Journal of Economic Dynamics & Control 94 (2018) 225–241

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

The distribution of cross sectional momentum returns[☆]

Oh Kang Kwon[a,*], Stephen Satchell[a,b]

[a] Discipline of Finance, Codrington Building (H69), The University of Sydney, NSW 2006, Australia
[b] Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK

ARTICLE INFO

Article history:
Received 28 November 2017
Revised 2 June 2018
Accepted 6 June 2018
Available online 15 June 2018

JEL classification:
C40
G11

Keywords:
Cross sectional momentum
Return distribution
Moments

ABSTRACT

Although there is a vast empirical literature on cross sectional momentum (CSM) returns, there are no known analytical results on their distributional properties due, in part, to the mathematical complexity associated with their determination. In this paper, we derive the density of CSM returns in analytic form, along with moments of all orders, under the assumption that underlying asset returns are multivariate normal. The resulting expressions are highly non-trivial in general and involve truncated normal distributions. The distribution of CSM returns can be formally described as a mixture of the unified skew-normal family of distributions. However, if the asset returns are independent, then the density of the CSM returns is shown to be a mixture of univariate normals. In order to shed light on the general case, we present a detailed analysis of the case of two underlying assets, which is shown to explain many of the key features of CSM returns reported in the empirical literature.

1. Introduction

Momentum based trading strategies rely on the persistence in the relative performances of assets over successive time periods called the ranking and holding periods respectively. For example, a cross sectional momentum (CSM) strategy, considered in Jegadeesh and Titman (1993), sorts the returns from n assets over the ranking period, and constructs a portfolio over the holding period consisting of an equally weighted long position in the m_+ best performing assets ("winners") and an equally weighted short position in the m_- worst performing assets ("losers"). Such strategies have become popular in practice, due mainly to the fact that they tend to exhibit positive returns. Another reason for their popularity is due to Carhart (1997), who pioneered momentum as a key component in the factor based risk analysis of investment returns.

Since it has been argued that excess returns from momentum based trading strategies would indicate a violation of the assumption of market efficiency, the returns generated by these strategies have been the subject of considerable empirical research spanning extensive asset classes, jurisdictions, and investment periods. Various authors, including Jegadeesh and Titman (1993, 2001), Asness (1994) and Israel and Moskowitz (2013), found that momentum strategies are profitable in US equities markets over different time periods dating back to 1927. Analogous results were found for country equity indices by Richards (1997), Asness et al. (1997), Chan et al. (2000) and Hameed and Yuanto (2002), for emerging markets by Rouwenhorst (1998), for exchange rate markets in Okunev and White (2003) and Menkhoff et al. (2012), for commodities by

☆ We thank the Editor and the reviewers for the helpful comments on earlier versions of this paper that have significantly improved the readability of the final version.

* Corresponding author.
E-mail addresses: ohkang.kwon@sydney.edu.au (O. Kwon), ses999gb@yahoo.co.uk (S. Satchell).

https://doi.org/10.1016/j.jedc.2018.06.002
0165-1889/

Erb and Harvey (2006), for futures contracts in Moskowitz et al. (2012), and in industries by Sefton and Scowcroft (2004). Similar results were also found by Asness et al. (2013) and Daniel and Moskowitz (2016) for markets in the European Union, Japan, United Kingdom and United States, and across asset classes including fixed income, commodities, foreign exchange and equity from 1972 through 2013.

In contrast to the extensive literature on the empirical properties of momentum based returns, there are relatively few that consider the distributional properties of these returns from a theoretical viewpoint, and none of these address the CSM returns as defined in this paper. Most of the known theoretical results, obtained for example by Lo and MacKinlay (1990), Jegadeesh and Titman (1993), Lewellen (2002) and Moskowitz et al. (2012), are concerned only with the expected values and the first order autocorrelations of the returns from the so-called weighted relative strength strategy in which the portfolio over the holding period is constructed from all underlying assets weighted, essentially, in proportion to their absolute or relative returns over the ranking period.

By assuming that underlying asset returns are Gaussian, we derive in this paper the distribution and the moments of CSM returns in the general case, and in a number of special cases under which resulting expressions simplify significantly. Anticipating our results, the densities obtained involve truncated normal distributions, which is a result partially discussed in Grundy and Martin (2001) Appendix A. The results generalize naturally to arbitrary fixed weight portfolios, albeit with added complexity in notation.

The remainder of this paper is organized as follows: Section 2 introduces the notation and the key results on multivariate normal distributions, and Section 3 provides a mathematically precise definition of CSM returns. Although the expressions for the CSM return density and moments are quite complex in general, they simplify considerably in the case of two assets with one winner and one loser, and this special case is examined in detail in Section 4. Implications of the results in the 2-asset case are considered in Section 5, where they are used to explain many of the empirical features reported in the literature. Numerical examples illustrating the different CSM return distributions that can be generated using parameters estimated from market data are given in Section 6, the distributional properties of CSM returns in the general case are derived in Section 7, and the paper concludes with Section 8.

2. Notation and preliminary discussion

For the convenience of the reader, we introduce in this section the notation that will be used throughout the paper, and present some preliminary discussion on cross sectional momentum returns to motivate the framework under which we develop the theory in later sections.

For any $\boldsymbol{x} \in \mathbb{R}^n$, we will write x_i for the i-th coordinate of $\boldsymbol{x}$, and given $\boldsymbol{y} \in \mathbb{R}^n$ write $\boldsymbol{x} \prec \boldsymbol{y}$ if and only if $x_i < y_i$ for all $1 \le i \le n$. Similarly, given a matrix $M \in \mathbb{R}^{m \times k}$, we will write $M_{i,j}$ for the (i, j)-th entry of M, and the transpose of a vector or a matrix will be denoted by the superscript $'$. The density of an n-dimensional normal distribution, with mean $\boldsymbol{\mu}$ and covariance Σ, at $\boldsymbol{x} \in \mathbb{R}^n$ will be denoted $\phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)$ so that

$$\phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}, \tag{1}$$

and for any $\boldsymbol{a} \in \mathbb{R}^n$ we define

$$\Phi_n[\boldsymbol{a}; \boldsymbol{\mu}, \Sigma] = \int_{\boldsymbol{x} \prec \boldsymbol{a}} \phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)\, d\boldsymbol{x}, \tag{2}$$

where $d\boldsymbol{x} = dx_1, \ldots, dx_n$. In general, given random variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, their joint probability density function will be denoted $f_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}$.

For any $n \in \mathbb{N}$, let S_n be the group of permutations of $\{1, 2, \ldots, n\}$, and given any $\tau \in S_n$, denote by $\tau_i = \tau(i)$ the image of $1 \le i \le n$ under τ, so that, for example,

$$S_3 = \{(1\,2\,3), (1\,3\,2), (2\,1\,3), (2\,3\,1), (3\,1\,2), (3\,2\,1)\}. \tag{3}$$

We now provide a brief description of CSM returns and some explanation of the relevance of our assumptions. Let $l_r \in \mathbb{N}$ be the length of the ranking period and $l_h \in \mathbb{N}$ the length of the holding period. Then the construction of a cross sectional momentum strategy can be described as follows. Firstly, in month t, all underlying assets are ranked and sorted into quantiles, for example quintiles or deciles, based on their returns over the past l_r-month ranking period from month $t - l_r - 1$ to $t - 1$. One month is omitted to avoid short-term reversals and similar phenomena. At this time, the top quantile portfolio, consisting of “winners”, is purchased and the worst performing quantile, consisting of “losers”, is shorted for the l_h-month holding period from month t to $t + l_h$. The two portfolios may be equally weighted, value weighted, or more generally weighted arbitrarily using fixed weights.

The problem of determining the distributional properties of the CSM returns can be regarded as a 2-period problem. At any time t, the first period is the ranking period from $t - l_r - 1$ to $t - 1$ over which the asset returns are described by the vector $\boldsymbol{r}_t$, and the second period is the holding period from $t + 1$ to $t + l_h$ over which the corresponding asset returns are $\boldsymbol{r}_{t+1}$. The vector $(\boldsymbol{r}_t', \boldsymbol{r}_{t+1}')$ is assumed to be Gaussian, and since we do not make any assumptions on stationarity the joint distribution of $\boldsymbol{r}_t$ and $\boldsymbol{r}_{t+1}$ is a $2n$-dimensional vector of normal returns with means and variances depending on t and $t + 1$ as in Assumption 1. This flexibility means that the exclusion of the one month period from $t - 1$ to t or the assumption $l_r \neq l_h$, for example, do not pose any difficulties.

However, when we do assume stationarity so that $\mathbb{E}[\boldsymbol{r}_t] = \mathbb{E}[\boldsymbol{r}_{t+1}]$ with corresponding restrictions on covariances, the condition $l_r = l_h$ needs to be imposed, unless there is a countervailing parametric change. Fortunately, many momentum strategies are of this form where the length of the ranking and holding periods are equal. In this case, the exclusion of a period is still consistent with strict stationarity, and since it is assumed that the vector return process is Gaussian this is equivalent to weak stationarity.

It should be noted that there is a certain advantage in having full flexibility to choose the lengths of ranking and holding periods. Certainly, most theoretical time-series models of momentum set the holding period length to one. One exception is He and Li (2015) who model time-series momentum in a continuous-time framework and allow both ranking and holding period lengths to be arbitrary.

3. Cross sectional momentum returns

We begin this section with a general framework under which to investigate the distribution of cross sectional momentum (CSM) returns, and show how well-known models for asset returns, such as VARMA(p,q), are special cases of this framework.

Fix $0 < m_+, m_-, n \in \mathbb{N}$ such that $m_+ + m_- \le n$, and for each $1 \le i \le n$ denote by $r_{i,\,t}$ the return on asset i at time t. Let

$$\mathbf{r}_t = (r_{1,t}, \dots, r_{n,t})' \in \mathbb{R}^n, \tag{4}$$

and for any $\tau \in S_n$, where S_n is the set of permutations on the set $\{1, 2, \dots, n\}$ as described in (3), define

$$r_{\tau,m_\pm,t} = \frac{1}{m_+}\sum_{i=1}^{m_+} r_{\tau_i,t} - \frac{1}{m_-}\sum_{i=1}^{m_-} r_{\tau_{n-m_-+i},t} \in \mathbb{R}, \tag{5}$$

$$x_{\tau_i,t} = r_{\tau_{i+1},t} - r_{\tau_i,t} \in \mathbb{R}, \quad 1 \le i \le n-1, \tag{6}$$

$$\boldsymbol{x}_{\tau,t} = (x_{\tau_1,t}, \dots, x_{\tau_{n-1},t})' \in \mathbb{R}^{n-1}, \tag{7}$$

$$\boldsymbol{z}_{\tau,t} = \left(r_{\tau,m_\pm,t+1}, \boldsymbol{x}'_{\tau,t}\right)' \in \mathbb{R}^n. \tag{8}$$

Note that any given $\tau \in S_n$ defines an ordering, $r_{t,\tau_1} > r_{t,\tau_2} > \dots > r_{t,\tau_n}$, of the components of $\boldsymbol{r}_t$. So $r_{\tau,m_\pm,t}$ represents the return on a portfolio where the top m_+ ranked assets are equally weighted and held long while the bottom m_- assets are equally weighted and held short. The assumption of equal weighting is for notational simplicity only, and not crucial for the general theoretical results. Note also that $\boldsymbol{x}_{\tau,\,t}$ is defined to allow the ranking of the components of $\boldsymbol{r}_t$ corresponding to $\tau \in S_n$ to be written succinctly as $\boldsymbol{x}_{\tau,t} \prec \mathbf{0}_{n-1}$.

The next assumption on the distribution of asset returns is made to permit the derivation of explicit expressions for the distributional properties of CSM returns. It is possible to work under more general assumptions, such as asset returns being multivariate Student's t, but this would be at the expense of additional complexity and we limit ourselves to the multivariate normal case in this paper.

Assumption 1. The vector $(\boldsymbol{r}'_{t+1}, r'_t)'$ has a multivariate normal distribution for all $t \in \mathbb{N}$.

Note that if $(\boldsymbol{r}'_{t+1},' r_t)'$ satisfies Assumption 1, then we may write

$$\begin{bmatrix} \boldsymbol{r}_{t+1} \\ \boldsymbol{r}_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{t+1} \\ \boldsymbol{\mu}_t \end{bmatrix}, \begin{bmatrix} \Lambda_{t+1,t+1} & \Lambda_{t+1,t} \\ \Lambda_{t,t+1} & \Lambda_{t,t} \end{bmatrix}\right), \tag{9}$$

with $\boldsymbol{\mu}_u \in \mathbb{R}^n$, $\Lambda_{u,v} \in \mathbb{R}^{n\times n}$, and $\Lambda_{u,\,u}$ positive definite for $u, v \in \{t, t+1\}$. The next result shows that returns following a vector autoregressive moving average process, VARMA(p,q), satisfy the above assumption.

Proposition 1. *Suppose* $\boldsymbol{r}_t \sim$ VARMA(p,q), *so that*

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \sum_{i=1}^{p} A_i \boldsymbol{r}_{t-i} + \boldsymbol{\varepsilon}_t + \sum_{i=1}^{q} M_i \boldsymbol{\varepsilon}_{t-i}, \tag{10}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n\times n}$ *for* $1 \le i \le p$, $M_i \in \mathbb{R}^{n\times n}$ *for* $1 \le i \le q$, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_n, \Sigma)$ *with* $\Sigma \in \mathbb{R}^{n\times n}$ *positive definite for all* $t \in \mathbb{N}$, $\mathbb{E}[\boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}'_t] = \delta_{s,t}\Sigma$ *for all* $s, t \in \mathbb{N}$, *and* $\delta_{s,\,t}$ *denotes the Kronecker delta so that*

$$\delta_{s,t} = \begin{cases} 1, & s = t, \\ 0, & s \ne t. \end{cases}$$

If $(\boldsymbol{r}'_0, \dots, \boldsymbol{r}'_{p-1})'$ *is multivariate normal and independent of* $\boldsymbol{\varepsilon}_t$, *or the* VARMA$(p,q)$ *process satisfies conditions for weak stationarity, then* $(\boldsymbol{r}'_{t+1}, \boldsymbol{r}'_t)$ *satisfies Assumption 1.*

Proof. This follows from the well-known properties of the VARMA(p, q) process. □

We now provide a mathematically precise representation of the cross sectional momentum return which was defined verbally following Eq. (3).

Definition 1. The (m_+, m_-)-*cross sectional momentum return*, $r_{m_\pm,t+1}$, is defined by

$$r_{m_\pm,t+1} = \sum_{\tau \in S_n} \mathbb{I}_{\{\boldsymbol{x}_{\tau,t} \prec \mathbf{0}_{n-1}\}} r_{\tau,m_\pm,t+1}, \tag{11}$$

where $\mathbb{I}_A$, for any subset $A \subset \mathbb{R}^n$, denotes the indicator function on the set A.

For intuition behind the definition of $r_{m_\pm,t+1}$, note that the components of $\boldsymbol{r}_t$, representing asset returns over the ranking period, can be arranged in any of the $n!$ orderings corresponding to the permutations $\tau \in S_n$. For each such ranking $r_{t,\tau_1} > r_{t,\tau_2} > \cdots > r_{t,\tau_n}$, the m_+ winner returns over the holding period are $r_{t+1,\tau_1}, \ldots, r_{t+1,\tau_{m_+}}$ while the m_- loser returns are $r_{t+1,\tau_{n-m_+1}}, \ldots, r_{t+1,\tau_n}$. Equally weighting the returns in the winner and the loser portfolios gives $r_{\tau,m_\pm,t+1}$, and since the ranking of components of $\boldsymbol{r}_t$ determined by $\tau \in S_n$ is equivalent to the condition $\boldsymbol{x}_{\tau,t} \prec \mathbf{0}_{n-1}$, summing over all possible $r_{\tau,m_\pm,t+1}$ and prefixing by the matching indicator function gives the expression for $r_{m_\pm,t+1}$ in (11).

For example, if $n = 3$, then the distinct rankings of the components of $\boldsymbol{r}_t \in \mathbb{R}^3$ partitions $\mathbb{R}^3$ into regions corresponding to the six permutations in S_3, viz.

$$\begin{aligned}
\mathcal{R}_{(1\,2\,3)} &= \{r_{1,t} > r_{2,t} > r_{3,t}\}, \quad \mathcal{R}_{(1\,3\,2)} = \{r_{1,t} > r_{3,t} > r_{2,t}\},\\
\mathcal{R}_{(2\,1\,3)} &= \{r_{2,t} > r_{1,t} > r_{3,t}\}, \quad \mathcal{R}_{(2\,3\,1)} = \{r_{2,t} > r_{3,t} > r_{1,t}\},\\
\mathcal{R}_{(3\,1\,2)} &= \{r_{3,t} > r_{1,t} > r_{2,t}\}, \quad \mathcal{R}_{(3\,2\,1)} = \{r_{3,t} > r_{2,t} > r_{1,t}\}.
\end{aligned}$$

Inequalities defining, say, the first region $\mathcal{R}_{(1\,2\,3)}$ is evidently equivalent to the condition

$$\boldsymbol{x}_{(1\,2\,3),t} = (r_{2,t} - r_{1,t}, r_{3,t} - r_{2,t}) \prec (0, 0),$$

with remaining regions permitting analogous equivalent definitions in terms of vectors $\boldsymbol{x}_{\tau,t} \in \mathbb{R}^2$ defined in (7). In this way, the six vectors $\boldsymbol{x}_{\tau,\,t}$, with $\tau \in S_3$, determine a mutually exclusive and exhaustive covering, $\{\mathcal{R}_\tau \mid \tau \in S_3\}$, of $\mathbb{R}^3$ and allows $r_{m_\pm,t+1}$ to be written in the mathematically precise form (11).

Since the expressions for the distributional properties of CSM returns in the general case are quite complex, we begin by considering the special case of 2 assets in the next section that admits simpler expressions that are easier to interpret.

4. Special case of two assets

As alluded to previously, potential computational difficulties associated with practical implementation of the expressions derived in this paper require some discussion. In particular, looking ahead at Theorem 2 shows that the expression for the density of CSM returns will consist of

$$\binom{n}{n - m_- - m_+,\ m_-,\ m_+} = \frac{n!}{(n - m_- - m_+)!m_-!m_+!}$$

distinct terms in the general case. For the CSM strategy that takes a long position in the top decile and a short position in the bottom decile from the S&P500 index, this corresponds to $500!/(400!(50!)^2)$ distinct terms which is very large. However, in many of the non-equity applications, the number of terms is much smaller. For example, the empirical study of the US industry return data in Sefton and Scowcroft (2004) correspond to $n = 10$, $m_+ = 2$, and $m_- = 2$ and results in 1,260 distinct terms in the expression for the CSM return density, and the study in Foltice and Langer (2015) correspond to $m_+ = 1$ and $m_- = 1$ in a universe of n stocks giving $n(n-1)$ distinct terms. In any case, it is shown in Section 6 that the returns generated by Monte Carlo simulation provide an accurate approximation to the CSM return densities, and is computationally feasible for practical applications, while analytic expressions enable theoretical investigation of the distributional properties of CSM returns.

It is difficult to infer much about the distributional properties of CSM returns in the general case without, for example, assuming (a) independence between $\boldsymbol{r}_t$ and $\boldsymbol{r}_{t+1}$, or (b) letting l_r or l_h become large, or (c) assuming very specific forms for the means and covariances. These special cases will be considered in detail in Theorems 3, 4, and Corollary 2, respectively. In this section, we consider, in the absence of any simplifying assumptions, the properties of the 2 asset CSM portfolio returns corresponding to the case $n = 2$ and $m_\pm = 1$. This special case exhibits many of the key features of general CSM portfolios, and provides useful insights into their behaviour.

Throughout this section, we denote by $(1\,2) \in S_2$ the identity permutation, and use the notation, following (6),

$$\begin{aligned}
&x_{(1\,2),u} = r_{2,u} - r_{1,u}, \qquad \mu_u = \mu_{2,u} - \mu_{1,u}, \qquad \sigma_{i,u}^2 = \operatorname{var}(r_{i,u}),\\
&\rho_u = \frac{\operatorname{cov}(r_{1,u}, r_{2,u})}{\sigma_{1,u}\sigma_{2,u}}, \qquad \rho_{i,j} = \frac{\operatorname{cov}\left(r_{i,t}, r_{j,t+1}\right)}{\sigma_{i,t}\sigma_{j,t+1}},\\
&\varsigma_u^2 = \sigma_{1,u}^2 + \sigma_{2,u}^2 - 2\rho_u\sigma_{1,u}\sigma_{2,u},
\end{aligned}$$

$$\varsigma_{t,t+1} = \rho_{1,1}\sigma_{1,t}\sigma_{1,t+1} + \rho_{2,2}\sigma_{2,t}\sigma_{2,t+1} - \rho_{1,2}\sigma_{1,t}\sigma_{2,t+1} - \rho_{2,1}\sigma_{2,t}\sigma_{1,t+1},$$
$$\varrho_{t,t+1} = \frac{\varsigma_{t,t+1}}{\varsigma_t\varsigma_{t+1}},$$

where $1 \le i, j \le 2$ and $u \in \{t, t+1\}$. The main motivation behind these definitions is that in the case of 2 assets, the ordering of the returns over the ranking period and the CSM return over the holding period, at least up to sign, are determined by the differences $r_{2,t} - r_{1,t}$ and $r_{2,t+1} - r_{1,t+1}$ respectively. In this light, note that μ_t and μ_{t+1} are the expected values of these differences in the returns of two assets over ranking and holding periods, and ς_t^2 and ς_{t+1}^2 are the corresponding variances. Moreover, $\rho_{i,j}$ represents the correlation between the asset i return over the ranking period and the asset j return over the holding period, $\varsigma_{t,t+1}$ is the covariance of $r_{2,t} - r_{1,t}$ and $r_{2,t+1} - r_{1,t+1}$, and $\varrho_{t,t+1}$ is the corresponding correlation. We begin by deriving the CSM return density, $f_{r_{m\pm,t+1}}$, in a more explicit form.

Proposition 2. *Let $n = 2$, $m_\pm = 1$, and suppose $(\mathbf{r}'_{t+1}, \mathbf{r}'_t)'$ satisfies Assumption 1. Then*

$$f_{r_{1\pm,t+1}}(r) = \phi_1\left(r; -\mu_{t+1}, \varsigma_{t+1}^2\right)\Phi_1\left[0; \mu_t - \frac{\varrho_{t,t+1}\varsigma_t}{\varsigma_{t+1}}(r + \mu_{t+1}), \varsigma_t^2\left(1 - \varrho_{t,t+1}^2\right)\right] + \phi_1\left(r; \mu_{t+1}, \varsigma_{t+1}^2\right)\Phi_1\left[0; -\mu_t - \frac{\varrho_{t,t+1}\varsigma_t}{\varsigma_{t+1}}(r - \mu_{t+1}), \varsigma_t^2\left(1 - \varrho_{t,t+1}^2\right)\right]. \quad (12)$$

Proof. Refer to A.8. □

It follows that the density of $r_{1\pm,t+1}$ is essentially a weighted sum of two univariate normals, but with weights depending on r, which are known to be from the unified skew-normal family of distributions[1] studied in Arellano-Valle and Azzalini (2006). It is worth noting that both weights either simultaneously approach 0 or 1 as $r \to \pm\infty$, depending on the sign of $\varrho_{t,t+1}$. Next we derive the expressions for the first four moments of $r_{1\pm,t+1}$.

Theorem 1. *If $n = 2$, $m_\pm = 1$, and $(\mathbf{r}'_{t+1}, \mathbf{r}'_t)'$ satisfies Assumption 1, then the first four central moments of $r_{1\pm,t+1}$ are as follows:*

$$\mu_1(r_{1\pm,t+1}) = \mu_{t+1}\left(2\Phi_1\left[\frac{\mu_t}{\varsigma_t}\right] - 1\right) + \frac{2\varsigma_{t,t+1}}{\varsigma_t}\phi_1\left(\frac{\mu_t}{\varsigma_t}\right),$$
$$\mu_2(r_{1\pm,t+1}) = \mu_{t+1}^2 + \varsigma_{t+1}^2,$$
$$\mu_3(r_{1\pm,t+1}) = \mu_{t+1}\left(\mu_{t+1}^2 + 3\varsigma_{t+1}^2\right)\left(2\Phi_1\left[\frac{\mu_t}{\varsigma_t}\right] - 1\right) + \frac{2\varsigma_{t,t+1}\left(\varsigma_{t,t+1}^2(\mu_t^2 - \varsigma_t^2) + 3\varsigma_t^2(\varsigma_t^2(\mu_{t+1}^2 + \varsigma_{t+1}^2) - \varsigma_{t,t+1}\mu_t\mu_{t+1})\right)}{\varsigma_t^5}\phi_1\left(\frac{\mu_t}{\varsigma_t}\right),$$
$$\mu_4(r_{1\pm,t+1}) = \mu_{t+1}^4 + 6\mu_{t+1}^2\varsigma_{t+1}^2 + 3\varsigma_{t+1}^4.$$

Proof. Refer to A.9. □

Direct calculation using the moments from Theorem 1 gives the following for the mean, variance, skewness, and the kurtosis of $r_{1\pm,t+1}$:

$$\mu_{r_{1\pm,t+1}} = \mu_{t+1}\left(2\Phi_1\left[\frac{\mu_t}{\varsigma_t}\right] - 1\right) + \frac{2\varsigma_{t,t+1}}{\varsigma_t}\phi_1\left(\frac{\mu_t}{\varsigma_t}\right), \quad (13)$$

$$\sigma_{r_{1\pm,t+1}}^2 = \varsigma_{t+1}^2 + \left[1 - (2\Phi_1[\mu_t/\varsigma_t] - 1)^2\right]\mu_{t+1}^2 \quad (14)$$

$$- \frac{4\varsigma_{t,t+1}\phi(\mu_t/\varsigma_t)}{\varsigma_t^2}(\varsigma_t\mu_{t+1}(2\Phi_1[\mu_t/\varsigma_t] - 1) + \phi_1(\mu_t/\varsigma_t)\varsigma_{t,t+1}),$$

$$\text{skew}_{r_{1\pm,t+1}} = \frac{8\Phi_1[\mu_t/\varsigma_t](\Phi_1[\mu_t/\varsigma_t] - 1)(2\Phi_1[\mu_t/\varsigma_t] - 1)\mu_{t+1}^3}{\sigma_{r_{1\pm,t+1}}^3} + \frac{12\varsigma_{t,t+1}\mu_{t+1}^2\phi_1(\mu_t/\varsigma_t)}{\sigma_{r_{1\pm,t+1}}^3\varsigma_t}(2\Phi_1[\mu_t/\varsigma_t] - 1)^2 \quad (15)$$

[1] Refer also to the discussion following Corollary 1 on the relationship between the CSM return density and unified skew-normal densities.

Table 1
Means, covariances, and auto-covariances in the monthly returns for MSFT and XOM over the period January 2013 to February 2018.

	Mean	Covariance		Autocorrelation	
		MSFT	XOM	MSFT	XOM
MSFT	0.024949	0.003666	0.000366	−0.281551	0.107471
XOM	0.003033	0.000366	0.001717	−0.004105	−0.199592

$$
\begin{aligned}
&+\frac{6\varsigma^2_{t,t+1}\mu_{t+1}\phi_1(\mu_t/\varsigma_t)}{\sigma^3_{r_{1\pm,t+1}}\varsigma_t^3}(4\varsigma_t\phi_1(\mu_t/\varsigma_t)(2\Phi_1[\mu_t/\varsigma_t]-1)-\mu_t)\\
&+\frac{2\varsigma^3_{t,t+1}\phi_1(\mu_t/\varsigma_t)}{\sigma^3_{r_{1\pm,t+1}}\varsigma_t^5}\left(\varsigma_t^2(8\phi_1^2(\mu_t/\varsigma_t)-1)+\mu_t^2\right),
\end{aligned}
$$

$$
\text{kurt}_{r_{1\pm,t+1}}=\frac{\mu^4_{t+1}+6\mu^2_{t+1}\varsigma^2_{t+1}+3\varsigma^4_{t+1}}{\sigma^4_{r_{1\pm,t+1}}}. \tag{16}
$$

In particular, these expressions imply that if expected returns μ_t and μ_{t+1} are positive, and autocorrelation $\rho_{t,t+1}$ is also positive, then $\mu_{r_{1\pm,t+1}} > \mu_{t+1}$ so that the CSM portfolio provides higher return than simply holding the asset with higher expected return long and the asset with lower expected return short. A somewhat unexpected result is that unlike other quantities, $\text{kurt}_{r_{1\pm,t+1}}$ is independent of the CSM autocorrelation term $\varrho_{t,t+1}$. Economic implications of these expressions are discussed in the next section.

However, before concluding this section, we provide a brief discussion of the case where $n=3$ and $m_+=m_-=1$. In this case, we would order the returns of the three assets over the ranking period, then take a long position in the asset with the highest return and short the one with the lowest return. As we see from (3), and the formula at the start of this section, the CSM return density will comprise of six terms in comparison to the two in (12). Moreover, each of these six terms is a product of a univariate density and a bivariate integral, as opposed to the latter being a univariate integral in (12). Already, similar intuitive analyses in this case are not immediately available, but we note that the CSM return density is still a weighted sum of univariate densities but, again, with weights depending on *r*. Fixed weights will only occur in special cases, for example when $\boldsymbol{r}_t$ and $\boldsymbol{r}_{t+1}$ are independent. These results in full generality are presented in Section 7.

5. Empirical implications

In this section, we apply the results from the previous section to explain some of the stylized features of CSM returns reported in the empirical literature, and in order to do this analytically, we restrict to the tractable 2-asset case. Firstly, we note that $\mu_{r_{1\pm,t+1}}$ is increasing in self-autocorrelations, $\rho_{1,1}$ and $\rho_{2,2}$, and decreasing in cross lagged correlations $\rho_{1,2}$ and $\rho_{2,1}$, often referred to as time-series momentum, as can be seen from (13). This result is consistent with the discussion in Moskowitz et al. (2012) p241. Furthermore, our result allows for the possibility that $\mu_{r_{1\pm,t+1}}$ may be positive when cross lagged correlations are negative, a result attributed to Lewellen (2002). Next, we show that the reversal in the expected CSM return with increasing holding period length can be accommodated by our results.

Proposition 3. *Let $\bar{\mu}_{t+1}=\mu_{t+1}/l_h$ and $\bar{\varsigma}_{t+1}=\varsigma_{t+1}/l_h$, where l_h is the length of the holding period. If $(2\Phi_1(\mu_t/\varsigma_t)-1)\bar{\mu}_{t+1}<0$ and $\varrho_{t,t+t}>0$, then the expected CSM return, $\mu_{r_{1\pm,t+1}}$, increases with holding period length for small l_h and decreases for large l_h with turning point occurring at*

$$
l_h^\star=\frac{\varrho^2_{t,t+t}\phi_1^2(\mu_t/\varsigma_t)\bar{\varsigma}^{-2}_{t+1}}{(2\Phi_1(\mu_t/\varsigma_t)-1)^2\bar{\mu}^2_{t+1}}. \tag{17}
$$

Proof. Refer to A.10. □

So, if there is a reversal in the ordering of returns from the ranking period to the holding period, and the autocorrelation, $\varrho_{t,t+t}$, is positive, then a reversal in the expected CSM return will occur. This is a pattern observed by many authors, with a turning point occurring at around 12 months, reported for example in Table 1 by Sefton and Scowcroft (2004).

As observed, *inter alia*, in Ramchmand and Susmel (2007), Longing and Solnik (2001), Ang and Bekaert (2002), Ang and Chen (2002) and Sancetta and Satchell (2007), a well-documented feature in market crashes is that the contemporaneous correlations tend to 1, so that in the 2-asset case $\rho_{t+1}\to 1$. We shall explore the implications of this to skewness and kurtosis, and see if these implications are consistent with the empirical literature on momentum return crashes. In particular, Daniel and Moskowitz (2016) find that momentum strategies with negative skewness and large kurtosis tend to crash.

Analytically, we note from the expression for $\text{skew}_{r_{1\pm,t+1}}$ in (15) that as $\rho_{t+1}\to 1$, the variance, ς^2_{t+1}, of $r_{2,t+1}-r_{1,t+1}$ will decrease which, *ceteris paribus*, will increase the magnitude of the skewness of CSM returns. If the skewness is already negative, it will become more negative as $\rho_{t+1}\to 1$, consistent with the findings in Daniel and Moskowitz (2016). We note that changes to ρ_{t+1} do not affect the numerators of the terms that appear in the expression for the skewness in (15).

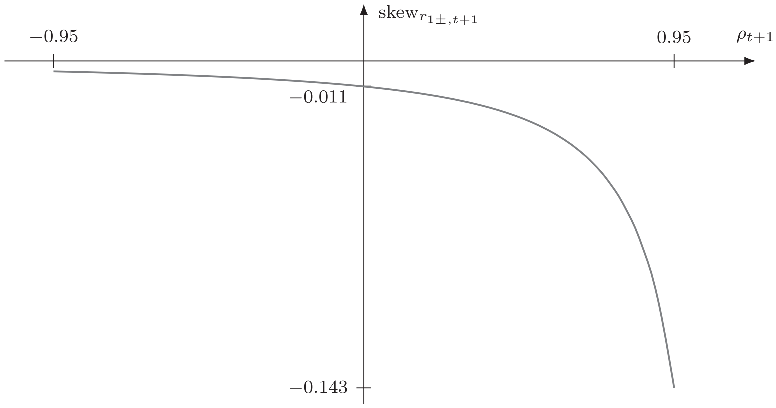

Fig. 1. Skewness of the CSM returns as a function of asset correlation in the holding period.

Next, we consider a numerical example using the parameters estimated from monthly returns on Microsoft (MSFT) and Exxon Mobil (XOM) from January 2013 to February 2018. The stationary means, covariances, and autocorrelations for the two assets are shown in Table 1. The impact on skewness of $\rho_{t+1} \to 1$, which occurs in the event of a market crash, is shown in Fig. 1. All parameters other than the asset correlation, ρ_{t+1}, in the holding period were kept fixed at values shown in Table 1, and ρ_{t+1} was varied from −0.95 to 0.95. The figure shows that the skewness of CSM return, which is negative in this case, becomes more pronounced as $\rho_{t+1} \to 1$. In terms of economic intuition, Sancetta and Satchell (2007) present a theory linking increases in market volatility to increases in ρ_{t+1} by assuming that returns follow a Sharpe 1-factor model with uncorrelated errors. Combined with our above discussion, this suggests that increases in market volatility can lead to increased negative skewness in CSM returns, which is a result familiar to practitioners.

We now consider the properties of kurtosis in the limiting cases as $\varsigma_{t+1} \to 0$ which correspond to $\sigma_{1,t+1} = \sigma_{2,t+1}$ and $\rho_{t+1} \to 1$, and $\varsigma_{t+1} \to \infty$ which corresponds to $\sigma_{1,t+1}, \sigma_{2,t+1} \to \infty$ and $|\rho_{t+1}| < 1$.

Proposition 4. *Let $n = 2$, and suppose $\sigma_{1,t+1} = \sigma_{2,t+1}$. Then the kurtosis given in (16) satisfies the following:*

$$\lim_{\rho_{t+1}\to 1} \text{kurt}_{r_{1\pm,t+1}} = \frac{1}{16\Phi_1^2[\mu_t/\varsigma_t](1-\Phi_1[\mu_t/\varsigma_t])^2}. \tag{18}$$

Proof. Refer to A.11. □

It is interesting to note from (18) that the CSM return distribution can exhibit both mesokurtosis and leptokurtosis, for example, as $\mu_t \to 0$ and $\mu_t \to \pm\infty$ respectively.

Moreover, in a market crash situation under the assumptions of Proposition 4, we see that kurtosis can become arbitrarily large if $\mu_{1,\,t} \neq \mu_{2,\,t}$ since the denominator of $\text{kurt}_{r_{1\pm,t+1}}$ tends to zero as $\rho_t \to 1$. It is interesting to note also that

$$\lim_{\varsigma_{t+1}\to 0} \text{kurt}_{r_{1\pm,t+1}} = \frac{3}{1-4\varrho_{t,t+t}^2\phi_1^2(\mu_t/\varsigma_t)} \geq \frac{3}{1-2\varrho_{t,t+t}^2/\pi}, \tag{19}$$

so that $r_{1\pm,t+1}$ is leptokurtic in general as $\varsigma_{t+1} \to \infty$, and the leptokurtosis increases as $\varsigma_t \to \infty$. That is, the distribution of $r_{1\pm,t+1}$ becomes progressively more leptokurtic with increasing asset volatility. This is consistent with existing literature on volatility dependence of CSM returns reported, for example, in Wang and Xu (2015).

In order to tie our theoretical implications with empirical observations, we consider the CSM returns based on the Fama-French "10 Portfolios Formed on Momentum" data over the period January 1927 to December 2017. CSM portfolios were constructed using the top and bottom deciles from this data, which can be considered as a special case with $n = 10$ and $m_\pm = 1$ in our framework, and the mean, standard deviation, skewness and the excess kurtosis computed from the resulting portfolio returns over various periods are shown in Table 2. Consistent with findings reported in the empirical literature, CSM portfolios provide a small positive expected return over all periods considered.[2] It is interesting to note that skewness is negative for all periods, indicating that this may be a characteristic feature of CSM returns, and that skewness is more negative in periods with significant financial turmoil such as the the Great Depression (contained in period 1927–1957) and the Global Financial Crisis (contained in period 1988–2017), which is consistent with our theoretical implications. Moreover, the data show that CSM returns are highly leptokurtic in general, and leptokurtosis is higher in periods of significant market uncertainty, which is, once again, consistent with our theoretical implications. Similar results, viz. negative skewness and leptokurtosis, were also found on a much smaller scale from forming CSM portfolios with $n = 3$ and $m_\pm = 1$ using returns on Microsoft (MSFT), Exxon Mobil (XOM), and General Electric (GE) over the period April 1986 to February 2018.

[2] It should be noted that the periods considered were quite long with each being approximately 30 years in length.

Table 2
Mean, standard deviation, skewness and excess kurtosis on CSM returns computed using the Fama-French “10 Portfolios Formed on Momentum” data from January 1927 to December 2017. All periods in the table begin in January and end in December.

Period	1927 – 1957	1958 – 1987	1988 – 2017	**1927 – 2017**
Mean	0.0053	0.0128	0.0064	**0.0081**
Standard dev.	0.1006	0.0473	0.0705	**0.0763**
Skewness	−4.3015	−1.3551	−2.9045	**−4.2553**
	32.2000	6.2651	19.4015	**39.0622**

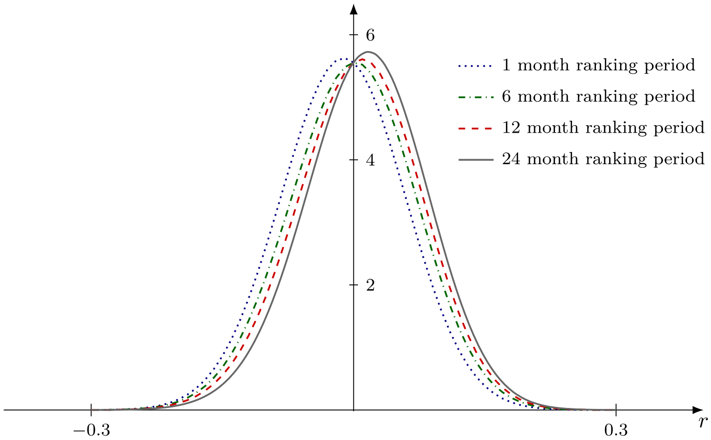

Fig. 2. Densities of 2-asset CSM returns with holding period fixed at 1 month and ranking period ranging from 1 month to 24 months. Returns over all holding periods were normalized to represent monthly returns.

Finally, although we would expect that analogous explicit expressions for the mean, variance, skewness and kurtosis are not readily available in the general case, the empirical results examined in this section should generalize given that the broad structure of the CSM return density and its moments remain unchanged.

6. Numerical examples

We begin this section by examining the different shapes of CSM return densities that can be obtained by considering the impact of increasing either the ranking or the holding period while keeping the other fixed. To ensure realistic parameter values, we use the stationary means, covariances, and autocorrelations for Microsoft and Exxon Mobil from Table 1.

In Fig. 2, the holding period was set to 1 month and the ranking period varied from 1 month to 24 months. It shows that the CSM return density is approximately normal and becomes progressively more normal with increasing ranking period. In Fig. 3, the ranking period was set to 1 month while the holding period was varied from 1 month to 24 months, and shows that the CSM return density is approximately normal for short holding periods, but progressively becomes more skewed and bimodal with increasing holding period. It will be established, in Theorem 4, that both of these observations are consistent with the theoretical framework of this paper. Although the former case, where the ranking period is longer relative to the holding period, is more common in practice, the latter is equivalent to a situation where there is a sharp divergence in expected returns from ranking to holding periods, and serves to illustrate some of the interesting shapes for CSM return densities that can arise. Since the CSM return density is effectively a sum of univariate skew-normal densities,[3] when the means of the component densities are sufficiently displaced relative to their standard deviations, the resulting combined density will be bimodal, and multimodal in general.

As discussed in Section 4, it will follow from Theorem 2 that the analytic expression for CSM return density becomes computationally infeasible as the number of assets, n, becomes large, and so we demonstrate a computationally efficient method for computing the densities using Monte Carlo simulation. A comparison of the densities, computed analytically and by Monte Carlo simulation, in the case of the 2-asset example with Microsoft and Exxon Mobil with 1 month ranking period and 24 month holding period is shown in Fig. 4. Monte Carlo simulation used 262,143 paths and the time taken was approximately 79.7 ms on a PC with Core i5-4750 CPU @ 3.20 GHz and 8 GB of RAM. As the figure shows, the density obtained from simulation closely mirrors the analytical counterpart, and could be used for computing densities and moments in practical situations.

[3] Refer to the discussion following Corollary 1 on the relationship between CSM return density and unified skew-normal densities.

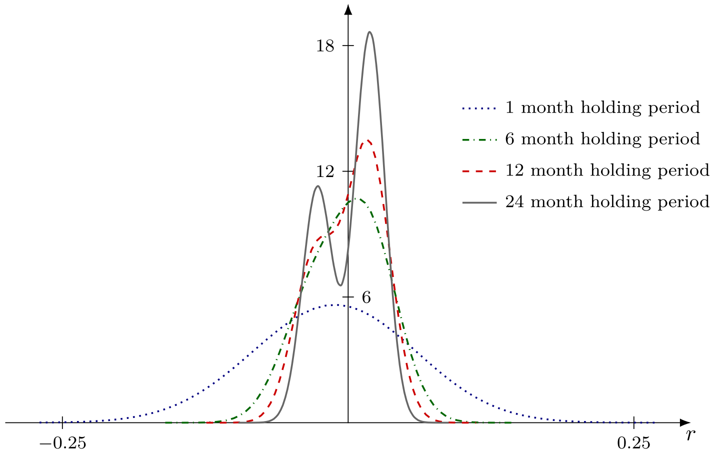

Fig. 3. Densities of 2-asset CSM returns with ranking period fixed at 1 month and holding period ranging from 1 month to 24 months. Returns over all holding periods were normalized to represent monthly returns.

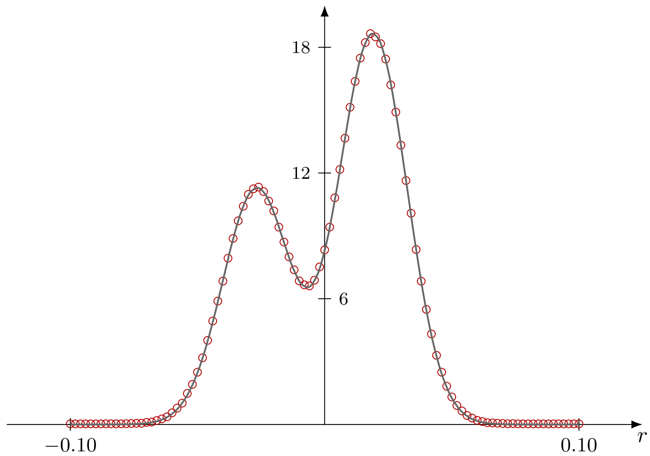

Fig. 4. Comparison of analytical and Monte Carlo densities for 2 assets. Solid curve is the analytic density, and the curve traced out by ◦ is the density computed by Monte Carlo.

7. Distribution of CSM returns in the general case

In this section, we derive the distributional properties of CSM returns in the general case. We begin with a theorem that expresses the density of CSM returns, based on m_+ equally weighted long positions and m_- equally weighted short positions from a universe of n assets, as the sum over the permutations in S_n of terms that are essentially the product of a univariate normal density function and an $(n-1)$-dimensional cumulative normal distribution function. As noted earlier, this could be generalized to two arbitrary fixed-weight portfolios each of whose weights add up to one.

For any $\tau \in S_n$, let $P_\tau \in \mathbb{R}^{n\times n}$ be the permutation matrix corresponding to τ, and let $D_{n-1} \in \mathbb{R}^{(n-1)\times n}$ be the matrix defined by

$$D_{n-1} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}. \tag{20}$$

Moreover, for any $m_-, m+ \in \mathbb{N}$ such that $m_- + m_+ \leq n$, let

$$\iota_{m_\pm} = \left(\frac{1}{m_+}\mathbf{1}'_{m_+}, 0, \ldots, 0, -\frac{1}{m_-}\mathbf{1}'_{m_-}\right)' \in \mathbb{R}^n \tag{21}$$

where $\mathbf{1}_{m_\pm} = (1, 1, \ldots, 1)' \in \mathbb{R}^{m_\pm}$

Theorem 2. *Let $f_{r_{m\pm,t+1}}(r)$ be the probability density of the (m_+, m_-)-cross sectional momentum return $r_{m\pm,t+1}$. If $(\mathbf{r}'_{t+1}, \mathbf{r}'_t)'$ satisfies Assumption 1, then we have*

$$f_{r_{m\pm,t+1}}(r) = \sum_{\tau \in S_n} \phi_1\left(r; \mu_{\tau,m_\pm,t+1}, \varsigma^2_{r,\tau,t+1}\right) \Phi_{n-1}\left[\mathbf{0}; \boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}\left(r - \mu_{\tau,m_\pm,t+1}\right)\boldsymbol{\lambda}_{\tau,t}, \Lambda_{\boldsymbol{x}|r,\tau,t}\right], \tag{22}$$

where $\mu_{\tau,m_\pm,t+1} = \boldsymbol{\iota}'_{m_\pm} P_\tau \boldsymbol{\mu}_{t+1}$, $\boldsymbol{\mu}_{\tau,t} = D_{n-1} P_\tau \boldsymbol{\mu}_t$,

$$\varsigma^2_{r,\tau,t+1} = \boldsymbol{\iota}'_{m_\pm} P_\tau \Lambda_{t+1,t+1} P'_\tau \boldsymbol{\iota}_{m_\pm},$$

$$\boldsymbol{\lambda}_{\tau,t} = D_{n-1} P_\tau \Lambda_{t,t+1} P'_\tau \boldsymbol{\iota}_{m_\pm},$$

$$\Lambda_{\boldsymbol{x}|r,\tau,t} = D_{n-1} P_\tau \Lambda_{t,t} P'_\tau\ D'_{n-1} - \varsigma^{-2}_{r,\tau,t+1} \boldsymbol{\lambda}_{\tau,t} \boldsymbol{\lambda}'_{\tau,t}.$$

Proof. Refer to A.2. □

The intuition behind the expression in (22) is that there are $n!$ possible orderings of asset returns over the ranking period, corresponding to the elements of S_n, and these rankings determine a partition of $\mathbb{R}^n$ into $n!$ mutually disjoint subsets. The summand in (22) corresponding to $\tau \in S_n$ is then the CSM return density at r conditional on $\boldsymbol{r}_t$ being in the element of the partition of $\mathbb{R}^n$ corresponding to the permutation τ.

Although compact in form, expression (22) is not convenient for the purposes of computing the moments of $r_{m\pm,t+1}$, since the variable r appears in both the univariate normal densities and their associated weights. Using an alternative decomposition of $f_{r_{\tau,m_\pm,t+1}, \boldsymbol{x}_{\tau,t}}(r, \boldsymbol{x})$, by conditioning $r_{m_\pm,t+1}$ on $\boldsymbol{x}_{\tau,\,t}$, allows the elimination of the dependence of r from the weights and gives the following equivalent expression for $f_{r_{m_\pm,t+1}}(r)$.

Corollary 1. *If the assumption in Theorem 2 is satisfied, then*

$$f_{r_{\tau,m\pm,t+1}}(r) = \sum_{\tau \in S_n} \int_{\boldsymbol{x} \in \mathbb{R}^{n-1}_-} \phi_1\left(r; \mu_{\tau,m\pm,t+1|\boldsymbol{x}}, \varsigma^2_{r|\boldsymbol{x},\tau,t+1}\right) \phi_{n-1}\left(\boldsymbol{x}; \boldsymbol{\mu}_{\tau,t}, \Lambda_{\boldsymbol{x},\tau,t}\right) d\boldsymbol{x}, \tag{23}$$

where $\Lambda_{\boldsymbol{x},\tau,t} = D_{n-1} P_\tau \Lambda_{t,t} P'_\tau\ D'_{n-1}$, $\varsigma^2_{r|\boldsymbol{x},\tau,t+1} = \varsigma^2_{r,\tau,t+1} - \boldsymbol{\lambda}'_{\tau,t} \Lambda^{-1}_{\boldsymbol{x},\tau,t} \boldsymbol{\lambda}_{\tau,t}$, *and*

$$\mu_{\tau,m_\pm,t+1|\boldsymbol{x}} = \mu_{\tau,m_\pm,t+1} + \boldsymbol{\lambda}'_{\tau,t} \Lambda^{-1}_{\boldsymbol{x},\tau,t} (\boldsymbol{x} - \boldsymbol{\mu}_{\tau,t})$$

Proof. Refer to A.3. □

What insights can we glean from these rather formidable formulae, especially the expression for $f_{r_{m\pm,t+1}}(r)$ in (22)? Firstly, we can consider special cases of these expressions that simplify considerably, as we do in Proposition 2, Theorem 3 and Corollary 2. However, can we also get insights from the general formulae directly? We can answer this by an intuitive argument that allows us to decompose, for example, the expression in (22) as follows. Since $f_{r_{m\pm,t+1}}(r)$ is a probability density function (pdf), integrating the left-hand side of (22) over $\mathbb{R}$ gives 1. Next, all $n!$ terms on the right-hand side are non-negative by inspection, and so the integral of each term is non-negative and bounded by 1. Hence we can re-scale each term to convert it into a pdf and the overall result will be a mixture of these pdfs. Each term then bears a strong resemblance to a skew-normal, and so we can assert that the pdf of CSM returns is a mixture of pdfs closely related to skew-normals, thus, intuitively the general formula gives us insights into how the pdf of CSM return density is built up from such densities, where each component arises from an ordering of the underlying asset returns over the ranking period. Indeed, we can say more. We see that the individual components of our mixture distribution are members of what is known as the unified skew-normal (SUN) family of distributions as defined in Arellano-Valle and Azzalini (2006). Thus we can identify our CSM distribution as a mixture of SUN densities. Much is known about these densities which have been introduced in the literature to capture multivariate non-normality. Intuitively, this tells us that we would expect skewness and excess kurtosis in CSM returns, which is consistent with well-established empirical findings, but have not previously been established analytically. Finally, the very fact that $f_{r_{m\pm,t+1}}(r)$ is a mixture tells us that multi-modality can be expected.

These results, although useful as theoretical descriptions of the CSM return density, are not particularly suited for the understanding of empirical results. This is for a number of reasons. Firstly, for a reasonable size, n, of underlying assets, the density consists of a large number of terms as discussed in Section 4. Secondly, the representation of the density is not in terms of the more fundamental parameters that would be estimated in practice. However, in special cases, we can derive simpler and practically meaningful expressions such as those obtained in Section 4, and analysed in Section 5, for the 2-asset case, which also provide useful insights.

We therefore consider other special cases that will shed light on Theorem 2. Firstly, the density, $f_{r_{m\pm,t+1}}(r)$, simplifies significantly when the asset returns have the same mean and pairwise covariances. This result is of some interest since it gives conditions under which the CSM return density is normal.

Corollary 2. *If $(\mathbf{r}'_{t+1}, \mathbf{r}'_t)'$ satisfies Assumption 1, and the means and covariances in (9) take the degenerate form*

$$\boldsymbol{\mu}_u = \mu_u \mathbf{1}_n, \quad \Lambda_{u,v} = \delta_{u,v}\sigma_u\sigma_v(1 - \rho_{u,v}) I_{n\times n} + \rho_{u,v}\sigma_u\sigma_v \mathbf{1}_n \mathbf{1}'_n, \tag{24}$$

where $\mu_u \in \mathbb{R}$, $\sigma_u \in \mathbb{R}_+$, $\rho_{u,v} \in (-1, 1)$, and $u, v \in \{t, t+1\}$, then in the notation of Theorem 2 we have

$$f_{r_{m\pm,t+1}}(r) = \phi_1\left(r; 0, \varsigma^2_{r,\mathrm{id},t+1}\right),$$

where id $\in S_n$ *is the identity element that leaves all asset indices unchanged.*

Proof. Refer to A.4. □

Another special case in which $f_{r_{m\pm,t+1}}(r)$ simplifies considerably is when $\mathbf{r}_t$ and $\mathbf{r}_{t+1}$ are independent.

Theorem 3. *If* $(\mathbf{r}'_{t+1}, \mathbf{r}'_t)'$ *satisfies Assumption 1, and* $\mathbf{r}_t$ *and* $\mathbf{r}_{t+1}$ *are independent, then*

$$f_{r_{m\pm,t+1}}(r) = \sum_{\tau\in S_n} \phi_1\left(r;\, \mu_{\tau,m_\pm,t+1},\, \varsigma^2_{r,\tau,t+1}\right)\Phi_{n-1}\left[\mathbf{0};\, \boldsymbol{\mu}_{\tau,t},\, D_{n-1}P_\tau \Lambda_{t,t}P'_\tau\ D'_{n-1}\right], \tag{25}$$

and $\sum_{\tau\in S_n}\Phi_{n-1}\left[\mathbf{0};\, \boldsymbol{\mu}_{\tau,t},\, D_{n-1}P_\tau\Lambda_{t,t}P'_\tau\ D'_{n-1}\right] = 1.$

Proof. Since $\boldsymbol{\lambda}_{\tau,t} = \mathbf{0}_{n-1}$ in this case, (25) follows immediately from (22), and the final statement follows from integrating both sides of (25) over $r \in \mathbb{R}$. □

The decomposition in (25) is of particular interest, since it expresses $f_{r_{m\pm,t+1}}(r)$ as a mixture of univariate normal densities with weights summing to unity. The next result establishes the impact of the lengths of the ranking and holding periods, l_r and l_h respectively, on the distribution of the CSM returns.

Theorem 4. *Let* l_r *and* l_h *and the lengths of ranking and holding periods respectively. Suppose* $\mu_{i,\,t} \neq \mu_{j,\,t}$ *if* $i \neq j$*, and by reordering the indices if necessary assume* $\mu_{1,t} > \mu_{2,t} > \ldots > \mu_{n,t}$.

(a) *If* l_h *is fixed, and* $(1, 2, \ldots, n) \in S_n$ *denotes the identity permutation, then*

$$\lim_{l_r\to\infty} f_{r_{m\pm,t+1}}(r) \to \phi_1\left(r;\, \mu_{(1,2,\ldots,n),m_\pm,t+1},\, \varsigma^2_{r,\mathrm{id},t+1}\right). \tag{26}$$

That is, the distribution of $r_{m\pm,t+1}$ *tends to univariate normal as* $l_r \to \infty$.

(b) *If* l_r *is fixed, then*

$$\lim_{l_h\to\infty} f_{r_{m\pm,t+1}}(r) \to \sum_{\tau\in U}\phi_1\left(r;\, \mu_{\tau,m_\pm,t+1},\, \varsigma^2_{r,\tau,t+1}\right) + \sum_{\tau\in V}\phi_1\left(r;\, \mu_{\tau,m_\pm,t+1},\, \varsigma^2_{r,\tau,t+1}\right)\Phi_{n-1}\left[\mathbf{0};\, \boldsymbol{\mu}_{\tau,t},\, \Lambda_{\mathbf{x}|r,\tau,t}\right], \tag{27}$$

where $U = \{\tau \in S_n \mid \mu_{\tau,m_\pm,t+1}\boldsymbol{\lambda}_{\tau,t} \succ \mathbf{0}\}$ *and* $V = \{\tau \in S_n \mid \mu_{\tau,m_\pm,t+1}\boldsymbol{\lambda}_{\tau,t} = \mathbf{0}\}$. *In particular, if* $\mathbf{r}_t$ *and* $\mathbf{r}_{t+1}$ *are independent and* $\mu_{\tau,m\pm,t+1} \neq 0$ *for some* $\tau \in S_n$*, then* $f_{r_{m\pm,t+1}}(r)$ *becomes multimodal with one mode approaching* ∞ *and another approaching* $-\infty$.

Proof. Refer to A.5. □

There are numerous examples in the literature on ranking periods of up to three years and similar lengths of time for holding periods, including, for example, Jegadeesh and Titman (1993), Chan et al. (1996) and Sefton and Scowcroft (2004). There may be some interesting simplifications if we used one or two factor models as in Grundy and Martin (2001). Their structure assumes cross-sectional randomness in parameters which would extend and complicate Theorem 2, but the idea needs exploring further and in future work we hope to look at the implications of a Sharpe 1-factor model in analysing Theorem 2.

The moments of CSM returns in the general case are now presented in Proposition 5. For any $p \in \mathbb{N}$, denote by $\mu_p(r_{m\pm,t+1})$ the p-th moment of $r_{m\pm,t+1}$. Then $\mu_p(r_{m\pm,t+1})$ can be obtained in terms of the moments of truncated multivariate normal distributions.

Proposition 5. *If* $(\mathbf{r}'_{t+1}, \mathbf{r}'_t)'$ *satisfies Assumption 1, then*

$$\mu_p(r_{m\pm,t+1}) = \sum_{\tau\in S_n}\sum_{\substack{k=0,\\ k\text{ even}}}^{p}\binom{p}{k}(k-1)!!\,\varsigma^k_{r|\mathbf{x},\tau,t+1}\sum_{l=0}^{p-k}\binom{p-k}{l}\mu^{p-k-l}_{\tau,m\pm,t+1}$$
$$\int_{\mathbf{x}\in\mathbb{R}^{n-1}_-}\left(\boldsymbol{\lambda}'_{\tau,t}\Lambda^{-1}_{\mathbf{x},\tau,t}\left(\mathbf{x}-\boldsymbol{\mu}_{\tau,t}\right)\right)^l\phi_{n-1}\left(\mathbf{x};\,\boldsymbol{\mu}_{\tau,t},\,\Lambda_{\mathbf{x},\tau,t}\right)d\mathbf{x}, \tag{28}$$

where $\Lambda_{\mathbf{x},\,\tau,\,t}$ *and* $\varsigma^2_{r|\mathbf{x},t+1,\tau}$ *are as defined in Corollary 1.*

Proof. Refer to A.6. □

In the special case where $\mathbf{r}_t$ and $\mathbf{r}_{t+1}$ are independent, a more explicit expression for $\mu_p(r_{m\pm,t+1})$ can be obtained.

Corollary 3. *If* $(\mathbf{r}'_{t+1}, \mathbf{r}'_t)'$ *satisfies Assumption 1, and* $\mathbf{r}_t$ *and* $\mathbf{r}_{t+1}$ *are independent, then*

$$\mu_p(r_{m\pm,t+1}) = \sum_{\tau\in S_n}\Phi_{n-1}\left[\mathbf{0};\,\boldsymbol{\mu}_{\tau,t},\,\Lambda_{\mathbf{x},\tau,t}\right]\sum_{\substack{k=0,\\ k\text{ even}}}^{p}\binom{p}{k}(k-1)!!\,\varsigma^k_{r,\tau,t+1}\mu^{p-k}_{\tau,m_\pm,t+1}. \tag{29}$$

Proof. Refer to A.7. □

8. Conclusion

In this paper, analytic expressions for the density and the moments of cross sectional momentum (CSM) returns were derived under the assumption that underlying asset returns are multivariate normal. The general case is quite involved, and can be described as a mixture of the unified skew-normal family of distributions. The weights of the mixture correspond to different regime probabilities and could be analysed empirically although it would be a computationally challenging task. Whilst the general case sheds light on some features of CSM returns, we can analyse the density under certain special conditions such as those considered in Corollary 1, Theorems 3, and 4.

In the case of two assets, one long and one short, it is possible to analyse the problem fairly succinctly, and we can identify many of the features reported in the empirical literature. Generalizations of the results to other return processes, such as elliptical distributions, are possible. Linkages between momentum returns to market states, which has been the subject of empirical research, for example by Chordia and Shivakumar (2002) and Cooper et al. (2004), could also be considered using an extension of the approach introduced in this paper.

Appendix A. Proofs

A1. Preliminary results

We begin with some known results on the multivariate normal distribution.

Theorem 5. *Let $n_1, n_2 \in \mathbb{N}$ and suppose $\boldsymbol{x} \sim \mathcal{N}_{n_1+n_2}(\boldsymbol{\mu}, \Sigma)$, where*

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}, \tag{A.1}$$

with $\boldsymbol{x}_i, \boldsymbol{\mu}_i \in \mathbb{R}^{n_i}$ and $\Sigma_{i,j} \in \mathbb{R}^{n_i \times n_j}$ for $1 \le i, j \le 2$, and Σ positive definite. Then the conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2$ is normal with mean and covariance

$$\boldsymbol{\mu}_{\boldsymbol{x}_1|\boldsymbol{x}_2} = \boldsymbol{\mu}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \tag{A.2}$$

$$\Sigma_{\boldsymbol{x}_1|\boldsymbol{x}_2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}, \tag{A.3}$$

respectively, and $\phi_{n_1+n_2}(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)$ decomposes as

$$\phi_{n_1+n_2}(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = \phi_{n_1}(\boldsymbol{x}_1; \boldsymbol{\mu}_{\boldsymbol{x}_1|\boldsymbol{x}_2}, \Sigma_{\boldsymbol{x}_1|\boldsymbol{x}_2})\phi_{n_2}(\boldsymbol{x}_2; \boldsymbol{\mu}_2, \Sigma_{2,2}). \tag{A.4}$$

Proof. Refer to Muirhead (1982) Theorem 1.2.11. □

Recall that the permutation group S_n acts naturally on the set of polynomials, $\mathbb{R}[x_1, \ldots, x_n]$ by the rule

$$\tau p(x_1, \ldots, x_n) = p(x_{\tau_1}, \ldots, x_{\tau_n})$$

for any polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ and $\tau \in S_n$. Now, let $n_1, n_2 \in \mathbb{N}$ be fixed, and let $p_{n_1,2n_2} \in \mathbb{R}[x_1, \ldots, x_{n_1+2n_2}]$ be the polynomial

$$p_{n_1,2n_2}(x_1, \ldots, x_{n_1+2n_2}) = \left(\prod_{i=1}^{n_1} x_i\right)\left(\prod_{i=1}^{n_2}(x_{n_1+2i} - x_{n_1+2i-1})^2\right). \tag{A.5}$$

Denote by $Z(n_1, 2n_2)$ the stabilizer of $p_{n_1,2n_2}$ under the action of $S_{n_1+2n_2}$ so that

$$Z(n_1, 2n_2) = \{\tau \in S_{n_1+2n_2} \mid \tau p_{n_1,2n_2} = p_{n_1,2n_2}\}, \tag{A.6}$$

and let $Q(n_1, 2n_2) = S_{n_1+2n_2}/Z(n_1, 2n_2)$ be the quotient group,[4] with elements of $Q(n_1, 2n_2)$ identified with their coset representatives $\tau \in S_{n_1+2n_2}$. Finally, define

$$\{1, 2, \ldots, n\}^m = \prod_{i=1}^{m}\{1, 2, \ldots, n\} = \{(i_1, i_2, \ldots, i_m) \mid 1 \le i_j \le n, \ 1 \le j \le m\} \tag{A.7}$$

as the m-fold Cartesian product of $\{1, 2, \ldots, n\}$.

Theorem 6. *Let $\boldsymbol{x} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, where $n \in \mathbb{N}$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ is positive definite. Given $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_m) \in \{1, 2, \ldots, n\}^m$, if the $\boldsymbol{\kappa}$-th moment of $\boldsymbol{x}$ is defined by*

$$\mu_{\boldsymbol{\kappa}}(\boldsymbol{x}) = \mathbb{E}\left[\prod_{i=1}^{m} x_{\kappa_i}\right], \tag{A.8}$$

[4] Refer to Rotman (1995) Chapter 2 for the details on quotient groups.

then

$$\mu_{\boldsymbol{\kappa}}(\boldsymbol{x}) = \sum_{\substack{k,l\in\mathbb{N}\\ k+2l=m}} \sum_{\tau\in Q(k,2l)} \left(\prod_{i=1}^{k}\mu_{\kappa_{\tau_i}}\right)\left(\prod_{i=1}^{l}\Sigma_{\kappa_{\tau_{k+2i-1}},\kappa_{\tau_{k+2i}}}\right). \tag{A.9}$$

Proof. Refer to Withers (1985) Theorem 1.1. □

Corollary 4. *Let $x \sim \mathcal{N}_1(\mu, \sigma^2)$. Then for any $m \in \mathbb{N}$, the m-th moment of x is given by*

$$\mu_m(x) = \sum_{\substack{i=0\\ i\,even}}^{m} \binom{m}{i}(i-1)!!\,\sigma^i\mu^{m-i}, \tag{A.10}$$

where $k!! = \prod_{i=0}^{\lfloor \frac{k}{2}\rfloor}(k-2i)$ is the double factorial of $k \in \mathbb{N}$.

Proof. Follows from Theorem 6, since the inner sum in (A.9), corresponding to $2l = i$, consists of $\binom{m}{i}(i-1)!!$ identical terms all equal to $\sigma^i\mu^{m-i}$. □

Processes that play a key role in the investigation of CSM returns are $\boldsymbol{z}_{\tau,t}$ defined in (8), and the next result provides the distribution of these processes under Assumption 1.

Lemma 1. *If the return process $\boldsymbol{r}_t$ satisfies Assumption 1, then $\boldsymbol{z}_{\tau,t}$ follows a multivariate normal distribution for all $\tau \in S_n$ and $t \in \mathbb{N}$. In particular, $\boldsymbol{z}_{\tau,t}$ is multivariate normal if $\boldsymbol{r}_t \sim$ VARMA(p, q), and either $(\boldsymbol{r}'_0, \dots, \boldsymbol{r}'_{p-1})'$ is multivariate normal and independent of $\boldsymbol{\varepsilon}_t$, or the VARMA$(p, q)$ process satisfies conditions for weak stationarity.*

Proof. For any $\tau \in S_n$, let P_τ be the $n\times n$ permutation matrix corresponding to τ, let $\boldsymbol{\iota}_{m_\pm}$ be as defined in (21), and let $D_{n-1} \in \mathbb{R}^{(n-1)\times n}$ be as defined in (20). Then $\boldsymbol{x}_{\tau,t} = D_{n-1}P_\tau\boldsymbol{r}_t$ and $r_{\tau,m_\pm,t+1} = \boldsymbol{\iota}'_{m_\pm}P_\tau\boldsymbol{r}_{t+1}$, and so $\boldsymbol{z}_{\tau,t}$ can be written as

$$\boldsymbol{z}_{\tau,t} = \begin{bmatrix} r_{\tau,m_\pm,t+1}\\ \boldsymbol{x}_{\tau,t}\end{bmatrix} = \begin{bmatrix} \boldsymbol{\iota}'_{m_\pm}P_\tau & O_{1\times n}\\ O_{(n-1)\times n} & D_{n-1}P_\tau\end{bmatrix}\begin{bmatrix}\boldsymbol{r}_{t+1}\\ \boldsymbol{r}_t\end{bmatrix}. \tag{A.11}$$

Hence, $\boldsymbol{z}_{\tau,t}$ is a linear transformation of $\left(\boldsymbol{r}'_{t+1}, \boldsymbol{r}'_t\right)'$, which is assumed to be multivariate normal, and so $\boldsymbol{z}_{\tau,t}$ is itself multivariate normal. □

Recall that under Assumption 1, $(\boldsymbol{r}'_{t+1}, \boldsymbol{r}'_t)'$ is distributed according to (9) and, as discussed in Section 2, the 2-period model with such returns is quite general since models with longer ranking and holding periods can be transformed to a model of this type. The decomposition of $\boldsymbol{z}_{\tau,t}$ in (A.11) then establishes the next result.

Proposition 6. *Let $(\boldsymbol{r}'_{t+1}, \boldsymbol{r}'_t)'$ be given by (9). Then for any $\tau \in S_n$ and $t \in \mathbb{N}$,*

$$\boldsymbol{z}_{\tau,t} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\iota}'_{m_\pm}P_\tau\boldsymbol{\mu}_{t+1}\\ D_{n-1}P_\tau\boldsymbol{\mu}_t\end{bmatrix}, \begin{bmatrix}\boldsymbol{\iota}'_{m_\pm}P_\tau\Lambda_{t+1,t+1}P'_\tau\boldsymbol{\iota}_{m_\pm} & \boldsymbol{\iota}'_{m_\pm}P_\tau\Lambda_{t+1,t}P'_\tau D'_{n-1}\\ D_{n-1}P_\tau\Lambda_{t,t+1}P'_\tau\boldsymbol{\iota}_{m_\pm} & D_{n-1}P_\tau\Lambda_{t,t}P'_\tau D'_{n-1}\end{bmatrix}\right). \tag{A.12}$$

A2. Proof of Theorem 2

For each permutation $\tau \in S_n$, define a subset $\mathcal{R}_\tau \subset \mathbb{R}^n$ by

$$\mathcal{R}_\tau = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_{\tau_{i+1}} - x_{\tau_i} < 0\right\}. \tag{A.13}$$

Then $\mathcal{R}_\tau \cap \mathcal{R}_\upsilon = \varnothing$ for $\tau \neq \upsilon \in S_n$, and $\mathbb{R}^n = \cup_{\tau\in S_n}\mathcal{R}_\tau$ up to a set of measure zero, and so

$$\begin{aligned} f_{r_{m_\pm,t+1}}(r) &= \int_{\boldsymbol{r}\in\mathbb{R}^n} f_{r_{m_\pm,t+1},\boldsymbol{r}_t}(r,\boldsymbol{r})\,d\boldsymbol{r}\\ &= \sum_{\tau\in S_n}\int_{\boldsymbol{r}\in\mathcal{R}_\tau} f_{r_{m_\pm,t+1},\boldsymbol{r}_t}(r,\boldsymbol{r})\,d\boldsymbol{r} = \sum_{\tau\in S_n}\int_{\boldsymbol{r}\in\mathcal{R}_\tau} f_{r_{\tau,m_\pm,t+1},\boldsymbol{r}_t}(r,\boldsymbol{r})\,d\boldsymbol{r}, \end{aligned}$$

where the final equality follows from the fact that the summands of $r_{m_\pm,t+1}$ in (11), other than the term $\mathbb{I}_{\{\boldsymbol{x}_{\tau,t}\prec\boldsymbol{0}_{n-1}\}}r_{\tau,m_\pm,t+1}$, vanish on $\mathcal{R}_\tau$. Now, for each $\tau \in S_n$ consider the integral

$$\mathcal{I}_\tau(r) = \int_{\boldsymbol{r}\in\mathcal{R}_\tau} f_{r_{\tau,m_\pm,t+1},\boldsymbol{r}_t}(r,\boldsymbol{r})\,d\boldsymbol{r},$$

and let $g_\tau : \mathbb{R}^n \to \mathbb{R}^n$ be defined by

$$g_\tau(\boldsymbol{r}) = \begin{bmatrix} 1 & O_{1\times(n-1)}\\ O_{(n-1)\times 1} & D_{n-1}\end{bmatrix}P_\tau\boldsymbol{r} = \begin{bmatrix} r_{\tau_1}\\ \boldsymbol{x}_\tau\end{bmatrix},$$

where $\boldsymbol{x}_\tau = \left(x_{\tau_1}, \ldots, x_{\tau_{n-1}}\right)'$ and $x_{\tau_i} = r_{\tau_{i+1}} - r_{\tau_i}$. Then g_τ is invertible, and the Jacobian of g_τ is 1. So making the change of variables from $\boldsymbol{r}$ to $(r_{\tau_1}, \boldsymbol{x}'_\tau)'$, we obtain

$$\mathcal{I}_\tau(r) = \int_{(r_{\tau_1}, \boldsymbol{x}'_\tau)' \in \mathbb{R}\times\mathbb{R}_-^{n-1}} f_{r_{\tau,m_\pm,t+1}, r_{\tau_1}, \boldsymbol{x}_{\tau,t}}(r, s, \boldsymbol{x}_\tau)\, ds\, d\boldsymbol{x}_\tau,$$

since $x_{\tau_i} < 0$ on $\mathcal{R}_\tau$ for $1 \le i \le n-1$. Applying the properties of conditional densities gives

$$\begin{aligned}\mathcal{I}_\tau(r) &= \int_{\boldsymbol{x}_\tau \in \mathbb{R}_-^{n-1}} \int_{-\infty}^{\infty} f_{r_{\tau_1}, \boldsymbol{x}_{\tau,t} | r_{\tau,m_\pm,t+1}}(s, \boldsymbol{x}_\tau \,|\, r)\, f_{r_{\tau,m_\pm,t+1}}(r)\, ds\, d\boldsymbol{x}_\tau \\ &= f_{r_{\tau,m_\pm,t+1}}(r) \int_{\boldsymbol{x}_\tau \in \mathbb{R}_-^{n-1}} \mathcal{J}_\tau(\boldsymbol{x}_\tau, r) f_{\boldsymbol{x}_{\tau,t} \,|\, r_{\tau,m_\pm,t+1}}(\boldsymbol{x}_\tau \,|\, r)\, d\boldsymbol{x}_\tau,\end{aligned}$$

where

$$\mathcal{J}_\tau(\boldsymbol{x}_\tau, r) = \int_{-\infty}^{\infty} f_{r_{\tau_1} | \boldsymbol{x}_{\tau,t}, r_{\tau,m_\pm,t+1}}(s \,|\, \boldsymbol{x}_\tau, r)\, ds.$$

But the conditional density of r_{τ_1} given $\left(\boldsymbol{x}_{\tau,t}, r_{\tau,m_\pm,t+1}\right)$ is univariate normal by Theorem 5, and so $\mathcal{J}_\tau(\boldsymbol{x}_\tau, r) = 1$ for all $(\boldsymbol{x}_\tau, r) \in \mathbb{R}_-^{n-1} \times \mathbb{R}$ and

$$\mathcal{I}_\tau(r) = f_{r_{\tau,m_\pm,t+1}}(r) \int_{\boldsymbol{x}_\tau \in \mathbb{R}_-^{n-1}} f_{\boldsymbol{x}_{\tau,t} \,|\, r_{\tau,m_\pm,t+1}}(\boldsymbol{x}_\tau \,|\, r)\, d\boldsymbol{x}_\tau. \tag{A.14}$$

Now, by Proposition 6, we have

$$\boldsymbol{z}_{\tau,t} = \begin{bmatrix} r_{\tau,m_\pm,t+1} \\ \boldsymbol{x}_{\tau,t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\tau,m_\pm,t+1} \\ \boldsymbol{\mu}_{\tau,t} \end{bmatrix}, \begin{bmatrix} \varsigma^2_{r,\tau,t+1} & \boldsymbol{\lambda}'_{\tau,t} \\ \boldsymbol{\lambda}_{\tau,t} & \Lambda_{\boldsymbol{x},\tau,t} \end{bmatrix}\right), \tag{A.15}$$

where $\Lambda_{\boldsymbol{x},\tau,t} = D_{n-1} P_\tau \Lambda_{t,t} P'_\tau\, D'_{n-1}$, and so setting $\boldsymbol{x}_1 = \boldsymbol{x}_{\tau,t}$ and $\boldsymbol{x}_2 = r_{\tau,m_\pm,t+1}$ in Theorem 5 gives

$$f_{\boldsymbol{x}_{\tau,t} \,|\, r_{\tau,m_\pm,t+1}}(\boldsymbol{x}_\tau \,|\, r) = \phi_{n-1}\left(\boldsymbol{x};\, \boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}\left(r - \mu_{\tau,m_\pm,t+1}\right)\boldsymbol{\lambda}_{\tau,t},\, \Lambda_{\boldsymbol{x}|r,\tau,t}\right). \tag{A.16}$$

Computing the integral of $f_{\boldsymbol{x}_{\tau,t} \,|\, r_{\tau,m_\pm,t+1}}(\boldsymbol{x}_\tau \,|\, r)$ over $\boldsymbol{x}_\tau \in \mathbb{R}_-^{n-1}$ gives

$$\mathcal{I}_\tau(r) = f_{r_{\tau,m_\pm,t+1}}(r)\, \Phi_{n-1}\left[\boldsymbol{0};\, \boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}\left(r - \mu_{\tau,m_\pm,t+1}\right)\boldsymbol{\lambda}_{\tau,t},\, \Lambda_{\boldsymbol{x}|r,\tau,t}\right],$$

and since $r_{\tau,m_\pm,t+1} \sim \mathcal{N}\left(\bar{r}_{\tau,m_\pm,t+1}, \varsigma^2_{r,\tau,t+1}\right)$, we have

$$\mathcal{I}_\tau(r) = \phi_1\left(r;\, \mu_{\tau,m_\pm,t+1}, \varsigma^2_{r,\tau,t+1}\right)\Phi_{n-1}\left[\boldsymbol{0};\, \boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}\left(r - \mu_{\tau,m_\pm,t+1}\right)\boldsymbol{\lambda}_{\tau,t},\, \Lambda_{\boldsymbol{x}|r,\tau,t}\right].$$

Summing $\mathcal{I}_\tau(r)$ over $\tau \in S_n$ gives (22).

A.3. Proof of Corollary 1

It follows from the definition of conditional densities that

$$f_{\boldsymbol{x}_{\tau,t} \,|\, r_{\tau,m_\pm,t+1}}(\boldsymbol{x}_\tau \,|\, r) f_{r_{\tau,m_\pm,t+1}}(r) = f_{r_{\tau,m_\pm,t+1} | \boldsymbol{x}_{\tau,t}}(r \,|\, \boldsymbol{x}_\tau) f_{\boldsymbol{x}_{\tau,t}}(\boldsymbol{x}_\tau),$$

and substituting the right-hand side into the expression for $\mathcal{I}_\tau(r)$ in (A.14) gives

$$\mathcal{I}_\tau(r) = \int_{\boldsymbol{x}_\tau \in \mathbb{R}_-^{n-1}} f_{r_{\tau,m_\pm,t+1} | \boldsymbol{x}_{\tau,t}}(r \,|\, \boldsymbol{x}_\tau) f_{\boldsymbol{x}_{\tau,t}}(\boldsymbol{x}_\tau)\, d\boldsymbol{x}_\tau.$$

Setting $\boldsymbol{x}_1 = r_{\tau,m_\pm,t+1}$ and $\boldsymbol{x}_2 = \boldsymbol{x}_{\tau,t}$ in Theorem 5 implies

$$f_{r_{\tau,m_\pm,t+1} | \boldsymbol{x}_{\tau,t}}(r, \boldsymbol{x}_\tau) = \phi_1\left(r;\, \mu_{\tau,m\pm,t+1|\boldsymbol{x}}, \varsigma^2_{r|\boldsymbol{x},\tau,t+1}\right),$$

and since $f_{\boldsymbol{x}_{\tau,t}}(\boldsymbol{x}_\tau) = \phi_{n-1}\left(\boldsymbol{x}_\tau;\, \boldsymbol{\mu}_{\tau,t}, \Lambda_{\boldsymbol{x},\tau,t}\right)$, we have

$$\mathcal{I}_\tau(r) = \int_{\boldsymbol{x}_\tau \in \mathbb{R}_-^{n-1}} \phi_1\left(r;\, \mu_{\tau,m\pm,t+1|\boldsymbol{x}}, \varsigma^2_{r|\boldsymbol{x},\tau,t+1}\right)\phi_{n-1}\left(\boldsymbol{x}_\tau;\, \boldsymbol{\mu}_{\tau,t}, \Lambda_{\boldsymbol{x},\tau,t}\right) d\boldsymbol{x}_\tau$$

and summing over $\tau \in S_n$ gives (23).

A.4. Proof of Corollary 2

If $\boldsymbol{\mu}_u$ and $\Lambda_{u,v}$ are as given in (24), then $\bar{r}_{m\pm,\tau,t+1} = 0$, $\bar{\boldsymbol{x}}_{\tau,t} = \boldsymbol{0}$, and $\boldsymbol{\lambda}_{\tau,t} = \boldsymbol{0}$ for all $\tau \in S_n$, and the remaining terms that appear in (22) are independent of the permutation $\tau \in S_n$. So the summands are identical, and since there are $n!$ summands, we have

$$f_{r_{m\pm,t+1}}(r) = n!\, \phi_1\left(r;\, 0, \varsigma^2_{r,\mathrm{id},t+1}\right)\Phi_{n-1}\left[\boldsymbol{0};\, \boldsymbol{0}, \Lambda_{\boldsymbol{x}|r,\mathrm{id},t}\right].$$

Integrating both sides over $r \in \mathbb{R}$, and noting that the cumulative distribution term on the right-hand side is independent of r, gives $\Phi_{n-1}\left[\boldsymbol{0};\, \boldsymbol{0}, \Lambda_{\boldsymbol{x}|r,\mathrm{id},t}\right] = 1/n!$ and so the result follows.

A.5. Proof of Theorem 4

Using the notation from Theorem 2, we have $\boldsymbol{\mu}_{\tau,t} = O(l_r)$ and $\boldsymbol{\lambda}_{\tau,t} = O(\sqrt{l_r})$ as $l_h \to \infty$, and $\varsigma^2_{r,\tau,t+1} = O(l_h)$, $\mu_{\tau,m_\pm,t+1} = O(l_h)$, and $\boldsymbol{\lambda}_{\tau,t} = O(\sqrt{l_h})$ as $l_r \to \infty$. It follows that $\boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}(r - \mu_{\tau,m_\pm,t+1}\boldsymbol{\lambda}_{\tau,t}) \to \boldsymbol{\mu}_{\tau,t}$ as $l_h \to \infty$. Since $\mu_{1,t} > \mu_{2,t} > \ldots > \mu_{n,t}$ by assumption, we have $\lim_{l_h\to\infty}(\mu_{i+1,t} - \mu_{i,t}) \to -\infty$ for all $1 \le i \le n-1$, and so $\lim_{l_r\to\infty}\boldsymbol{\mu}_{\text{id},t} \to (-\infty, -\infty, \ldots, -\infty)$. For id $\neq \tau \in S_n$, we have $\lim_{l_h\to\infty}(\mu_{\tau_{i+1},t} - \mu_{\tau_i,t}) \to \infty$ for some $1 \le i \le n-1$, and since $\Lambda_{\boldsymbol{x}|r,\tau,t} = O(l_r)$,

$$\lim_{l_r\to\infty} \Phi_{n-1}\left[\mathbf{0};\, \boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}(r - \mu_{\tau,m_\pm,t+1})\boldsymbol{\lambda}_{\tau,t},\, \Lambda_{\boldsymbol{x}|r,\tau,t}\right] \to \delta_{\tau,\text{id}}.$$

Hence the only non-zero summand in (22), as $l_r \to \infty$, is the term corresponding to the identity permutation, and this establishes (a). Next, note that as $l_h \to \infty$,

$$\boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}(r - \mu_{\tau,m_\pm,t+1})\boldsymbol{\lambda}_{\tau,t} \to -\varsigma^{-2}_{r,\tau,t+1}\mu_{\tau,m_\pm,t+1}\boldsymbol{\lambda}_{\tau,t} = O(\sqrt{l_h}), \tag{A.17}$$

so that $\lim_{l_h\to\infty}\Phi_{n-1}\left[\mathbf{0};\, \boldsymbol{\mu}_{\tau,t} + \varsigma^{-2}_{r,\tau,t+1}(r - \mu_{\tau,m_\pm,t+1})\boldsymbol{\lambda}_{\tau,t},\, \Lambda_{\boldsymbol{x}|r,\tau,t}\right] \to 0$ if and only if $-\mu_{\tau,m_\pm,t+1}\boldsymbol{\lambda}_{\tau,t} \succ \mathbf{0}$, which implies (27). For the final statement in (b), note that $\boldsymbol{\lambda}_{\tau,t} = \mathbf{0}$ if $\boldsymbol{r}_t$ and $\boldsymbol{r}_{t+1}$ are independent so that $V = S_n$ and (27) becomes

$$\lim_{l_h\to\infty} f_{r_{m\pm,t+1}}(r) \to \sum_{\tau\in S_n}\phi_1\left(r;\, \mu_{\tau,m_\pm,t+1},\, \varsigma^2_{r,\tau,t+1}\right)\Phi_{n-1}\left[\mathbf{0};\, \boldsymbol{\mu}_{\tau,t},\, \Lambda_{\boldsymbol{x}|r,\tau,t}\right],$$

which is a sum of univarite normals. Now, since $\mu_{\tau,m\pm,t+1} \neq 0$ for some $\tau \in S_n$ by assumption, we may suppose, without loss of generality, that $\mu_{\tau,m\pm,t+1} > 0$. But then $\mu_{\tau^\star,m\pm,t+1} = -\mu_{\tau,m\pm,t+1}$ for some $\tau^\star \in S_n$, and since $\mu_{\tau,m_\pm,t+1} = O(l_h)$, we have $\mu_{\tau,m\pm,t+1} \to \infty$ and $\mu_{\tau^\star,m\pm,t+1} \to -\infty$ as $l_h \to \infty$, which shows that there are constituent normal densities on the right-hand side of (A.17) with means that tend to $-\infty$ and $+\infty$. This completes the proof.

A.6. Proof of Proposition 5

Using the density, $f_{r_{m_\pm,t+1}}(r)$, from (23) allows $\mu_p(r_{m\pm,t+1})$ to be expressed as the sum

$$\sum_{\tau\in S_n}\int_{\boldsymbol{x}\in\mathbb{R}^{n-1}_-}\phi_{n-1}\left(\boldsymbol{x};\, \boldsymbol{\mu}_{\tau,t},\, \Lambda_{\boldsymbol{x},\tau,t}\right)\int_{r\in\mathbb{R}} r^p\phi_1\left(r;\, \mu_{\tau,m\pm,t+1|\boldsymbol{x}},\, \varsigma^2_{r|\boldsymbol{x},\tau,t+1}\right)dr\, d\boldsymbol{x}.$$

Now, applying the expression for the univariate moments from (A.10) and the definition of $\bar{r}_{\tau,m\pm,t+1|\boldsymbol{x}}$ from Corollary 1, inner integrals can be computed as follows:

$$\begin{aligned}
&\int_{r\in\mathbb{R}} r^p\phi_1\left(r;\, \mu_{\tau,m\pm,t+1|\boldsymbol{x}},\, \varsigma^2_{r|\boldsymbol{x},\tau,t+1}\right)dr \\
&\quad = \sum_{\substack{k=0\\ k\text{ even}}}^{p}\binom{p}{k}(k-1)!!\,\varsigma^k_{r|\boldsymbol{x},\tau,t+1}\left(\mu_{\tau,m\pm,t+1} + \boldsymbol{\lambda}'_{\tau,t}\Lambda^{-1}_{\boldsymbol{x},\tau,t}(\boldsymbol{x} - \boldsymbol{\mu}_{\tau,t})\right)^{p-k} \\
&\quad = \sum_{\substack{k=0\\ k\text{ even}}}^{p}\binom{p}{k}(k-1)!!\,\varsigma^k_{r|\boldsymbol{x},\tau,t+1}\sum_{l=0}^{p-k}\binom{p-k}{l}\mu^{p-k-l}_{\tau,m\pm,t+1}\left(\boldsymbol{\lambda}'_{\tau,t}\Lambda^{-1}_{\boldsymbol{x},\tau,t}(\boldsymbol{x} - \boldsymbol{\mu}_{\tau,t})\right)^l.
\end{aligned}$$

Substituting back into the expression for $\mu_p(r_{m\pm,t+1})$ gives (28).

A.7. Proof of Corollary 3

In this case, we have $\boldsymbol{\lambda}_{\tau,t} = \mathbf{0}_{n-1}$, and the only non-zero term in the innermost sum in (28) is where $l = 0$. The corresponding integral reduces to

$$\int_{\boldsymbol{x}\in\mathbb{R}^{n-1}_-}\phi_{n-1}\left(\boldsymbol{x};\, \boldsymbol{\mu}_{\tau,t},\, \Lambda_{\boldsymbol{x},\tau,t}\right)d\boldsymbol{x} = \Phi_{n-1}\left[\mathbf{0};\, \boldsymbol{\mu}_{\tau,t},\, \Lambda_{\boldsymbol{x},\tau,t}\right],$$

and substituting back into (28) gives (29).

A.8. Proof of Proposition 2

Let $(1\,2) \in S_2$ be the identity permutation, and denote by $(2\,1) \in S_2$ the non-trivial permutation. Then since $x_{(1,2),u} = r_{2,u} - r_{1,u}$ and $x_{(2\,1),u} = r_{1,u} - r_{2,u}$ for $u \in \{t, t+1\}$, in the notation introduced in Theorem 2

$$\begin{aligned}
\varsigma^2_{r,\tau,t+1} &= \varsigma^2_{t+1}, & \lambda_{\tau,t} &= -\varrho_{t,t+1}\varsigma_t\varsigma_{t+1}, \\
\varsigma^2_{r,\tau,t} &= \varsigma^2_t, & \Lambda_{\boldsymbol{x}|r,\tau,t} &= \varsigma^2_{t+1}\left(1 - \varrho^2_{t,t+1}\right)
\end{aligned}$$

for all $\tau \in S_2$, and

$$\mu_{(12),1_\pm,t+1} = -\mu_{t+1}, \quad \mu_{(21),1_\pm,t+1} = \mu_{t+1},$$
$$\mu_{(12),t} = \mu_t, \quad \mu_{(21),t} = -\mu_t,$$

and substituting these terms into (22) gives the density in (12).

A.9. Proof of Theorem 1

Using the notation from Proposition 2 and computing directly the terms $\lambda_{\tau,t}$, $\Lambda_{x,\tau,t}$, $\varsigma^2_{r|x,t+1,\tau}$, and $\mu_{\tau,1\pm,t+1|x}$ defined in Corollary 1 gives

$$\varsigma_t^2 = \sigma_{1,t}^2 + \sigma_{2,t}^2 - 2\rho_t\sigma_{1,t}\sigma_{2,t} = \Lambda_{x,\tau,t}, \tag{A.18}$$

$$\varsigma_{t,t+1} = \rho_{1,1}\sigma_{1,t}\sigma_{1,t+1} + \rho_{2,2}\sigma_{2,t}\sigma_{2,t+1} - \rho_{1,2}\sigma_{1,t}\sigma_{2,t+1} - \rho_{2,1}\sigma_{2,t}\sigma_{1,t+1} = -\lambda_{\tau,t}, \tag{A.19}$$

$$\varsigma^2_{r|x,t+1,\tau} = \varsigma^2_{t+1} - \frac{\varsigma^2_{t,t+1}}{\varsigma^2_t}, \tag{A.20}$$

$$\mu_{\tau,1\pm,t+1} = -\varepsilon_\tau\mu_{t+1} = -\mu_{\tau,t+1}, \tag{A.21}$$

where $\varepsilon_{(12)} = 1$ and $\varepsilon_{(21)} = -1$. The integral appearing in (28) can be computed explicitly in this case and gives

$$\begin{aligned} I_{l,\tau,t} &= \int_{\boldsymbol{x}\in\mathbb{R}_-} \left(\lambda_{\tau,t}\Lambda^{-1}_{x,\tau,t}(x-\mu_{\tau,t})\right)^l \phi_1(x;\mu_{\tau,t},\Lambda_{x,\tau,t})dx \\ &= \lambda^l_{\tau,t}\Lambda^{-\frac{l}{2}}_{x,\tau,t}\int_{-\infty}^0 \left(\frac{x-\mu_{\tau,t}}{\sqrt{\Lambda_{x,\tau,t}}}\right)^l \phi_1(x;\mu_{\tau,t},\Lambda_{x,\tau,t})dx \\ &= \lambda^l_{\tau,t}\Lambda^{-\frac{l}{2}}_{x,\tau,t}\int_{-\infty}^{\frac{-\mu_{\tau,t}}{\sqrt{\Lambda_{x,\tau,t}}}} z^l\phi_1(z)dz. \end{aligned}$$

If we define $\eta_{\tau,t} = -\mu_{\tau,t}/\sqrt{\Lambda_{x,\tau,t}}$ and $J_{l,\tau,t} = \int_{-\infty}^{\eta_{\tau,t}} z^l\phi_1(z)\,dz$, then direct calculations establish that $J_{0,\tau,t} = \Phi_1[\eta_{\tau,t}]$, $J_{1,\tau,t} = -\phi_1(\eta_{\tau,t})$, and for $l \geq 2$

$$J_{l,\tau,t} = -\eta^{l-1}_{\tau,t}\phi_1(\eta_{\tau,t}) + (l-1)J_{l-2,\tau,t}.$$

Substituting $I_{l,\tau,t}$ back into (28) and simplifying the resulting expressions give the four moments of $r_{1\pm,t+1}$.

A.10. Proof of Proposition 3

It follows from the expression for $\mu_{r_{1\pm,t+1}}$ in (13) that

$$\mu_{r_{1\pm,t+1}} = \bar{\mu}_{t+1}(2\Phi_1[\mu_t/\varsigma_t]-1)l_h + 2\varrho_{t,t+1}\bar{\varsigma}_{t+1}\phi_1(\mu_t/\varsigma_t)\sqrt{l_h},$$

which is a quadratic in $\sqrt{l_h}$. Note that $\mu_{r_{1\pm,t+1}}$, as a function of $\sqrt{l_h}$, is concave down if and only if the leading coefficient is negative, and the maximum occurs at positive $\sqrt{l_h}$ if and only if the coefficients are of opposite signs, which are equivalent to $(2\Phi_1(\mu_t/\varsigma_t)-1)\bar{\mu}_{t+1} < 0$ and $\varrho_{t,t+t} > 0$. Solving for the turning point, and squaring, gives the expression for $l_h^\star$ in (17).

A.11. Proof of Proposition 3

Since $\sigma_{1,t+1} = \sigma_{2,t+1}$ by assumption, we have $\lim_{\rho_{t+1}\to 1}\varsigma_{t+1} = 0$, and it follows from (16) that

$$\lim_{\rho_{t+1}\to 1}\sigma^2_{r_{1\pm,t+1}} = \left(1-(2\Phi_1[\mu_t/\varsigma_t]-1)^2\right)\mu^2_{t+1},$$

so that substituting into (14) gives

$$\lim_{\rho_{t+1}\to 1}\mathrm{kurt}_{r_{1\pm,t+1}} = \lim_{\rho_{t+1}\to 1}\frac{\mu^4_{t+1}}{\sigma^4_{r_{1\pm,t+1}}} = \frac{1}{\left(1-(2\Phi_1[\mu_t/\varsigma_t]-1)^2\right)^2},$$

which simplifies to (18).

References

Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. Rev. Financ. Stud. 15, 1137–1187.
Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. J. Financ. Econ. 63, 443–494.
Arellano-Valle, R.B., Azzalini, A., 2006. On the unification of families of skew-normal distributions. Scand. J. Stat. 33, 561–574.
Asness, A., 1994. Variables that explain stock returns. University of Chicago (Ph.D. Dissertation).
Asness, C.S., Liew, J.M., Stevens, J.M., 1997. Parallels between the cross-sectional predictability of stock and country returns. J. Portf. Manag. 23, 79–87.
Asness, C.S., Moskowitz, T.J., Pedersen, R.L., 2013. Value and momentum everywhere. J. Finance 58, 929–985.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52 (1), 57–82.
Chan, K., Hameed, A., Tong, W., 2000. Profitability of momentum strategies in the international equity markets. J. Financ. Quant. Anal. 35, 153–172.
Chan, L.K., Jegadeesh, N., Lakonishok, J., 1996. Momentum strategies. J. Finance 51 (5), 1681–1713.
Chordia, T., Shivakumar, L., 2002. Momentum, business cycle, and time-varying expected returns. J. Finance 57, 985–1019.
Cooper, M., Gutierrez, R., Hameed, A., 2004. Market states and momentum. J. Finance 59, 1345–1365.
Daniel, K., Moskowitz, T.J., 2016. Momentum crashes. J. Financ. Econ. 122 (2), 221–247.
Erb, C.B., Harvey, C.R., 2006. The strategic and tactical value of commodity futures. Financ. Anal. J. 62, 69–97.
Foltice, B., Langer, T., 2015. Profitable Momentum Trading Strategies for Individual Investors. Scholarship and Professional Work – Business. Paper 264. Butler University.
Grundy, B., Martin, J.S., 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. Rev. Financ. Stud. 14, 29–78.
Hameed, A., Yuanto, K., 2002. Momentum strategies: evidence from the pacific basin stock markets. J. Financ. Res. 25 (3), 383–397.
He, X., Li, K., 2015. Profitability of time series momentum. J. Bank. Finance 53, 140–157–546.
Israel, R., Moskowitz, T.J., 2013. The role of shorting, firm size, and time on market anomalies. J. Financ. Econ. 108, 275–301.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. J. Finance 48 (1), 65–91.
Jegadeesh, N., Titman, S., 2001. Profitability and momentum strategies: an evaluation of alternative explanations. J. Finance 56, 699–720.
Lewellen, J., 2002. Momentum and autocorrelation in stock returns. Rev. Financ. Stud. 15, 533–563.
Lo, A., MacKinlay, C., 1990. When are contrarian profits due to stock market overreaction? Rev. Financ. Stud. 3 (2), 175–205.
Longing, F., Solnik, B., 2001. Extreme correlation of international equity markets. J. Finance 56, 649–678.
Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2012. Currency momentum strategies. J. Financ. Econ. 106, 660–684.
Moskowitz, T., Ooi, Y.H., Pedersen, L.H., 2012. Time series momentum. J. Financ. Econ. 104 (2), 228–250.
Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. John Wiley & Sons, New Jersey.
Okunev, J., White, D., 2003. Do momentum-based strategies still work in foreign currency markets? J. Financ. Quant. Anal. 38, 425–447.
Ramchmand, L., Susmel, R., 2007. Volatility and cross correlation across major stock markets. J. Empir. Finance 5, 379–416.
Richards, A., 1997. Winner–loser reversals in national stock market indices: can they be explained? J. Finance 52, 2129–2144.
Rotman, J., 1995. An Introduction to the Theory of Groups. Springer-Verlag, New York.
Rouwenhorst, K.G., 1998. International momentum strategies. J. Finance 53, 267–284.
Sancetta, A., Satchell, S.E., 2007. Changing correlation and equity portfolio diversification failure for linear factor models during market declines. Appl. Math. Finance 14 (3), 227–242.
Sefton, J., Scowcroft, A., 2004. A Decomposition of Portfolio Momentum Returns. Tanaka Business School, Imperial College London Discussion Paper TBS/DP04/9.
Wang, K., Xu, J., 2015. Market volatility and momentum. J. Empir. Finance 30, 79–91.
Withers, C.S., 1985. The moments of the multivariate normal. Bull. Aust. Math. Soc. 32, 103–107.