	0.134	5.0	−0.101	0.066	5.2	−0.101	0.065	4.6
ϕ_{d22}	0.3	0.296	0.132	4.4	0.297	0.064	4.1	0.299	0.066	3.0
ϱ_v	0.6	0.601	0.029	6.6	0.599	0.014	4.3	0.599	0.014	4.4
ϱ_d	0.6	0.602	0.055	5.8	0.600	0.029	6.6	0.599	0.028	6.4
ϱ_1	0.2	0.204	0.056	5.1	0.202	0.029	5.2	0.199	0.028	4.4
ϱ_2	0.2	0.202	0.057	5.0	0.202	0.028	4.7	0.200	0.029	5.6
ϱ_{vd}	0.2	0.202	0.057	4.6	0.201	0.028	5.8	0.200	0.028	5.8
ϱ_{dv}	0.2	0.203	0.056	5.0	0.202	0.029	6.2	0.200	0.029	6.4

Notes: The tables present the Monte Carlo mean estimate (mean), its standard deviation (stdev) and the empirical size of the t-test (t) for each individual parameter at the 5% significance level. The first block contains parameters of volatility equations, the second block contains parameters of direction equations, the third block contains component dependence parameters. The number of Monte Carlo replications is 1,000 and the sample sizes are $n = 500$ and $n = 2000$; 'symm' and 'asymm' stand for the cases when the shape parameters of absolute value distribution are equal or different, respectively.

Table 2. Estimation results for the univariate and multivariate linear model.

	Univariate linear			Multivariate linear	
parameter	LT_{LT}	IT_{IT}	parameter	LT_{LT+IT}	IT_{LT+IT}
$\omega_{\ell i}$	0.00229 (0.0012)	0.00283 (0.00068)	$\omega_{\ell i}$	0.00120 (0.00111)	0.00287 (0.00064)
$\alpha_{\ell i}$	0.040 (0.054)	0.132 (0.049)	$\alpha_{\ell ii}$	−0.170 (0.105)	0.109 (0.09782)
$\xi_{\ell i}$	−0.0905 (0.0311)	−0.0588 (0.0157)	$\xi_{\ell i}$	−0.0889 (0.0301)	−0.0593 (0.0151)
$\zeta_{\ell i}$	0.438 (0.094)	0.253 (0.051)	$\zeta_{\ell i}$	0.399 (0.095)	0.253 (0.051)
			$\alpha_{\ell ij}$	0.493 (0.179)	0.014 (0.049)
$\sigma_{\ell i}$	0.0272 (0.0011)	0.0139 (0.0007)	$\sigma_{\ell i}$	0.0269 (0.0010)	0.0139 (0.0007)
			ρ	0.819 (0.017)	

Notes: Estimation based on the Gaussian quasi-density. Standard errors are reported in parentheses below the estimate. ρ is the correlation coefficient between the innovations of the two equations.

Table 3. Estimation results for the univariate and multivariate volatility submodels.

	Univariate MEM			Multivariate MEM			
	Decomposition			Volatility only		Decomposition	
parameter	LT_{LT}	IT_{IT}	parameter	LT_{LT+IT}	IT_{LT+IT}	LT_{LT+IT}	IT_{LT+IT}
ω_{vi}	-0.024 (0.035)	-0.095 (0.089)	ω_{vi}	-0.032 (0.052)	-0.388 (0.149)	-0.026 (0.056)	-0.351 (0.136)
β_{vi}	0.886 (0.016)	0.870 (0.029)	β_{vi}	0.877 (0.016)	0.764 (0.042)	0.882 (0.017)	0.783 (0.036)
α_{vi}	0.087 (0.012)	0.083 (0.015)	α_{vii}	0.071 (0.016)	0.070 (0.020)	0.066 (0.016)	0.061 (0.020)
γ_{vi}	-0.079 (0.029)	-0.100 (0.038)	γ_{vii}	-0.120 (0.035)	-0.003 (0.045)	-0.124 (0.035)	0.027 (0.051)
ξ_{vi}	-0.958 (0.326)	-1.100 (0.348)	ξ_{vi}	-0.888 (0.310)	-1.274 (0.440)	-0.851 (0.302)	-1.146 (0.414)
ζ_{vi}	0.642 (0.525)	-0.237 (0.811)	ζ_{vi}	0.485 (0.508)	2.231 (1.260)	0.423 (0.525)	1.877 (1.172)
ς_i	1.207 (0.037)	1.114 (0.033)	ς_i	1.248 (0.037)	1.159 (0.033)	1.249 (0.037)	1.159 (0.033)
			α_{vij}	0.021 (0.014)	0.063 (0.021)	0.022 (0.014)	0.066 (0.020)
			γ_{vij}	0.047 (0.035)	-0.120 (0.045)	0.053 (0.038)	-0.129 (0.046)
			ϱ_v	0.604 (0.024)		0.603 (0.024)	
ED	-0.49	-0.42	ED	-0.80	1.11	0.22	0.49

Notes: Standard errors are reported in parentheses below the estimate. ED is excess dispersion test statistics for validity of conditionally Weibull marginal.

Table 4. Estimation results for the univariate and multivariate direction submodels.

	Univariate probit			Multivariate probit			
	Decomposition			Direction only		Decomposition	
parameter	LT_{LT}	IT_{IT}	parameter	LT_{LT+IT}	IT_{LT+IT}	LT_{LT+IT}	IT_{LT+IT}
ω_{di}	−0.023 (0.078)	0.256 (0.089)	ω_{di}	−0.023 (0.087)	0.176 (0.088)	0.049 (0.082)	0.206 (0.089)
ϕ_{di}	0.156 (0.096)	0.245 (0.103)	ϕ_{dii}	0.137 (0.122)	−0.011 (0.133)	0.135 (0.115)	−0.021 (0.141)
ξ_{di}	−2.53 (1.15)	−4.44 (1.19)	ξ_{di}	−2.43 (1.14)	−4.57 (1.17)	−2.33 (1.123)	−4.36 (1.11)
ζ_{di}	18.66 (4.42)	−10.99 (4.58)	ζ_{di}	18.28 (4.31)	11.59 (4.50)	15.66 (4.30)	9.24 (4.45)
			ϕ_{dij}	0.025 (0.126)	0.445 (0.127)	−0.015 (0.105)	0.434 (0.130)
			ϱ_d	0.845 (0.026)		0.835 (0.026)	

Notes: Standard errors are reported in parentheses below the estimate.

Table 5. Estimates of degrees of dependence from the univariate and multivariate decomposition models.

	Decomposition models	
parameter	Univariate	Bivariate
ϱ_1	0.136 (0.046)	0.123 (0.045)
ϱ_2	0.152 (0.048)	0.203 (0.045)
ϱ_v		0.603 (0.024)
ϱ_d		0.835 (0.026)
ϱ_{vd}		-0.022 (0.047)
ϱ_{dv}		0.300 (0.042)

Notes: Standard errors are reported in parentheses below the estimate.

Table 6. Mean log-likelihoods ℓ and BIC.

Model	Linear			Bivariate standalone		Decomposition		
	Univariate		Bivariate	Volatility	Direction	Univariate		Bivariate
k	6	6	16	19	11	12	12	34
Partial ℓ	2.1840	2.8599		6.8378	−1.0702	2.3337	2.9895	
Total ℓ	5.0439		5.6099	5.7676		5.3232		5.8085
BIC	−7305.1		−8107.4	−8245.9		−7634.9		**−8279.39**

Notes: BIC is computed as $BIC = -2T\ell + k \ln T$.

Table 7. Equation-by-equation quality of fit measures.

	Univariate		Bivariate	
	LT_{LT}	IT_{LT}	LT_{LT+IT}	IT_{LT+IT}
	pseudo-R^2 for returns			
Linear model	5.15%	8.95%	7.21%	8.91%
Decomposition model	5.10%	9.70%	5.34%	8.22%
	pseudo-R^2 for absolute returns			
Linear model	3.67%	8.51%	5.29%	8.33%
Decomposition model	14.38%	12.35%	16.11%	12.39%
	proportion of successful signs			
Linear model	59.3%	67.5%	59.2%	67.8%
Decomposition model	59.7%	68.2%	64.3%	69.5%

Notes: The pseudo-R^2 for returns and pseudo-R^2 for absolute values are computed as

$$\text{pseudo-}R^2_r = \frac{\left(\sum_t \left(\hat{r}_{t|t-1} - \bar{\hat{r}}_{t|t-1}\right)\left(\hat{r}_t - \overline{r}_t\right)\right)^2}{\sum_t \left(\hat{r}_{t|t-1} - \bar{\hat{r}}_{t|t-1}\right)^2 \sum_t \left(\hat{r}_t - \overline{r}_t\right)^2},$$

$$\text{pseudo-}R^2_{|r|} = \frac{\left(\sum_t (\widehat{|r|}_{t|t-1} - \overline{\widehat{|r|}_{t|t-1}})(|\hat{r}_t| - \overline{|r_t|})\right)^2}{\sum_t (\widehat{|r|}_{t|t-1} - \overline{\widehat{|r|}_{t|t-1}})^2 \sum_t (|\hat{r}_t| - \overline{|r_t|})^2},$$

where $\widehat{|r|}_{t|t-1} = |\hat{r}_{t|t-1}|$ for the linear model and $\widehat{|r|}_{t|t-1} = \varphi_t$ for the decomposition model. The proportions of successful sign predictions are computed as

$$S = \frac{1}{T}\sum_t \left(I_t^+ \mathbb{I}_{\{r_t>0\}} + \left(1 - I_t^+\right)\mathbb{I}_{\{r_t\leq 0\}}\right),$$

where $I_t^+ = \mathbb{I}_{\{\hat{r}_{t|t-1}>0\}}$ for the linear model and $I_t^+ = \mathbb{I}_{\{p_t^C>0.5\}}$ for the decomposition model.

Figure 1: Histogram of pseudo-R^2 for returns, for linear and decomposition models in Monte Carlo simulations.

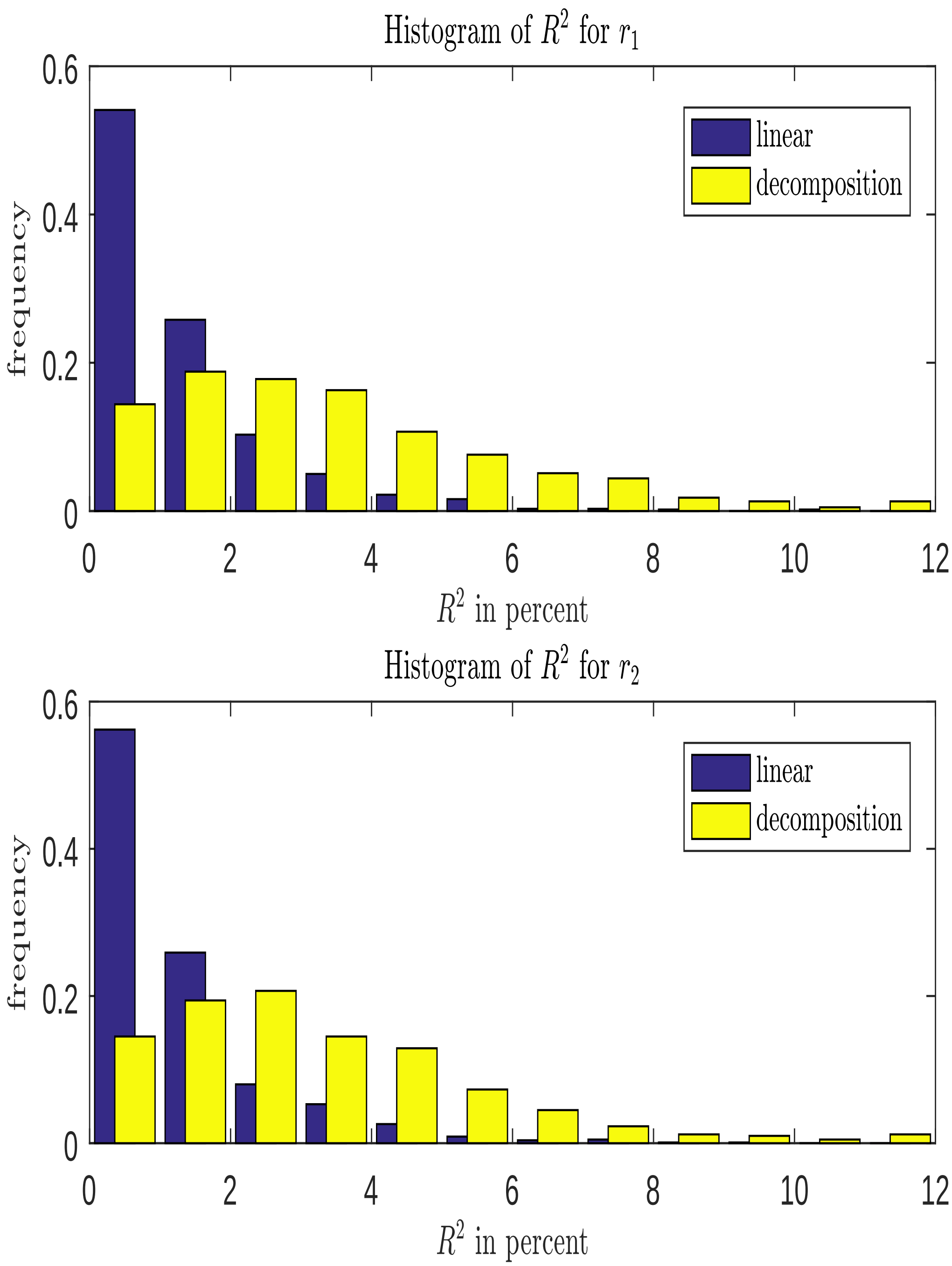


Figure 2: Histogram of pseudo-R^2 for absolute values of returns, for linear and decomposition models in Monte Carlo simulations.

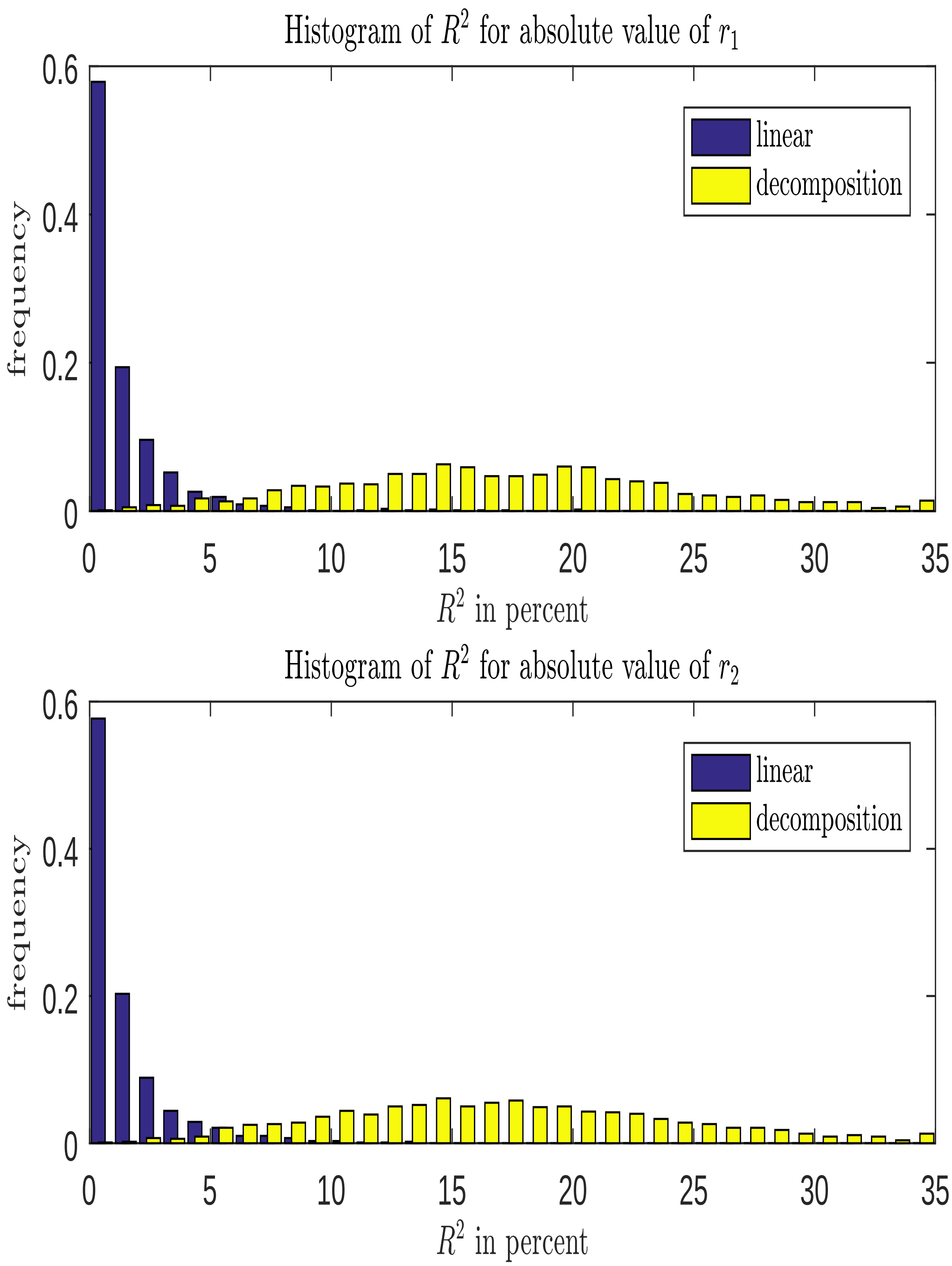


Figure 3: Histogram of successful return sign predictions, for linear and decomposition models in Monte Carlo simulations.

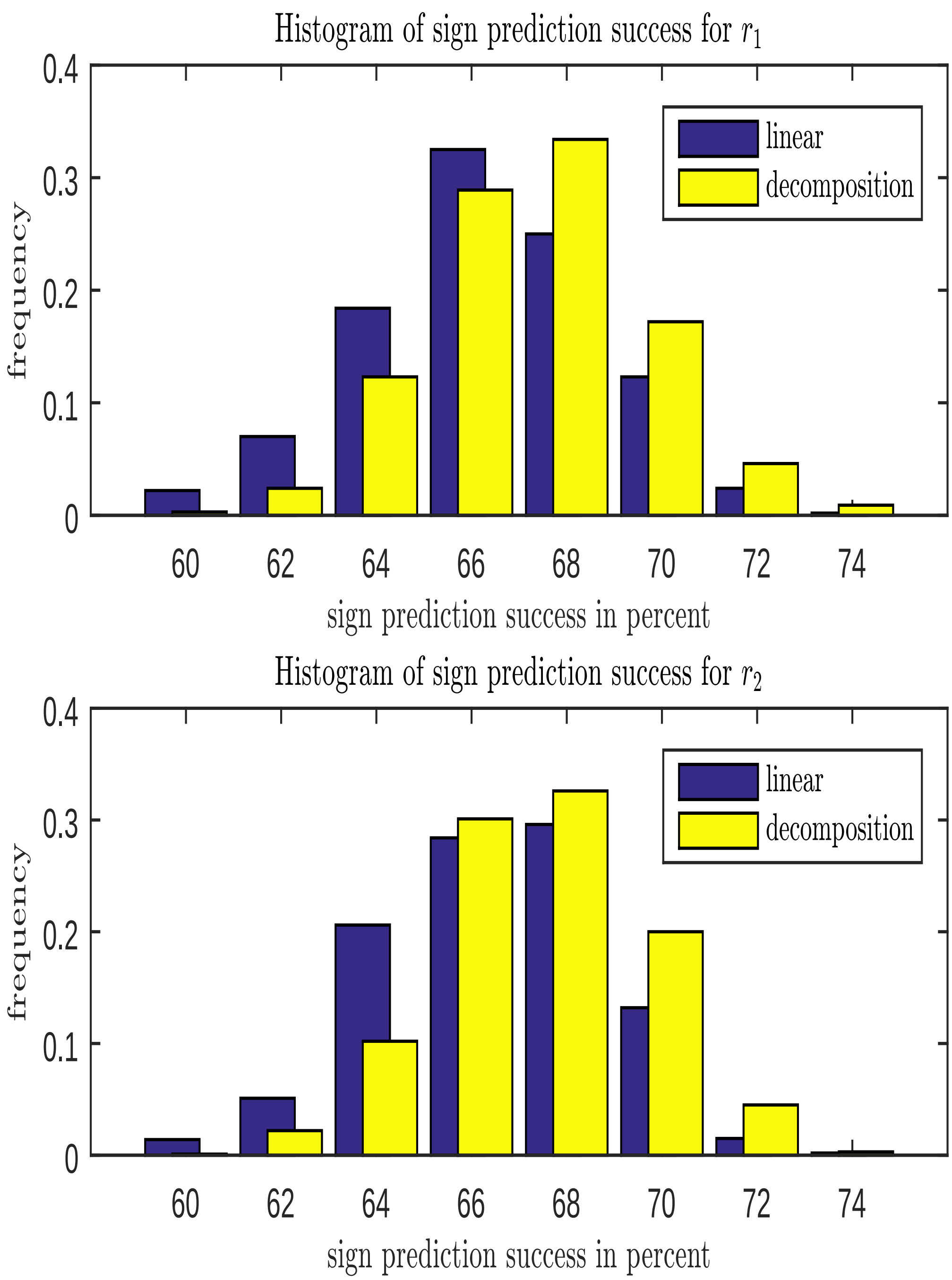
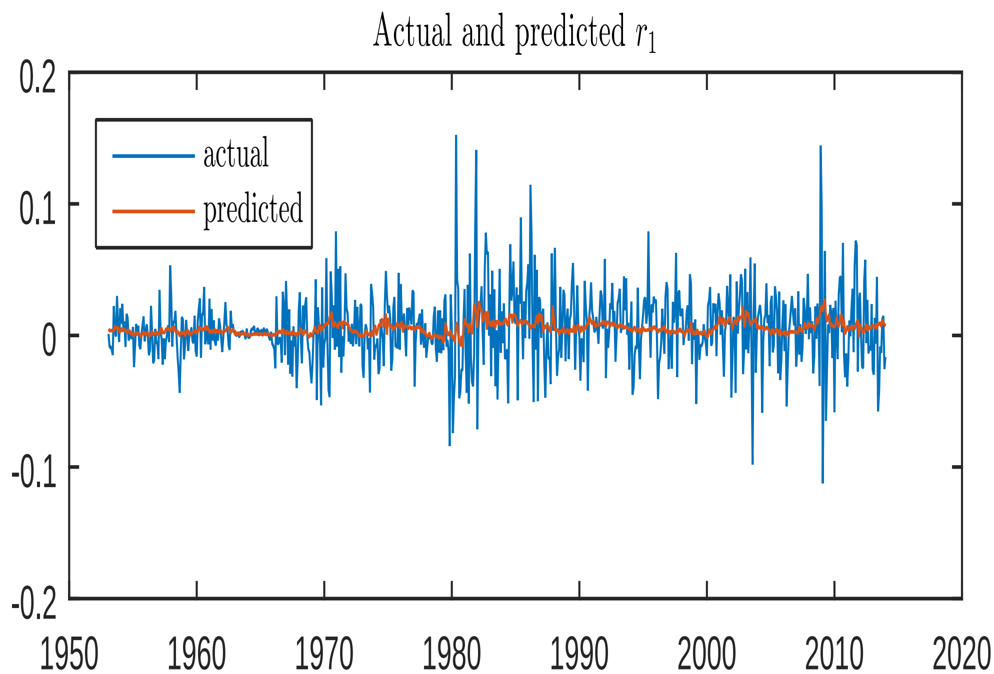


Figure 4: Actual and predicted long-term (r_1) and intermediate-term (r_2) government bond returns.



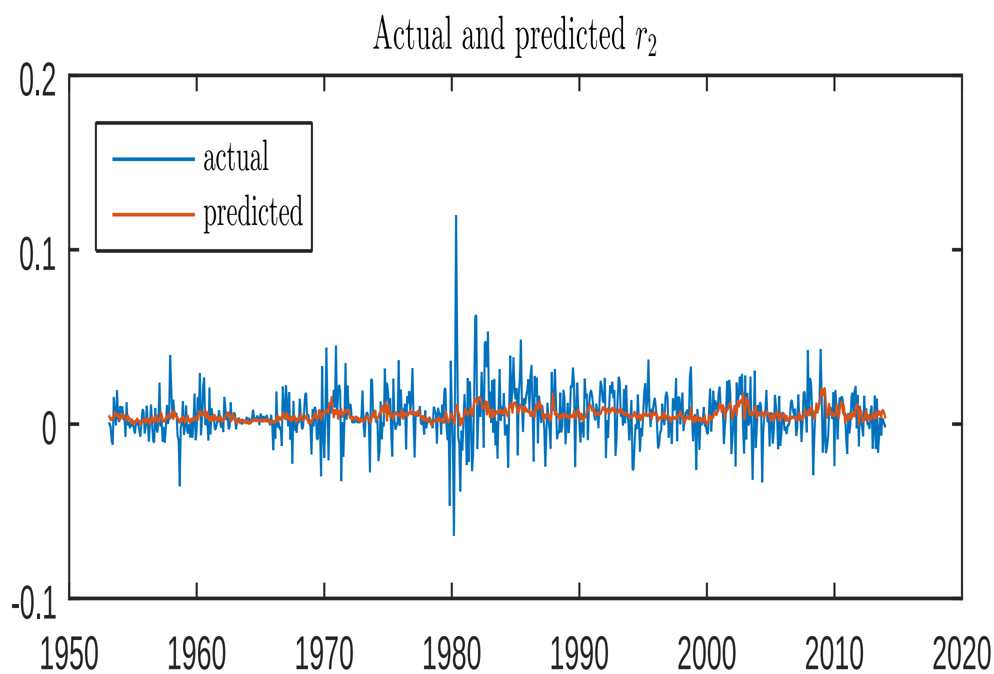


Figure 5: Conditional variances and correlation of long-term (r_1) and intermediate-term (r_2) government bond returns.