dxFeed Benchmark Orthogonal-Collinear Index Family Methodology

dxFeed Index Management Team <im@dxfeed.com>

December 28, 2022

Contents

1 Description

dxFeed Benchmark Orthogonal-Collinear™ (BORC™) Index Family Methodology aims to construct an asset portfolio with a weak (or alternatively strong) correlation with some reference asset (e.g., broad market index, gold ETF, etc.). In particular a numerical optimization procedure is utilized to find a suitable set of index parameters. Care is taken to avoid bias (e.g., outlier removal is performed) and ensure the index's statistical properties do not change quickly after the end of the parameter estimation period.

The resulting instrument can be considered an investment alternative to the selected reference asset and may be used in portfolio diversification scenarios.

2 Index Model

Let C be a set of index components, e.g., $C = \{\mathrm{ETH}, \mathrm{ADA}, \ldots\}$ for a cryptocurrency-based index. Denote the amount of component c as Share_c and the set of all such amounts for the corresponding components in C as Share_C. Let Symbol_C be a set of symbols for obtaining market data for all the index components. The index parameters are then $\Theta = \{\mathrm{Share}_C, \mathrm{Symbol}_C\}$.

Suppose $\mathrm{Price}_c(t \mid \mathrm{Symbol}_C)$ is a function for obtaining a price of a component c at time t; then the index value at time t is the price of the corresponding portfolio:

$$\mathrm{Index}(t \mid \Theta) = \sum_{c \in C} \mathrm{Price}_c(t \mid \mathrm{Symbol}_C) \cdot \mathrm{Share}_c \tag{1}$$

3 Component Selection

The set of index components C for each index within the BORC family is constructed by sequentially applying selected filters to the target asset class, e.g., keeping only the components that exceed a certain market capitalization or ADTV threshold. The particular filters are specific for each index within the family. In addition if the parameter derivation procedure below yields zero weights for some components, they are removed from the final set C.

This procedure aims to select predictable and stable components so their statistical properties and the properties of the resulting index remain the same after the end of the parameter estimation period.

4 Parameter Derivation

The index value at each time t is the portfolio price of selected assets C, as computed by (1). The Share_c parameter values are chosen for each $c \in C$, so that the resulting time series of index values has the minimum (or maximum) possible correlation with the reference asset, e.g., the broad market index on the parameter estimation period $\mathcal{T}$, comprised of T days, $\mathcal{T} = \{t_1, \dots, t_T\}$. Each Share_c corresponds to the weight of the component c in the portfolio at the end of the period t_T, namely

$$\text{Share}_c = \frac{\text{Index}(t_T)}{\text{Price}_c(t_T \mid \text{Symbol}_C)} \text{Weight}_c, \quad c \in C. \tag{2}$$

In turn, Weight_c (for all $c \in C$) is derived using an optimization procedure described below:

1. Let $n = |C|$; enumerate C (e.g., alphabetically), and fix a weight space

   $$\mathcal{W} = \left\{ \mathbf{w} \in \mathbb{R}^n : \|\mathbf{w}\|_1 = 1, w_i \in [0, w^{(\max)}] \; \forall i \in 1:n \right\}.$$

   $w^{(\max)}$ caps the maximum weight allowed per component (e.g., 30%). Setting $w^{(\max)} < 1$ serves as a regularization procedure. Otherwise the optimizer may assign 100% weight to a single component based on its performance in the parameter estimation period, which may degrade quickly for all $t > t_T$.

2. Define the objective function as follows:

   (a) Let $\mathbf{r}(\mathbf{w}) = (r_1(\mathbf{w}), \dots, r_T(\mathbf{w}))^\intercal$ be a vector of simple returns of the portfolio for the given weight vector:

   $$r_t(\mathbf{w}) = \frac{\text{Index}(t \mid \Theta')}{\text{Index}(t-1 \mid \Theta')} - 1, \quad t \in 1:T,$$

   where $\text{Share} \in \Theta'$ is computed using (2) with $\text{Index}(t_T)$ set to 1 (as the normalization constant is irrelevant for the optimization) and $\text{Weight} = \mathbf{w}$ (suitably mapped), i.e. as

   $$\text{Share}_c = \frac{\text{Weight}_c}{\text{Price}_c(t_T \mid \text{Symbol}_C)}.$$

   Component prices at time t_T are used to obtain the most relevant value of the Share parameter at the effective date t_{T+1}.

   Similarly, let $\mathbf{r}^*$ be a vector of simple returns of the reference asset for the parameter estimation period $\mathcal{T}$.

   (b) Index return outliers are winsorized to decrease their influence on the correlation. In particular, let

$$\gamma_{\text{lower}} = \text{quantile}_{1/4}(\mathbf{r}) - \alpha \cdot \text{IQR}(\mathbf{r})$$
$$\gamma_{\text{upper}} = \text{quantile}_{3/4}(\mathbf{r}) + \alpha \cdot \text{IQR}(\mathbf{r})$$

be the lower and upper return thresholds, where α is an IQR margin (e.g., 3). All returns not in the range $[\gamma_{\text{lower}}, \gamma_{\text{upper}}]$ are then capped at the corresponding (lower or upper) endpoints.

(c) Optionally, assign m power larger weights v_t to more recent observations as

$$v_t = \frac{\tilde{v}_t}{\sum_{s=1}^{T} \tilde{v}_s}, \quad \tilde{v}_t = t^m, \quad t \in 1:T.$$

If $m = 0$, equal weights $v_t = 1/T$ are obtained, of course.

(d) Compute the weighted sample means as

$$\bar{\mathbf{r}}(\mathbf{w} \mid \mathbf{v}) = \sum_{t=1}^{T} v_t r_t(\mathbf{w}), \quad \bar{\mathbf{r}}^*(\mathbf{v}) = \sum_{t=1}^{T} v_t r_t^*.$$

(e) The objective function is then the weighted sample correlation of $\mathbf{r}(\mathbf{w})$ and $\mathbf{r}^*$ computed the usual way:

$$f(\mathbf{w}) = \text{corr}(\mathbf{r}(\mathbf{w}), \mathbf{r}^* \mid \mathbf{v}) = \frac{\text{cov}(\mathbf{r}(\mathbf{w}), \mathbf{r}^* \mid \mathbf{v})}{\text{sd}(\mathbf{r}(\mathbf{w}) \mid \mathbf{v})\,\text{sd}(\mathbf{r}^* \mid \mathbf{v})},$$

where

$$\text{cov}(\mathbf{r}(\mathbf{w}), \mathbf{r}^* \mid \mathbf{v}) = \sum_{t=1}^{T} v_t \left(r_t(\mathbf{w}) - \bar{\mathbf{r}}(\mathbf{w} \mid \mathbf{v})\right)\left(r_t^* - \bar{\mathbf{r}}^*(\mathbf{v})\right)$$
$$\text{sd}(\mathbf{r}(\mathbf{w}) \mid \mathbf{v}) = \sqrt{\sum_{t=1}^{T} v_t \left(r_t(\mathbf{w}) - \bar{\mathbf{r}}(\mathbf{w} \mid \mathbf{v})\right)^2}$$
$$\text{sd}(\mathbf{r}^* \mid \mathbf{v}) = \sqrt{\sum_{t=1}^{T} v_t \left(r_t^* - \bar{\mathbf{r}}^*(\mathbf{v})\right)^2}$$

3. Numerically minimize (or maximize) the objective function $f(\mathbf{w})$ on $\mathcal{W}$:

$$\mathbf{w}^* = \arg\min_{\mathbf{w} \in \mathcal{W}} f(\mathbf{w}), \quad \text{or} \quad \mathbf{w}^* = \arg\max_{\mathbf{w} \in \mathcal{W}} f(\mathbf{w}).$$

The obtained vector $\mathbf{w}^*$ is used to uniquely define the set of shares using (2).

Components with zero weights, as determined by the optimization procedure, are removed from the index.

On the very first rebalancing, $\text{Index}(t_T)$ is set to be equal to the close value of the benchmark asset on t_T.

5 Lifecycle & Maintenance

5.1 Rebalancing

The index's composition is reviewed periodically—see the accompanying factsheet for details. The rebalancing might occur as a result of this review.

To facilitate a more stable index composition, up to k previous index components are added to the component set yielded by the selection procedure. However, following the general rule, if their weights are determined to be zero as a result of optimization, they are discarded.

Suppose the parameter estimation procedure yields a new set of weights and symbols—Weight' and Symbol' respectively—corresponding to a new component set C', effective at day $t^* = t_T + 1$, where t_T is the last day of the parameter estimation period $\mathcal{T}$. Then the new set of shares is computed as

$$\mathrm{Share}'_c = \frac{\mathrm{Index}(t^* - 1)}{\mathrm{Price}_c(t^* - 1)} \mathrm{Weight}'_c, \quad c \in C',$$

where $\mathrm{Index}(t^* - 1)$ is the close index value and $\mathrm{Price}_c(t^* - 1)$ is the close price of component c on the day preceding the effective date t^*.

5.2 Symbol Removal

If no rates can be fetched for component c^* due to the removal of the corresponding symbol, the default behavior is to use its last known value to compute the index ("last observation carried forward"). The index composition will be updated before or at the next scheduled rebalancing, at the discretion of the steering committee as follows:

1. Set $C' = C \setminus \{c^*\}$.
2. Renormalize the weights as $\mathrm{Weight}'_{c'} = \mathrm{Weight}_{c'} / \sum_{c \in C'} \mathrm{Weight}_c$, for all $c' \in C'$.
3. Rebalance the index as in section 5.1 using $\mathrm{Weight}'_{C'}$, $\mathrm{Symbol}_{C'}$.

Disclaimer

Devexperts Inc. ("dxFeed") might receive compensation for licensing its indices to third parties and providing custom calculation services. dxFeed products are governed by the terms and conditions of the agreements under which they are provided. A license is required from dxFeed to display, create derivative works of, and/or distribute any product or service that uses, is based upon, and/or refers to any dxFeed index data.

It is not possible to invest directly in an index. Exposure to an asset class represented by an index is available through investable instruments based on that index. dxFeed makes no assurance that investment products based on the index will accurately track index performance or provide positive investment returns. dxFeed is not an investment advisor and makes no representation regarding the advisability of investing in any investment fund or other investment vehicle, including those based on dxFeed products or any documents and statements found in this document. Prices for dxFeed indices are calculated by dxFeed based on the prices of the index's individual constituents as set by their primary exchange or source. Prices are received by dxFeed either directly or from one of its third-party vendors. Vendors receive the prices from the primary exchanges.

Charts and graphs are provided for illustrative purposes only. The charts and graphs may reflect hypothetical historical performance. All information presented before the launch date is back-tested. Back-tested performance is hypothetical, not actual performance. The back-test calculations are based on the same methodology in effect after the indices' official launch. Back-tested performance reflects the application of an index's methodology and selection of index constituents. It adds the benefit of hindsight and knowledge of factors that may have positively affected its performance. Back-testing cannot account for all financial risks affecting results and may be considered to reflect survivor/look-ahead bias. Actual returns may differ significantly from, and be lower than, back-tested returns. Past performance is not an indication or guarantee of future results. Except for certain custom index calculation services, all information provided by dxFeed is impersonal. It is not tailored to the needs of any person, entity, or group of persons.

The materials contained in this document have been prepared solely for informational purposes based upon information generally available to the public, retrieved from sources believed to be reliable. No content within these materials (including index data, ratings, analysis, research, valuations, model, software, or other application or output therefrom) or any part thereof ("Content") may be modified, reverse-engineered, reproduced, or distributed in any form or by any means, nor copied or stored in a database or retrieval system, without dxFeed's prior written permission.

dxFeed does not assume any obligation to update the Content following publication in any form or format. The Content shall not be used for any unlawful or unauthorized purposes. dxFeed and its third-party data providers and licensors (collectively "dxFeed Parties") do not guarantee the Content's accuracy, completeness, timeliness, or availability. dxFeed is not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content.

The Content is provided on an "as is" basis. dxFeed disclaims all express or implied warranties, including, but not limited to, any warranties of merchantability or fitness for a particular purpose or use, freedom from bugs, software errors or defects, that the Content's functioning will be uninterrupted, or that the Content will operate with any software or hardware configuration.

In no event shall dxFeed be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special, or consequential damages, costs, expenses, legal fees, or losses (including, but not limited to, lost income or lost profits and opportunity costs) in connection with any use of the Content, even if advised of the possibility of such damages.