Pooling Dynamic Conditional Correlation Models

Bram van Os* and Dick van Dijk†

Econometric Institute, Erasmus University Rotterdam

September 16, 2021

Abstract The Dynamic Conditional Correlation (DCC) model by Engle (2002) has become an extremely popular tool for modeling the time-varying dependence of asset returns. However, applications to large cross-sections have been found to be problematic, due to the curse of dimensionality. We propose a novel DCC model with Conditional LInear Pooling (CLIP-DCC) which endogenously determines an optimal degree of commonality in the correlation innovations, allowing a part of the update to be of reduced dimension. In contrast to existing approaches such as the Dynamic EquiCOrrelation (DECO) model, the CLIP-DCC model does not restrict long-run behavior, thereby naturally complementing target correlation matrix shrinkage approaches. Empirical findings suggest substantial benefits for a minimum-variance investor in real-time. Combining the CLIP-DCC model with target shrinkage yields the largest improvements, confirming that they address distinct parts of uncertainty of the conditional correlation matrix.

*vanos@ese.eur.nl
†djvandijk@ese.eur.nl

1 Introduction

The conditional covariance matrix of asset returns is of crucial importance for portfolio construction and risk management. Multivariate ARCH-type models, such as the Dynamic Conditional Correlation (DCC) model by Engle (2002), have become a popular tool to model and forecast time-varying conditional covariance matrices. However, for many of these models the quality of the covariance matrix estimates is known to deteriorate as the number of assets grows. A key contributor to the estimation uncertainty of the DCC model, particularly in large cross-sections, is the unstructured nature of the information used to update the correlations. As noted by Engle and Kelly (2012), this causes the individual correlations of the DCC model to evolve independently, leaving the information on the correlation between other assets untapped.

To better handle a large number of assets, we propose a DCC model with Conditional LInear Pooling (CLIP-DCC). Inspired by the Dynamic EquiCOrrelation (DECO) model of Engle and Kelly (2012), the CLIP-DCC model draws from the entire cross-section to model the conditional correlations such that the value of any correlation pair depends on the history of all pairs instead of just its own. The CLIP-DCC model differs in two important aspects, though. First, the CLIP-DCC model does not completely synchronise the conditional correlations and also allows for pair-specific dynamics to acknowledge that each correlation may partly show idiosyncratic movements. The level of commonality is governed by a parameter that is estimated alongside the other parameters during likelihood estimation, such that an optimal amount of structure is determined endogenously. Second, a re-centering step is used to avoid (implicit) restrictions on the long-run correlation matrix. As a result, the CLIP-DCC model may, for example, still use the sample correlation matrix as the long-run target, as in the original DCC model.

Furthermore, by preserving long-run dynamics the CLIP-DCC model naturally complements approaches that shrink the correlation targeting matrix, where a combined approach essentially allows one to separately deal with uncertainty of the long-run correlation matrix

and the movement around it. This separation is favorable, as these two likely warrant very different levels of structure from a bias-variance perspective.

A Monte Carlo simulation experiment shows that the parameters of the CLIP-DCC model may be effectively estimated using the Composite Likelihood (CL) method, similar to the standard DCC model, see Pakel et al. (2021). Furthermore, to empirically evaluate the performance of our dynamic correlation models, we construct global minimum variance (GMV) portfolios in real-time and consider their out-of-sample performance. Specifically, we consider daily US large-cap stock returns for the period February 1981 until December 2020 for a wide selection of portfolio sizes ranging from 10 to 500 stocks. We find that the CLIP-DCC model is clearly favored, with significant reductions in out-of-sample portfolio variance compared to the DCC model, whereby the relative improvements increase as the dimension grows. In addition, we find that the CLIP-DCC model outperforms the DCC model greatly around the 2008 financial crisis. Moreover, we find in a combined approach of the CLIP-DCC model with non-linear shrinkage (NLS) of the target, using the method of Ledoit and Wolf (2020), that improvements are additive. For example, we find for the portfolio universe size $N = 500$ an out-of-sample GMV portfolio annualized standard deviation of 6.400(6.621) and 6.122(6.375) for the DCC and the CLIP-DCC model with(out) NLS of the target, respectively. This confirms the notion that these methods address different sources of uncertainty and are in fact complementary.

This paper is related to and builds upon a rich literature of multivariate volatility modeling, see e.g. Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009) for an overview. In particular, this paper is closely connected to various extensions of the DCC model that help accommodate large cross-sections. This includes composite likelihood estimation, linear and non-linear shrinkage of the target and the usage of intraday high and low prices, see Pakel et al. (2021), Hafner and Reznikova (2012), Engle et al. (2019) and De Nard et al. (2020), respectively. Our paper is most closely related to the (Block) DECO models introduced by Engle and Kelly (2012). While the information pooling aspect of the DECO model

is attractive, it is likely to impose too much structure by assuming that the correlations are identical across all pairs. The Block-DECO model relaxes this assumption by instead imposing a block structure. However, determining the optimal block size and group allocation may be difficult, requiring an exogenous notion of group membership. In addition, both the DECO and the Block-DECO model also implicitly heavily structure the long-run correlation matrix, which may be suboptimal. Precisely these concerns are addressed in the CLIP-DCC framework. Additionally, we extend the CLIP-DCC model to a block-based version, which can incorporate group structure information if available. Empirically, we find that the Block-DECO model improves upon a DECO-type model, but find no additional advantage of a Block-CLIP-DCC model over the CLIP-DCC model. This indicates that the industry memberships used to impose the block structure are not particularly informative for the movement of the conditional correlations around the long-run.

The outline of this paper is as follows. Section 2 develops the CLIP-DCC model, showing how it can be obtained from the combination of a DCC model with an appropriately scaled DECO-type model. A simulation study to assess the consistency of the CL estimator for the CLIP-DCC model is carried out in Section 3. Section 4 presents a real-time trading application to daily stock data and Section 5 concludes. Finally, proofs and additional results can be found in the Appendix.

2 Methodology

2.1 The DCC and DECO Models

Let r_t denote the $N \times 1$ vector of zero mean asset returns at time $t = 1, \ldots, T$. Extensions to a non-zero and possibly time-varying mean are straightforward, but not considered here for simplicity. We assume that the conditional covariance matrix of r_t is time-varying and denoted by $\Sigma_t := \mathbb{E}[r_t r_t' | \mathcal{I}_{t-1}]$, where $\mathcal{I}_{t-1}$ denotes the set containing all information available

at time $t-1$. Following Engle (2002) and many others, we decompose Σ_t as

$$\Sigma_t = D_t R_t D_t, \tag{1}$$

where D_t denotes the diagonal matrix of square root conditional variances that in practice may be modelled by any standard GARCH-type model and R_t is the conditional correlation matrix. The $N \times 1$ vector of standardized returns is denoted by z_t and given as

$$z_t = D_t^{-1} r_t. \tag{2}$$

Using (1) we find that $\mathbb{E}[z_t z_t' | \mathcal{I}_{t-1}] = R_t$, such that the outer product of the devolatized returns $z_t z_t'$ provides an unbiased ex-post proxy of the true conditional correlation matrix R_t.

The Dynamic Conditional Correlation (DCC) model by Engle (2002) for R_t is given as

$$Q_t^{DCC} = C(1-a-b) + a z_{t-1} z_{t-1}' + b Q_{t-1}^{DCC}, \tag{3}$$

$$R_t^{DCC} = (\tilde{Q}_t^{DCC})^{-1/2} \, Q_t^{DCC} \, (\tilde{Q}_t^{DCC})^{-1/2}, \tag{4}$$

where $\tilde{Q}_t^{DCC}$ is a diagonal matrix containing the diagonal elements of Q_t^{DCC}, a and b are non-negative scalar parameters such that $a+b<1$ and C is a (symmetric) positive definite matrix of parameters. Note that $\mathbb{E}(R_t) \approx C$, such that C can be interpreted as the long-run correlation matrix, whereby the approximate nature stems from the standardization step in (4) to go from Q_t^{DCC} to R_t^{DCC}. It is standard practice to employ a correlation targeting approach, using the sample covariance of the standardized returns z_t as an estimator for C.

Finally, we remark that Aielli (2013) finds that there is an issue with the consistency of this target estimator for C due to the standardization. He therefore proposes a corrected DCC (cDCC) specification. However, Engle et al. (2019) among others note that practically there are hardly any differences between this cDCC specification and the original DCC model.

For simplicity and with the empirical application in mind, which also considers shrinkage of the target matrix, we therefore follow Engle et al. (2019) and consider the original DCC specification.

Because an unrestricted correlation matrix contains $\mathcal{O}(N^2)$ parameters, large cross-sections pose a problem for the accuracy of the conditional correlation matrix estimates of the DCC model. The Dynamic Equicorrelation (DECO) model by Engle and Kelly (2012) dramatically restricts the correlation matrix by imposing that all pairwise conditional correlations are equal. Practically, this restriction is implemented by applying a transformation to the conditional correlation matrix provided by a DCC-type recursion as follows

$$\rho_t = \frac{1}{N(N-1)} \iota_N' (R_t^{DCC} - I_N) \iota_N, \tag{5}$$

$$R_t^{DECO} = \rho_t J_{N \times N} + (1 - \rho_t) I_N, \tag{6}$$

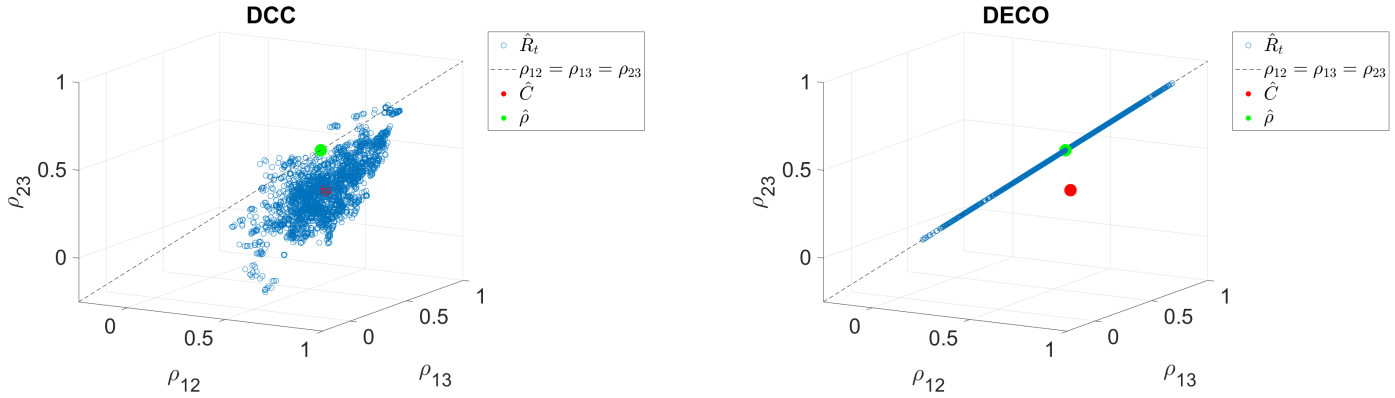
where ρ_t is the average correlation at time t, ι_N is a $N \times 1$ vector of ones, I_N is the $N \times N$ identity, $J_{N \times N}$ is an $N \times N$ matrix of ones, R_t^{DCC} is obtained using (3) and (4) and R_t^{DECO} is the equicorrelation matrix output. From a practical standpoint, the main allure of this model is the computational ease of estimation due to the simplicity of the likelihood, see Engle and Kelly (2012) for details.

To visually illustrate the differences and limitations of the DCC and DECO models, we consider a representative empirical example. Namely, we estimate both models on daily return data of the 30 industry portfolios constructed by Kenneth French, from January 2000 until December 2009[1]. These 30 industry portfolios are formed by assigning all NYSE, AMEX and NASDAQ stocks to one of 30 portfolios based on their SIC code. Figure 1 depicts the dynamic correlations between the first three portfolios of the data-set, which are Food, Beer and Smoke, for the two correlation models.

In Figure 1, we find that all DCC correlation estimates lie within the positive domain,

[1]Data is obtained from: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Figure 1: Selection of dynamic correlation estimates of the DCC and DECO models using thirty industry portfolios, January 2000 until December 2009.



Note: This figure contains the dynamic correlation estimates of the Food (1), Beer (2) and Smoke (3) industry portfolios for the DCC and DECO models from January 3, 2000 until December 9, 2009. The models are estimated using Composite Likelihood on the devolatized returns obtained from GJR-GARCH(1,1) specifications and with target $\hat{C} = T^{-1}\sum_{t=1}^{T} z_t z_t'$.

reflecting positive co-movement of the considered industries. In addition, we observe that due to the equicorrelation structure that the dynamic correlation estimates of the DECO model lie on the 3-dimensional '45 degree' line. Further comparing the DCC and the DECO model, we make two important observations. First, we note that the DECO model also implicitly warps the unconditional correlation matrix as a direct consequence of the structure imposed on the conditional correlations. That is, we observe that the DCC model is centered around an unrestricted long-run estimated by $\hat{C} = T^{-1}\sum_{t=1}^{T} z_t z_t'$ with $[\hat{C}_{12}\ \hat{C}_{13}\ \hat{C}_{23}] = [0.621\ 0.436\ 0.344]$. In contrast, the DECO model is centered around an equicorrelation matrix with estimated equicorrelation parameter $\hat{\rho} = 0.537$. To make this warping explicit, note that

$$\mathbb{E}[R_t^{DECO}] = \mathbb{E}[\rho_t J_{N\times N} + (1-\rho_t)I_N] = \bar{\rho} J_{N\times N} + (1-\bar{\rho})I_N, \tag{7}$$

$$\bar{\rho} = \mathbb{E}[\rho_t] = \frac{1}{N(N-1)}\iota_N'(\mathbb{E}[R_t^{DCC}] - I_N)\iota_N, \tag{8}$$

where $\bar{\rho}$ denotes the long-run average correlation. Due to the linearity of the transformation from R_t^{DCC} to R_t^{DECO} as outlined in (5) and (6), we have that the long-run equicorrelation matrix $\mathbb{E}[R_t^{DECO}]$ can be obtained by applying the same transformation to $\mathbb{E}[R_t^{DCC}]$.

Intuitively, however, it would be attractive if we could decouple the structure imposed on the conditional correlations from the structure imposed on the long-run, as these two likely require a different level of structure from a bias-variance standpoint. This stems from the differences in effective sample size between the two. Namely, estimation of the long-run correlation matrix can draw from the entire sample. Conversely, the conditional correlations can only tap from a much smaller selection of recent observations. This is reflected in the estimation scheme, where $\hat{C}$ uses the entire sample while the DCC recursion in (3) amounts to an exponentially weighted moving average. For example, if the concentration ratio N/T is relatively small, the sample target estimator $\hat{C}$ will provide a fairly accurate estimate of the long-run correlation matrix, demanding little additional structure. In practice, one may also find that $\hat{C}$ is far from an equicorrelation matrix, directly revealing ex-ante that the DECO model is unlikely to render a reasonable description of this aspect of the data. On the other hand, if N itself is not small and the autocorrelation of the conditional correlations is low, then the conditional correlation matrix can likely benefit a lot from structure. Decoupling the structure imposed on the conditional and unconditional correlations would essentially allow us to separately deal with uncertainty of the long-run correlation matrix and the deviations from it.

The second observation we draw from Figure 1 is that the DCC and DECO models are extremes on the bias-variance spectrum. This is a direct consequence of the fact that the DCC model is innovated using $z_t z_t'$, an unbiased, yet highly noisy, ex-post proxy of the true conditional correlation matrix R_t. Therefore, at least from a cross-sectional perspective, no structure is imposed whatsoever. In contrast, the DECO model imposes a maximal amount of structure from a cross-sectional perspective by assuming that all pairwise correlations of the conditional correlation matrix are equal. While this structure will undoubtedly greatly reduce estimation uncertainty, it is likely to also come with a substantial bias. As noted by Engle and Kelly (2012), the DCC model essentially updates the correlations independently, while on the other hand the DECO maximally draws from the entire cross-section by con-

sidering the average correlation at each point in time. Therefore, it might be worthwhile exploring a setup that can fill the gap and provide a better bias-variance trade-off. Put differently, we would prefer a model with an 'optimal' level of commonality in the update-mechanism.

The main contribution of this paper, namely the DCC model with Conditional LInear Pooling (CLIP-DCC) is the result of addressing these two concerns. Specifically, we introduce a scaled DECO-type model to undo the warping of the long-run dynamics in Section 2.2 and consider a mixture setup to provide a more appropriate bias-variance trade-off in Section 2.3.

2.2 The Scaled Direct DECO Model

In this section, we introduce a model similar to the DECO model, but with the structure applied at the pseudocorrelation level Q_t. Moreover, the model involves a re-centering step to undo the warping of the long-run correlation matrix. We argue that it is generally more convenient to apply structure in the pseudocorrelation space Q_t than at the correlation level R_t. In this way, one can consider transformations that preserve positive definiteness but do not necessarily maintain the unit diagonal, without having to standardize twice. The re-centering step is an example of such a transformation. Additionally, we show that imposing structure at the pseudocorrelation level can come with large computational gains.

To facilitate discussion we first introduce the concept of a compound symmetric (CS) matrix, which is closely related to the equicorrelation matrix.

Definition 1 A square matrix S of size $N \times N$ is a compound symmetric matrix if it can written be as

$$S = oJ_{N\times N} + (d - o)I_N, \tag{9}$$

with diagonal element $d \in \mathbb{R}$ and off-diagonal element $o \in \mathbb{R}$.

Lemma 1 A CS matrix S of size $N \times N$ with $N \geq 2$, diagonal element $d \in \mathbb{R}$ and off-diagonal element $o \in \mathbb{R}$ is positive definite if and only if $d > 0$ and $\frac{o}{d} \in (\frac{-1}{N-1}, 1)$. Additionally, if S is positive definite then standardization, i.e. $(\tilde{S})^{-1/2}\, S\, (\tilde{S})^{-1/2}$, yields a valid equicorrelation matrix, where $\tilde{S}$ is a diagonal matrix containing the diagonal elements of S.

Second, we introduce a mapping that can be used to turn a square matrix into a CS matrix. For any $A \in \mathbb{R}^{N \times N}$ with $N \geq 2$ we define the mapping $\theta(A)$ as

$$\theta(A) := \theta^O(A) J_{N \times N} + [\theta^D(A) - \theta^O(A)] I_N, \tag{10}$$

where the scalars $\theta^D(A)$ and $\theta^O(A)$ are the diagonal and off-diagonal averages of A, that is,

$$\theta^D(A) := \frac{1}{N} \sum_{i=1}^{N} a_{ii}, \tag{11}$$

$$\theta^O(A) := \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i}^{N} a_{ij}, \tag{12}$$

where a_{ij} with $i, j \in \{1, \ldots, N\}$ are the elements of A. It follows that $\theta(A)$ is a CS matrix of size $N \times N$ with diagonal element $\theta^D(A)$ and off-diagonal element $\theta^O(A)$. For completeness, we define $\theta(\cdot)$ to be the identity mapping if $N = 1$. Note that $\theta(\cdot)$ yields the same output as the DECO transformation in (5) and (6) if the input has a unit diagonal, such that $R_t^{DECO} = \theta(R_t^{DCC})$. Therefore, $\theta(\cdot)$ can be interpreted as an intuitive extension of the DECO transformation in (5) and (6) that accommodates a non-unit diagonal input. The following lemma summarizes several useful properties of $\theta(\cdot)$.

Lemma 2 $\theta(\cdot)$ is a linear mapping that preserves positive (semi-)definiteness, that is for $A \in \mathbb{R}^{N \times N}$ we have that

1. $\theta(A + B) = \theta(A) + \theta(B),\ \forall B \in \mathbb{R}^{N \times N}$,

2. $\theta(kA) = k\theta(A),\ \forall k \in \mathbb{R}$,

3. If A is positive (semi-)definite, then $\theta(A)$ is positive (semi-)definite.

The direct DECO (dDECO) model is obtained by applying the compound symmetry transformation $\theta(\cdot)$ to Q_t^{DCC} as given in (3). That is, we define $Q_t^{dDECO} := \theta(Q_t^{DCC})$ and using Lemma 2 we obtain

$$Q_t^{dDECO} = \theta(C)(1-a-b) + a\theta(z_{t-1}z_{t-1}') + bQ_{t-1}^{dDECO}, \tag{13}$$

where Q_t^{dDECO} is positive definite if Q_t^{DCC} is positive definite. Standardization of Q_t^{dDECO} by dividing by its diagonal elements will yield a valid equicorrelation matrix (see again Lemma 1). This model is therefore highly similar to the DECO model as it is based on the same equal pairwise correlations assumption, with the difference between the two models being a slightly different implementation of this constraint. Specifically, the dDECO model reverses the order of standardization and pooling compared to the DECO model, note again that $R_t^{DECO} = \theta(R_t^{DCC})$.

Because Q_t^{dDECO} has only two unique elements, namely its diagonal element $c_t := \theta^D(Q_t^{DCC})$ and its off-diagonal element $q_t := \theta^O(Q_t^{DCC})$, we do not need to track Q_t^{DCC} in full. That is, we only need to consider

$$c_t = \theta^D(C)(1-a-b) + a\theta^D(z_{t-1}z_{t-1}') + bc_{t-1}, \tag{14}$$

$$q_t = \theta^O(C)(1-a-b) + a\theta^O(z_{t-1}z_{t-1}') + bq_{t-1}, \tag{15}$$

which admit an intuitive form as they are driven by the diagonal and off-diagonal mean of the DCC driving information $z_{t-1}z_{t-1}'$, respectively. Afterwards, Q_t^{dDECO} can be obtained from c_t and q_t as in (9). Because (14) and (15) are two scalar processes, computational demands are much lower than the DECO model, which requires tracking the $N \times N$ matrix R_t^{DCC}.

Due the linearity of $\theta(\cdot)$, we have that both the DECO model and the dDECO model have (approximately) displaced the center of movement from C to $\theta(C)$, an equicorrelation matrix if C has a unit diagonal. First noting that a direct translation, i.e. additive adjustment, is not feasible due to positive definiteness requirements, we propose a multiplicative re-centering step to undo the warping of the long-run correlation matrix. The scaled dDECO (sdDECO) model is constructed from the dDECO model as follows

$$Q_t^{sdDECO} := [C^{1/2}\theta(C)^{-1/2}]Q_t^{dDECO}[\theta(C)^{-1/2}C^{1/2}], \tag{16}$$

where $C^{1/2}$ and $\theta(C)^{-1/2}$ denote the symmetric square-root of C and $\theta(C)$ obtained using the eigenvalue decomposition. Symmetry and positive definiteness of Q_t^{sdDECO} is inherited from the positive definiteness of Q_t^{dDECO} and immediately apparent.

The update recursion for Q_t^{sdDECO} can be written directly as

$$Q_t^{sdDECO} = C(1-a-b) + aZ_{t-1} + bQ_{t-1}^{sdDECO}, \tag{17}$$

where Z_{t-1} is a positive semi-definite innovation term (see again Lemma 2) that pools the information in $z_{t-1}z_{t-1}'$ and is scaled in order to preserve long-run behavior. Specifically, we have that Z_{t-1} is given as

$$Z_{t-1} = [C^{1/2}\theta(C)^{-1/2}]\theta(z_{t-1}z_{t-1}')[\theta(C)^{-1/2}C^{1/2}], \tag{18}$$

where due to the linearity of $\theta(\cdot)$ we have that $\mathbb{E}(\theta(z_{t-1}z_{t-1}')) = \theta(\mathbb{E}(z_{t-1}z_{t-1}')) \approx \theta(C)$, such that $\mathbb{E}(Z_{t-1}) \approx C$. Here the approximate nature stems from minor differences arising from standardization. From there it straightforward to see that the sdDECO model, by construction, has again center of movement C. Because the sdDECO model leaves the unconditional expectation unaltered, the pooling can be interpreted to be 'conditional'.

2.3 The CLIP-DCC Model

Although the sdDECO model preserves long-run dynamics, it still imposes a large amount of structure on the conditional correlations. Namely, all time-variation is generated by two scalar processes, see again (14)-(16), which could introduce a substantial bias. Therefore, to overcome our second concern and provide a more 'optimal' bias-variance trade-off, we now present the DCC model with Conditional LInear Pooling (CLIP-DCC). Specifically, we assume that our pseudocorrelation process $Q_t^{CLIP-DCC}$ is a convex combination of Q_t^{DCC} and Q_t^{sdDECO}, that is

$$Q_t^{CLIP-DCC} = (1-w)Q_t^{DCC} + wQ_t^{sdDECO}, \tag{19}$$

where $w \in [0,1]$ is the mixture weight. Using (3) and (17), we may also directly write the update recursion for $Q_t^{CLIP-DCC}$ as

$$Q_t^{CLIP-DCC} = C(1-a-b) + a[(1-w)z_{t-1}z_{t-1}' + wZ_{t-1}] + bQ_{t-1}^{CLIP-DCC}, \tag{20}$$

where Z_{t-1} is given by (18). It is straightforward to show that $Q_t^{CLIP-DCC}$ is positive definite if Q_t^{DCC} is positive definite and has center of movement C. Furthermore, re-parameterization with $a_1 = a(1-w)$ and $a_2 = aw$ yields,

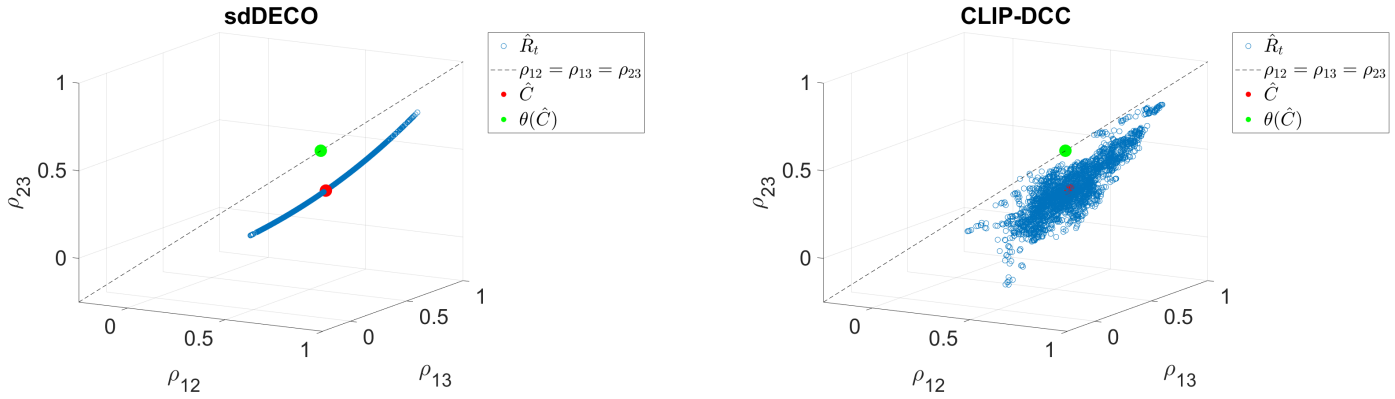
$$Q_t^{CLIP-DCC} = C(1-a_1-a_2-b) + a_1 z_{t-1}z_{t-1}' + a_2 Z_{t-1} + bQ_{t-1}^{CLIP-DCC}, \tag{21}$$

which reveals that, effectively, we are simply adding a pooling term Z_{t-1} to the DCC recursion as an additional explanatory variable. As noted by Engle and Kelly (2012), the correlations of the DCC model evolve essentially independently while those of the (d)DECO model co-move perfectly. In contrast, the CLIP-DCC model allows for a more nuanced level of cross-dependence, with magnitude determined by w.

To visually illustrate the appeal of the CLIP-DCC framework, we again consider our

empirical example. Figure 2 depicts the dynamic correlations between the Food, Beer and Smoke industries for the sdDECO and CLIP-DCC models. We observe in Figure 2 that, like

Figure 2: Selection of dynamic correlation estimates of the sdDECO and CLIP-DCC models using thirty industry portfolios, January 2000 until December 2009.



Note: This figure contains the dynamic correlation estimates of the Food (1), Beer (2) and Smoke (3) industry portfolios for the sdDECO and CLIP-DCC models from January 3, 2000 until December 9, 2009. The models are estimated using Composite Likelihood on the devolatized returns obtained from GJR-GARCH(1,1) specifications and with target $\hat{C} = T^{-1} \sum_{t=1}^{T} z_t z_t'$.

the DECO model, the sdDECO model shows a line-type movement pattern. However, the estimates of the sdDECO model have center of movement $\hat{C}$, similar to the DCC model, as result of the scaling step. It is worthwhile noting that the induced curvature is essentially the result of the standardization step to go from Q_t^{sdDECO} to R_t^{sdDECO}. Moreover, we observe that $R_t^{CLIP-DCC}$ suggests a more appropriate bias-variance trade-off than R_t^{DCC}, see again Figure 1, and R_t^{sdDECO}, while also preserving the location of the long-run correlation matrix. The estimated mixture weight $\hat{w}$ is equal to 0.514, such that the CLIP-DCC model relies roughly equally on the unstructured DCC innovation $z_{t-1}z_{t-1}'$ and the pooled sdDECO innovation Z_{t-1} to update the conditional correlations, see again (20).

Finally, we note that although intuitively one may think of the sdDECO component in the CLIP-DCC model as a shrinkage target, our approach differs from traditional shrinkage methods because we estimate the mixture weight w simultaneously with the DCC model parameters (based on the likelihood, as discussed in detail in Section 2.5). As a direct consequence, separation of the structure imposed on the conditional and unconditional correlation

matrix is crucial. This is because the sample correlation matrix $\hat{C}$ used for targeting already provides the best estimate of the long-run C in-sample, such that its optimal pooling intensity is near zero. As a result, a setup where a dDECO component is used instead of sdDECO component in (19), yields very small estimates of the mixture weight w, effectively reducing the model to the base DCC model. Structuring the long-run correlation matrix therefore demands a different approach, which will be outlined in Section 2.5.

2.4 The Block-CLIP-DCC Model

In the presence of a clear group structure of the assets, one would like to use this information in the estimation of the conditional correlation matrix. While the CLIP-DCC model does not need any such information, we may extend our framework to exploit it when available. This section therefore describes how the CLIP-DCC model can be extended to incorporate group-membership information.

Given a K-group partition $n_1, n_2, \ldots, n_K$ of the cross-section of N assets such that $n_j \in \mathbb{N}^+, j = 1, 2 \ldots, K$ and $\sum_{j=1}^{K} n_j = N$, we assume without loss of generality that the data is already 'sorted' such that the assets $i = 1, \ldots, n_1$ are in the first group, the assets $i = n_1 + 1, \ldots, n_1 + n_2$ are in the second group and so forth. For notational convenience, we summarize this K-group information in the $K \times 1$ vector $G := [n_1, n_2, \ldots, n_K]'$. We now introduce a mapping $\theta^{BL}(\cdot, G)$ that can be used to turn a square matrix into a 'K-block compound symmetric matrix' with group structure G. That is, for $A \in \mathbb{R}^{N \times N}$ with $N \geq 2$, we define

$$\theta^{BL}(A, G) = \begin{bmatrix} \theta(A^*_{11}) & \tau(A^*_{12}) & \cdots & \tau(A^*_{1K}) \\ \tau(A^*_{21}) & \theta(A^*_{22}) & \cdots & \tau(A^*_{2K}) \\ \vdots & \vdots & \ddots & \vdots \\ \tau(A^*_{K1}) & \tau(A^*_{K2}) & \cdots & \theta(A^*_{KK}), \end{bmatrix}, \tag{22}$$

$$\tau(V) = (\frac{1}{m_1 m_2} \iota'_{m_1} V \iota_{m_2}) J_{m_1 \times m_2}, \; \forall V \in \mathbb{R}^{m_1 \times m_2}, \tag{23}$$

where A^*_{ij} for $i, j = 1, 2 \ldots, K$ is the ij-th block of A using the block partition from G. Here we observe that the diagonal blocks of $\theta^{BL}(A, G)$ are obtained by applying the compound symmetry transformation $\theta(\cdot)$ to the diagonal blocks of A. For the off-diagonal blocks, we use the function $\tau(\cdot)$, which returns a constant matrix with the mean of the input matrix everywhere. The following lemma summarizes some useful properties of $\theta^{BL}(\cdot, G)$, matching the properties of $\theta(\cdot)$.

Lemma 3 $\theta^{BL}(\cdot, G)$, with $G := [n_1, n_2, \ldots, n_K]'$, $n_1, n_2, \ldots, n_K$ such that $n_j \in \mathbb{N}^+, j = 1, 2, \ldots, K$ and $\sum_{j=1}^{K} n_j = N$, is a linear mapping that preserves positive (semi-)definiteness, that is, for $A \in \mathbb{R}^{N \times N}$ with $N \geq 2$ we have that

1. $\theta^{BL}(A + B, G) = \theta^{BL}(A, G) + \theta^{BL}(B, G), \ \forall B \in \mathbb{R}^{N \times N}$,
2. $\theta^{BL}(kA, G) = k\theta^{BL}(A, G), \ \forall k \in \mathbb{R}$,
3. If A is positive (semi-)definite, then $\theta^{BL}(A, G)$ is positive (semi-)definite.

Using Lemma 3, we may repeat the steps in the previous sections and obtain a Block-dDECO, a Block-sdDECO and a Block-CLIP-DCC model by replacing $\theta(\cdot)$ with $\theta^{BL}(\cdot, G)$. Note that the dDECO, sdDECO and CLIP-DCC models are nested cases with $K = 1$ and $G = N$.

In practice, we may be unsure of the 'best' group structure and have multiple candidate structures. Therefore, we may in theory also allow for multiple (distinct) block structures simultaneously. That is, we can consider a mixture setup of the DCC model with L Block-sdDECO models, each with a distinct group structure G_l, $l = 1, 2, \ldots, L$. This then boils down to having L additional explanatory terms in the pseudo-correlation update recursion of the form

$$B_{l,t-1} = [C^{1/2}\theta^{BL}(C, G_l)^{-1/2}]\theta^{BL}(z_{t-1}z'_{t-1}, G_l)[\theta^{BL}(C, G_l)^{-1/2}C^{1/2}], \tag{24}$$

where $B_{l,t-1}$ is a positive semi-definite (by Lemma 3) explanatory term with group structure G_l, suitably scaled to preserve long-run dynamics (i.e. we have $\mathbb{E}(B_{l,t-1}) \approx C$).

Our purpose here is not to determine an optimal block-structure but merely to illustrate that the mixture approach of the CLIP-DCC model and the Block-DECO approach can be seen as different solutions to the same bias-variance trade-off problem, which are not mutually exclusive.

2.5 Parameter Estimation and Target Shrinkage

The parameters of the sdDECO and CLIP-DCC models can be estimated using maximum likelihood. In particular, we follow the standard three-step estimation procedure from the literature as suggested by Engle (2002). This entails that we first estimate univariate GARCH-type models for each asset and use the devolatized returns to estimate the dynamic correlation models. Second, in order to prevent the likelihood estimation of $\mathcal{O}(N^2)$ parameters contained in the intercept matrix C, see again (3), we employ a targeting approach using the sample covariance of the devolatized returns. However, it is well established that the quality of the sample correlation matrix degrades as the concentration ratio N/T grows, see e.g. Ledoit and Wolf (2004). Therefore, while it may be that the deviations or movement around the long-run correlation matrix are more noisy and require more structure, the sample covariance $\hat{C}$ used for targeting may also benefit from shrinkage for larger values of the concentration ratio N/T. Our models are easily combined with methods that shrink the target by replacing $\hat{C}$ with its shrinkage estimator.

For example, one could consider the non-linear shrinkage (NLS) estimator by Ledoit and Wolf (2012) which shrinks the sample correlation matrix at the eigenvalue level and has both theoretical and empirical advantages when compared to linear shrinkage estimators. Engle et al. (2019) use this method to shrink the target of the DCC model, yielding significant improvements for a minimum-variance investor in high dimensional settings and outperforming simpler shrinkage methods. Ledoit and Wolf (2020) provide an analytical version

of this approach, greatly reducing computational load. To compare and evaluate the effects of target shrinkage, deviation shrinkage as in the CLIP-DCC framework and their possibly synergy, we also consider this NLS estimator for the target and use the code made available by Michael Wolf[2].

Third, we use Gaussian quasi-maximum likelihood (QML) to estimate the scalar parameters a, b and w. However, two main problems arise for large N when using traditional full maximum likelihood (FML) estimation for DCC models. First, this concerns estimation feasibility. That is, evaluating the likelihood can quickly become very computationally intensive as N grows. This is because one is required to perform a multitude of operations involving matrices of size $N \times N$, including multiplications, determinants and inverses. Second, Pakel et al. (2021) find that for realistic sample sizes the parameter estimates of a and b of the DCC model become meaningfully biased as N becomes large. In particular, they find that the estimates of a and b tend to 0 as N increases.

To tackle both problems at once, i.e. to prevent this parameter estimation bias and greatly reduce computational load, Pakel et al. (2021) propose to estimate the model using Composite Likelihood (CL). This method approximates the full log likelihood using an average of a selection of bivariate log likelihoods constructed from pairs of asset returns. Because the CLIP-DCC model nests the DCC model it stands to reason that its parameters also suffer a similar fate if estimated using FML. Indeed, in unreported Monte Carlo simulations, we find that the mixture weight w appears to be overestimated when using FML estimation when N becomes large. Of course, the CLIP-DCC model also inherits the computational requirements of the DCC model. We therefore propose to also estimate the CLIP-DCC model using CL. Throughout, we shall make use of CL estimation based on contiguous pairs, that is we pair asset 1 and 2, 2 and 3 and so forth. This results in $N - 1$ pairs whose bivariate log likelihoods will be averaged to obtain the CL to be maximized. Pakel et al. (2021) find this yields highly similar parameter estimates as compared to CL based on all pairs, which is

[2]https://www.econ.uzh.ch/dam/jcr:11d24ab0-7ec2-4b3f-8ef4-7affaa727d25/analytical_shrinkage.m.zip

much more computationally intensive. To assess the consistency of the parameter estimates by the CL method for the CLIP-DCC model, particularly in view of the mixture weight w, we conduct a Monte Carlo study using the setup of Pakel et al. (2021) in the next section.

3 Monte Carlo Simulations

We simulate data from the CLIP-DCC model assuming a conditional multivariate Gaussian distribution for $N \in \{10, 30, 100\}$, $T = 2000$, $a = 0.05$, $b = 0.93$ and $w \in \{0, 0.25, 0.5, 0.75, 1\}$. Note that this selection of mixture weights also includes the DCC model ($w = 0$) and the sdDECO model ($w = 1$). For simplicity, we set all $\sigma_{i,t} = 1$ and do not consider estimation of the univariate GARCH models. Furthermore, we set the intercept matrix C equal to $C_{i,j} = \pi_i \pi_j$ for $i \neq j$ and $C_{i,i} = 1$ for $i = 1, \ldots, N$ and $j = 1, \ldots, N$, where the π_i are drawn from a truncated normal distribution with mean 0.5 and standard deviation 0.1 and truncation interval $[0.1, 0.9]$.

Table 1 contains the Monte Carlo means and standard deviations of the parameter estimates of a, b and w, obtained by estimating the CLIP-DCC model on the simulated data using CL estimation based on 500 replications. We observe for all considered settings and for all three parameters that the average estimate is very close to the true parameter value of the data-generating process. In particular, even for $w = 0$ and $w = 1$, when the CLIP-DCC model collapses to the DCC model and the sdDECO model, respectively, the CL approach performs satisfactory. Additionally, we find that the Monte Carlo standard deviations decrease as N increases. This suggests that the potential efficiency loss of CL decreases as N becomes large, in line with the results of Pakel et al. (2021). Furthermore, we find in unreported additional simulations for different values of T that bias tends to decrease as T grows. Finally, we remark that the amount of skewness and excess kurtosis of the parameter estimates is mostly mild except when the true mixture weight w is at either of the bounds ($w = 0$ or $w = 1$), which naturally compresses the distribution of $\hat{w}$. These findings assure us

Table 1: Monte Carlo means and standard deviations of the parameters of the CLIP-DCC model estimated using CL.

		$w = 0$	$w = 0.25$	$w = 0.5$	$w = 0.75$	$w = 1$
	a	0.052	0.050	0.051	0.050	0.051
		(0.004)	(0.006)	(0.008)	(0.010)	(0.011)
$N = 10$	b	0.928	0.927	0.926	0.925	0.925
		(0.005)	(0.008)	(0.012)	(0.017)	(0.022)
	w	0.037	0.240	0.502	0.761	0.986
		(0.051)	(0.097)	(0.087)	(0.068)	(0.026)
	a	0.051	0.050	0.050	0.051	0.050
		(0.003)	(0.005)	(0.006)	(0.008)	(0.009)
$N = 30$	b	0.928	0.927	0.927	0.926	0.927
		(0.003)	(0.005)	(0.007)	(0.012)	(0.016)
	w	0.028	0.242	0.500	0.763	0.994
		(0.040)	(0.081)	(0.064)	(0.040)	(0.012)
	a	0.051	0.050	0.050	0.051	0.050
		(0.002)	(0.004)	(0.005)	(0.006)	(0.008)
$N = 100$	b	0.928	0.928	0.926	0.925	0.927
		(0.002)	(0.003)	(0.005)	(0.010)	(0.014)
	w	0.026	0.243	0.503	0.763	0.999
		(0.034)	(0.069)	(0.047)	(0.029)	(0.004)

Note: This table contains the average parameter estimates (over the replications) of the CLIP-DCC model estimated using CL. The standard deviations of the parameter estimates are displayed in parentheses below the averages. The data is simulated from a CLIP-DCC model with $a = 0.05$, $b = 0.93$, $w \in \{0, 0.25, 0.5.0.75, 1\}$, $N \in \{10, 30, 100\}$ and $T = 2000$ using 500 replications.

that the CLIP-DCC model parameters may be effectively estimated using the CL approach.

4 Empirical Application

4.1 Data

For the empirical application we collect daily stock prices from the Center for Research in Security Prices (CRSP) database for the period from February 6, 1981 until December 31, 2020. We select a different investment universe at each estimation date by selecting the N stocks with the highest market capitalization at that date. This data selection is similar to the one used by Engle et al. (2019). We consider a wide range of portfolio sizes with

$N \in \{10, 30, 50, 100, 300, 500\}$. Note that the investment universes for the portfolio sizes $N = 30$, $N = 100$ and $N = 500$ correspond roughly to the constituents of the Dow Jones Industrial Average (DJIA), the S&P 100 and the S&P 500, respectively. For computational purposes the correlation models are re-estimated every 21 trading days with an estimation window of 2500 days. At each estimation date, we use the N largest stocks with the additional requirement that closing prices are available for both the entire estimation window as well as 21 days ahead for evaluation purposes. This yields exactly 360 estimation times and 7560 trading days for evaluation. To investigate the effects of a smaller estimation sample, we also perform the analysis using the same selection of stocks with an estimation window of 1250 days. Finally, we sort the data alphabetically based on the tickers at the estimation date, similar to Pakel et al. (2021). Although the models presented are invariant to permutations of the data, it does matter for the CL estimation procedure based on contiguous pairs. Differences for all but the block-based models are negligible though.

4.2 Evaluation

Because it is difficult to assess the quality of correlation matrix forecasts directly in the absence of high quality ex-post measures, we follow the common practice of indirect evaluation by using the conditional correlation matrix estimates to construct portfolios, see e.g. Engle and Kelly (2012) and Engle et al. (2019). Specifically, we consider the global minimum variance (GMV) portfolio. This portfolio is popular due to its simplicity and its independence of the mean return, which empirically is often poorly estimated, see e.g. Michaud (1989). The analytical solution for the GMV portfolio weight vector u_t is given as

$$u_t = \frac{\Sigma_t^{-1} \iota_N}{\iota_N' \Sigma_t^{-1} \iota_N}, \tag{25}$$

where Σ_t denotes the covariance matrix at time t. For our empirical application we use the covariance matrix forecasts from our models at the investment day, which can be obtained

from the correlation matrix forecasts from the DCC models and the diagonal matrix of volatilities obtained from the univariate GARCH-type models. Filling in this covariance matrix forecast in (25) will then yield a feasible estimator of the GMV portfolio weights. More specifically, we consider a daily re-balancing approach whereby we make new GMV portfolios every day using one-step ahead forecasts of the covariance matrix. Note that for computational considerations, we still estimate the model parameters only every 21 days. This set-up is similar to the approach used by Engle and Kelly (2012).

In terms of evaluation, we consider the (annualized) average (AV) out-of-sample daily log returns, the (annualized) standard deviation (SD) of the out-of-sample daily log returns and the information ratio (IR), which is obtained by dividing the average by the standard deviation. Naturally, since the objective of the GMV portfolio is to minimize variance, we are mostly interested in the out-of-sample standard deviation. To assess whether the differences in out-of-sample standard deviations are significant, we employ the test by Ledoit and Wolf (2011). Because of the large sample size (7560 out-of-sample days), we refrain from bootstraping but use their test statistic based on heteroskedasticity-and-autocorrelation corrected (HAC) standard errors.

4.3 Results

Table 2 contains the average and min-max range of the parameter estimates of the different DCC models for the investment universe $N = 100$. Findings for the other portfolio universe sizes are highly similar and not displayed for brevity. For the marginals, we follow Pakel et al. (2021) and use GJR-GARCH(1,1) specifications estimated using Gaussian QML to devolatize the returns. They find that this model produces the best volatility forecasts using similar data. Specifically, we have that $\sigma_{i,t}^2 = \omega_i + (\alpha_i + \gamma_i \mathbf{1}[y_{i,t-1} < 0])y_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2$ for each asset $i = 1, \ldots, N$, with ω_i, α_i, γ_i and β_i the corresponding parameters.

We observe in Table 2 that the average parameter estimates $\hat{a}$ and $\hat{b}$ and the corresponding min-max range are fairly standard. That is, small values of $\hat{a}$, large values of $\hat{b}$ and a sum close

Table 2: Average parameter estimates for the different dynamic conditional correlation models for $N = 100$, December 1990 until December 2020.

	DCC	dDECO	sdDECO	CLIP-DCC
$\hat{a}$	0.015 [0.006, 0.050]	0.039 [0.007, 0.112]	0.043 [0.008, 0.138]	0.040 [0.009, 0.136]
$\hat{b}$	0.969 [0.858, 0.991]	0.945 [0.794, 0.991]	0.944 [0.795, 0.991]	0.948 [0.800, 0.989]
$\hat{w}$				0.653 [0.474, 0.767]
$\hat{a}_1$				0.012 [0.004, 0.035]
$\hat{a}_2$				0.028 [0.005, 0.102]

Note: This table contains for $N = 100$ the average parameter estimates of the DCC, dDECO, sdDECO and CLIP-DCC models across the different estimation windows. Specifically, we estimate the model every 21-days using a moving window of length $T = 2500$ for a total of 360 estimation moments. Furthermore, the minimum and and maximum parameter estimates are displayed in brackets behind the average.

to 1. Comparing the parameter estimates of the DCC model to the estimates of the pooled models, we note that the pooled models admit a larger $\hat{a}$ and a smaller $\hat{b}$. This reveals that the pooled models require on average less smoothing over time as a consequence of the imposed cross-sectional structure. In particular, for the CLIP-DCC model, we note that $\hat{a}_1 = \hat{a}(1-\hat{w})$, the parameter for the DCC innovation term $z_{t-1}z_{t-1}'$, see (21), is fairly comparable to the parameter estimate $\hat{a}$ of the DCC model. Therefore, despite the addition of the pooled innovation term Z_{t-1}, the contribution of the DCC innovation term is largely maintained in the CLIP-DCC model, while the autoregressive parameter estimate $\hat{b}$ is reduced somewhat. This highlights the simultaneous bias-variance trade-off in the time dimension and the cross-sectional dimension as a result of the joint estimation of the smoothing parameters a and b and the pooling parameter w.

In Table 2, we also observe that the average estimate of the mixture weight $\hat{w}$ is equal to 0.653, indicating that the sdDECO component of the CLIP-DCC model is found to dominate the DCC component, see (19). In line with this, we have that $\hat{a}_2 = \hat{a}\hat{w}$ is more than twice as large as $\hat{a}_1 = \hat{a}(1 - \hat{w})$ on average, indicating that the pooled innovation term Z_{t-1} is found to be more informative than the unstructured DCC innovation $z_{t-1}z_{t-1}'$, see again (21). Although time-variation of the parameter estimates is not necessarily insightful due to the changing investment universe, we do remark a gradual increase of the mixture weight $\hat{w}$ over time combined with a sharper decrease in the persistence $\hat{a} + \hat{b}$ after the 2008 Financial

Crisis. The former observation suggests a mild increase in correlation co-movement around the long-run.

Table 3 presents the summary statistics of the daily out-of-sample log returns of the GMV portfolios constructed using the different models and the $1/N$ portfolio for comparison purposes. First and foremost, we find that the CLIP-DCC model has the lowest out-of-sample SD for all considered portfolio sizes. The SDs of the GMV portfolios decline as the portfolio dimension N grows, in line with the increasing possibilities for diversification. Moreover, we observe that the relative improvements of the CLIP-DCC model compared to the DCC model increase as well. That is, for $N = 10$ the improvements are minor with a SD of 15.392 for the CLIP-DCC model compared to 15.473 for the DCC model. By contrast the improvements for $N = 500$ are much larger with a SD of 6.375 versus 6.621 for the CLIP-DCC and the DCC models, respectively. These decreases in SDs of the CLIP-DCC model compared to the DCC model are found to be highly significant for all but the smallest portfolio size $N = 10$, which is only significant at the ten percent level. This suggests that allowing for a level of commonality in the movement of the conditional correlations around the long-run becomes more important as the dimension grows.

Second, we observe that the dDECO model performs poorly compared to the other models in all metrics and that the sdDECO model performs much better. This suggests that an equicorrelation structure on the long-run correlation matrix indeed incurs a too high bias for the reduction in variance obtained. Interestingly, we find that the sdDECO model offers comparable performance to the DCC model. This indicates that the variance reduction by the structure of the sdDECO model roughly cancels against the bias incurred relative to the DCC model. The outperformance of the CLIP-DCC model is then the result of allowing for an in-between solution, providing a superior bias-variance trade-off. Third, we find that even though the $1/N$ portfolio often possesses a relatively high AV, it also performs poorly in terms of SD. This indicates that correlation modeling is certainly a worthwhile endeavour for a minimum-variance investor in this setting.

Table 3: Daily out-of-sample GMV porfolio performance constructed using different DCC models for $N \in \{10, 30, 50, 100, 300, 500\}$, December 1990 until December 2020.

		DCC	dDECO	sdDECO	CLIP-DCC	1/N
	AV	4.856	3.758	5.778	5.272	7.283
$N = 10$	SD	15.473	15.745	15.480	**15.392***	19.167
	IR	0.314	0.239	0.373	0.343	0.380
	AV	6.050	5.272	6.340	6.243	8.041
$N = 30$	SD	13.901	14.846	13.841	**13.660*****	18.457
	IR	0.435	0.355	0.458	0.457	0.436
	AV	4.506	3.198	3.791	4.221	8.458
$N = 50$	SD	13.060	14.448	13.151	**12.842*****	18.271
	IR	0.345	0.221	0.288	0.329	0.463
	AV	2.981	1.356	1.317	2.553	8.273
$N = 100$	SD	10.892	13.134	10.989	**10.726*****	18.225
	IR	0.274	0.103	0.120	0.238	0.454
	AV	6.239	1.654	2.516	5.418	9.034
$N = 300$	SD	7.987	11.163	7.877	**7.751*****	17.975
	IR	0.781	0.148	0.319	0.699	0.503
	AV	5.108	0.820	1.673	4.554	9.558
$N = 500$	SD	6.621	9.137	6.476	**6.375*****	17.961
	IR	0.772	0.090	0.258	0.714	0.532

Note: This table contains the annualized average (AV), standard deviation (SD) and information ratio (IR) of the out-of-sample daily log returns for the GMV portfolios constructed using different dynamic correlations models and the $1/N$ portfolio. The lowest SD per dimension size is highlighted in bold. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 2500 days and re-estimation of the parameters every 21 days. A significant decrease of the (logarithmic squared) SD of the CLIP-DCC model compared to the DCC model is indicated with a *,** and *** for a $p - value$ below 0.1, 0.05 and 0.01, respectively, using the two-sided test by Ledoit and Wolf (2011) with HAC standard errors.

For robustness, we also consider two mean-variance (MV) portfolios and the quasi-likelihood (QLIKE) loss, see Patton and Sheppard (2009). For the MV portfolios the variance is minimized subject to a return constraint. Here we mimic the strategies of Engle and Kelly (2012) and Engle et al. (2019), which use the sample mean and a momentum signal for the return constraint, respectively. Results and details can be found in Appendix Table B.1 and B.2. There we also find the CLIP-DCC model to significantly reduce variance compared to

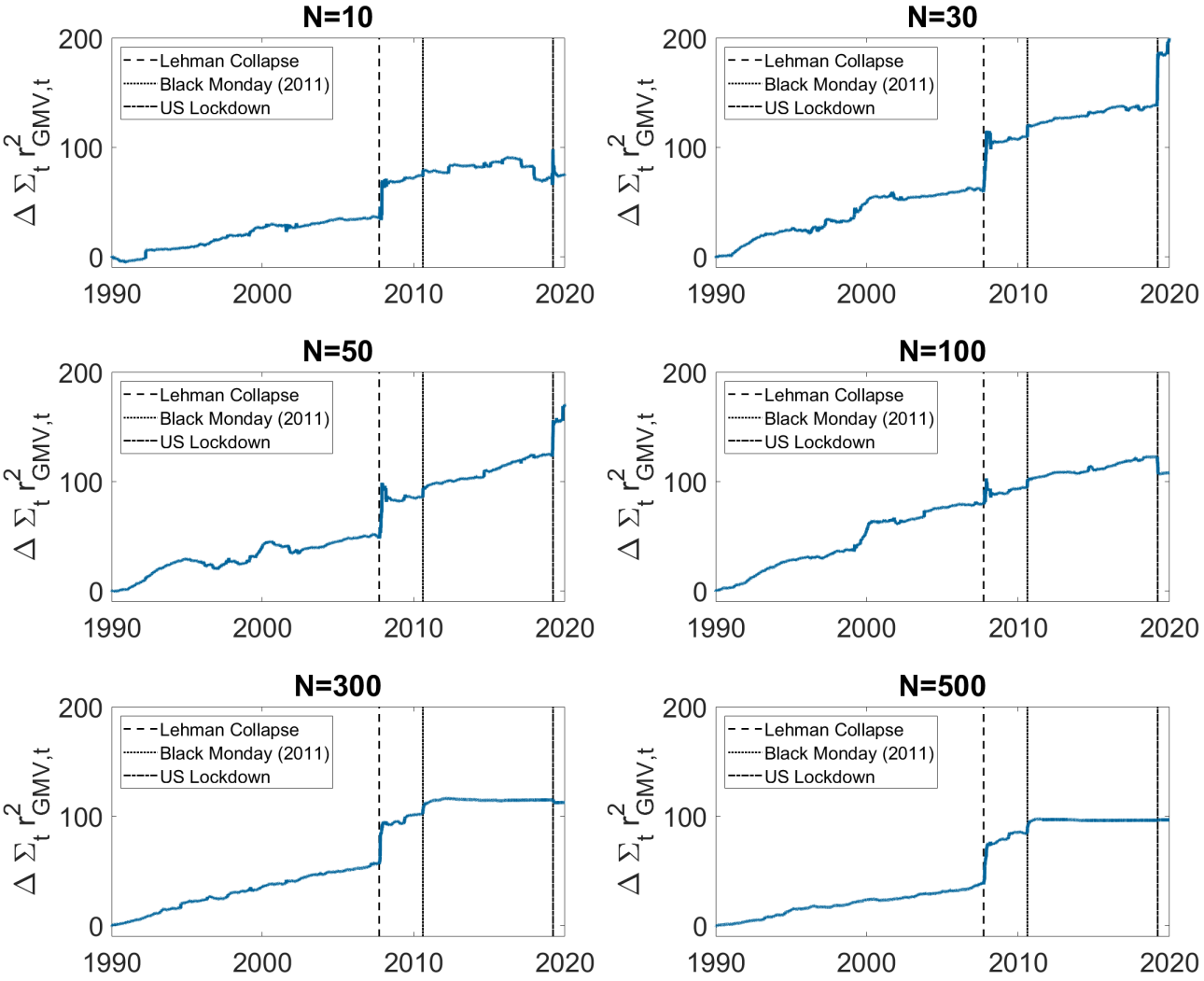
the DCC model, although the effect on the mean differs across portfolio sizes. Maximum IR portfolios or portfolios with leverage constraints are left for future research. Appendix Table B.3 summarizes the QLIKE performance of the models, where we also find the CLIP-DCC model to outperform the DCC model.

To further compare the DCC and CLIP-DCC models, we investigate the time course of their difference in performance. Figure 3 presents the difference in cumulative out-of-sample GMV portfolio variance between the DCC and CLIP-DCC models. Specifically, we subtract the cumulative out-of-sample variance, as proxied by the sum of the GMV portfolio squared logarithmic returns, of the CLIP-DCC model from the DCC model. A higher value therefore reflects a larger gain from the use of the CLIP-DCC framework relative to the base DCC model.

In Figure 3, we mostly find a steadily increasing cumulative benefit from using the CLIP-DCC model. However, during the peak of the 2008 Financial Crisis, we observe a large sudden increase in the difference in the cumulative variance of the DCC model and the CLIP-DCC model. Although the movement may visually appear to be near instantaneous, it spans about a one to three month period. Furthermore, a second similar upward jump of a smaller magnitude is located around Black Monday, 2011, following the downgrading of US sovereign debt by Standard and Poor's. This highlights the beneficial effects of the CLIP-DCC framework around periods of financial turmoil. A possible explanation is an increase in equity correlation co-movement associated with market downturns, see e.g Ang and Chen (2002). This would reduce the bias incurred by the cross-sectional structure of the CLIP-DCC model. A second possible explanation may be an increase in uncertainty of the conditional correlations during these periods of high volatility, favoring the pooling aspect of the CLIP-DCC framework.

Furthermore, for $N = 300$ and $N = 500$ we find that the CLIP-DCC model offers comparable performance to the DCC model between 2011 and 2020. This suggests that for large cross-sections the pooling function $\theta(\cdot)$ may be overly restrictive, as it effectively

Figure 3: Difference in cumulative out-of-sample GMV portfolio variance between the DCC and CLIP-DCC model, December 1990 until December 2020.



Note: This figure contains for different investment universe sizes the evolution of the difference in cumulative out-of-sample GMV portfolio variance between the DCC and CLIP-DCC model, denoted by $\Delta \sum_t r^2_{GMV,t} = \sum_{t=1}^{T} [r^2_{DCC,t} - r^2_{CLIP-DCC,t}]$, where $r_{DCC,t}$ and $r_{CLIP-DCC,t}$ denote the out-of-sample logarithmic returns of the GMV portfolio constructed by the DCC and CLIP-DCC model, respectively. The vertical lines reflect economically relevant dates.

imposes a one-factor type structure on the sdDECO component. Allowing the degrees of freedom of the pooled output to increase with the dimension may be beneficial. Finally, we find an effect of the COVID-19 pandemic around the date of the US lockdown in March 2020.

The effect of this pandemic recession is however more ambiguous, likely due to its different nature. In particular, we have that the CLIP-DCC framework provides benefits for the smaller investment universes $N = \{10, 30, 50\}$, losses for the medium sizes $N = \{100, 300\}$ and has almost no effect for $N = 500$.

Next, we investigate the effects of target shrinkage and its interplay with the CLIP-DCC approach. Table 4 presents summary statistics of the daily out-of-sample log returns of the GMV portfolios of the DCC, dDECO, sdDECO and CLIP-DCC models using the NLS approach of Ledoit and Wolf (2020) for the target. Comparing Table 3 and Table 4, we observe that NLS of the target decreases portfolio SD in all cases. We find that the benefits of NLS are minor for the small portfolio sizes up to $N = 100$, but increase for the large portfolio sizes $N = 300$ and $N = 500$. Furthermore, also when using target shrinkage the CLIP-DCC model achieves the lowest out-of-sample SD among all models. In fact, we find that the improvements of NLS and the CLIP-DCC model over the base DCC model are additive. This confirms their theoretical complementary nature, acting on different components of the uncertainty of the conditional correlation matrix. Here we find that the CLIP-DCC approach provides the largest benefits, although for $N = 500$ the improvements due to NLS of the target are close.

Appendix Tables B.4 and B.5 contain results of the GMV portfolio analysis using a shorter estimation window of 1250 days. Qualitatively, we find similar results. We note however that the relative benefits of the CLIP-DCC model and NLS of the target depend on the concentration ratio N/T. For $T = 1250$, we find that the break-even point is at $N = 100$, where both techniques provide roughly equal value. For the smaller portfolio dimensions below $N = 100$ the CLIP-DCC model reduces the SD the most, while for the large portfolio sizes $N = 300$ and $N = 500$ NLS of the target is most useful. In sum, the NLS method appears not necessary for 'small' N/T, but its benefits increase as this ratio increases. This is directly in line with intuition, because as the concentration ratio N/T rises the quality of the sample covariance matrix deteriorates, see again Ledoit and Wolf

(2004). On the other hand, benefits of the CLIP-DCC model start earlier, for example even for $N = 10$ and $T = 2500$ small improvements are found, but relative improvements over the base DCC model grow at slower rate with N/T.

Table 4: Daily out-of-sample GMV porfolio performance constructed using different DCC models for $N \in \{10, 30, 50, 100, 300, 500\}$ and NLS of the target, December 1990 until December 2020.

		DCC	dDECO	sdDECO	CLIP-DCC	1/N
	AV	4.858	3.759	5.787	5.277	7.283
$N = 10$	SD	15.471	15.745	15.474	**15.388***	19.167
	IR	0.314	0.239	0.374	0.343	0.380
	AV	6.064	5.271	6.377	6.261	8.041
$N = 30$	SD	13.891	14.845	13.826	**13.649*****	18.457
	IR	0.437	0.355	0.461	0.459	0.436
	AV	4.501	3.199	3.882	4.247	8.458
$N = 50$	SD	13.035	14.446	13.115	**12.812*****	18.271
	IR	0.345	0.221	0.296	0.331	0.463
	AV	2.979	1.356	1.358	2.517	8.273
$N = 100$	SD	10.845	13.132	10.956	**10.682*****	18.225
	IR	0.275	0.103	0.124	0.236	0.454
	AV	6.198	1.665	2.791	5.313	9.034
$N = 300$	SD	7.841	11.161	7.746	**7.608*****	17.975
	IR	0.790	0.149	0.360	0.698	0.503
	AV	4.865	0.820	1.779	4.305	9.558
$N = 500$	SD	6.400	9.137	6.206	**6.122*****	17.961
	IR	0.760	0.090	0.287	0.703	0.532

Note: This table contains the annualized average (AV), standard deviation (SD) and information ratio (IR) of the out-of-sample daily log returns for the GMV portfolios constructed using different dynamic correlations models and the $1/N$ portfolio. The lowest SD per dimension size is highlighted in bold. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 2500 days and re-estimation of the parameters every 21 days. A significant decrease of the (logarithmic squared) SD of the CLIP-DCC model compared to the DCC model is indicated with a *,** and *** for a $p - value$ below 0.1, 0.05 and 0.01, respectively, using the two-sided test by Ledoit and Wolf (2011) with HAC standard errors.

Finally, we examine the performance of the block-based correlation models. We follow Engle and Kelly (2012) and use industry group membership based on SIC codes to impose

a block structure. Specifically, we use the $K = 5$ and $K = 10$ Fama-French industry categorisation and estimate the BL-dDECO, BL-sdDECO and BL-CLIP-DCC models for $N \in \{100, 300, 500\}$, also employing NLS of the target. Table 5 summarizes the out-of-sample performance of the GMV portfolios constructed using these models.

Table 5: Daily out-of-sample GMV porfolio performance constructed using different industry block-based DCC models for $N \in \{100, 300, 500\}$ and NLS of the target, December 1990 until December 2020.

$K = 5$		BL-dDECO	BL-sdDECO	BL-CLIP-DCC	1/N
	AV	1.112	1.063	2.233	8.273
$N = 100$	SD	12.116	10.890	**10.675**	18.225
	IR	0.092	0.098	0.209	0.454
	AV	1.594	2.818	5.250	9.034
$N = 300$	SD	10.130	7.711	**7.590**	17.975
	IR	0.157	0.365	0.692	0.503
	AV	1.023	1.939	4.328	9.558
$N = 500$	SD	8.394	6.208	**6.109**	17.961
	IR	0.122	0.312	0.708	0.532

$K = 10$		BL-dDECO	BL-sdDECO	BL-CLIP-DCC	1/N
	AV	2.028	1.842	2.694	8.273
$N = 100$	SD	11.657	10.872	**10.694**	18.225
	IR	0.174	0.169	0.252	0.454
	AV	2.890	3.084	5.336	9.034
$N = 300$	SD	9.080	7.731	**7.598**	17.975
	IR	0.318	0.399	0.702	0.503
	AV	1.628	2.096	4.339	9.558
$N = 500$	SD	7.347	6.234	**6.122**	17.961
	IR	0.222	0.336	0.709	0.532

Note: This table contains the annualized average (AV), standard deviation (SD) and information ratio (IR) of the out-of-sample daily log returns for the GMV portfolios constructed using different block-based dynamic correlations models and the $1/N$ portfolio. The lowest SD per dimension size is highlighted in bold. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 2500 days and re-estimation of the parameters every 21 days. Five ($K = 5$) and ten ($K = 10$) Fama-French industry membership is used as a block-structure.

In Table 5, we observe large improvements of the BL-dDECO model over the dDECO

model, some improvements of the BL-sdDECO model over the sdDECO model and nearly no benefit of the BL-CLIP-DCC model over the CLIP-DCC model. Comparing the block structures, we find $K = 10$ to be superior for the BL-dDECO and BL-sdDECO models, conversely $K = 5$ is slightly better for the BL-CLIP-DCC model. This confirms that the poor performance of the dDECO model is mainly the result of imposing too much structure on the long-run correlation matrix and loosening it increases performance. For example, for $N = 500$ we find a SD of 9.137 for the dDECO model, a SD of 8.394 and 7.347 for the $K = 5$ and $K = 10$ BL-dDECO models and a SD of 6.400 for the DCC model. Furthermore, the fact that the BL-CLIP-DCC model appears to offer no benefit over the CLIP-DCC model suggests that correlation movement information within industries is not much more informative than information from assets in other industries. It would be interesting to consider an empirical application with data that admits a very clear group structure, e.g. when considering many assets from a few differently behaving asset classes. This is left for future research.

5 Conclusion

To better accommodate large cross-sections, we propose to augment the Dynamic Conditional Correlation (DCC) model of Engle (2002) by allowing for a degree of commonality in the update information. Specifically, we propose the DCC model with Conditional LInear Pooling (CLIP-DCC) where the degree of commonality is controlled by a parameter, such that the optimal level of structure is determined endogenously. Additionally, the CLIP-DCC model preserves long-run dynamics, thereby naturally complementing methods that shrink the target. Consequently, the CLIP-DCC model differs from the Block-DECO model in two key ways, as the latter requires an exogenous group allocation and also restricts long-run behavior. We show in a Monte Carlo study that the parameters of the CLIP-DCC model, including the pooling intensity parameter, can be effectively estimated using composite like-

lihood. A real-time empirical application to a large selection of US stocks from 1981 until 2020 indicates that the CLIP-DCC framework provides significant benefits for a minimum variance investor. We conclude that it is theoretically as well as economically interesting to linearly pool the movement of the DCC model around the long-run correlation matrix and that a combined approach with target shrinkage may be particularly rewarding.

References

Aielli, G.P. (2013). Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics* **31**, 282–299.

Ang, A. and J. Chen (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics* **63**, 443–494.

Bauwens, L., S. Laurent, and J.V. Rombouts (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics* **21**, 79–109.

De Nard, G., R.F. Engle, O. Ledoit, and M. Wolf (2020). Large dynamic covariance matrices: enhancements based on intraday data. *University of Zurich, Department of Economics, Working Paper*.

Engle, R.F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* **20**, 339–350.

Engle, R.F. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics* **30**, 212–228.

Engle, R.F., O. Ledoit, and M. Wolf (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics* **37**, 363–375.

Hafner, C.M. and O. Reznikova (2012). On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis* **56**, 3533–3545.

Ledoit, O. and M. Wolf (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management* **30**, 110–119.

— (2011). Robust performances hypothesis testing with the variance. *Wilmott* **2011**, 86–89.

— (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* **40**, 1024–1060.

— (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics* **48**, 3043–3065.

Michaud, R.O. (1989). The Markowitz optimization enigma: Is 'optimized'optimal? *Financial Analysts Journal* **45**, 31–42.

Pakel, C., N. Shephard, K. Sheppard, and R.F. Engle (2021). Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics* **39**, 652–668.

Patton, A.J. and K. Sheppard (2009). Evaluating volatility and correlation forecasts. In: T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (eds). *Handbook of Financial Time Series*. Springer Verlag, 801–838.

Roustant, O. and Y. Deville (2017). On the validity of parametric block correlation matrices with constant within and between group correlations. *Working paper, available at https://arxiv.org/abs/1705.09793*.

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In: T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (eds). *Handbook of Financial Time Series*. Springer Verlag, 201–229.

A Proofs

A.1 Lemma 1

First, we note that $\frac{1}{d}S$ with $d > 0$ is positive definite if and only if S is positive definite. Next, we may write $\frac{1}{d}S = \rho J_{N\times N} + (1-\rho)I_N$ with $\rho = \frac{o}{d}$, which may be shown to be positive definite if and only if $\frac{o}{d} \in (\frac{-1}{N-1}, 1)$. This follows from the fact that $\frac{1}{d}S$ has two distinct eigenvalues, $1 + (N-1)\rho$ and $1 - \rho$, which can be seen to both be positive if and only if $\frac{o}{d} \in (\frac{-1}{N-1}, 1)$. This also reveals that standardizing a positive definite compound symmetric matrix S yields a valid equicorrelation matrix $\frac{1}{d}S$, see also Definition 2.1 and Lemma 2.1 of Engle and Kelly (2012).

A.2 Lemma 2

For the case that $N = 1$ we define $\theta(\cdot)$ to be the identity mapping, such that the properties of Lemma 2 trivially hold. We now proceed with the case $N \geq 2$. By noting that $\theta(\cdot)$ is a linear function of $\theta^D(\cdot)$ and $\theta^O(\cdot)$, which are clearly linear themselves, we have that $\theta(\cdot)$ is a linear mapping and properties 1 and 2 immediately follow.

Using Lemma 1, it suffices for the positive definiteness part of property 3 to show that $\theta^D(A) > 0$ and $\frac{\theta^O(A)}{\theta^D(A)} \in (\frac{-1}{N-1}, 1)$. By definition we have that if A is positive definite that $x'Ax > 0, \forall x \in \mathbb{R}^N$ with $x \neq 0_N$. By selecting x to be a vector containing a single 1 at the i-th position and 0 elsewhere, we observe that $a_{ii} > 0$ for all $i \in 1, \ldots, N$. This implies that all diagonal elements are strictly larger than 0, such that clearly $\theta^D(A) > 0$. In addition, by selecting x to be a vector with 1 on the i-th position, -1 on the j-th position and 0 elsewhere, we obtain $a_{ii} + a_{jj} > 2a_{ij}$ for all $i, j \in 1, \ldots, N$, where $i \neq j$. From there is straightforward to show that $\theta^D(A) > \theta^O(A)$, such that $\frac{\theta^O(A)}{\theta^D(A)} < 1$. Finally, by selecting $x = \iota_N$ we obtain $N\theta^D(A) + N(N-1)\theta^O(A) > 0$ which in turn may be rewritten to show that $\frac{\theta^O(A)}{\theta^D(A)} > \frac{-1}{N-1}$. Together, this entails that $\frac{\theta^O(A)}{\theta^D(A)} \in (\frac{-1}{N-1}, 1)$ which concludes the proof.

For the positive semi-definite case of property 3, we note that if A is positive semi-

definite then by definition we have that $x'Ax \geq 0, \forall x \in \mathbb{R}^N$. Therefore we may use the same arguments as for the positive definite case but lose the strictness of the inequalities. Here we note that if $\theta^D(A) = 0$, this implies that A and also $\theta(A)$ are a $N \times N$ matrix of zeros, which is positive semi-definite (all eigenvalues are 0). Therefore for arguments that utilize $\frac{\theta^O(A)}{\theta^D(A)}$ we can consider the case that $\theta^D(A) > 0$. Because the eigenvalues of a CS matrix are $1 + (N-1)\frac{\theta^O(A)}{\theta^D(A)}$ and $1 - \frac{\theta^O(A)}{\theta^D(A)}$, at either of the bounds one of the eigenvalues is equal to 0, such that we are in the positive semi-definite case.

A.3 Lemma 3

From Lemma 2 we have that $\theta(\cdot)$ is a linear function. It can be straightforwardly verified that also the matrix averaging transformation $\tau(\cdot)$ is a linear function. Therefore, since $\theta^{BL}(\cdot, G)$ is composed of $\theta(\cdot)$ and $\tau(\cdot)$ operations, it is straightforward to show that properties 1 and 2 of Lemma 3 hold.

With regards to property 3, we follow the proof structure of Theorem 1 from Roustant and Deville (2017). Specifically, we first assume that A is positive semi-definite and define the block averaging function $\tau^{BL}(A, G)$ for $A \in \mathbb{R}^{N \times N}$ with $N \geq 2$ and block structure G,

$$\tau^{BL}(A, G) = \begin{bmatrix} \tau(A^*_{11}) & \tau(A^*_{12}) & \cdots & \tau(A^*_{1K}) \\ \tau(A^*_{21}) & \tau(A^*_{22}) & \cdots & \tau(A^*_{2K}) \\ \vdots & \vdots & \ddots & \vdots \\ \tau(A^*_{K1}) & \tau(A^*_{K2}) & \cdots & \tau(A^*_{KK}), \end{bmatrix}, \tag{A.1}$$

which may also be written as

$$\tau^{BL}(A, G) = L(G)AL(G)', \tag{A.2}$$

$$L(G) = \begin{bmatrix} 1/n_1 J_{n_1 \times n_1} & O_{n_1 \times n_2} & \cdots & O_{n_1 \times n_K} \\ O_{n_2 \times n_1} & 1/n_2 J_{n_2 \times n_2} & \cdots & O_{n_2 \times n_K} \\ \vdots & \vdots & \ddots & \vdots \\ O_{n_K \times n_1} & O_{n_K \times n_2} & \cdots & 1/n_K J_{n_K \times n_K}, \end{bmatrix}, \tag{A.3}$$

where $O_{n_j \times n_k}$ is a $n_j \times n_k$ matrix of zeros. Here $L(G)$ can be seen to be positive semi-definite as it is a block diagonal matrix with positive semi-definite diagonal blocks. From there it can be seen that $\tau^{BL}(A, G)$ is positive semi-definite if A is positive semi-definite. We then consider the difference $\Delta(A, G) = \theta^{BL}(A, G) - \tau^{BL}(A, G)$ which admits the following form

$$\Delta(A, G) = \begin{bmatrix} \theta(A^*_{11}) - \tau(A^*_{11}) & O_{n_1 \times n_2} & \cdots & O_{n_1 \times n_K} \\ O_{n_2 \times n_1} & \theta(A^*_{22}) - \tau(A^*_{22}) & \cdots & O_{n_2 \times n_K} \\ \vdots & \vdots & \ddots & \vdots \\ O_{n_K \times n_1} & O_{n_K \times n_2} & \cdots & \theta(A^*_{KK}) - \tau(A^*_{KK}), \end{bmatrix}, \tag{A.4}$$

that is, $\Delta(A, G)$ is a block diagonal matrix with diagonal matrices $\theta(A^*_{jj}) - \tau(A^*_{jj})$ for $j = 1, \ldots, K$. These diagonal matrices may be rewritten as

$$\begin{aligned} \theta(A^*_{jj}) - \tau(A^*_{jj}) &= \theta(A^*_{jj}) - \frac{1}{n_j^2} \iota'_{n_j} A^*_{jj} \iota_{n_j} J_{n_j \times n_j} \\ &= \theta(A^*_{jj}) - \frac{1}{n_j^2} [n_j(n_j - 1)\theta^O(A^*_{jj}) + n_j \theta^D(A^*_{jj})] J_{n_j \times n_j} \\ &= \theta^O(A^*_{jj}) J_{n_j \times n_j} + [\theta^D(A^*_{jj}) - \theta^O(A^*_{jj})] I_{n_j} - \frac{1}{n_j^2} [n_j(n_j - 1)\theta^O(A^*_{jj}) + n_j \theta^D(A^*_{jj})] J_{n_j \times n_j} \\ &= [\theta^D(A^*_{jj}) - \theta^O(A^*_{jj})] I_{n_j} + \frac{1}{n_j} [\theta^O(A^*_{jj}) - \theta^D(A^*_{jj})] J_{n_j \times n_j} \\ &= [\theta^D(A^*_{jj}) - \theta^O(A^*_{jj})][I_{n_j} - \frac{1}{n_j} J_{n_j \times n_j}], \end{aligned} \tag{A.5}$$

where $\theta^D(A^*_{jj}) - \theta^O(A^*_{jj})$ is a non-negative scalar (see again the proof of Lemma 2) and $I_{n_j} - \frac{1}{n_j} J_{n_j \times n_j}$ a positive semi-definite matrix. The latter fact may be derived from Lemma 1 by noting that it is a CS matrix or from noticing that is in fact a projection matrix. We

now have that $\Delta(A, G)$ is positive semi-definite because all its diagonal matrices are positive semi-definite. Since $\theta^{BL}(A, G) = \Delta(A, G) + \tau^{BL}(A, G)$, it follows that $\theta^{BL}(A, G)$ is also positive semi-definite as it is the sum of two positive semi-definite matrices.

Finally, we show that if A is positive definite that $\theta^{BL}(A, G)$ is also positive definite. For $x \in \mathbb{R}^N$ we consider $h(x) = x'\theta^{BL}(A, G)x$, which may also be written as

$$h(x) = x'\Delta(A, G)x + x'\tau^{BL}(A, G)x, \tag{A.6}$$

where $h(x)$ is 0 if and only if $x'\Delta(A, G)x$ and $x'\tau^{BL}(A, G)x$ are both equal to 0 (as neither can be negative due to positive semi-definiteness). First, we have for $x'\tau^{BL}(A, G)x$ that

$$x'\tau^{BL}(A, G)x = x'L(G)AL(G)'x = x^m(G)Ax^m(G), \tag{A.7}$$

$$x^m(G) = \left[(\tfrac{1}{n_1}\iota'_{n_1}x_1^*)\iota'_{n_j} \quad (\tfrac{1}{n_2}\iota'_{n_2}x_2^*)\iota'_{n_j} \quad \cdots \quad (\tfrac{1}{n_K}\iota'_{n_K}x_K^*)\iota'_{n_K}\right]', \tag{A.8}$$

where x_j^* for $j = 1, \ldots, K$ is the j-th subvector of x based on the group structure G. Because A is assumed positive definite we have that $x'\tau^{BL}(A, G)x$ is 0 if and only if $x^m(G) = 0_N$. We also observe that $x^m(G)$ can be viewed as a vector of group means, such that $x^m(G) = 0_N$ if and only if the group means of the vector x based on the group structure G are all 0. That is $x'\tau^{BL}(A, G)x = 0$ if and only if $\frac{1}{n_j}\iota'_{n_j}x_j^* = 0$ for are $j = 1, \ldots, K$. Note that if $n_j = 1$, i.e. for groups with only one member, this directly implies that that $x_j^* = 0$.

Second, we have for $x'\Delta(A, G)x$ that

$$x'\Delta(A, G)x = \sum_{j=1}^{K}[\theta^D(A_{jj}^*) - \theta^O(A_{jj}^*)]x_j^{*\prime}[I_{n_j} - \frac{1}{n_j}J_{n_j\times n_j}]x_j^*, \tag{A.9}$$

which is 0 if and only if $x_j^{*\prime}[I_{n_j} - \frac{1}{n_j}J_{n_j\times n_j}]x_j^* = 0$ for all $j = 1, \ldots, K$, because $[\theta^D(A_{jj}^*) - \theta^O(A_{jj}^*)] > 0$ for all $j = 1, \ldots, K$ with $n_j \geq 2$ if A is positive definite. Using the symmetric square root of $I_{n_j} - \frac{1}{n_j}J_{n_j\times n_j}$ we may show that $x_j^{*\prime}[I_{n_j} - \frac{1}{n_j}J_{n_j\times n_j}]x_j^* = 0$ implies that $[I_{n_j} - \frac{1}{n_j}J_{n_j\times n_j}]^{1/2}x_j^* = 0_{n_j}$, which in turn implies $[I_{n_j} - \frac{1}{n_j}J_{n_j\times n_j}]x_j^* = 0_{n_j}$. From there, we

observe that the condition $x_j^* - \iota_{n_j}(\frac{1}{n_j}\iota'_{n_j}x_j^*) = 0_{n_j}$ only holds if x_j^* is a scalar multiple of ι_{n_j}. If we combine this with the requirements for $x'\tau^{BL}(A,G)x = 0$, we then get that $h(x)$ is 0 if and only if $x_j^* = 0_{n_j}$ for all $j = 1,\ldots,K$, which is equivalent to $x = 0_N$. This means that $x'Ax > 0$ for all $x \in \mathbb{R}^N$ where $x \neq 0_N$. We conclude that if A is positive definite then $\theta^{BL}(A,G)$ is also positive definite.

B Additional Results

Table B.1: Daily out-of-sample MV portfolio (standard) performance constructed using different DCC models for $N \in \{10, 30, 50, 100, 300, 500\}$, December 1990 until December 2020.

		DCC	dDECO	sdDECO	CLIP-DCC	1/N
	AV	5.211	3.960	5.727	5.379	7.283
$N = 10$	SD	16.671	16.977	16.727	**16.630**	19.167
	IR	0.313	0.233	0.342	0.323	0.380
	AV	7.157	6.780	7.191	7.273	8.041
$N = 30$	SD	14.603	15.210	14.480	**14.353*****	18.457
	IR	0.490	0.446	0.497	0.507	0.436
	AV	5.522	5.098	4.693	5.268	8.458
$N = 50$	SD	13.609	14.507	13.647	**13.378*****	18.271
	IR	0.406	0.351	0.344	0.394	0.463
	AV	4.402	3.180	2.317	3.866	8.273
$N = 100$	SD	11.374	12.921	11.402	**11.178*****	18.225
	IR	0.387	0.246	0.203	0.346	0.454
	AV	7.396	3.530	3.527	6.646	9.034
$N = 300$	SD	8.601	11.025	8.442	**8.318*****	17.975
	IR	0.860	0.320	0.418	0.799	0.503
	AV	5.987	2.229	2.419	5.488	9.558
$N = 500$	SD	7.246	9.215	7.081	**6.953*****	17.961
	IR	0.826	0.242	0.342	0.789	0.532

Note: This table contains the annualized average (AV), standard deviation (SD) and information ratio (IR) of the out-of-sample daily log returns for the MV portfolios constructed using different dynamic correlations models and the $1/N$ portfolio. Specifically, we use the geometric sample mean and a 10 percent annual return target, similar to Engle and Kelly (2012). The lowest SD per dimension size is highlighted in bold. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 2500 days and re-estimation of the parameters every 21 days. A significant decrease of the (logarithmic squared) SD of the CLIP-DCC model compared to the DCC model is indicated with a *,** and *** for a $p - value$ below 0.1, 0.05 and 0.01, respectively, using the two-sided test by Ledoit and Wolf (2011) with HAC standard errors.

Table B.2: Daily out-of-sample MV portfolio (momentum) performance constructed using different DCC models for $N \in \{10, 30, 50, 100, 300, 500\}$, December 1990 until December 2020.

		DCC	dDECO	sdDECO	CLIP-DCC	1/N
	AV	5.028	4.261	5.497	5.201	7.283
$N = 10$	SD	16.317	16.666	16.374	**16.268**	19.167
	IR	0.308	0.256	0.336	0.320	0.380
	AV	6.300	5.822	6.584	6.492	8.041
$N = 30$	SD	14.016	15.176	13.983	**13.796*****	18.457
	IR	0.449	0.384	0.471	0.471	0.436
	AV	4.819	3.600	4.127	4.578	8.458
$N = 50$	SD	13.069	14.585	13.167	**12.853*****	18.271
	IR	0.369	0.247	0.313	0.356	0.463
	AV	3.765	2.117	2.226	3.413	8.273
$N = 100$	SD	11.012	13.178	11.129	**10.846*****	18.225
	IR	0.342	0.161	0.200	0.315	0.454
	AV	7.023	2.570	3.133	6.261	9.034
$N = 300$	SD	8.115	10.985	8.060	**7.895*****	17.975
	IR	0.865	0.234	0.389	0.793	0.503
	AV	5.450	1.756	1.806	4.950	9.558
$N = 500$	SD	6.789	9.149	6.697	**6.559*****	17.961
	IR	0.803	0.192	0.270	0.755	0.532

Note: This table contains the annualized average (AV), standard deviation (SD) and information ratio (IR) of the out-of-sample daily log returns for the MV portfolios constructed using different dynamic correlations models and the $1/N$ portfolio. Specifically, we use a momentum signal taking the geometric mean over the previous 252 days, excluding the most recent 21 days, similar to Engle et al. (2019). The target return is set to the arithmetic mean of this mean vector. The lowest SD per dimension size is highlighted in bold. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 2500 days and re-estimation of the parameters every 21 days. A significant decrease of the (logarithmic squared) SD of the CLIP-DCC model compared to the DCC model is indicated with a *,** and *** for a $p - value$ below 0.1, 0.05 and 0.01, respectively, using the two-sided test by Ledoit and Wolf (2011) with HAC standard errors.

Table B.3: Daily out-of-sample QLIKE using different DCC models for $N \in \{10, 30, 50, 100, 300, 500\}$, December 1990 until December 2020.

		DCC	dDECO	sdDECO	CLIP-DCC
	AV	14.449	15.109	14.432	14.357
$N = 10$	SD	13.216	13.045	12.897	12.890
	PI		0.308	0.460	0.481
	AV	41.882	46.497	41.326	41.136
$N = 30$	SD	34.644	32.843	32.897	33.286
	PI		0.208	0.495	0.578
	AV	71.932	78.882	69.883	69.905
$N = 50$	SD	51.532	47.935	47.933	48.777
	PI		0.218	0.560	0.661
	AV	153.837	164.419	145.324	146.550
$N = 100$	SD	97.050	86.360	86.837	89.782
	PI		0.279	0.683	0.796
	AV	515.756	517.575	452.134	464.607
$N = 300$	SD	307.942	237.952	253.684	270.684
	PI		0.465	0.892	0.945
	AV	973.402	905.116	823.696	851.357
$N = 500$	SD	550.633	378.781	434.648	464.157
	PI		0.611	0.949	0.974

Note: This table contains the average (AV) and standard deviation (SD) of the daily out-of-sample QLIKE score for the different DCC models. In addition, the proportion of improvement (PI) denotes the share of dates that the model has a lower QLIKE score than the DCC model. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 2500 days and re-estimation of the parameters every 21 days.

Table B.4: Daily out-of-sample GMV porfolio performance constructed using different DCC models for $N \in \{10, 30, 50, 100, 300, 500\}$ and $T = 1250$, December 1990 until December 2020.

		DCC	dDECO	sdDECO	CLIP-DCC	1/N
	AV	4.224	3.168	4.746	4.485	7.283
$N = 10$	SD	15.589	15.909	15.592	**15.521***	19.167
	IR	0.271	0.199	0.304	0.289	0.380
	AV	5.218	4.279	4.569	5.028	8.041
$N = 30$	SD	14.045	14.796	13.944	**13.854*****	18.457
	IR	0.372	0.289	0.328	0.363	0.436
	AV	2.890	2.068	1.990	2.559	8.458
$N = 50$	SD	13.373	14.499	13.308	**13.148** ***	18.271
	IR	0.216	0.143	0.150	0.195	0.463
	AV	3.239	0.853	0.976	2.725	8.273
$N = 100$	SD	11.142	13.104	11.173	**10.991*****	18.225
	IR	0.291	0.065	0.087	0.248	0.454
	AV	5.248	1.241	2.034	4.752	9.034
$N = 300$	SD	8.366	10.989	8.375	**8.205*****	17.975
	IR	0.627	0.113	0.243	0.579	0.503
	AV	4.928	0.883	1.919	4.496	9.558
$N = 500$	SD	7.167	9.022	7.148	**7.006*****	17.961
	IR	0.688	0.098	0.269	0.642	0.532

Note: This table contains the annualized average (AV), standard deviation (SD) and information ratio (IR) of the out-of-sample daily log returns for the GMV portfolios constructed using different dynamic correlations models and the $1/N$ portfolio. The lowest SD per dimension size is highlighted in bold. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 1250 days and re-estimation of the parameters every 21 days. A significant decrease of the (logarithmic squared) SD of the CLIP-DCC model compared to the DCC model is indicated with a *,** and *** for a $p - value$ below 0.1, 0.05 and 0.01, respectively, using the two-sided test by Ledoit and Wolf (2011) with HAC standard errors.

Table B.5: Daily out-of-sample GMV porfolio performance constructed using different DCC models for $N \in \{10, 30, 50, 100, 300, 500\}$, $T = 1250$ and NLS of the target, December 1990 until December 2020.

		DCC	dDECO	sdDECO	CLIP-DCC	1/N
	AV	4.227	3.169	4.731	4.476	7.283
$N = 10$	SD	15.583	15.908	15.582	**15.513***	19.167
	IR	0.271	0.199	0.304	0.289	0.380
	AV	5.282	4.280	4.735	5.143	8.041
$N = 30$	SD	14.010	14.791	13.897	**13.816*****	18.457
	IR	0.377	0.289	0.341	0.372	0.436
	AV	2.945	2.070	2.227	2.686	8.458
$N = 50$	SD	13.308	14.495	13.226	**13.081*****	18.271
	IR	0.221	0.143	0.168	0.205	0.463
	AV	3.181	0.855	1.085	2.734	8.273
$N = 100$	SD	11.028	13.100	11.056	**10.881*****	18.225
	IR	0.288	0.065	0.098	0.251	0.454
	AV	5.371	1.241	2.472	4.823	9.034
$N = 300$	SD	7.979	10.987	7.978	**7.813*****	17.975
	IR	0.673	0.113	0.310	0.617	0.503
	AV	4.571	0.883	2.034	4.054	9.558
$N = 500$	SD	6.590	9.021	6.492	**6.370*****	17.961
	IR	0.694	0.098	0.313	0.636	0.532

Note: This table contains the annualized average (AV), standard deviation (SD) and information ratio (IR) of the out-of-sample daily log returns for the GMV portfolios constructed using different dynamic correlations models and the $1/N$ portfolio. The lowest SD per dimension size is highlighted in bold. The out-of-sample periods ranges from December 1990 until December 2020 for a total of 7560 days, using an estimation window of 1250 days and re-estimation of the parameters every 21 days. A significant decrease of the (logarithmic squared) SD of the CLIP-DCC model compared to the DCC model is indicated with a *,** and *** for a $p - value$ below 0.1, 0.05 and 0.01, respectively, using the two-sided test by Ledoit and Wolf (2011) with HAC standard errors.