Multivariate Return Decomposition: Theory and Implications*

Stanislav Anatolyev† New Economic School

Nikolay Gospodinov‡ Federal Reserve Bank of Atlanta

August 2015

Abstract

In this paper, we propose a model based on multivariate decomposition of multiplicative – absolute values and signs – components of several 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.

*The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

†New Economic School, 100A Novaya Street, Skolkovo, Moscow, 143026, Russia. E-mail: sanatoly@nes.ru.

‡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. 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 Gaussian copula. This choice of copula is prompted by the flexibility and computability of the Gaussian copula whose parsimonious parameterization is readily interpretable. We show how the likelihood function is constructed from the data, how various conditional measures of interest (such as conditional means, variances, covariances and correlations, skewnesses and co-skewnesses, and so on) can be computed, and how the parameter estimates behave in finite samples. Finally, we work out an empirical application to two bond returns of different maturity.

For notational simplicity, we detail the multiplicative decomposition approach of bivariate processes, i.e. the case $m = 2$. This value of m helps keep the model parsimonious in order to avoid the curse of dimensionality. While higher values of m do not change materially the model's appearance and present any challenges to the underlying theory, in practice they imply a much higher risk of overparametrization. In addition, they bring in a need to consider a multivariate framework for binary directions-of-change. This does not often happens in economics and finance, with models like bivariate probit just beginning to gain attention recently (e.g., Nyberg, 2014); a rare exception is Anatolyev (2010).

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 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 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 a m-variate case, the ingredients of the decomposition model are m-variate positive 'marginal' for m absolute values, m-variate binary 'marginal' for m directions, and a $2m$-dimensional copula that links all components.

The marginals for absolute values have positive[1] support by their positivity. We assume that each of these m marginals are Weibull. One can use more flexible positive distributions (e.g., Gamma), but the use of Weibull appears to be sufficient in empirical applications in this paper and in Anatolyev and Gospodinov (2010). Each of m binary marginals is, of course, Bernoulli. 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) also used rotated Clayton, etc. In the multivariate setting, we suggest using the multivariate Gaussian copula, for a number of reasons. First, the Gaussian copula is fairly flexible in a multivariate context: it is parameterized by $m(2m-1)$ parameters, which in 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 only in this case is a multivariate MEM model with a Gaussian copula (as in Cipollini, Engle and Gallo, 2009), and the submodel for directions only is a multivariate probit model (Ashford and Sowden, 1970). Third, the Gaussian copula facilitates computations of various conditional distributions involved in the likelihood because the multivariate Gaussianity is very tractable in these terms. The general theory developed below, however, is applicable to other choices of the copula as

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

well.

2.1 Univariate decomposition

To illustrate the main idea of our approach, we first present the univariate decomposition method of Anatolyev and Gospodinov (2010). Let r_t be a time series of returns. It can be decomposed into two multiplicative components as

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

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, a linear transformation of sign $\mathrm{sign}(r_t)$.

Let $\psi_t = E(|r_t||\mathcal{F}_{t-1})$ be the conditional expectation of a conditionally Weibull distributed absolute value $|r_t|$ with a shape parameter ς, denoted as $|r_t||\mathcal{F}_{t-1} \sim \mathcal{W}(\psi_t, \varsigma)$. Let $p_t = \Pr\{r_t > 0|\mathcal{F}_{t-1}\} = \Phi(\theta_t)$ be the 'success' (i.e. the market's 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

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

where[2]

$$C(w, y) = \Phi_2(\Phi^{-1}(w), \Phi^{-1}(y), \varrho)$$

is a bivariate Gaussian copula with correlation parameter ϱ. The processes ψ_t and θ_t can be specified as functions of the variables in $\mathcal{F}_{t-1}$ adding to the set of parameters.

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. The following Proposition is proved in Appendix A.1.

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

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

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

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

[2]Here and elsewhere $\Phi_{2m}(\underbrace{\cdot, \cdot, ..., \cdot}_{2m \text{ times}}, R)$ is CDF of the standard $2m$-dimensional normal distribution with correlation matrix R.

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

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

where the series of 'distorted' probabilities is

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

for $t = 1, ..., T$.

2.2 Bivariate decomposition

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

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

$i = 1, 2$, where $I_{i,t} = \mathbb{I}\{r_{i,t} > 0\}$. The two absolute values are Weibull $\mathcal{W}(\psi_{i,t}, \varsigma_i)$, and the two directions are Bernoulli $\mathcal{B}(p_{i,t})$, where $\psi_{i,t} = E(|r_{i,t}||\mathcal{F}_{t-1})$ and $p_{i,t} = \Pr\{r_{i,t} > 0|\mathcal{F}_{t-1}\} = \Phi(\theta_{i,t})$. 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 two variables. Together, there are four components that are linked through a copula:

$$\begin{pmatrix} |r_{1,t}| \\ |r_{2,t}| \end{pmatrix}, \begin{pmatrix} I_{1,t} \\ I_{2,t} \end{pmatrix} |\mathcal{F}_{t-1} \sim C\left(\begin{pmatrix} \mathcal{W}(\psi_{1,t}, \varsigma_1) \\ \mathcal{W}(\psi_{2,t}, \varsigma_2) \end{pmatrix}, \begin{pmatrix} \mathcal{B}(p_{1,t}) \\ \mathcal{B}(p_{2,t}) \end{pmatrix}, R\right),$$

where

$$C(w_1, w_2, y_1, y_2) = \Phi_4(\Phi^{-1}(w_1), \Phi^{-1}(w_2), \Phi^{-1}(y_1), \Phi^{-1}(y_2), R)$$

is a quartivariate Gaussian copula with correlation matrix

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

Let us temporarily suppress the time indexing. Denote the marginal CDFs of volatility components by $F_1(u_1)$ and $F_2(u_2)$ and their marginal PDFs by $f_1(u_2)$ and $f_1(u_2)$. Denote the success probabilities of the direction components by p_1 and p_2. The following Proposition whose proof can be found in Appendix A.1 gives an expression for the quartivariate joint density/mass function.

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

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

where

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

is the bivariate PDF of the volatility submodel linked by the bivariate Gaussian copula $c(w_1, w_2, R_v)$ *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 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[3]

$$\begin{aligned} \pi_1(u_1, u_2) &= \Phi_2(\Phi^{-1}(1-p_1), \Phi^{-1}(1) | \Phi^{-1}(w_1), \Phi^{-1}(w_2))\big|_{w_1=F_1(u_1), w_2=F_2(u_2)} \\ \pi_2(u_1, u_2) &= \Phi_2(\Phi^{-1}(1), \Phi^{-1}(1-p_2) | \Phi^{-1}(w_1), \Phi^{-1}(w_2))\big|_{w_1=F_1(u_1), w_2=F_2(u_2)} \\ \pi_{12}(u_1, u_2) &= \Phi_2(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2) | \Phi^{-1}(w_1), \Phi^{-1}(w_2))\big|_{w_1=F_1(u_1), w_2=F_2(u_2)}. \end{aligned}$$

Note that the volatility-only submodel is the copula-based multivariate MEM (though with different marginals) from Cipollini, Engle and Gallo (2009). Recall that the density of the bivariate Gaussian copula is

$$c(w_1, w_2, \varrho_v) = \frac{1}{\sqrt{\det R_v}} \exp\left(-\frac{1}{2} \begin{pmatrix} \Phi^{-1}(w_1) \\ \Phi^{-1}(w_2) \end{pmatrix}' \left(R_v^{-1} - I_2\right) \begin{pmatrix} \Phi^{-1}(w_1) \\ \Phi^{-1}(w_2) \end{pmatrix} \right).$$

Note also that (see Appendix A.2), if there were no links to the volatility submodel,

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

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). The algorithm of computations of $\pi_1(u_1, u_2)$, $\pi_2(u_1, u_2)$ and $\pi_{12}(u_1, u_2)$ is described in Appendix A.3.

[3]The following expressions can be simplified using that $\Phi_2(y_1, \Phi^{-1}(1)) = \Phi(y_1)$ and $\Phi_2(\Phi^{-1}(1), y_2) = \Phi(y_2)$. However, we prefer not to do it for the sake of generality of further computations of conditional CDFs.

In our case, $f_i(u)$ and $F_i(u)$, $i = 1, 2$, are those of the Weibull distribution, and $p_i = \Phi(\theta_i)$, $i = 1, 2$, are probit success probabilities. Restoring time indexing, the joint log-likelihood equals

$$\begin{aligned} \ell_r &= \sum_{t=1}^{T} \sum_{i=1,2} \log f^{\mathcal{W}(\psi_{i,t},\varsigma_i)}(|r_{i,t}|) + \sum_{t=1}^{T} c(F^{\mathcal{W}(\psi_{1,t},\varsigma_1)}(|r_{1,t}|), F^{\mathcal{W}(\psi_{2,t},\varsigma_2)}(|r_{2,t}|), \varrho_v) \\ &\quad + \sum_{t=1}^{T} I_{1,t} I_{2,t} \log p_{11,t}^C + (1 - I_{1,t}) I_{2,t} \log p_{01,t}^C + I_{1,t}(1 - I_{2,t}) \log p_{10,t}^C + (1 - I_{1,t})(1 - I_{2,t}) \log p_{00,t}^C, \end{aligned}$$

where

$$p_{ij,t}^C = p_{ij}^C(|r_{1,t}|, |r_{2,t}|), \quad i, j \in \{0, 1\}, \quad t = 1, ..., T,$$

is a collection of the series of 'distorted' probabilities.

2.3 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 and r_2.

Suppose one is interested in the conditional expectation of $g(r_1, r_2)$ 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,

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

where $p_{ij}^C(u_1, u_2)$, $i, j \in \{0, 1\}$ are defined before as functions of $\pi_1(u_1, u_2)$, $\pi_2(u_1, u_2)$ and $\pi_{12}(u_1, u_2)$. 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:

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

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

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

As a consequence, the conditional means can be computed as

$$E_{t-1}\left[r_{1,t}\right] = \int_{u_1=0}^{+\infty}\int_{u_2=0}^{+\infty} u_1\left(1 - 2\pi_{1,t}(u_1,u_2)\right) f_{v,t}(u_1,u_2)du_1du_2,$$

and similarly for $E_{t-1}\left[r_{2,t}\right]$, or alternatively and more simply

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

where

$$\begin{aligned}
\xi_{1,t} &= \int_{u_1=0}^{+\infty}\int_{u_2=0}^{+\infty} u_1\left(1 - \pi_{1,t}(u_1,u_2)\right) f_{v,t}(u_1,u_2)du_1du_2 \\
&= \int_{u_1=0}^{+\infty} u_1 p_{1,t}^C(u_1)du_1 \int_{u_2=0}^{+\infty} f_{v,t}(u_1,u_2)du_2 \\
&= \int_{u_1=0}^{+\infty} u_1 p_{1,t}^C(u_1) f_{1,t}(u_1)du_1,
\end{aligned}$$

and similarly for $E_{t-1}\left[r_{2,t}\right]$. The conditional means can be used, among other things, for constructing the pseudo-R^2 measure.

The conditional variances are simply

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

where

$$E_{t-1}\left[|r_{1,t}|^2)\right] = \int_{u_1=0}^{+\infty} u_1^2 \int_{u_2=0}^{+\infty} f_{v,t}(u_1,u_2)du_1du_2,$$

and similarly for $\text{var}_{t-1}(r_{2,t})$. The conditional correlations are

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

where

$$E_{t-1}\left[r_{1,t}r_{2,t}\right] = \int_{u_1=0}^{+\infty}\int_{u_2=0}^{+\infty} u_1u_2\left(1 - 2\pi_{1,t}(u_1,u_2) - 2\pi_{2,t}(u_1,u_2) + 4\pi_{12,t}(u_1,u_2)\right) f_{v,t}(u_1,u_2)du_1du_2.$$

The two-dimensional integrals involved in these formulas are straightforward to compute using numerical methods. In our simulation and empirical work, we compute them 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 design is intentionally made symmetric across the two variables.

The volatility equations are specified as

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

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 $\varsigma_1 = \varsigma_2 = 1.2$.

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 the 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 the 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 also very close to the nominal level of the test. From standard deviation one can see 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.

To evaluate the accuracy of the predictions from the bivariate decomposition model, we compute the pseudo-R^2 and compare it to the R^2 from a linear model fitted to the simulated data. Figure 1 plots the histogram of the R^2 from the linear and decomposition models for $n = 500$. It turns out that for this DGP the R^2 figures are small, indicating that the noise by far exceeds the signal, and one should not expect big R^2 figures in a corresponding application. The decomposition model, being the true DGP, naturally produces a much higher R^2 than the linear model, and the corresponding distribution dominates that for the linear model. However, note that there is a non-trivial fraction of simulations when the R^2 from the linear model exceeds that of the decomposition model, that there are discrepancies in the distributions across the two variables that enter the model symmetrically, and that there are R^2 figures very close to zero. These facts indicate that the pseudo-R^2 is a very noisy measure of time-series fit in this application, 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 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.[4] 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). The data are monthly observations covering the period January 1953 – December 2013. The two series are denoted by r_1 (LT) and r_2 (IT), respectively.

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

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

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

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

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

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

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 $\rho = 0.82$.

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 $\psi_{i,t}$ 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 univariate model for absolute returns ('volatility submodel') is specified as

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

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

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

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

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

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

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

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

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

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

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

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

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

An additional parameter involved is the degree of conditional dependence ϱ_d between the two directions. 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 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. 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 2.3. The actual and predicted returns from the decomposition model are plotted in Figure 2. Table 7 provides information on the quality of the model predictions measured by the pseudo-R^2. Overall, the decomposition model tends to generate better predictions than linear model with the bivariate version offering noticeable improvements only for the long-term bond returns. 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 of predict any function of these components, while the pseudo-R^2 measures the fit of only a small subset of these functions (products of absolute values and signs). In contrast, the linear model is intrinsically tied to that objective and, as our simulations showed, the use of pseudo-R^2 places the decomposition model at a disadvantage. Nevertheless, we report the pseudo-R^2 given its popularity in applied work.

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 3 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 during the business cycle.

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 Gaussian copula which is parameterized by $m(2m-1)$ parameters. The computation of various conditional measures of interest are 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] Chernov, M., and P. Mueller, 2012, The term structure of inflation expectations, *Journal of Financial Economics* 106, 367–394.

[5] Cieslak, A., and P. Povala, 2015, Expected returns in Treasury bonds, *Review of Financial Studies*, forthcoming.

[6] Cipollini, F., R. F. Engle, and G.M. Gallo, 2009, A model for multivariate non-negative valued processes in financial econometrics, Working Paper, Universita di Firenze.

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

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

[9] 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.

[10] 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.

[11] 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.

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

[13] Liu, X. and R. Luger, 2015, Unfolded GARCH models, *Journal of Economic Dynamics and Control*, in press.

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

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

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

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

and hence

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

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

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

□

Proof of Proposition 2. We will suppress the time index throughout. Likewise, denote the marginal (Bernoulli) CDFs of direction components by $G_1(v_1)$ and $G_2(v_2)$ and their marginal success probabilities by p_1 and p_2. The joint CDF of the quadruple (u_1,u_2,v_1,v_2) is

$$F(u_1,u_2,v_1,v_2)=C(F_1(u_1),F_2(u_2),G_1(v_1),G_2(v_2)).$$

The joint PDF/PMF is derived by taking the second derivative with respect to the two continuous components and second-order difference with respect to the two discrete components:

$$\begin{aligned}
f(u_1,u_2,v_1,v_2) &= \frac{\partial^2 C}{\partial u_1 \partial u_2}(F_1(u_1),F_2(u_2),G_1(v_1),G_2(v_2)) \\
&\quad -\frac{\partial^2 C}{\partial u_1 \partial u_2}(F_1(u_1),F_2(u_2),G_1(v_1-1),G_2(v_2)) \\
&\quad -\frac{\partial^2 C}{\partial u_1 \partial u_2}(F_1(u_1),F_2(u_2),G_1(v_1),G_2(v_2-1)) \\
&\quad +\frac{\partial^2 C}{\partial u_1 \partial u_2}(F_1(u_1),F_2(u_2),G_1(v_1-1),G_2(v_2-1)) \\
&= f_1(u_1)f_2(u_2)f_{\partial\partial}(u_1,u_2,v_1,v_2),
\end{aligned}$$

where the last term is

$$\begin{aligned}
f_{\partial\partial}(u_1,u_2,v_1,v_2) &= \Bigg[\frac{\partial^2 C}{\partial w_1 \partial w_2}(w_1,w_2,G_1(v_1),G_2(v_2)) \\
&\quad -\frac{\partial^2 C}{\partial w_1 \partial w_2}(w_1,w_2,G_1(v_1-1),G_2(v_2)) \\
&\quad -\frac{\partial^2 C}{\partial w_1 \partial w_2}(w_1,w_2,G_1(v_1),G_2(v_2-1)) \\
&\quad +\frac{\partial^2 C}{\partial w_1 \partial w_2}(w_1,w_2,G_1(v_1-1),G_2(v_2-1))\Bigg]_{w_1=F_1(u_1),w_2=F_2(u_2)} \\
&= q_{11}(u_1,u_2)^{v_1 v_2} q_{01}(u_1,u_2)^{(1-v_1)v_2} q_{10}(u_1,u_2)^{v_1(1-v_2)} q_{00}(u_1,u_2)^{(1-v_1)(1-v_2)},
\end{aligned}$$

where

$$\begin{aligned}
q_{11}(u_1,u_2) &= \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1,1) - \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1-p_1,1) \\
&\quad -\frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1,1-p_2) + \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1-p_1,1-p_2), \\
q_{01}(u_1,u_2) &= \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1-p_1,1) - \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1-p_1,1-p_2) \\
q_{10}(u_1,u_2) &= \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1,1-p_2) - \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1-p_1,1-p_2) \\
q_{00}(u_1,u_2) &= \frac{\partial^2 C}{\partial w_1 \partial w_2}(F_1(u_1),F_2(u_2),1-p_1,1-p_2)
\end{aligned}$$

taking advantage of the fact that $C(w_1,w_2,y_1,0) = C(w_1,w_2,0,y_2) = C(w_1,w_2,0,0) = 0$. Using that

$$
\begin{aligned}
\frac{\partial^2\Phi_4(x_1,x_2,x_3,x_4)}{\partial x_1\partial x_2} &= \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\int_{-\infty}^{x_3}\int_{-\infty}^{x_4}\phi_4(t_1,t_2,t_3,t_4)dt_1dt_2dt_3dt_4 \\
&= \int_{-\infty}^{x_3}\int_{-\infty}^{x_4}\left(\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\phi_4(t_1,t_2,t_3,t_4)dt_1dt_2\right)dt_3dt_4 \\
&= \int_{-\infty}^{x_3}\int_{-\infty}^{x_4}\phi_4(x_1,x_2,t_3,t_4)dt_3dt_4 \\
&= \phi_2(x_1,x_2)\int_{-\infty}^{x_3}\int_{-\infty}^{x_4}\phi_2(t_3,t_4|x_1,x_2)dt_3dt_4 \\
&= \phi_2(x_1,x_2)\Phi_2(x_3,x_4|x_1,x_2),
\end{aligned}
$$

we have

$$
\begin{aligned}
\frac{\partial^2 C(w_1,w_2,y_1,y_2)}{\partial w_1\partial w_2} &= \frac{\partial^2\Phi_4(\Phi^{-1}(w_1),\Phi^{-1}(w_2),\Phi^{-1}(y_1),\Phi^{-1}(y_2))}{\partial w_1\partial w_2} \\
&= \left.\frac{\partial^2\Phi_4(x_1,x_2,x_3,x_4)}{\partial x_1\partial x_2}\right|_{x_1=\Phi^{-1}(w_1),x_2=\Phi^{-1}(w_2),x_3=\Phi^{-1}(y_1),x_4=\Phi^{-1}(y_2)} \\
&\quad \cdot\frac{\partial\Phi^{-1}(w_1)}{\partial w_1}\frac{\partial\Phi^{-1}(w_2)}{\partial w_2} \\
&= \phi_2(x_1,x_2)\Phi_2(x_3,x_4|x_1,x_2)|_{x_1=\Phi^{-1}(w_1),x_2=\Phi^{-1}(w_2),x_3=\Phi^{-1}(y_1),x_4=\Phi^{-1}(y_2)} \\
&\quad \cdot\left.\frac{1}{\phi\left(x_1\right)}\right|_{x_1=\Phi^{-1}(w_1)}\left.\frac{1}{\phi\left(x_2\right)}\right|_{x_2=\Phi^{-1}(w_2)} \\
&= \frac{\phi_2(\Phi^{-1}(w_1),\Phi^{-1}(w_2))}{\phi\left(\Phi^{-1}(w_1)\right)\phi\left(\Phi^{-1}(w_2)\right)}\Phi_2(\Phi^{-1}(y_1),\Phi^{-1}(y_2)|\Phi^{-1}(w_1),\Phi^{-1}(w_2)) \\
&= c_N(w_1,w_2,\varrho_v)\Phi_2(\Phi^{-1}(y_1),\Phi^{-1}(y_2)|\Phi^{-1}(w_1),\Phi^{-1}(w_2)),
\end{aligned}
$$

where

$$
c(w_1,w_2,\varrho_v)=\frac{1}{\sqrt{\det R_v}}\exp\left(-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(w_1)\\ \Phi^{-1}(w_2)\end{pmatrix}'\left(R_v^{-1}-I_2\right)\begin{pmatrix}\Phi^{-1}(w_1)\\ \Phi^{-1}(w_2)\end{pmatrix}\right)
$$

is bivariate Gaussian copula. Then,

$$
q_{ij}(u_1,u_2)=c(F_1(u_1),F_2(u_2),\varrho_v)\cdot p_{ij}^C(u_1,u_2),
$$

$i,j\in\{0,1\}$, where, using also that $\Phi_2(\Phi^{-1}(1),\Phi^{-1}(1)|\Phi^{-1}(w_1),\Phi^{-1}(w_2))=1$, we get

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

where

$$
\begin{aligned}
\pi_1(u_1,u_2) &= \Phi_2(\Phi^{-1}(1-p_1),\Phi^{-1}(1)|\Phi^{-1}(w_1),\Phi^{-1}(w_2))\big|_{w_1=F_1(u_1),w_2=F_2(u_2)}\\
\pi_2(u_1,u_2) &= \Phi_2(\Phi^{-1}(1),\Phi^{-1}(1-p_2)|\Phi^{-1}(w_1),\Phi^{-1}(w_2))\big|_{w_1=F_1(u_1),w_2=F_2(u_2)}\\
\pi_{12}(u_1,u_2) &= \Phi_2(\Phi^{-1}(1-p_1),\Phi^{-1}(1-p_2)|\Phi^{-1}(w_1),\Phi^{-1}(w_2))\big|_{w_1=F_1(u_1),w_2=F_2(u_2)}.
\end{aligned}
$$

Now, collecting the pieces,

$$
\begin{aligned}
f(u_1,u_2,v_1,v_2) &= f_1(u_2)f_1(u_2)\cdot(q_{11})^{v_1v_2}(q_{01})^{(1-v_1)v_2}(q_{10})^{v_1(1-v_2)}(q_{00})^{(1-v_1)(1-v_2)}\\
&= f_1(u_2)f_1(u_2)\cdot c_N(F_1(u_1),F_2(u_2),\varrho_v)\\
&\quad\cdot\left(p_{11}^C\right)^{v_1v_2}\left(p_{01}^C\right)^{(1-v_1)v_2}\left(p_{10}^C\right)^{v_1(1-v_2)}\left(p_{00}^C\right)^{(1-v_1)(1-v_2)}\\
&= f_v(u_1,u_2)\cdot f_d^C(u_1,u_2,v_1,v_2),
\end{aligned}
$$

where

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

is the normal copula-induced bivariate PDF of the volatility submodel, and

$$f_d^C(u_1,u_2,v_1,v_2)=p_{11}^C(u_1,u_2)^{v_1v_2}p_{01}^C(u_1,u_2)^{(1-v_1)v_2}p_{10}^C(u_1,u_2)^{v_1(1-v_2)}p_{00}^C(u_1,u_2)^{(1-v_1)(1-v_2)}$$

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

A.2 Special case: no link to volatility

When the bivariate direction model is separate from the volatility model, it reduces to the conventional bivariate probit model (Ashford and Sowden, 1970). Indeed,

$$
\begin{aligned}
p_{00}^C &= \Phi_{\varrho_d}(\Phi^{-1}(1-p_1),\Phi^{-1}(1-p_2))=\Phi_{\varrho_d}(-\theta_{1,t},-\theta_{2,t})\\
p_{01}^C &= 1-\Phi(\theta_{1,t})-\Phi_{\varrho_d}(-\theta_{1,t},-\theta_{2,t})=\Phi(-\theta_{1,t})-\left(\Phi(-\theta_{1,t})-\Phi_{-\varrho_d}(-\theta_{1,t},\theta_{2,t})\right)\\
&= \Phi_{-\varrho_d}(-\theta_{1,t},\theta_{2,t}),\\
p_{10}^C &= 1-\Phi(\theta_{2,t})-\Phi_{\varrho_d}(-\theta_{1,t},-\theta_{2,t})=\Phi(-\theta_{2,t})-\left(\Phi(-\theta_{2,t})-\Phi_{-\varrho_d}(\theta_{1,t},-\theta_{2,t})\right)\\
&= \Phi_{-\varrho_d}(\theta_{1,t},-\theta_{2,t})\\
p_{11}^C &= 1-(1-\Phi(\theta_{1,t}))-(1-\Phi(\theta_{2,t}))+\Phi_{\varrho_d}(-\theta_{1,t},-\theta_{2,t})\\
&= \Phi(\theta_{1,t})+\Phi(\theta_{2,t})-1+\left(\Phi(-\theta_{1,t})-\Phi_{-\varrho_d}(-\theta_{1,t},\theta_{2,t})\right)=\Phi(\theta_{2,t})-\Phi_{-\varrho_d}(-\theta_{1,t},\theta_{2,t})\\
&= \left(\Phi_{\varrho_d}(\theta_{1,t},\theta_{2,t})+\Phi_{-\varrho_d}(-\theta_{1,t},\theta_{2,t})\right)-\Phi_{-\varrho_d}(-\theta_{1,t},\theta_{2,t})\\
&= \Phi(\theta_{1,t},\theta_{2,t},\varrho_d)
\end{aligned}
$$

A.3 Computation of conditional distributions

Note that $\pi_1(u_1,u_2)$, $\pi_2(u_1,u_2)$ and $\pi_{12}(u_1,u_2)$ all have the form

$$\Phi_2\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} | \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$$

after plugging in a value of $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ from the set $\left\{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right\}$, and $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \Phi^{-1}(F_1(u_1)) \\ \Phi^{-1}(F_2(u_2)) \end{pmatrix}$. This is a CDF of a subvector of a quartivariate normal random variable and thus is (conditional) bivariate normal.

Represent this CDF by the from law of total probability as

$$\int_{-\infty}^{y_2} \Phi\left(y_1|\mathring{y}_2, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) \phi\left(\mathring{y}_2|\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) d\mathring{y}_2, \tag{A1}$$

where $\Phi\left(y_1|\mathring{y}_2, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$ is a CDF of a univariate conditional (on $\mathring{y}_2$, x_1 and x_2) normal random variable with the mean

$$\varrho'_{1|2v} R_{2v}^{-1} \begin{pmatrix} \mathring{y}_2 \\ x_1 \\ x_2 \end{pmatrix}$$

and variance $1 - \varrho'_{1|2v} R_{2v}^{-1} \varrho_{1|2v}$, where

$$R_{2v} = \begin{bmatrix} 1 & \varrho_{vd} & \varrho_2 \\ \varrho_{vd} & 1 & \varrho_v \\ \varrho_2 & \varrho_v & 1 \end{bmatrix}, \quad \varrho_{1|2v} = \begin{bmatrix} \varrho_d \\ \varrho_1 \\ \varrho_{dv} \end{bmatrix},$$

and $\phi\left(\mathring{y}_2|\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$ is a PDF of a univariate conditional (on x_1 and x_2) normal random variable with mean $e_2 R_{dv} R_v^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and variance $1 - e_2' R_{dv} R_v^{-1} R_{vd} e_2$, where e_i is i^{th} unit vector.

We evaluate the integral (A1) using the Gauss-Chebychev quadrature with 400 Chebychev quadrature nodes on $[-5, y_2]$. See formulas (7.2.5) and (7.2.7) in Judd (1998).

Table 1. Simulation results for the bivariate decomposition model.

		$n = 500$			$n = 2000$		
parameter	true value	mean	stdev	t	mean	stdev	t
ω_{v1}	0	−0.018	0.065	0.049	−0.004	0.029	0.048
β_{v1}	0.8	0.800	0.032	0.068	0.800	0.015	0.064
α_{v11}	0.1	0.095	0.024	0.078	0.099	0.011	0.051
γ_{v11}	−0.3	−0.302	0.054	0.074	−0.301	0.027	0.062
α_{v12}	0.05	0.049	0.017	0.075	0.050	0.008	0.047
γ_{v12}	0.2	0.203	0.052	0.072	0.200	0.025	0.055
ς_1	1.2	1.208	0.041	0.042	1.202	0.021	0.053
ω_{v2}	0	−0.017	0.064	0.051	−0.003	0.028	0.054
β_{v2}	0.8	0.801	0.031	0.081	0.801	0.014	0.053
α_{v21}	0.05	0.050	0.017	0.069	0.050	0.008	0.056
γ_{v21}	0.2	0.202	0.053	0.063	0.201	0.025	0.056
α_{v22}	0.1	0.094	0.024	0.084	0.099	0.011	0.057
γ_{v22}	−0.3	−0.303	0.054	0.065	−0.301	0.025	0.047
ς_2	1.2	1.207	0.041	0.040	1.203	0.020	0.051
ω_{d1}	0.3	0.308	0.117	0.057	0.302	0.056	0.047
ϕ_{d11}	0.3	0.294	0.133	0.050	0.299	0.063	0.035
ϕ_{d12}	−0.1	−0.106	0.136	0.052	−0.103	0.067	0.057
ω_{d2}	0.3	0.309	0.118	0.063	0.302	0.055	0.053
ϕ_{d21}	−0.1	−0.103	0.134	0.050	−0.101	0.066	0.052
ϕ_{d22}	0.3	0.296	0.132	0.044	0.297	0.064	0.041
ϱ_v	0.6	0.601	0.029	0.066	0.599	0.014	0.043
ϱ_d	0.6	0.602	0.055	0.058	0.600	0.029	0.066
ϱ_1	0.2	0.204	0.056	0.051	0.202	0.029	0.052
ϱ_2	0.2	0.202	0.057	0.050	0.202	0.028	0.047
ϱ_{vd}	0.2	0.202	0.057	0.046	0.201	0.028	0.058
ϱ_{dv}	0.2	0.203	0.056	0.050	0.202	0.029	0.062

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 $n = 2000$.

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)
$\delta_{\ell i}$	−0.0905 (0.0311)	−0.0588 (0.0157)	$\delta_{\ell i}$	−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)
δ_{vi}	−0.958 (0.326)	−1.100 (0.348)	δ_{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)
δ_{di}	-2.53 (1.15)	-4.44 (1.19)	δ_{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 pseudo-R^2.

	Univariate		Bivariate	
	LT_{LT}	IT_{LT}	LT_{LT+IT}	IT_{LT+IT}
Linear model	5.15%	8.95%	7.21%	8.91%
Decomposition model	5.10%	9.70%	5.34%	8.22%

Figure 1: Histogram of pseudo-R^2 of linear and decomposition model in Monte Carlo simulations.

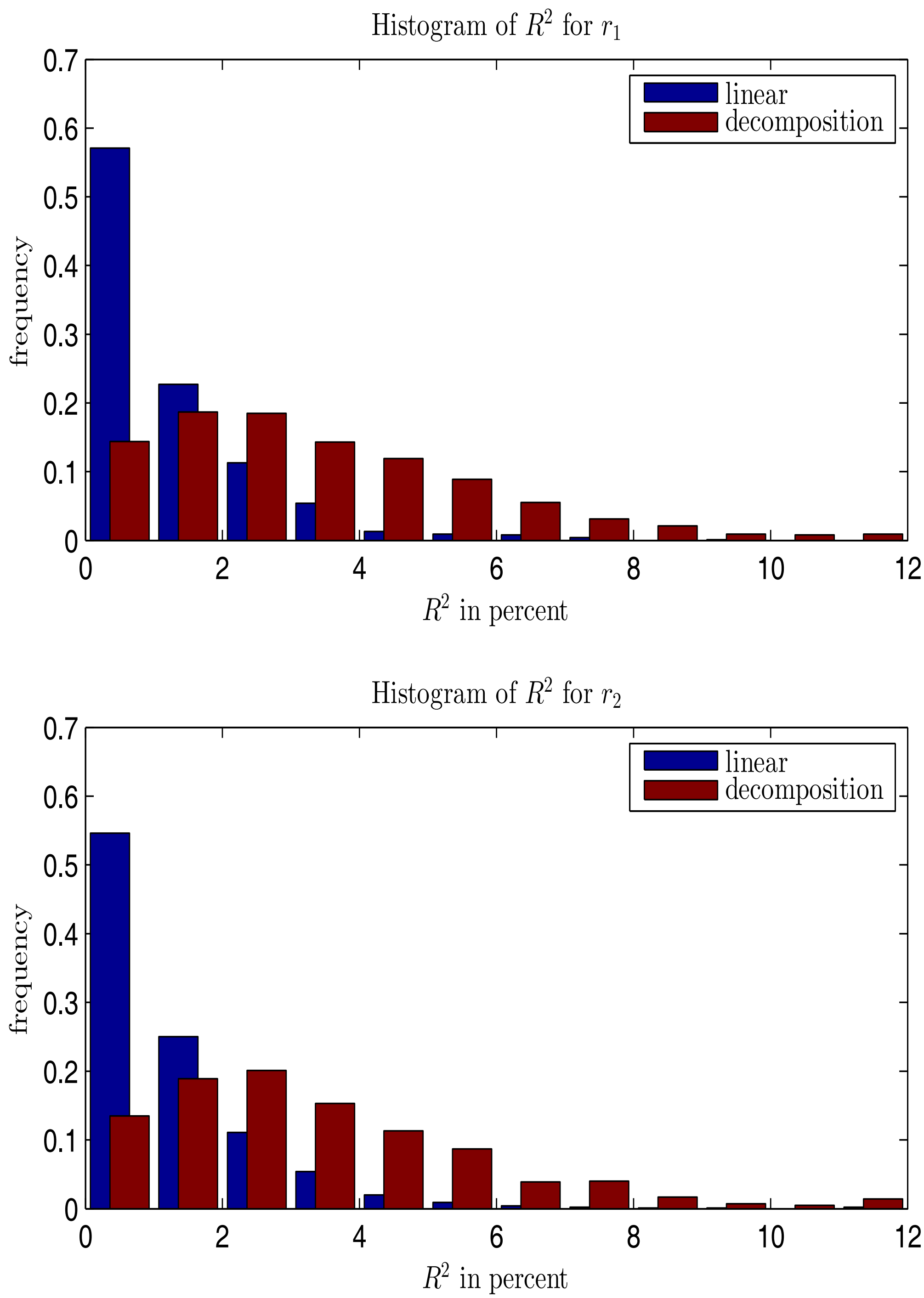

Figure 2: Actual and predicted long-term (r_1) and intermediate-term (r_2) government bond returns.

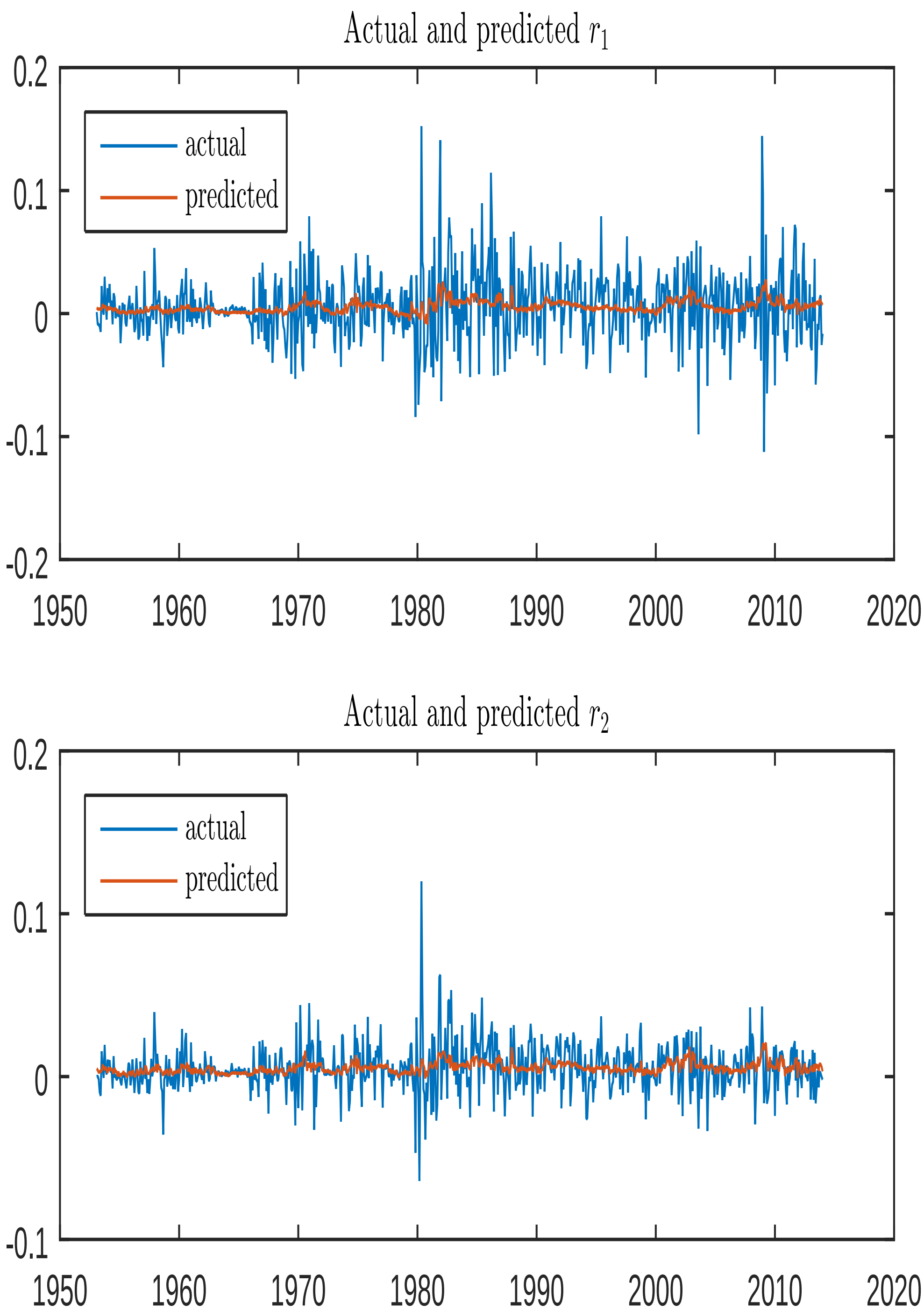

Figure 3: Conditional variances and correlation of long-term (r_1) and intermediate-term (r_2) government bond returns.