SUPPLEMENTARY MATERIAL FOR

Comparing Predictive Accuracy in the Presence of a Loss Function Shape Parameter

Sander Barendse and Andrew J. Patton

This Version: December 8, 2020

S.1 Equal predictive ability tests

In this section we consider tests of equal predictive ability, corresponding the null and alternative hypotheses (H_0' and H_1') in equations (4)-(5) of the main paper. We consider the test statistics:

$$\sup t_n^2 \equiv \sup_{\gamma \in \Gamma} t_n^2(\gamma) \tag{S.1}$$

$$\text{ave}\, t_n^2 \equiv \int_\Gamma t_n^2(\gamma) dJ(\gamma) \tag{S.2}$$

where J is a weight function on Γ, for example J may be Uniform on Γ.

As for the sup-t test statistic in the main paper, each of the above test statistics can be interpreted as functions $v(t_n)$, where v maps functionals on Γ to $\mathbb{R}$. Importantly, each of these functions is continuous with respect to the uniform metric, monotonic in the sense that if $Z_1(\gamma) \leq Z_2(\gamma)$ for all γ then $v(Z_1) \leq v(Z_2)$, and has the property that if $Z(\gamma) \to \infty$ for γ for some subset of Γ with positive mass under weight function J, then $v(Z) \to \infty$. Further, we can consider feasible versions of both of these test statistics based on S_n Monte Carlo samples of γ from distribution J on Γ, or based on a grid with K_n points in Γ.

The proofs presented in the next section are applicable to both the superior predictive ability tests presented in the main paper and the equal predictive ability tests outlined here. For the ave-t^2 we require absolute continuity of J in Proposition 1. This requirement is obviously met for J uniform on Γ.

S.2 Proofs

S.2.1 Proof of Theorem 1

Finite dimensional convergence of $\sqrt{n}\big(\bar{L}_n(\cdot) - E[L_1(\cdot)]\big)$ follows from a CLT for (centered) stationary mixing sequences (e.g. Theorem 4 in Doukhan et al. (1994)), and the Crámer-Wold device (Proposition 5.1 in White (2001)), under Assumptions 1, 2, and 4. The mixing condition of Theorem 4 in Doukhan et al. (1994)) is satisfied if $\lim_{T\to\infty}\sum_{t=1}^{T} t^{1/(r-1)}\alpha(t) < \infty$. It is easy to see that this holds for $\alpha(t) = O(t^{-A})$, with $A > r/(r-1)$. Notice that β-mixing implies α-mixing, with relation $\alpha(t) \le \frac{1}{2}\beta(t)$ between β-mixing and α-mixing coefficients (see, e.g., Doukhan et al. (1995, p. 397)). But under Assumption 1 $\beta(t)$ diminishes at a faster, geometric rate, such that the mixing condition is satisfied.

We apply Theorem 1 of Doukhan et al. (1995) to establish stochastic equicontinuity of $\sqrt{n}\big(\bar{L}_n(\cdot)-E[L_1(\cdot)]\big)$. First, notice from Application 1 in Doukhan et al. (1995) that the mixing condition is satisfied if $\lim_{T\to\infty}\sum_{t=1}^{T} t^{1/(r-1)}\beta(t) < \infty$, which was established in the preceding. Second, notice that under Assumption 2, the $L_{t+1}(\cdot)$ belong to $\mathcal{L}_{2r}$, where $\mathcal{L}_{2r}$ denotes the class of functions satisfying $\|f\|_{2r} < \infty$. From Application 1 in Doukhan et al. (1995) we then find that the entropy condition is satisfied if $\int_0^1 \sqrt{H_{[]}(\delta,\Gamma,\|\cdot\|_{2r})}du < \infty$, where $H_{[]}(\delta,\Gamma,\|\cdot\|_{2r})$ is defined as the natural logarithm of the $\mathcal{L}_{2r}$ bracketing numbers $N_{[]}(\delta,\Gamma,\|\cdot\|_{2r})$.

We can always choose N points in Γ, denoted γ_k, for $k = 1,\dots,N$, and collected in Γ_N, such that for each $\gamma \in \Gamma$, $\min_k|\gamma - \gamma_k| < GN^{-1/d}$, because Γ is a bounded subset of $\mathbb{R}^d$.

Assumption 3 implies that $\|L_{t+1}(\gamma) - L_{t+1}(\gamma')\|_{2r} \le \|L_{t+1}(\gamma) - L_{t+1}(\gamma')\|_{4r} \le C|\gamma - \gamma'|^{\lambda}$, for all $\gamma,\gamma' \in \Gamma$.

Setting $N(\delta) = \delta^{-d/\gamma}G^dC^{-d/\lambda}$, we therefore find that for all $\gamma \in \Gamma$ there exists a $\gamma_k \in \Gamma_N$ such that $\|L_{t+1}(\gamma) - L_{t+1}(\gamma_k)\|_{2r} \le C|\gamma - \gamma_k|^{\lambda} \le CG^{\lambda}N^{-\lambda/d} = \delta$. Hence, $N(\delta) = \delta^{-d/\lambda}G^dC^{-d/\lambda}$ satisfies the definition of the $\mathcal{L}_{2r}$-bracketing numbers. Moreover, the entropy condition $\int_0^1 H_{[]}(\delta,\Gamma,\|\cdot\|_{2r})du = \int_0^1 \log(C^{d/\lambda}G^d\delta^{-d/\lambda})d\delta = d\log(C^{1/\lambda}G)+\int_0^1 \delta^{-d/\lambda}d\delta = d\log(C^{1/\lambda}G)+\frac{1}{2}\sqrt{\pi d/\lambda} < \infty$ holds.

It follows from Theorem 1 in Doukhan et al. (1995) that $\sqrt{n}\big(\bar{L}_n(\cdot) - E[L_1(\cdot)]\big)$ is stochastically equicontinuous. Together with finite dimensional convergence this implies $\sqrt{n}\big(\bar{L}_n(\cdot) - E[L_1(\cdot)]\big) \Rightarrow Z(\cdot)$, with $Z(\cdot)$ a Gaussian process with covariance kernel $\Sigma(\cdot,\cdot)$.

Note that $\sigma_n^2(\cdot) \xrightarrow{a.s} \sigma^2(\cdot)$ uniformly over Γ under Assumption 4. That $v(\tau_n) \xrightarrow{d} v(\tilde{t})$ follows by application of the Continuous Mapping Theorem. □

S.2.2 Proof of Theorem 2

The result under H_0' follows from Theorem 1 and the distribution function of $v(\tilde{t})$ being absolutely continuous on $(0,\infty)$, and noting that $v(t_n) \leq v(\tau_n)$ for all n, since $E[L_1(\gamma)] \leq 0$ for all $\gamma \in \Gamma$. The absolute continuity of the distribution function of $v(\tilde{t})$ follows from $Z(\cdot)$ having a nondegenerate covariance kernel, and thus $\tilde{t}(\cdot)$ having nondegenerate covariance kernel under Assumption 4, and the particular functional forms of $v(\cdot)$ under consideration (see Theorem 11.1 of Davydov et al. (1998)).

The result under H_1' is established as follows. Under the assumptions of Theorem 1 it follows that $\bar{L}_n(\gamma) \xrightarrow{a.s} E\left[L_{t+1}(\gamma)\right] \equiv \Delta(\gamma)$, uniformly over Γ. Additionally, under Assumption 4 it follows that $\hat{\sigma}_n^2(\gamma) \xrightarrow{a.s.} \sigma_m^2(\gamma)$ uniformly over $\gamma \in \Gamma$, and $\inf_{\gamma\in\Gamma}\sigma^2(\gamma) > 0$.

By the Continuous Mapping Theorem $\sup_{\gamma\in\Gamma}\big(\bar{L}_n(\gamma)/\hat{\sigma}_n(\gamma)\big)^p \to \sup_{\gamma\in\Gamma}\big(E[L_1(\gamma)]/\sigma(\gamma)\big)^p \geq \Delta'$, for some $\Delta' > 0$, and $p = 1,2$. Hence, for $\sup t_n^p = n^{p/2}\sup_{\gamma\in\Gamma}\big(\bar{L}_n(\gamma)/\hat{\sigma}_n(\gamma)\big)^p$, $P[\sup t_n^p > c] \to 1$, for any constant $c \in \mathbb{R}$.

For the ave-t_n^2 test we use additionally that $\Gamma^\dagger \equiv \{\gamma : |\gamma - \gamma^\dagger|^\lambda < \Delta/C\}$ has positive mass under the J-measure, the constant C given in Assumption 3, since the density of J is assumed positive for all $\gamma \in \Gamma$. Then note that, for any $\gamma \in \Gamma$, $\big|E[L_{t+1}(\gamma^\dagger)]\big| - \big|E[L_{t+1}(\gamma^\dagger) - L_{t+1}(\gamma)]\big| \leq \big|E[L_{t+1}(\gamma)]\big|$ by the Triangle Inequality. Furthermore, from Jensen's inequality, Hölder's inequality and under Assumption 3, it follows that

$$\begin{aligned}\big|E[L_{t+1}(\gamma^\dagger) - L_{t+1}(\gamma)]\big| &\leq E\big[\big|L_{t+1}(\gamma^\dagger) - L_{t+1}(\gamma)\big|\big] \\ &\leq \big\|L_{t+1}(\gamma^\dagger) - L_{t+1}(\gamma)\big\|_{4r} \leq C|\gamma - \gamma^\dagger|^\lambda.\end{aligned}$$

Hence, there here exists a $\Delta'' > 0$ such that $|E\left[L_{t+1}(\gamma)\right]| = |\Delta(\gamma)| \geq \Delta''$, for all $\gamma \in \Gamma^\dagger$. It follows that $|\bar{L}_n(\gamma)| > \Delta''$, a.s., uniformly over $\Gamma^\dagger$. Hence, there exists a $\Delta''' > 0$ so that $n^{-1/2}|t_n(\gamma)| \xrightarrow{a.s.} \frac{|\bar{L}_n(\gamma)|}{\sigma(\gamma)} > \Delta'''$, a.s., uniformly over $\Gamma^\dagger$, and $P[\text{ave}t_n^2 > c] \to 1$, for any constant $c \in \mathbb{R}$. □

S.2.3 Proof of Theorem 3

That $\sqrt{n}\bar{L}_n^*(\cdot) \Rightarrow Z(\cdot)$ almost surely follows from Theorem 1 in Bühlmann (1995). Assumption A1, A2, and A3 in Bühlmann (1995) are satisfied under Assumptions 1, 2, and 5 respectively. Finally, Assumption A4 in that paper is established in the proof of Theorem 1, since $N(\delta)$ satisfies the definition of the $\mathcal{L}_{4r}$ bracketing numbers, and $N(\delta) = \delta^{-d/\lambda}G^dC^{-d/\lambda}$, for all $\delta > 0$.

Note that $\sigma_n^2(\cdot) \xrightarrow{a.s} \sigma^2(\cdot)$ uniformly over Γ under Assumption 4, such that $t_n^*(\cdot) \Rightarrow \tilde{t}(\cdot)$ almost surely under the Continuous Mapping Theorem.

That $v(t_n^*) \xrightarrow{d} v(\tilde{t})$ in probability follows by application of a Continuous Mapping Theorem for bootstrapped processes (see Theorem 10.8 in Kosorok (2008)), given that the bootstrap is consistent in probability, which is implied by $\sqrt{n}\bar{L}_n^*(\cdot) \Rightarrow Z(\cdot)$ almost surely. The result follows. □

S.2.4 Proof of Proposition 1

We show the result for $\text{ave}\, t_n^2$. The result for the other tests follows from similar steps. *Part 1:* The weak convergence of t_n^2 as established in Theorem 1 and the Continuous Mapping Theorem, implies stochastic equicontinuity (see, e.g., Proposition 1 in Andrews (1994)), i.e., for all $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} P\left(\sup_{|\gamma-\gamma'|<\delta} \left| t_n^2(\gamma) - t_n^2(\gamma') \right| > \varepsilon \right) = 0,$$

where we again use the Euclidean metric to metrize Γ.

From absolute continuity of J it follows that that $\int_\Gamma dJ(\gamma) = \sum_{i=1}^{K_n} \int_{\Gamma_n^i} dJ(\gamma)$. Hence,

$$\begin{aligned}
\left| \int_\Gamma t_n^2(\gamma) dJ(\gamma) - \sum_{i=1}^{K_n} t_n^2(\gamma_{n,i}) \int_{\Gamma_n^i} dJ(\gamma) \right| &\leq \sum_{i=1}^{K_n} \int_{\Gamma_n^i} \left| t_n^2(\gamma) - t_n^2(\gamma_{n,i}) \right| dJ(\gamma) \\
&\leq \sum_{i=1}^{K_n} \sup_{\gamma\in\Gamma_n^i} \left| t_n^2(\gamma) - t_n^2(\gamma_{n,i}) \right| \int_{\Gamma_n^i} dJ(\gamma) \\
&\leq \sup_{|\gamma-\gamma'|<\delta_n} \left| t_n^2(\gamma) - t_n^2(\gamma_{n,i}) \right| \sum_{i=1}^{K_n} \int_{\Gamma_n^i} dJ(\gamma) \\
&= \sup_{|\gamma-\gamma'|<\delta_n} \left| t_n^2(\gamma) - t_n^2(\gamma_{n,i}) \right|,
\end{aligned}$$

For any $\varepsilon > 0$ there exists a $\delta > 0$ (with $\delta_n < \delta$ eventually), such that

$$\begin{aligned}
\limsup_{n\to\infty} P\left(\left| \int_\Gamma t_n^2(\gamma) dJ(\gamma) - \sum_{i=1}^{K_n} t_n^2(\gamma_{n,i}) \int_{\Gamma_n^i} dJ(\gamma) \right| > \varepsilon \right) &\leq \limsup_{n\to\infty} P\left(\sup_{|\gamma-\gamma'|<\delta_n} \left| t_n^2(\gamma) - t_n^2(\gamma_{n,i}) \right| > \varepsilon \right) \\
&\leq \limsup_{n\to\infty} P\left(\sup_{|\gamma-\gamma'|<\delta} \left| t_n^2(\gamma) - t_n^2(\gamma_{n,i}) \right| > \varepsilon \right) \\
&< \varepsilon,
\end{aligned}$$

where the last display follows from the stochastic equicontinuity of $t_n^2(\gamma)$. Because δ is arbitrary, the result follows.

Part 2: We cover Γ with some hyperrectangle $\bar{\Gamma}$, which we can do because Γ is a bounded subset of Euclidian space. Consider the d-dimensional hyperrectangular grid of $\bar{\Gamma}$ with $\bar{K}_n$ elements $\{\bar{\Gamma}_n^i\}_{i=1}^{\bar{K}_n}$, such that $\sup_{\gamma,\gamma'\in\bar{\Gamma}_n^i} |\gamma-\gamma'| < \delta_n$, for all $i = 1,\dots,\bar{K}_n$.

Now let $\{\Gamma_i^n\}_{i=1}^{K_n}$ be the K_n elements of $\{\bar{\Gamma}_n^i\}_{i=1}^{\bar{K}_n}$, such that $\Gamma_i^n \cap \Gamma$ is nonempty, and choose the $\gamma_{n,i}$ such that $\gamma_{n,i} \in \Gamma$.

We can expand

$$\begin{aligned}
\widetilde{\operatorname{ave} t_n^2} - \widehat{\operatorname{ave} t_n^2} &= \frac{1}{S_n}\sum_{j=1}^{S_n} t_n^2(\gamma^{(j)}) - \sum_{i=1}^{K_n} t_n^2(\gamma_{n,i}) \int_{\Gamma_n^i} dJ(\gamma) \\
&= \sum_{i=1}^{K_n} t_n^2(\gamma_{n,i}) \left\{ \frac{1}{S_n}\sum_{j=1}^{S_n} \mathbb{1}\left(\gamma^{(j)} \in \Gamma_n^i\right) - \int_{\Gamma_n^i} dJ(\gamma) \right\} \\
&\quad + \sum_{i=1}^{K_n} \frac{1}{S_n}\sum_{j=1}^{S_n} \left(t_n^2(\gamma^{(j)}) - t_n^2(\gamma_{n,i}) \right) \mathbb{1}\left(\gamma^{(j)} \in \Gamma_n^i\right) \\
&= A_n + B_n.
\end{aligned}$$

Notice that

$$\begin{aligned}
|A_n| &\leq \sup_{\gamma\in\Gamma} t_n^2(\gamma) \cdot \sum_{i=1}^{K_n} \left| \frac{1}{S_n}\sum_{j=1}^{S_n} \mathbb{1}\left(\gamma^{(j)} \in \Gamma_n^i\right) - \int_{\Gamma_n^i} dJ(\gamma) \right| \\
&\leq K_n \sup_{\gamma\in\Gamma} t_n^2(\gamma) \sup_{\Gamma'\subset\Gamma} \left| \frac{1}{S_n}\sum_{j=1}^{S_n} \mathbb{1}\left(\gamma^{(j)} \in \Gamma'\right) - \int_{\Gamma'} dJ(\gamma) \right| \\
&= K_n O_p(1) C_n,
\end{aligned}$$

where the last line follows from Theorem 1.

Furthermore, we can show that $S_n^{1/2-\eta} C_n = o_p(1)$, for any $\eta \in (0,1/2)$, where the probability statement now holds under the J-measure, by a CLT for iid empirical processes. Notice that due to the hyperrectangular shape of the Γ_n^i, we have for each $\Gamma_n^i \subset \bar{\Gamma}$

$$\mathbb{1}\left(\gamma \in \Gamma_i^n\right) = \prod_{i=1}^{d} \mathbb{1}\left(\gamma_i \leq \overline{\Gamma_n^i}\right) \prod_{i=1}^{d}\left(1 - \mathbb{1}\left(\gamma_i \leq \underline{\Gamma_n^i}\right)\right), \tag{S.3}$$

with $\bar{\gamma}_i^n$ denotes the maximum of the ith coordinate of all points in Γ_n^i, and with $\underline{\gamma}_i^n$ denoting the mininum.

Indicator functions such as the factors in (S.3) are type I(b) functions in the definition of Andrews (1994), and by Theorem 3 in Andrews (1994) so is the product (S.3). A functional CLT follows from Theorem 1 and 2 in Andrews (1994)), and by application of the Continuous Mapping Theorem we find $\sup_{\Gamma'\subset\bar{\Gamma}}\left|\frac{1}{S_n}\sum_{j=1}^{S_n}\mathbb{1}\left(\gamma^{(j)}\in\Gamma'\right)-\int_{\Gamma'}dJ(\gamma)\right|=O_p(S_n^{-1/2})$. Hence, $S_n^{1/2-\eta}C_n=O_p(1)$.

Furthermore, notice that

$$\begin{aligned}
|B_n| &\leq \frac{1}{S_n}\sum_{j=1}^{S_n}\sum_{i=1}^{K_n}\left|t_n^2(\gamma^{(j)})-t_n^2(\gamma_{n,i})\right|\mathbb{1}\left(\gamma^{(j)}\in\Gamma_n^i\right)\\
&\leq 2^d\frac{1}{S_n}\sum_{j=1}^{S_n}\sup_{|\gamma^{(j)}-\gamma'|<\delta_n}\left|t_n^2(\gamma^{(j)})-t_n^2(\gamma')\right|\\
&\leq 2^d\sup_{|\gamma-\gamma'|<\delta_n}\left|t_n^2(\gamma)-t_n^2(\gamma')\right|=o_p(1),
\end{aligned}$$

by the stochastic equicontinuity of $t_n^2(\gamma)$ and where 2^d equals the maximum number of vertices shared amongst hyperrectangles in a hyperrectangular grid.

If we can choose $K_n=o(S_n^{-1/2+\eta})$ then $|A_n|=O_p(1)K_nC_n=O_p(1)o_p(1)=o_p(1)$. Hence, $|\widetilde{\text{ave}\,t_n^2}-\widehat{\text{ave}\,t_n^2}|=o_p(1)$. But we are free to choose the rate at which $K_n\to\infty$ as $n\to\infty$, so the result follows. □

S.3 Results for tests of equal predictive ability

In this section we present simulation and empirical results for tests of equal predictive ability, a two-sided counterpart to the one-sided test of superior predictive ability presented in the main paper.

S.3.1 Simulation results

In this section we present tables for the tests of equal predictive ability using the same three simulation designs as in the main paper.

For comparison with the proposed sup-t^2 and ave-t^2 tests, we consider a Bonferroni adjustment, as we do for the sup-t test considered in the main paper. We also study standard joint Wald tests. Consider some discrete parameter set $\Gamma_M=\{\gamma_1,\dots,\gamma_M\}\subset\Gamma$. The Wald test statistic is then obtained as:

$$\hat{Q}_n^h\equiv n\tilde{L}_n(\Gamma_M)'\hat{\Omega}_{M,n}^{-1}\tilde{L}_n(\Gamma_M), \tag{S.4}$$

where $\tilde{L}_n(\Gamma_M)\equiv\left(\bar{L}_n(\gamma_1)',\dots,\bar{L}_n(\gamma_M)'\right)'$, and $\hat{\Omega}_{M,n}$ is some HAC estimator of the asymptotic covariance matrix of $\sqrt{n}\tilde{L}_n(\Gamma_M)$, e.g., the estimator of Newey and West (1987).

A two-sided α-level test rejects the null hypothesis when $\hat{Q}_n^h > \chi^2_{M,1-\alpha}$, where $\chi^2_{M,1-\alpha}$ denotes the $(1-\alpha)$-quantile of a χ^2 distribution with M degrees of freedom. As M increases $\hat{\Omega}_{M,n}$ becomes near-singular, which can lead to erratic behavior in the test statistic; which we indeed observe in our simulation results below. A two-sided α-level test using the Bonferroni correction rejects the null hypothesis if, for at least one $\gamma \in \Gamma_M$, we find $n\bar{L}_n^2(\gamma)/\tilde{\sigma}_n^2(\gamma) > \chi^2_{1,1-\alpha/M}$, with $\tilde{\sigma}_n^2(\gamma)$ a HAC asymptotic covariance estimator of $\sqrt{n}\bar{L}_n(\gamma)$.

Tables S.1 to S.3 present the results for the three simulation designs, analogous to the results in Tables 1 to 3 of the main paper for the test of superior predictive ability. We find comparable size and power results to those in the main paper, although, as expected from the theory, the two-sided tests generally have size closer to the nominal value of 5% than the one-sided sup-t test. Regarding the joint Wald test, the near-singularity of the asymptotic covariance matrix $\hat{\Omega}_{M,n}$, which occurs with increasing K_n and S_n, impacts the test in such a way that we always reject for large K_n and S_n. Similar to the results in the main paper, the two-sided test based on the Bonferonni correction becomes conservative when we increase K_n and S_n.

S.3.2 Multi-step quantile forecasts of portfolio returns

In this section we study a multi-step forecast extension of the design in Section 4.2 of the main paper. We consider the h-period cumulative return vector $Y_{t+1:t+h} := \sum_{j=1}^{h} Y_{t+j}$, and consider forecasts derived from the GARCH-DCC and RiskMetrics models defined in Section 4.2.

The RiskMetrics forecast is given by

$$Q^{\mathrm{RM}}_{h|t,\alpha}(\gamma) = \Phi^{-1}(\alpha)\sqrt{\gamma'\hat{\Sigma}_{h|t}\gamma}, \tag{S.5}$$

where $\hat{\Sigma}_{h|t} = h \cdot \hat{\Sigma}_{t+1}$. This linear scaling of the h-step covariance matrix follows from the random walk structure of the RiskMetrics model.

Table S.1: Small sample rejection rates of equal expected utility tests on equal weighted and minimum-variance portfolio strategies

	n=500				**n=2,000**			
K_n	Wald	Bonf.	ave-t^2	sup-t^2	Wald	Bonf.	ave-t^2	sup-t^2
Panel A: Size properties								
1	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.06
10	0.51	0.02	0.07	0.05	0.57	0.02	0.06	0.04
50	0.50	0.01	0.08	0.05	0.54	0.01	0.05	0.04
100	0.50	0.01	0.08	0.06	0.50	0.01	0.07	0.04
250	0.53	0.01	0.09	0.07	0.56	0.00	0.08	0.05
Panel B: Power properties								
1	0.94	0.94	0.95	0.95	1.00	1.00	1.00	1.00
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50	0.81	0.99	1.00	1.00	0.83	1.00	1.00	1.00
100	0.64	0.99	1.00	1.00	0.72	1.00	1.00	1.00
250	0.42	0.95	1.00	1.00	0.53	1.00	1.00	1.00

Note: This table presents the rejection rates of the proposed two-sided tests (ave-t^2 and sup-t^2), as well as the benchmark tests (Wald and Bonferroni). The data is generated according to equation (18), and the equal-weighed and minimum-variance portfolio strategies are given in equation (17). The minimum-variance portfolio weights are estimated using a rolling window of $m = 120$ observations. The out-of-sample period consists of $n = 500$, and $2,000$ observations. We consider discrete grids of $\Gamma = [1, 10]$ formed using $K_n = 1, 10, 50, 100$, and 250 equally spaced grid points.

We compare this with the GARCH-DCC forecast:

$$Q^{\text{DCC}}_{h|t,\alpha}(\gamma) = \Phi^{-1}(\alpha)\sqrt{\gamma'\hat{\Omega}_{h|t}\gamma} \tag{S.6}$$

$$\begin{aligned}
\text{where} \quad \hat{\Omega}_{h|t} &= \sum_{j=1}^{h} \hat{H}^{1/2}_{t+j|t}\hat{C}_{t+j|t}\hat{H}^{1/2}_{t+j|t}, \\
\hat{H}_{t+j|t} &= \operatorname{diag}\left(\hat{h}_{t+j|t,1}, \ldots, \hat{h}_{t+j|t,N}\right), \\
\hat{h}_{t+j|t,i} &= \omega_0 \sum_{k=0}^{j-1}(\omega_1+\omega_2)^k + (\omega_1+\omega_2)^j h_{t,i}, \\
\hat{C}_{t+j|t} &= E_t[h_{t+j,i}]\big((1-\xi_1-\xi_2)\bar{C}\big)\sum_{k=0}^{j-1}(\xi_1+\xi_2)^k + (\xi_1+\xi_2)^j C_t,
\end{aligned}$$

We employ a simplification in generating the GARCH-DCC forecast to reduce computational

Table S.2: Small sample rejection rates of quantile forecast tests, for differences between multivariate GARCH-DCC and RiskMetrics models, two-sided tests

	n=500				n=2,000			
S_n	Wald	Bonf.	ave-t^2	sup-t^2	Wald	Bonf.	ave-t^2	sup-t^2
Panel A: Size properties								
31	0.36	0.03	0.03	0.05	0.03	0.01	0.01	0.02
50	0.93	0.03	0.03	0.04	0.13	0.01	0.02	0.01
100	1.00	0.02	0.03	0.04	0.81	0.00	0.01	0.01
250	1.00	0.01	0.03	0.04	1.00	0.00	0.02	0.01
500	-	0.01	0.03	0.04	1.00	0.00	0.02	0.01
1000	-	0.01	0.03	0.05	1.00	0.00	0.02	0.01
Panel B: Power properties								
31	0.42	0.10	0.20	0.11	0.33	0.50	0.81	0.52
50	0.91	0.07	0.16	0.10	0.38	0.42	0.66	0.52
100	1.00	0.05	0.13	0.10	0.82	0.30	0.52	0.50
250	1.00	0.03	0.11	0.10	1.00	0.20	0.46	0.50
500	-	0.02	0.12	0.10	1.00	0.14	0.46	0.51
1000	-	0.01	0.11	0.10	1.00	0.09	0.43	0.49

Note: This table presents the rejection rates of the proposed two-sided tests (ave-t^2 and sup-t^2) as well as the benchmark tests (Wald and Bonferroni). The quantile forecasts for the portfolio returns from the GARCH-DCC and multivariate RiskMetrics models are defined in equations (20) and (21). The data is generated as in equation (19) with $N = 30$. We test at 31 fixed portfolio weight vector being the equal weighted portfolio vector and the 30 basis vectors, as well as $S_n - 31$ weight vectors drawn uniformly from the unit simplex.

time: we assume that the cumulative return $Y_{t+1:t+h}$ is normal, whereas it can be shown that $Y_{t+1:t+h}$ has kurtosis greater than a normal random variable, if the GARCH-DCC model is correct. The correct multi-period forecast does not have a closed-form expression and can only be obtained via simulation, which is prohibitive in the context of a simulation study involving a bootstrap test. The misspecification that follows from the simplification is not a problem in our testing framework, since we do not require correct specification of the forecasts.

Table S.4 presents small sample rejection rates of the size and power experiments, for the 10-day forecasts. Compared to the results for the 1-day forecast given in Table 2 we note that the sup-t test has size closer to the nominal value 5%. Moreover, in the power experiment the outperformance of the sup-t test relative to the Bonferonni correction test is now larger, although the rejection rates are up to 19 percentage points smaller across tests.

Table S.3: Small sample rejection rates of Murphy diagram tests, for quantile forecast differences between GARCH, RiskMetrics, and Rolling Window models, two-sided tests

		Panel A: Size				Panel B: Power			
		n=500		n=2,000		n=500		n=2,000	
Test	K_n	RM	RW	RM	RW	RM	RW	RM	RW
DM	1	0.07	0.21	0.06	0.16	0.11	0.32	0.37	0.90
avg-t^2	50	0.04	0.14	0.04	0.11	0.04	0.26	0.10	0.76
avg-t^2	100	0.03	0.14	0.05	0.13	0.04	0.33	0.10	0.85
avg-t^2	250	0.02	0.15	0.04	0.14	0.03	0.30	0.12	0.89
sup-t^2	50	0.04	0.12	0.05	0.08	0.04	0.27	0.08	0.72
sup-t^2	100	0.03	0.10	0.05	0.09	0.04	0.30	0.08	0.77
sup-t^2	250	0.02	0.09	0.04	0.07	0.03	0.27	0.10	0.82

Note: This table presents the rejection rates of the proposed two-sided tests (ave-t^2 and sup-t^2) comparing forecasts from a GARCH model with those from the RiskMetrics (RM) and Rolling Window (RW) forecasts. The Diebold-Mariano test (DM) using the tick loss function is also presented. The quantile forecasts from the GARCH and RiskMetrics models are given in equations (20) and (21), with $N = 1$. We consider out-of-sample period lengths $n = 500$, and 2,000, and discrete grids of $\Gamma = [-20, 0]$ with K_n equally-spaced points.

Table S.4: Small sample rejection rates of 10-day quantile forecast tests, for differences between multivariate GARCH-DCC and RiskMetrics models

	Panel A: Size				Panel B: Power			
	n=500		n=2,000		n=500		n=2,000	
S_n	Bonf.	sup-t	Bonf.	sup-t	Bonf.	sup-t	Bonf.	sup-t
31	0.01	0.06	0.02	0.06	0.11	0.22	0.35	0.47
50	0.01	0.06	0.02	0.06	0.08	0.23	0.26	0.47
100	0.00	0.05	0.01	0.07	0.05	0.22	0.18	0.46
250	0.00	0.05	0.01	0.06	0.03	0.22	0.09	0.46
500	0.00	0.05	0.00	0.06	0.02	0.22	0.06	0.46
1000	0.00	0.06	0.00	0.07	0.01	0.21	0.04	0.45

Note: This table presents the rejection rates of the proposed one-sided test (sup-t), as well as the benchmark Bonferroni test. The 10-step-ahead quantile forecasts for the portfolio returns from the GARCH-DCC and multivariate RiskMetrics models are defined in equations (S.5) and (S.6). The data is generated as in equation (19) with $N = 30$. We test at 31 fixed portfolio weight vector being the equal weighted portfolio vector and the 30 basis vectors, as well as $S_n - 31$ weight vectors drawn uniformly from the unit simplex.