Supplement of Nonlin. Processes Geophys., 27, 349–371, 2020
https://doi.org/10.5194/npg-27-349-2020-supplement

Nonlinear Processes in Geophysics
Open Access
EGU

Supplement of

Simulation-based comparison of multivariate ensemble post-processing methods

Sebastian Lerch et al.

Correspondence to: Sebastian Lerch (sebastian.lerch@kit.edu)

The supplemental material provides additional figures with results for all parameter combination in all simulation settings considered in the main paper, as well as an additional simulation setting based on a multivariate truncated normal distribution (Setting S1).

Contents

1 Additional results for Setting 1 **2**
- 1.1 Results for different numbers of ensemble members 2
- 1.2 Results for different numbers of dimensions 5
- 1.3 Results for additional simulation parameter choices 8

2 Additional results for Setting 2 **21**
- 2.1 Additional results for $m = 50$ ensemble members 21
- 2.2 Results for different numbers of ensemble members 25

3 Additional results for Setting 3A **34**

4 Additional results for Setting 3B **36**

5 Additional results for a multivariate truncated normal distribution-based setting (Setting S1) **46**
- 5.1 Simulation setup of Setting S1 46
- 5.2 Results for Setting S1 46
- 5.3 Additional results for Setting S1 49

1 Additional results for Setting 1

1.1 Results for different numbers of ensemble members

Figures 1 and 2 show results for Setting 1 in terms of the ES and VS with simulation parameters identical to those discussed in Section 4.2.1 of the paper, but for ensemble sizes of $m = 5, 10, 20, 35, 50, 100$ instead.

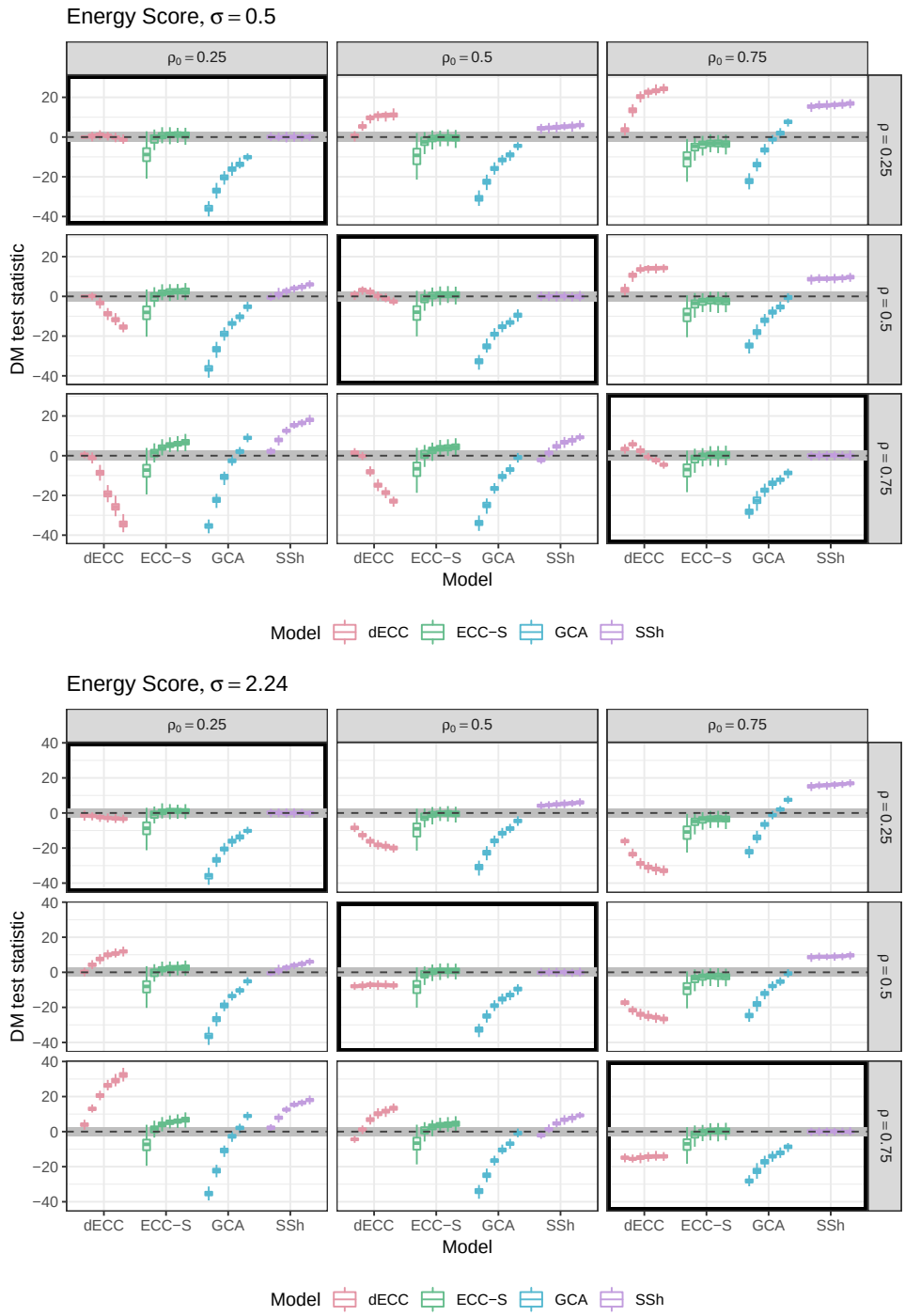


Figure 1: Summaries of DM test statistic values based on the ES for Setting 1 with $\epsilon = 1$, and $\sigma = 0.5$ (top), and $\sigma = \sqrt{5}$ (bottom). ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. For each model, the individual boxplots correspond to ensemble member numbers of $m = 5, 10, 20, 35, 50, 100$ from left to right (and light to dark shading). The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

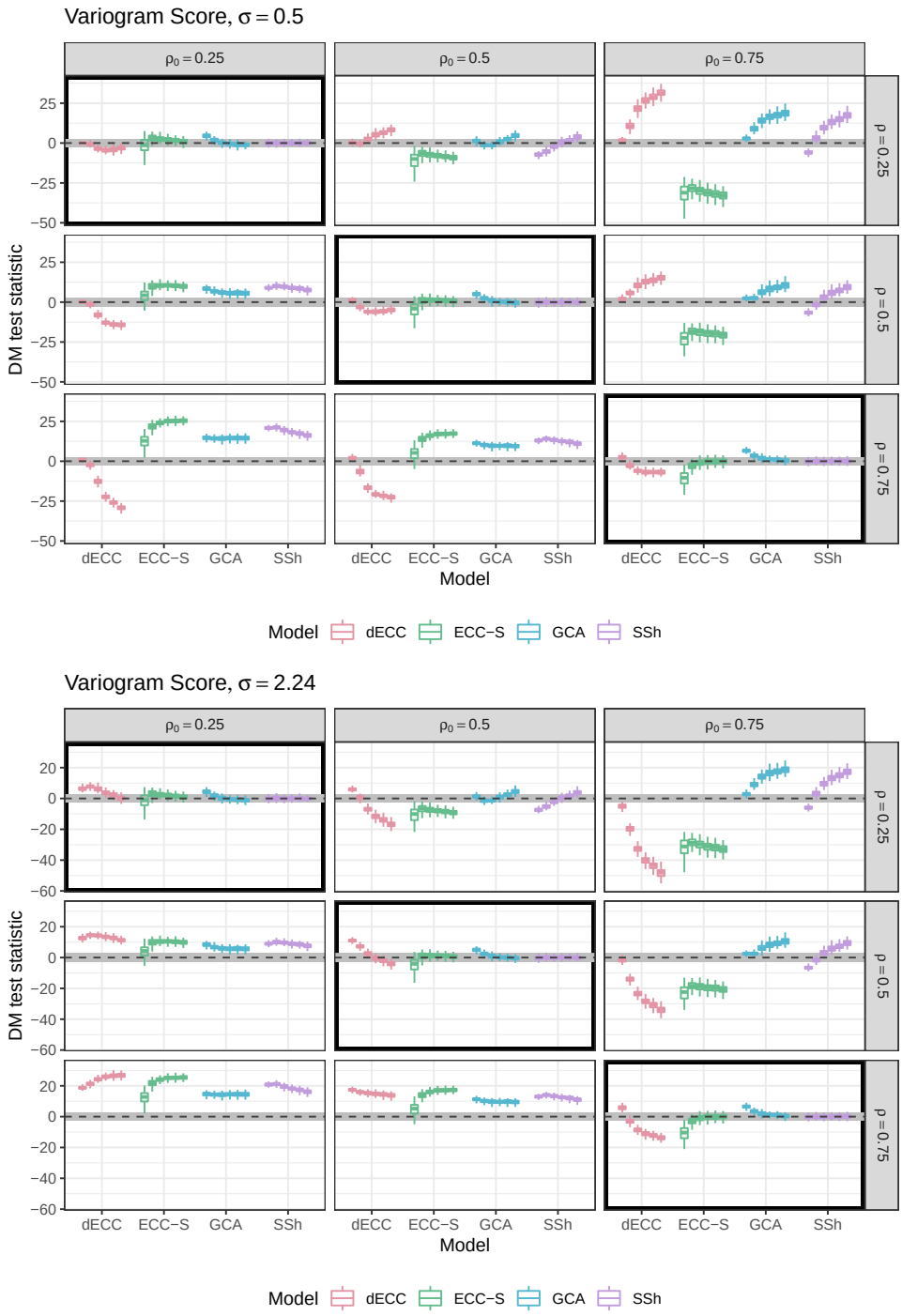


Figure 2: As Figure 1, but with results in terms of the VS.

1.2 Results for different numbers of dimensions

Figures 3 and 4 show results for Setting 1 in terms of the ES and VS with simulation parameters identical to those discussed in Section 4.2.1 of the paper, but for dimensions of $d = 2, 3, 4, 10, 20, 30, 50$ instead.

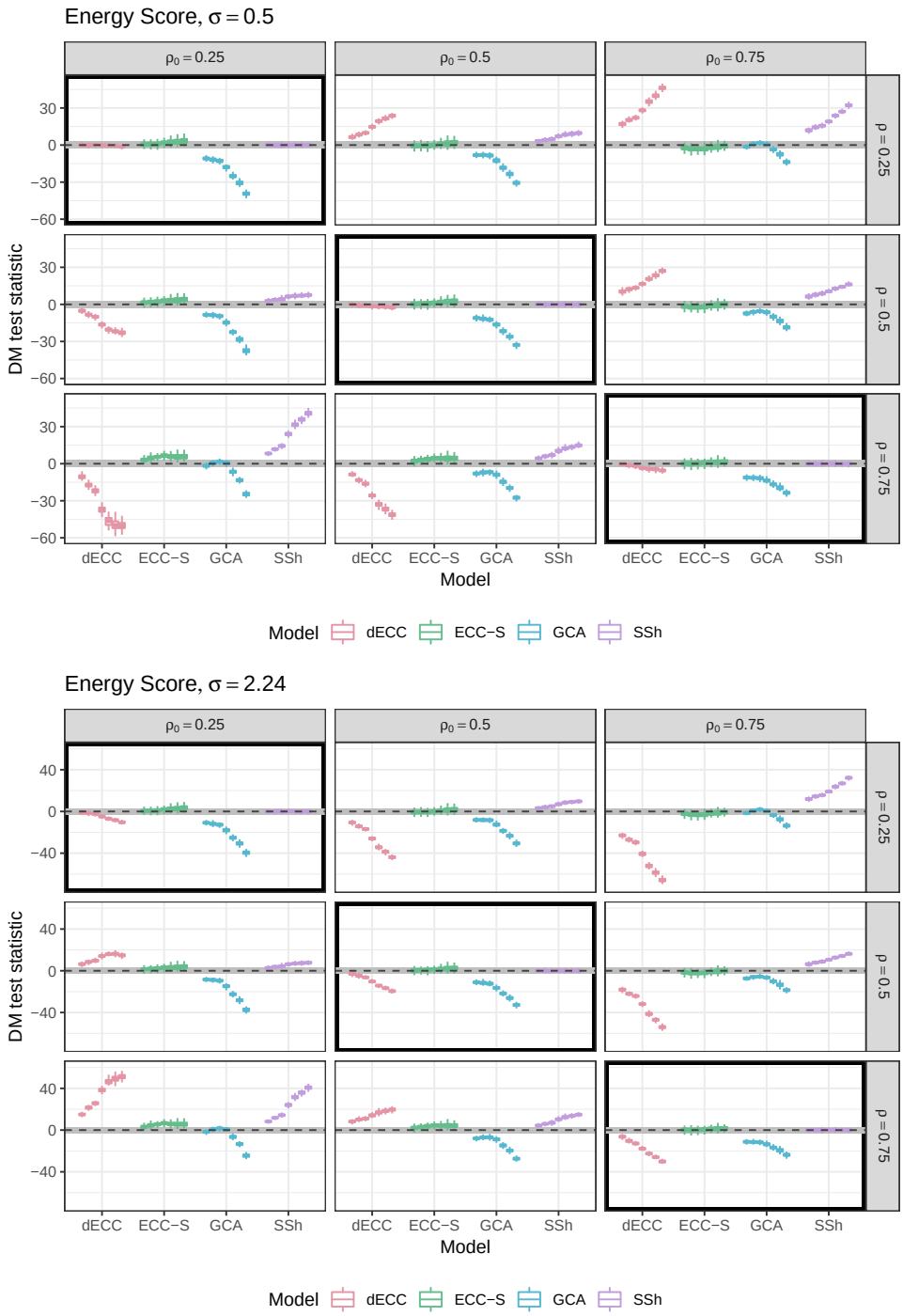


Figure 3: Summaries of DM test statistic values based on the ES for Setting 1 with $\epsilon = 1$, and $\sigma = 0.5$ (top), and $\sigma = \sqrt{5}$ (bottom). ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. For each model, the individual boxplots correspond to dimensions of $d = 2, 3, 4, 10, 20, 30, 50$ from left to right (and light to dark shading). The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

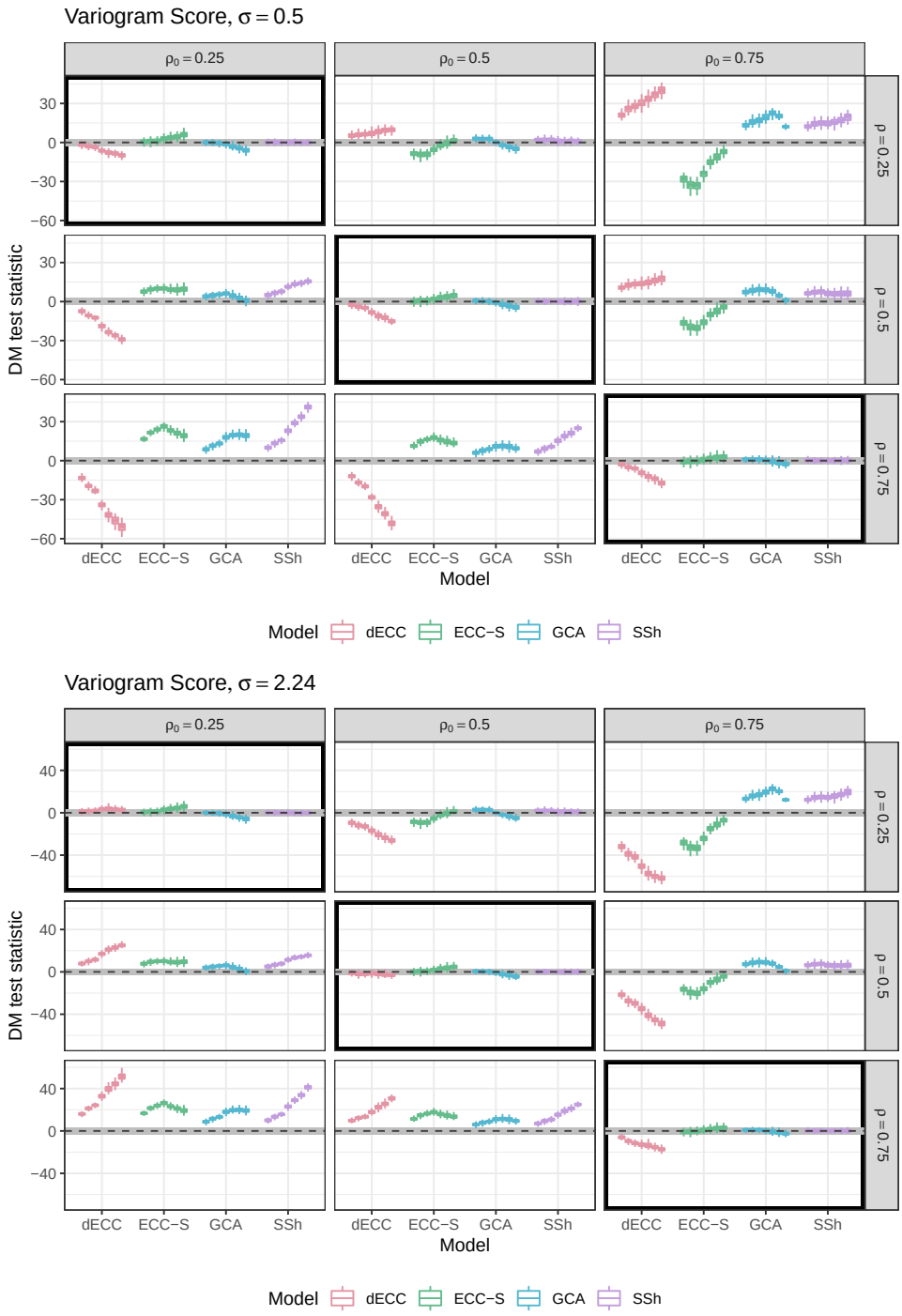


Figure 4: As Figure 3, but with results in terms of the VS.

1.3 Results for additional simulation parameter choices

An overview the additional results for Setting 1 is provided in Table 1.

Table 1: Overview of parameter combinations for Setting 1 with references to corresponding figures in this supplement.

ϵ	σ	Figure
0	0.5	5
0	1	6
0	$\sqrt{2}$	7
0	$\sqrt{5}$	8
1	0.5	9
1	1	10
1	$\sqrt{2}$	11
1	$\sqrt{5}$	12
3	0.5	13
3	1	14
3	$\sqrt{2}$	15
3	$\sqrt{5}$	16

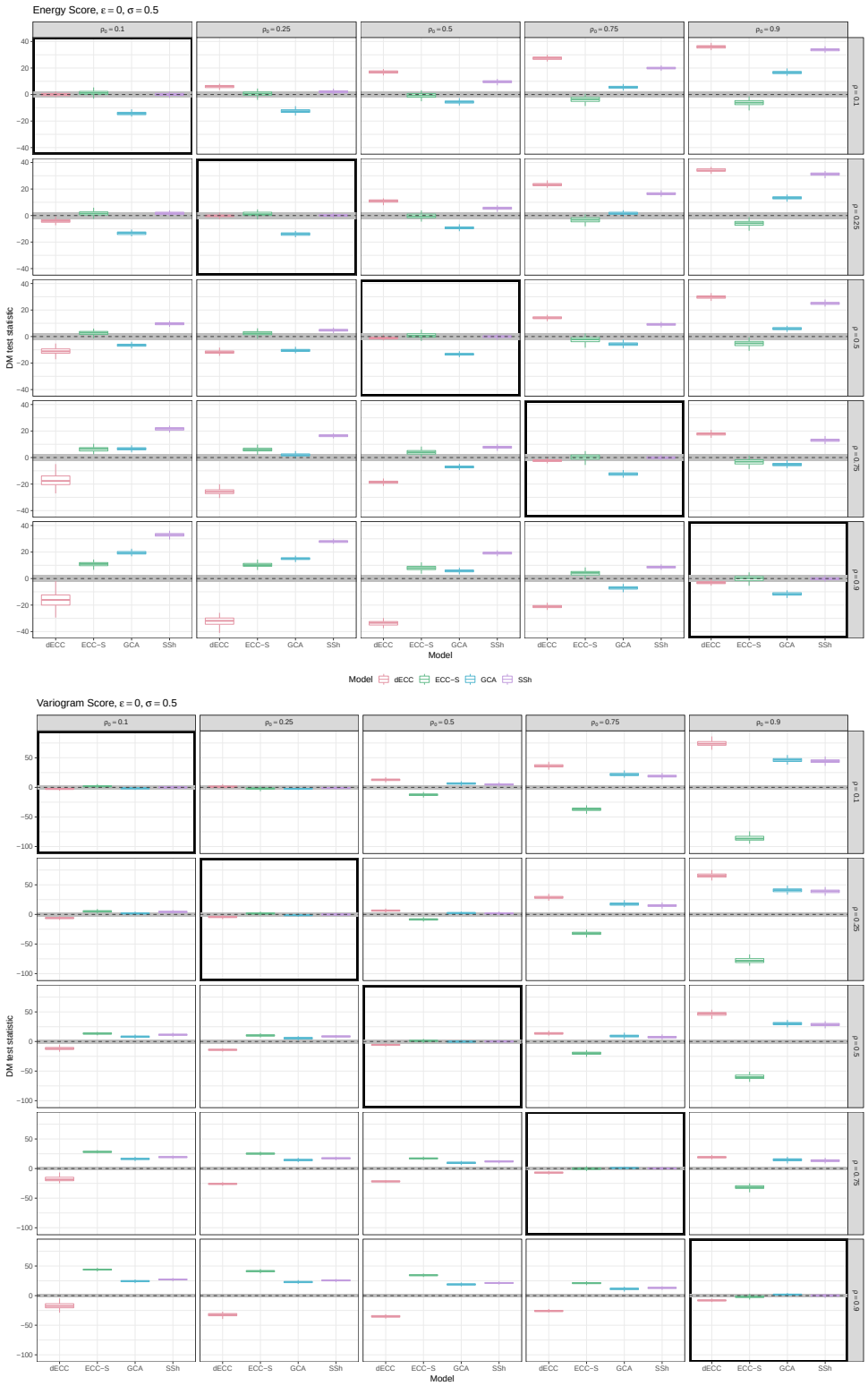
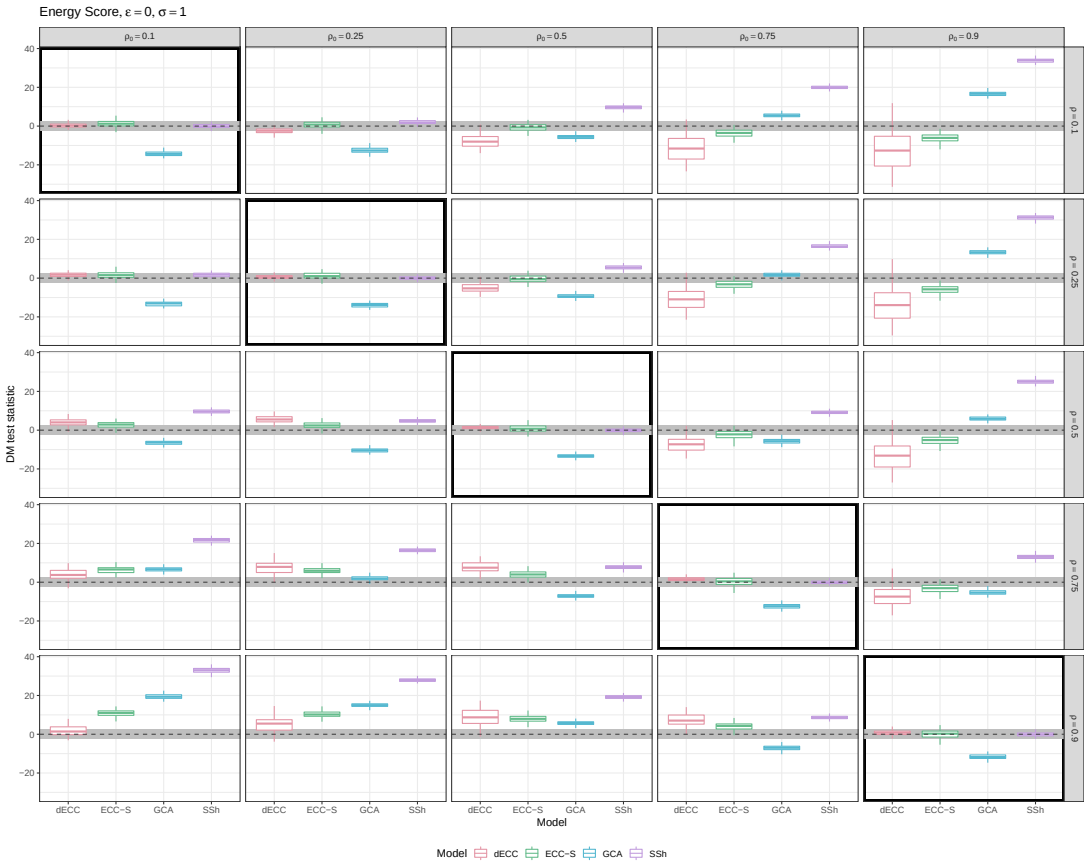


Figure 5: Summaries of DM test statistic values based on ES (top) and VS (bottom) for Setting 1 with $\epsilon = 0$, and $\sigma = 0.5$. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.



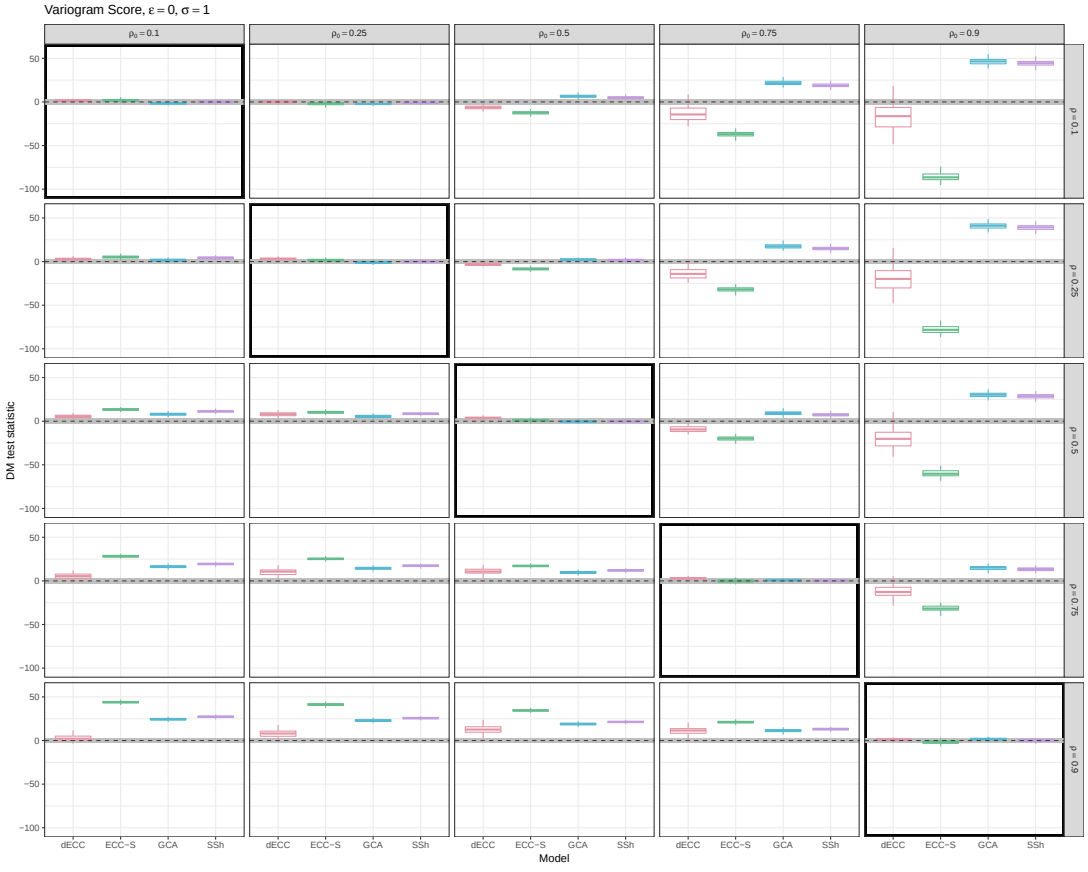
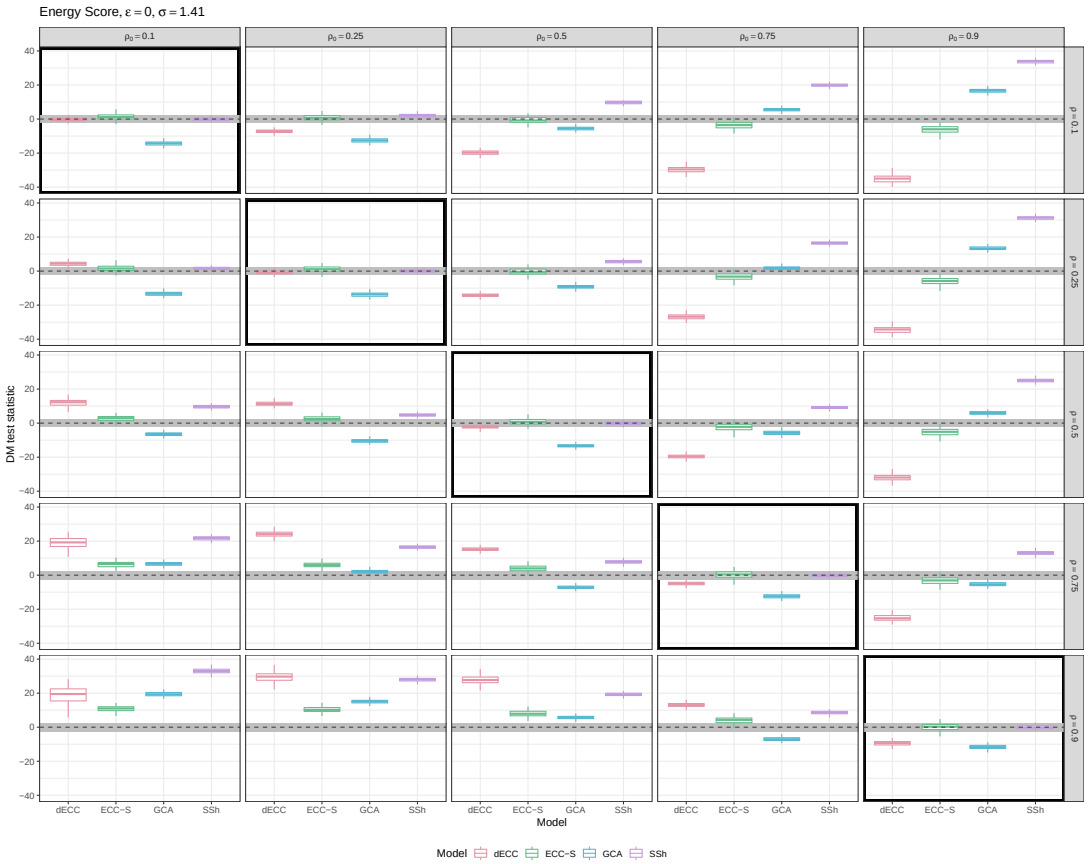


Figure 6: As Figure 5, but for $\epsilon = 0$, and $\sigma = 1$.



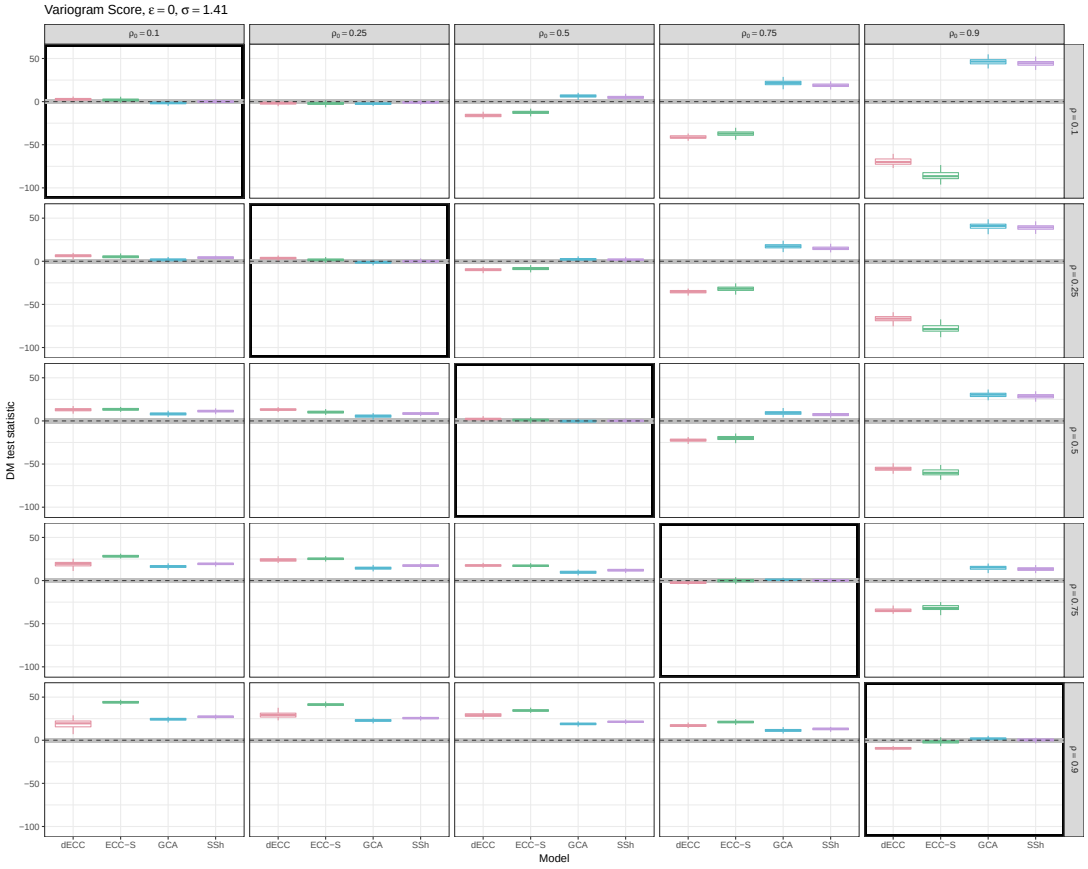
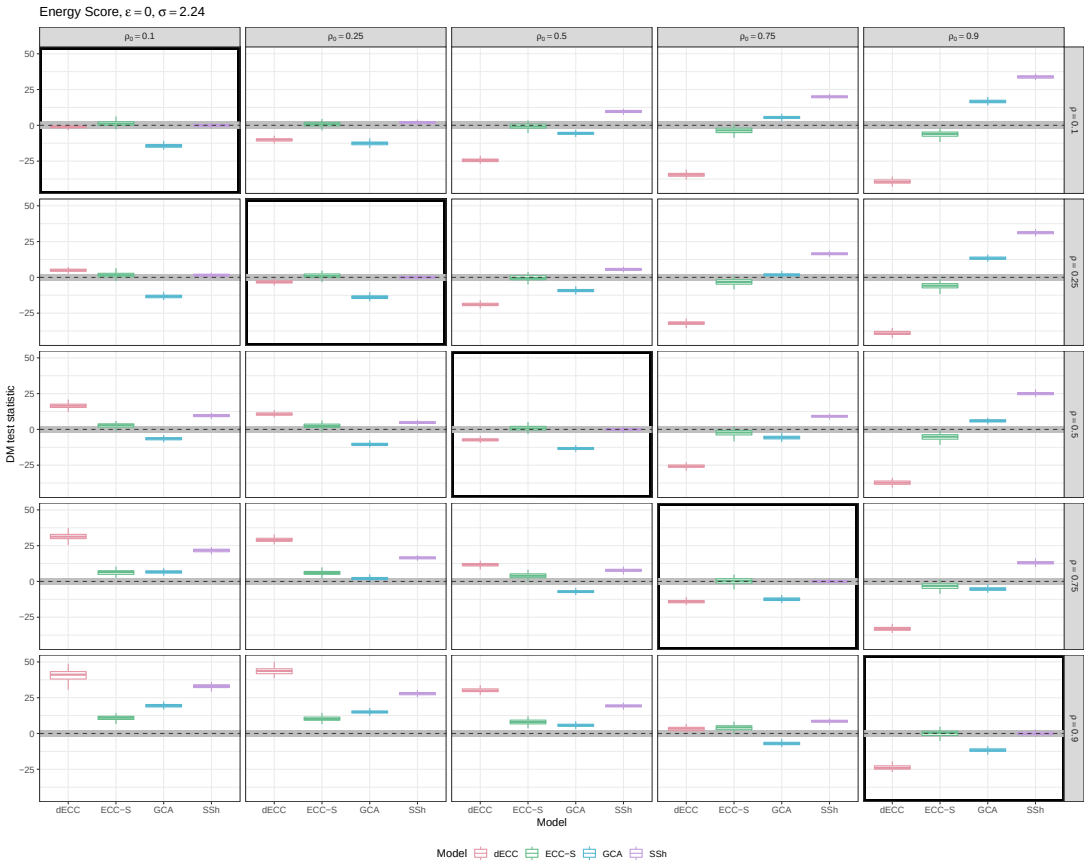


Figure 7: As Figure 5, but for $\epsilon = 0$, and $\sigma = \sqrt{2}$.



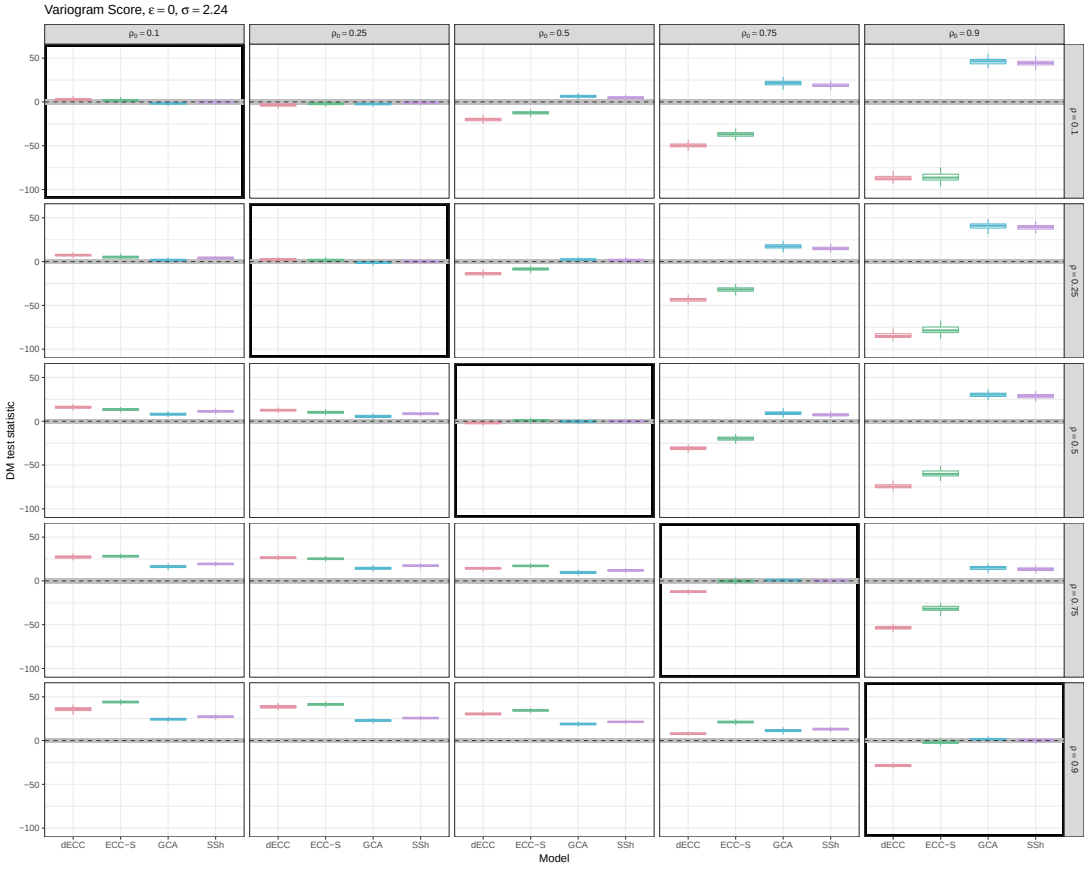


Figure 8: As Figure 5, but for $\epsilon = 0$, and $\sigma = \sqrt{5}$.

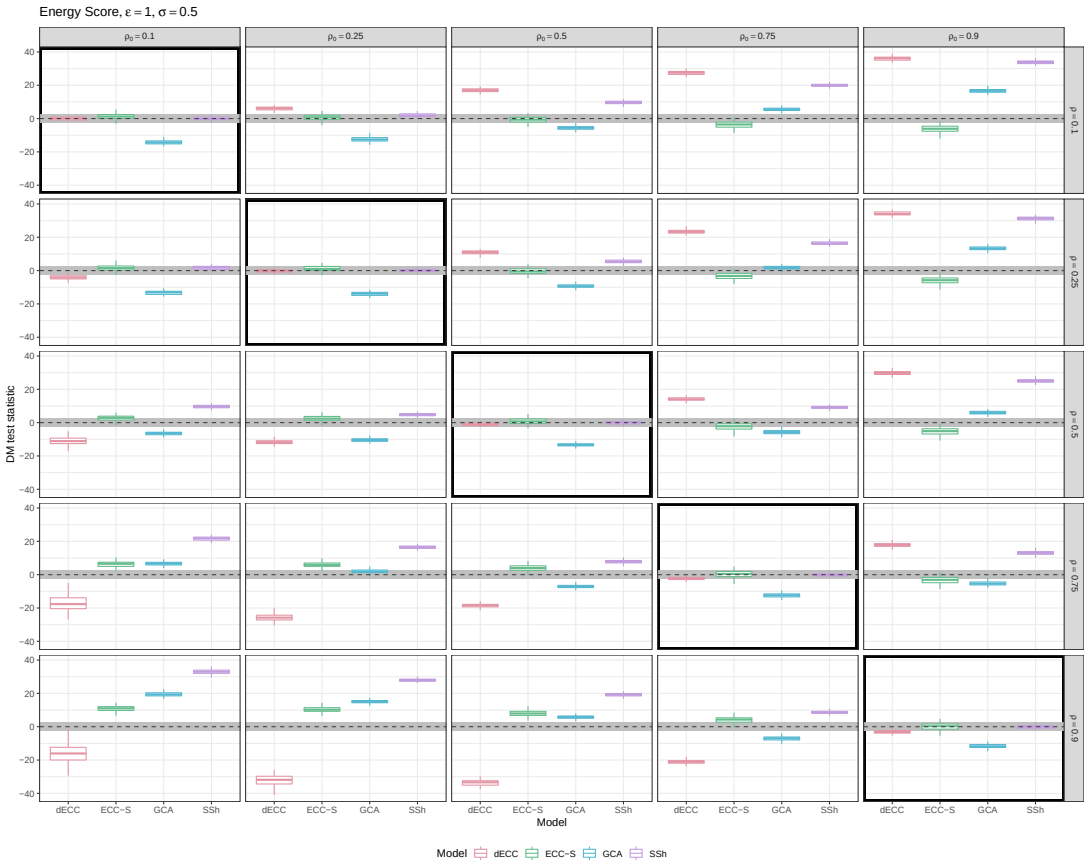


Figure 9: As Figure 5, but for $\epsilon = 1$, and $\sigma = 0.5$.

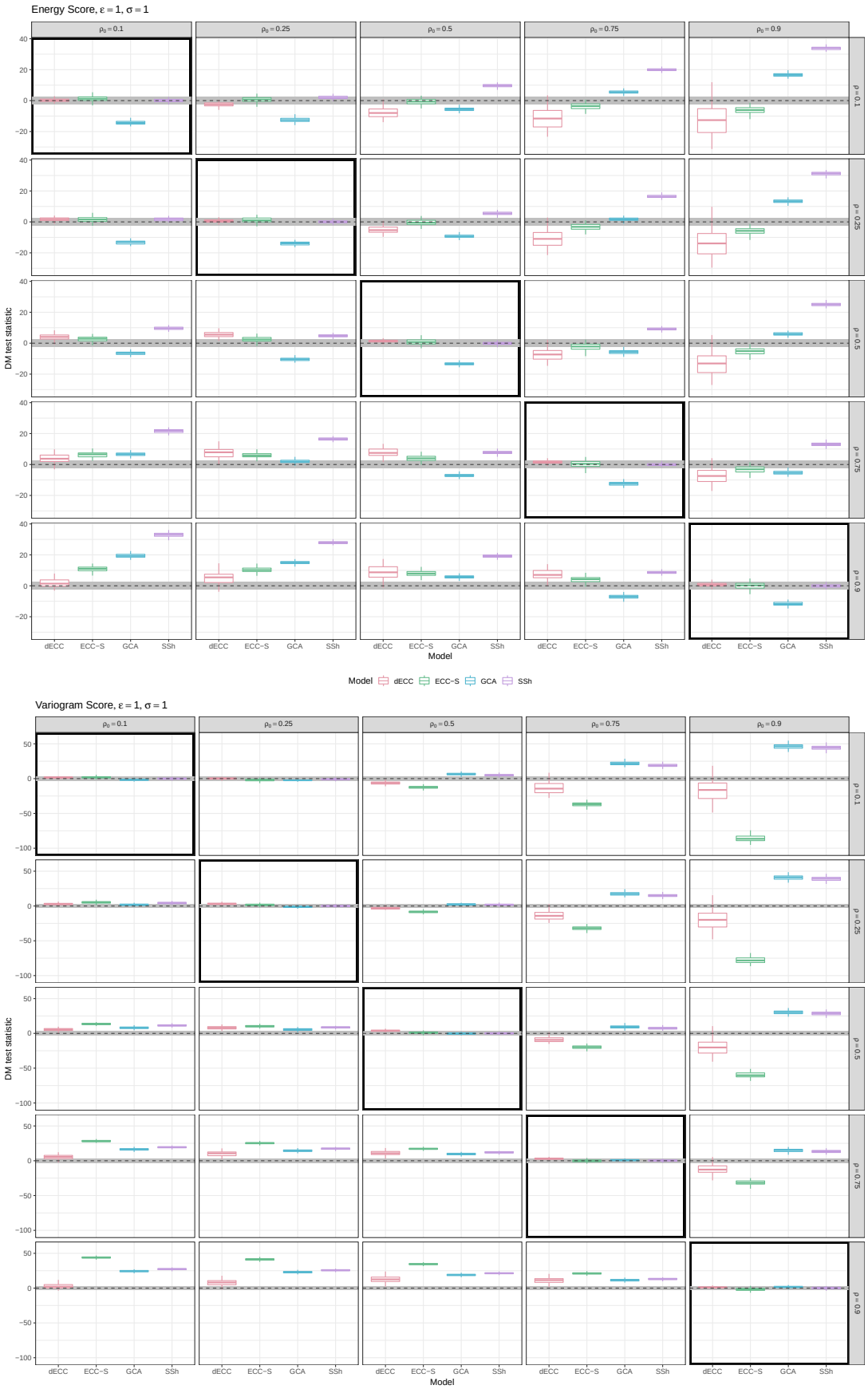
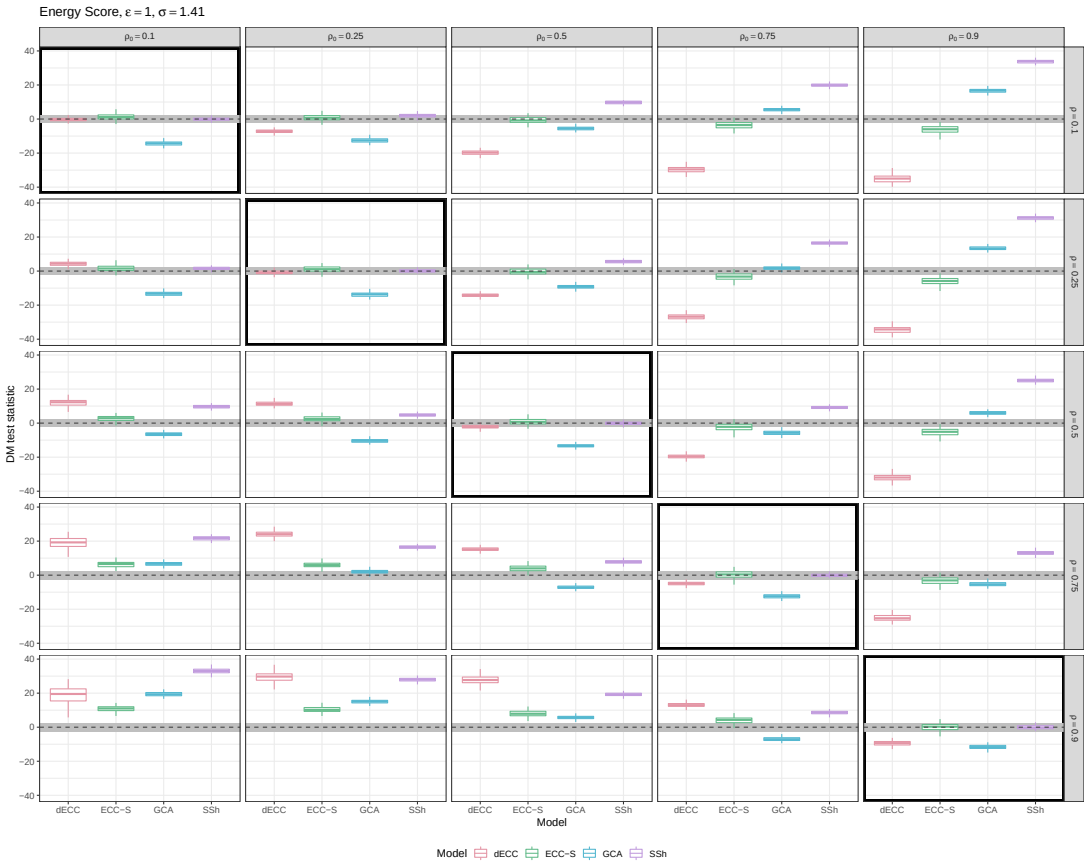


Figure 10: As Figure 5, but for $\epsilon = 1$, and $\sigma = 1$.



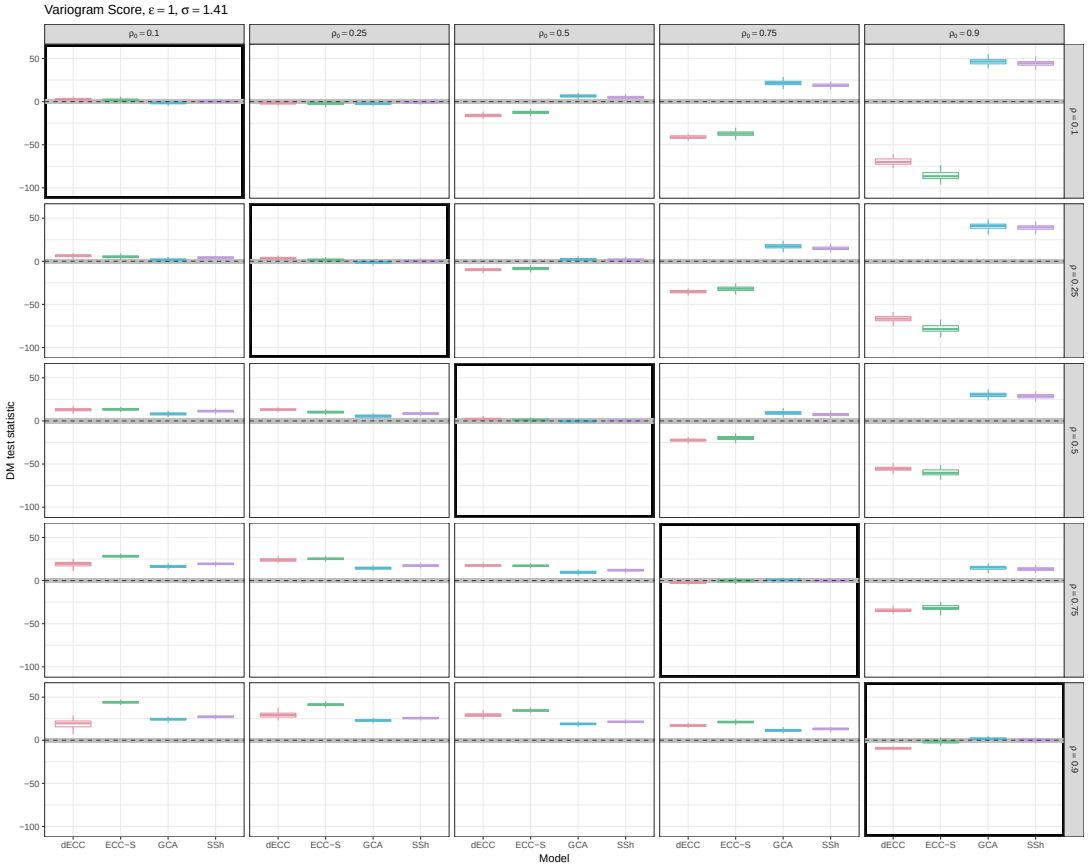
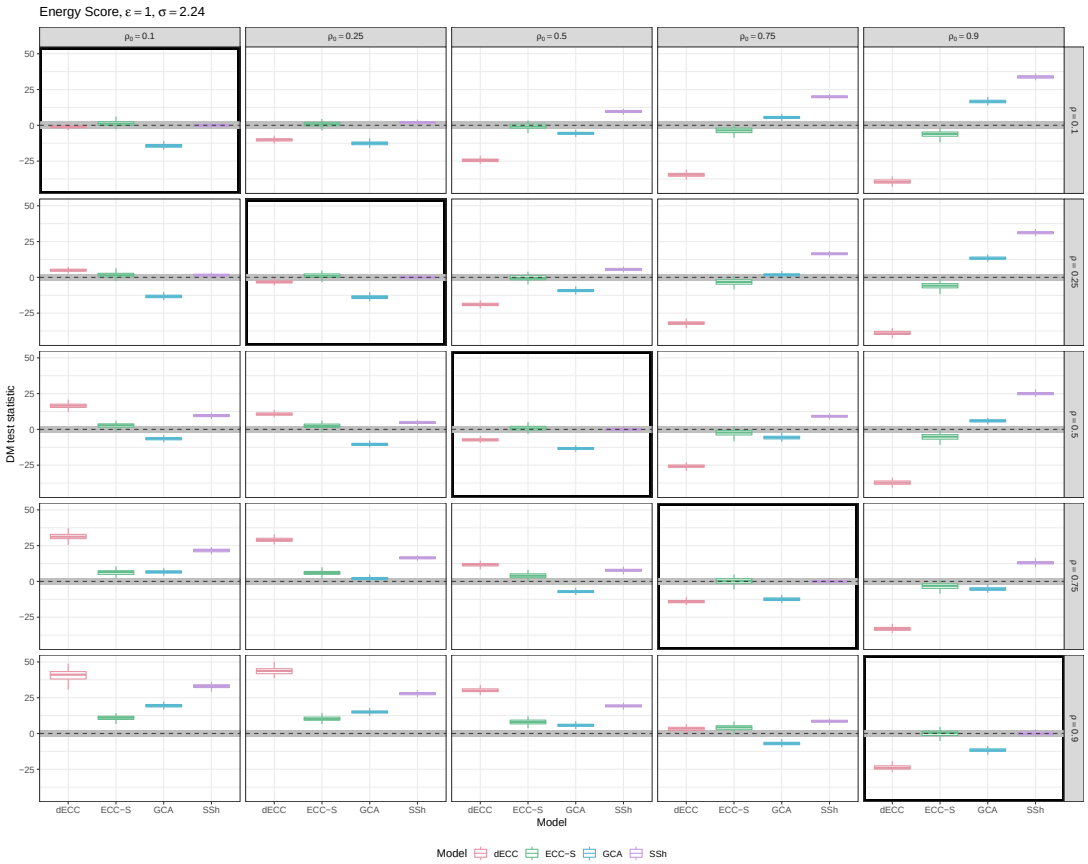


Figure 11: As Figure 5, but for $\epsilon = 1$, and $\sigma = \sqrt{2}$.



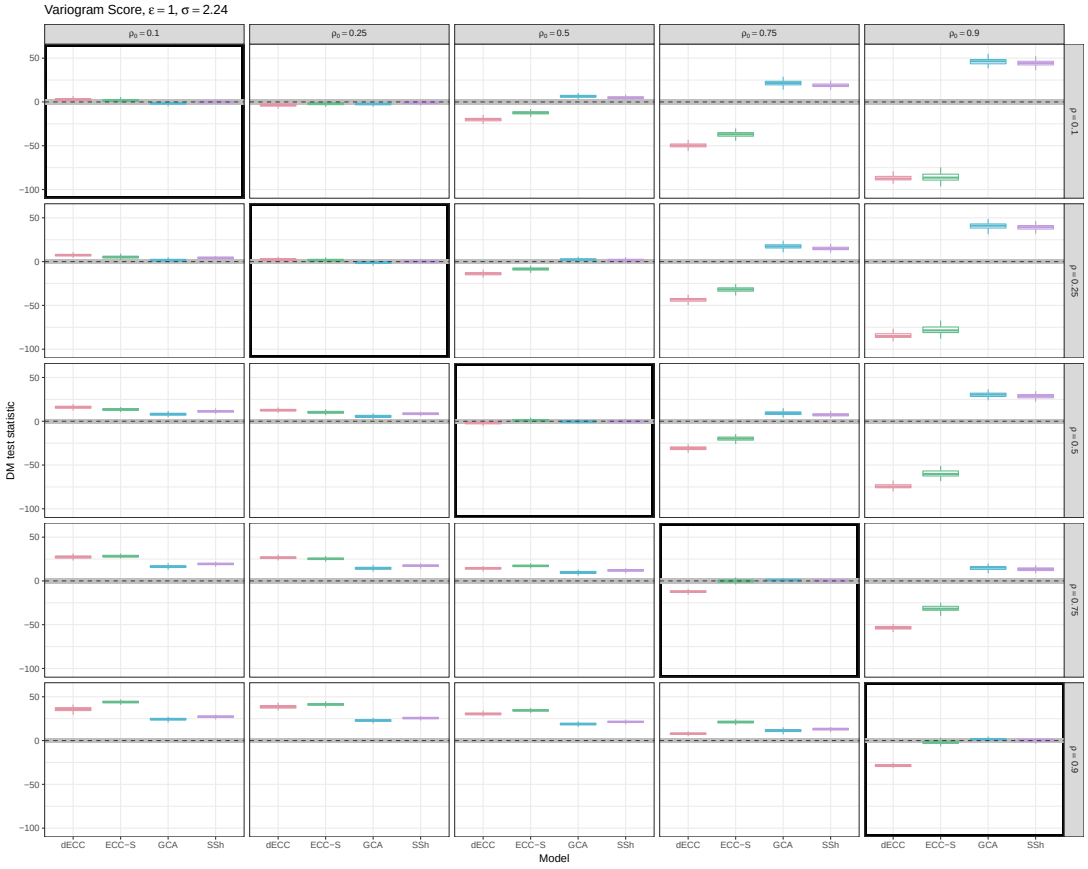


Figure 12: As Figure 5, but for $\epsilon = 1$, and $\sigma = \sqrt{5}$.

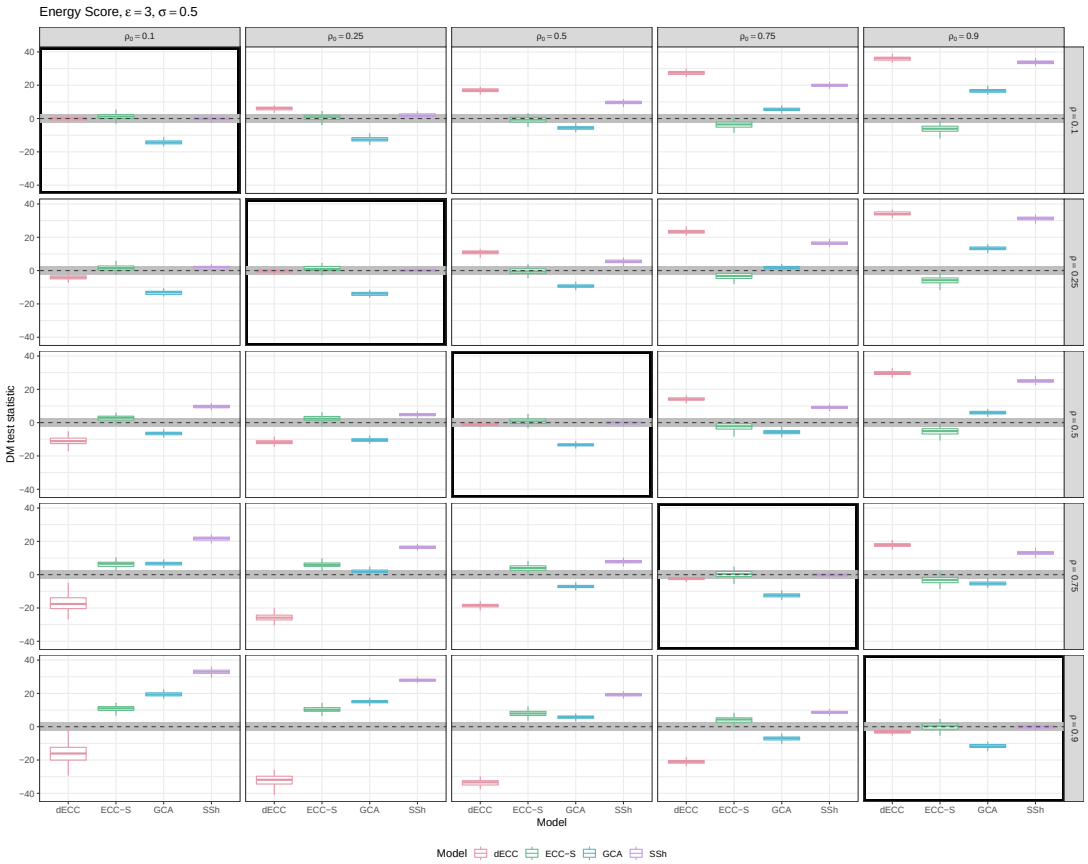
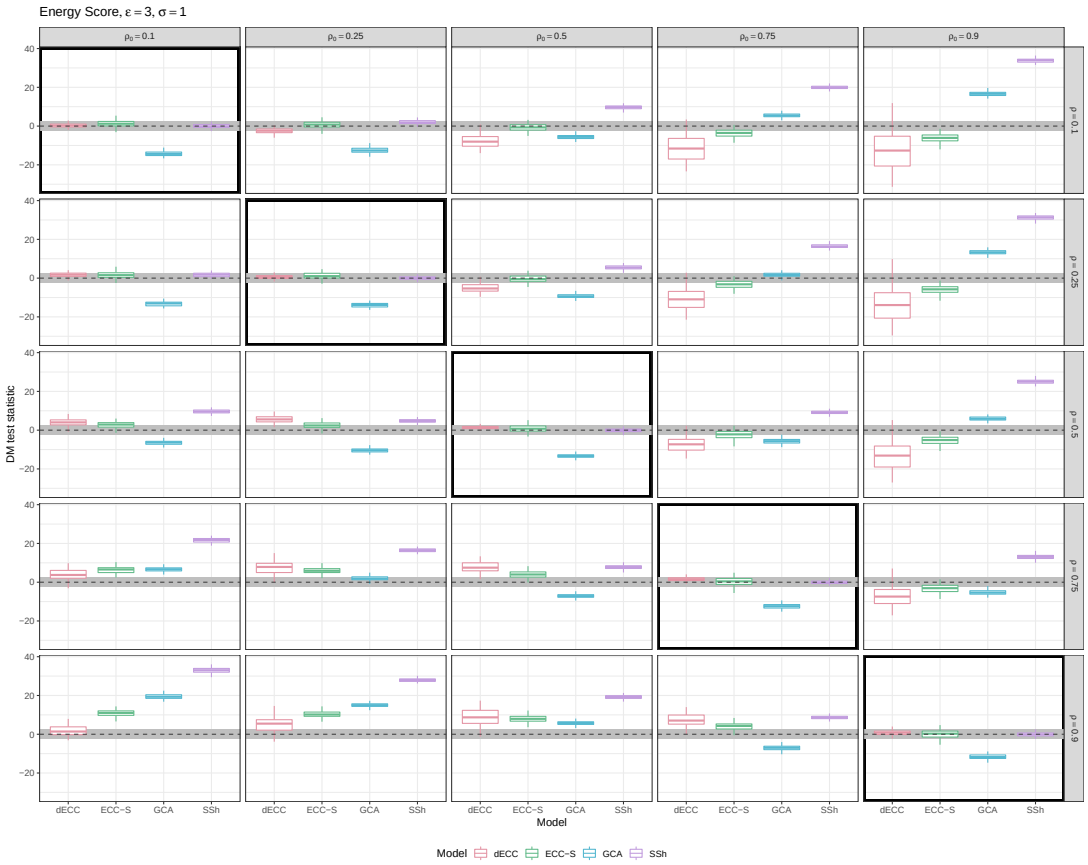


Figure 13: As Figure 5, but for $\epsilon = 3$, and $\sigma = 0.5$.



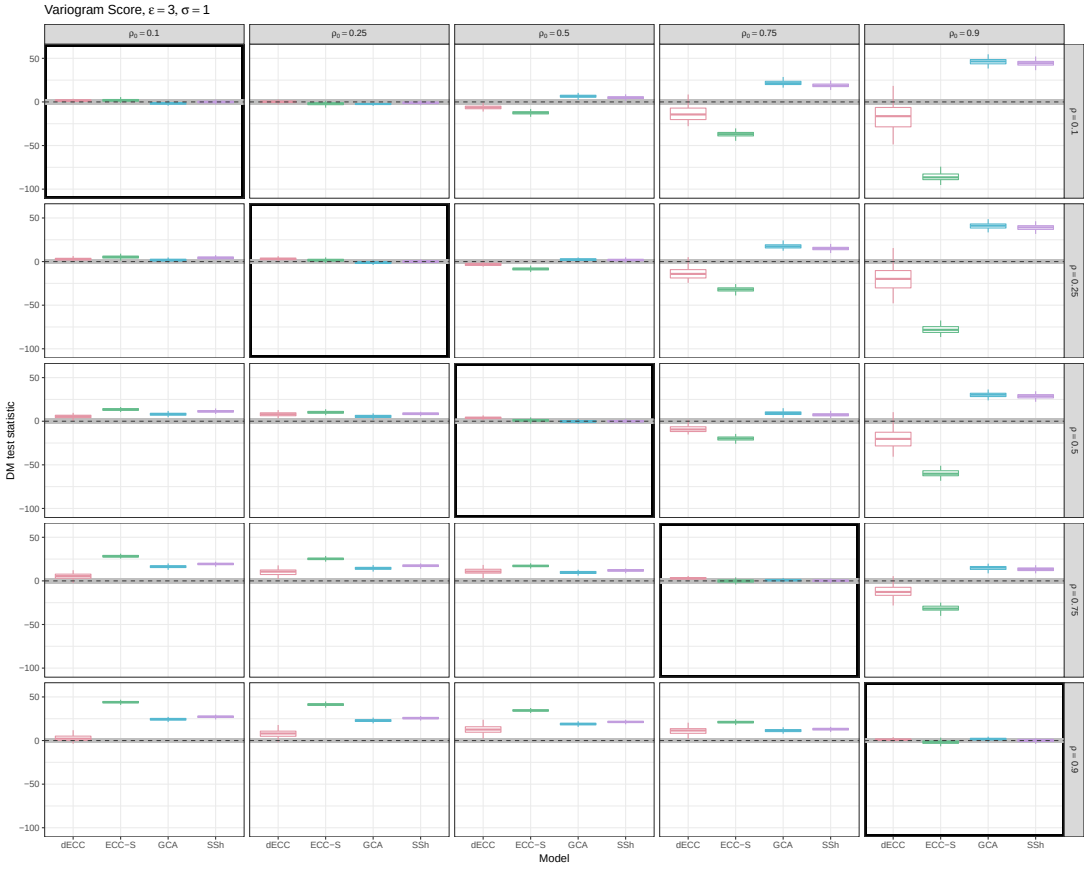
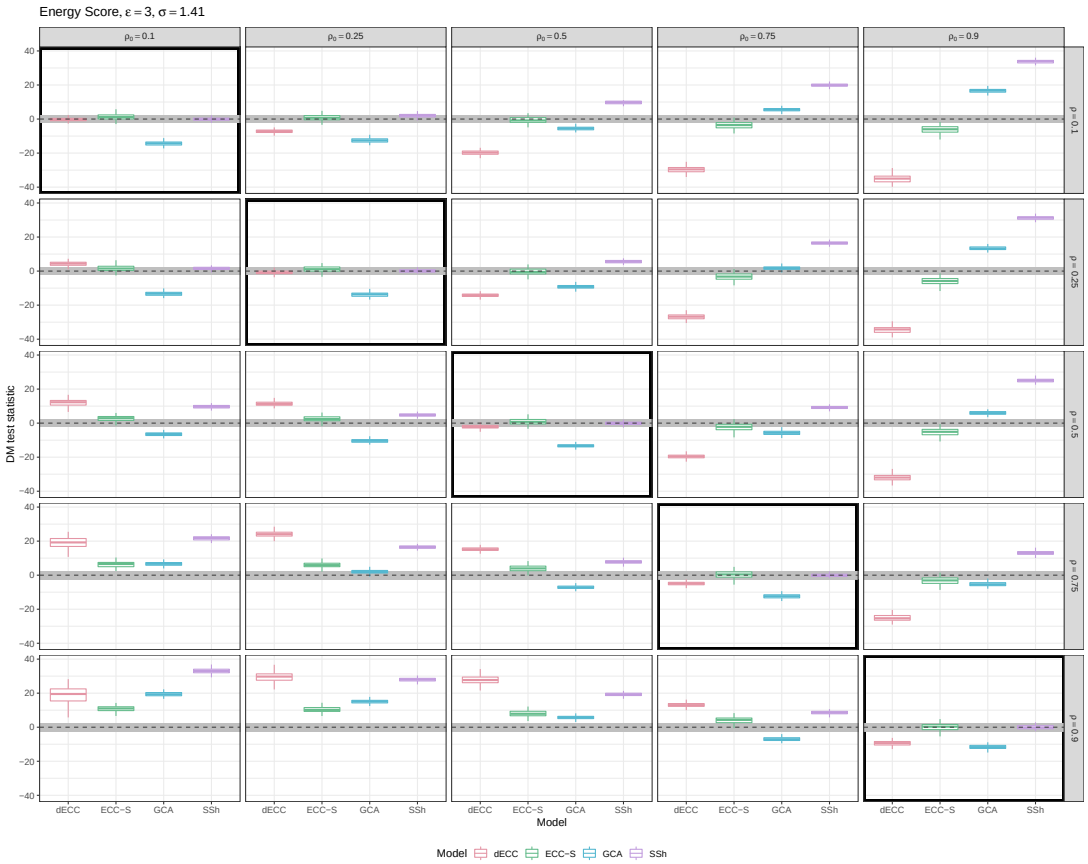


Figure 14: As Figure 5, but for $\epsilon = 3$, and $\sigma = 1$.



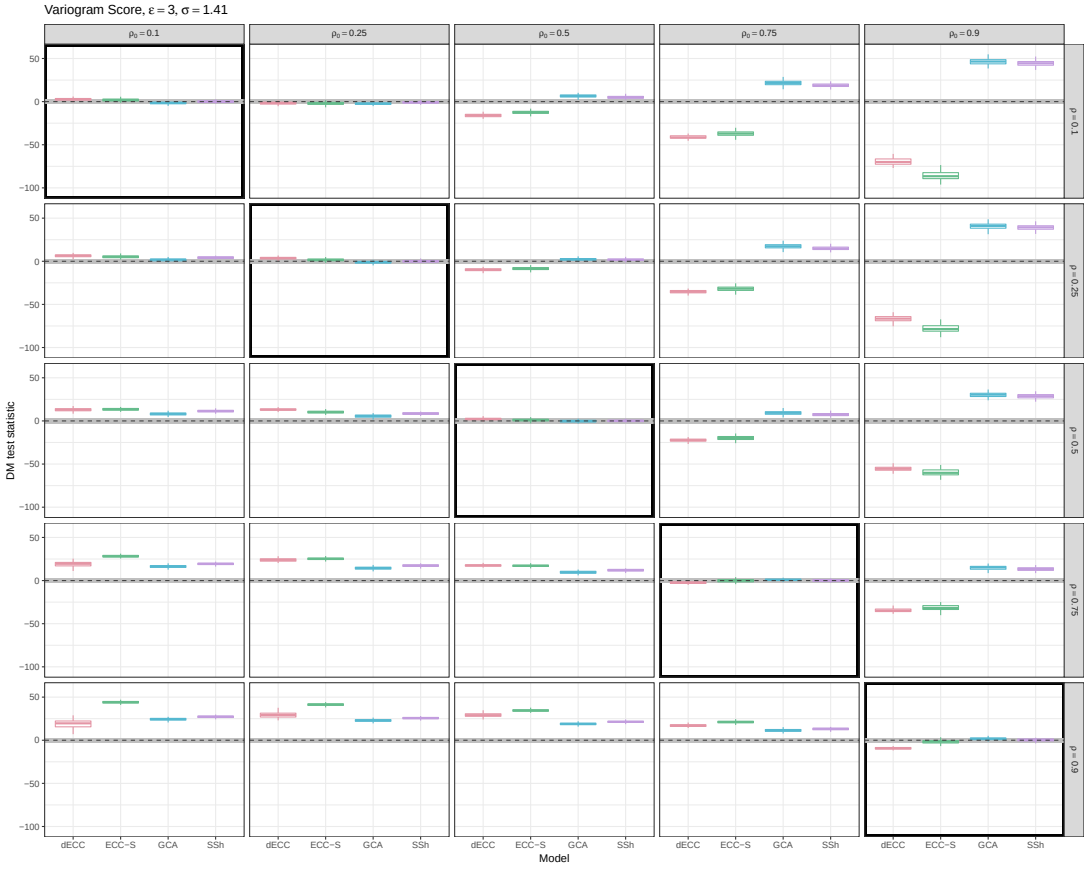
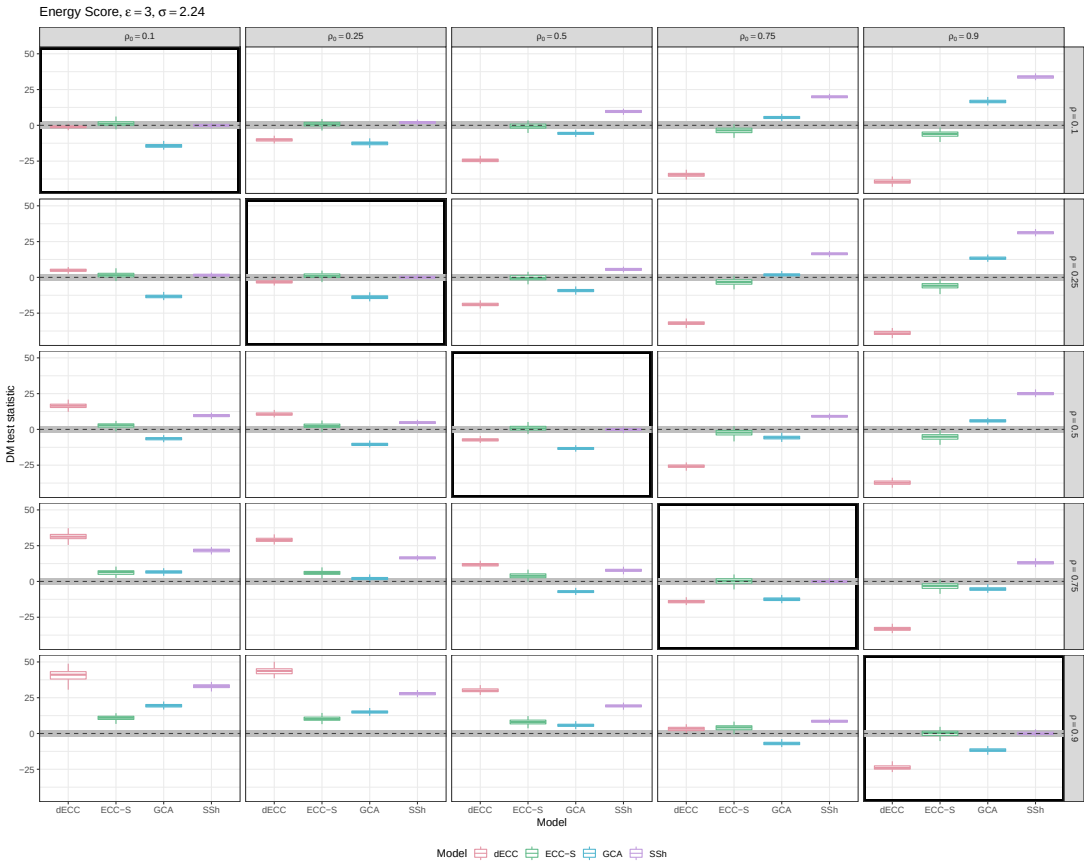


Figure 15: As Figure 5, but for $\epsilon = 3$, and $\sigma = \sqrt{2}$.



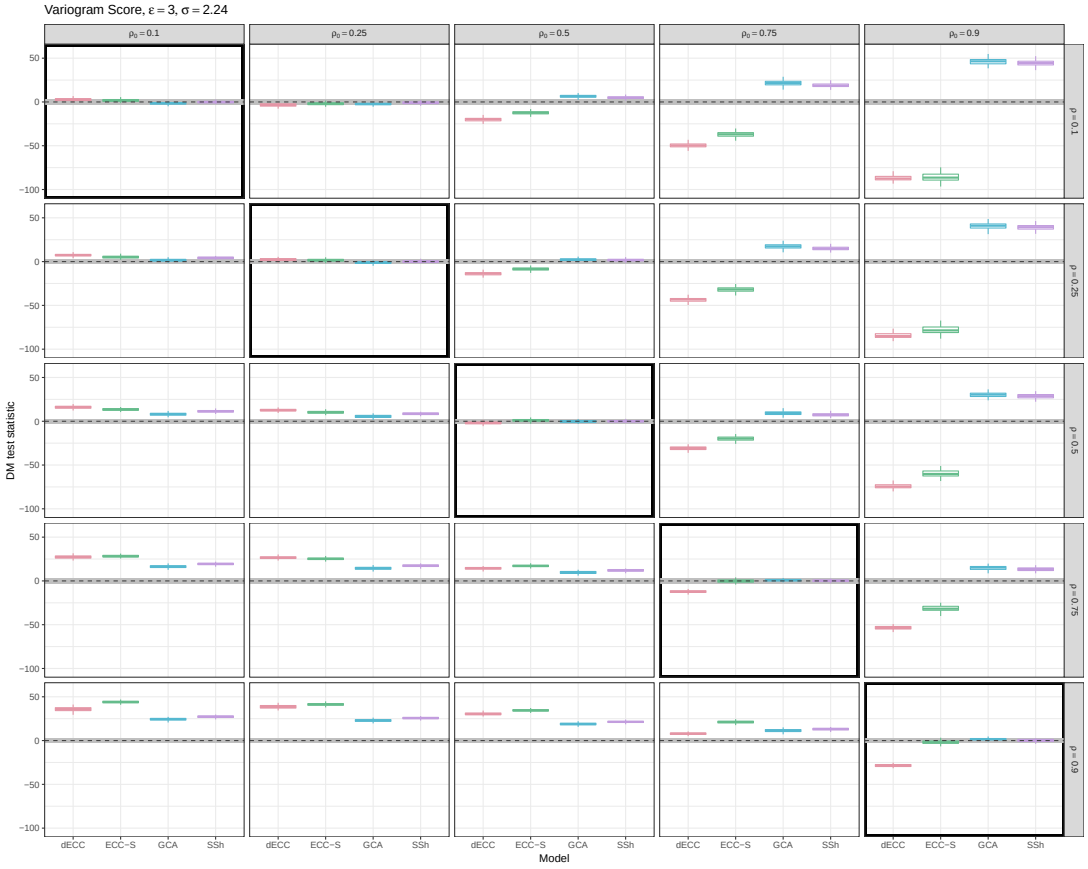


Figure 16: As Figure 5, but for $\epsilon = 3$, and $\sigma = \sqrt{5}$.

2 Additional results for Setting 2

2.1 Additional results for $m = 50$ ensemble members

Results for scenarios A, C, D from the following table.

	μ_0	ξ_0	σ_0	μ	ξ	σ
A	0.0	-0.1	1.0	1.0	0.0	0.2
B	0.0	-0.1	1.0	0.0	0.0	2.0
C	1.0	0.3	1.0	0.0	0.0	2.0
D	0.0	0.0	1.0	0.0	0.0	1.0

Table 2: Different simulation scenarios for Setting 2.

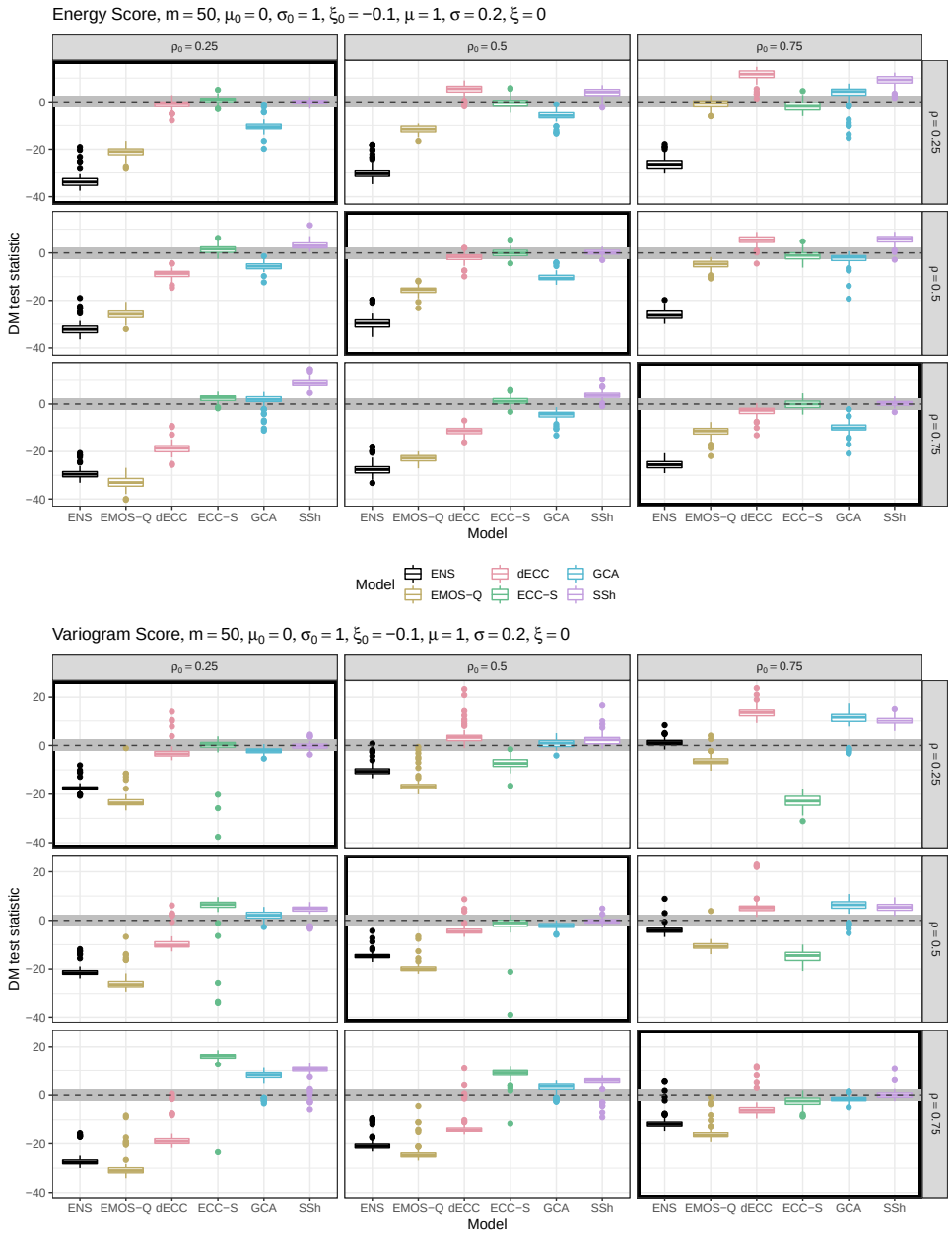


Figure 17: Summaries of DM test statistic values based on ES (top) and VS (bottom) for Setting 2, scenario A from Table 2 with $m = 50$. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

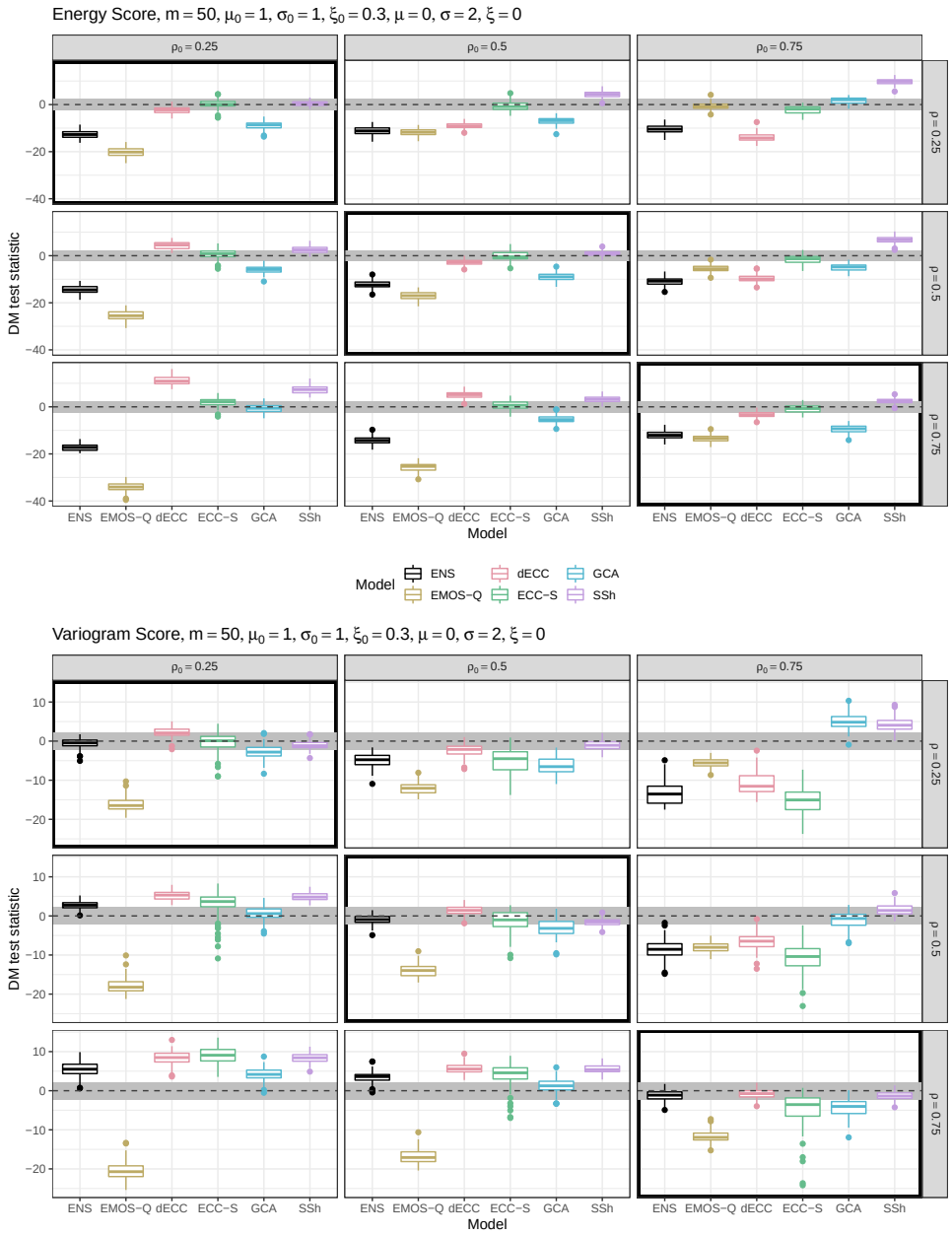


Figure 18: Summaries of DM test statistic values based on ES (top) and VS (bottom) for Setting 2, scenario C from Table 2 with $m = 50$. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

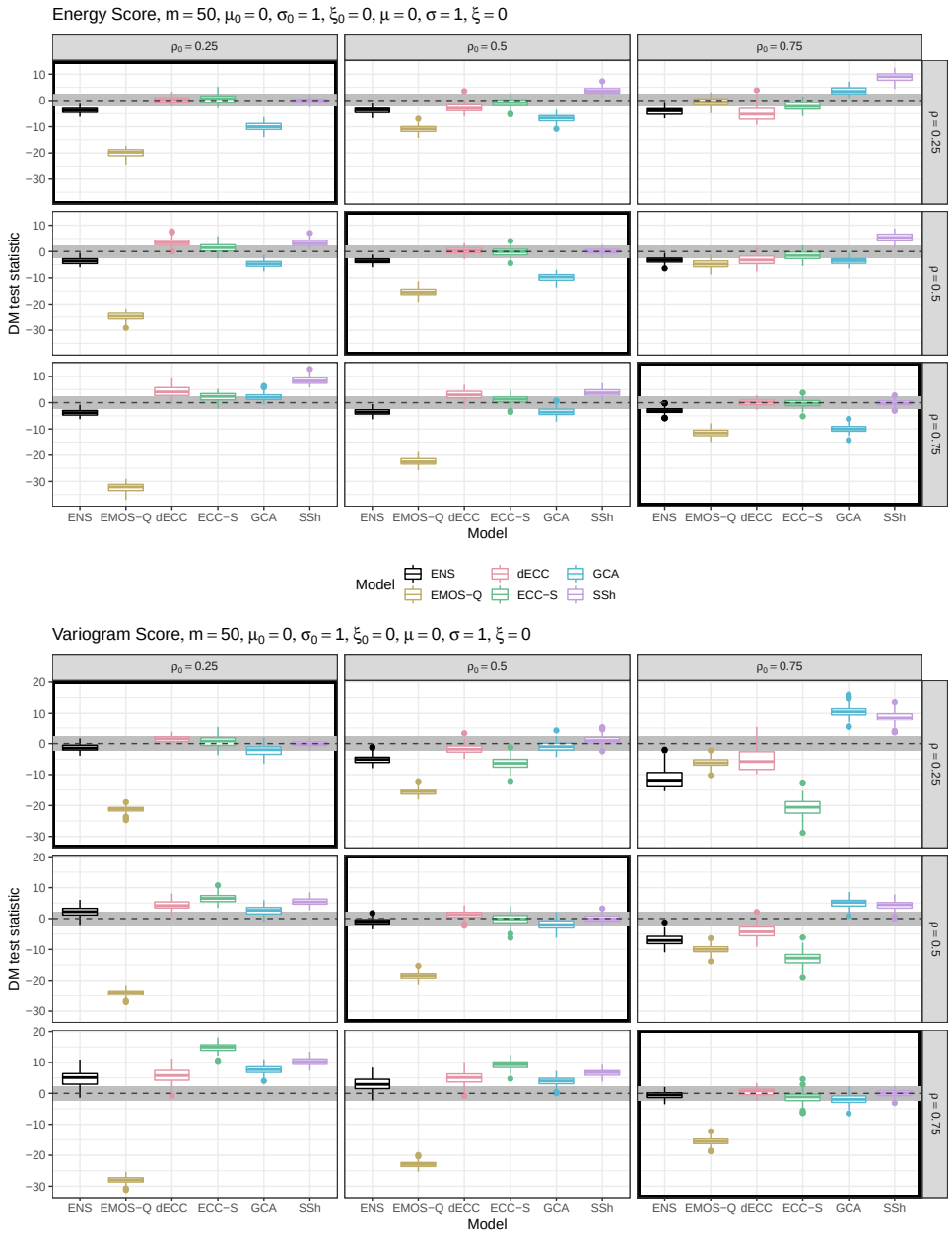


Figure 19: Summaries of DM test statistic values based on ES (top) and VS (bottom) for Setting 2, scenario D from Table 2 with $m = 50$. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

2.2 Results for different numbers of ensemble members

Figures 20 – 23 show results in terms of the ES for Scenarios A, B, C, and D of Setting 2 defined in Table 2 and discussed in Section 4.2.2 of the paper. The general structure of the Figures is the same as for those in Supplement Section 2.1 and Section 4.2.2 of the paper, however, every panel shows grouped boxplots for each methods comparing ensemble sizes of 5, 20, 50 and 100. Similarly, Figures 24 – 27 show member comparative results in terms of the VS.

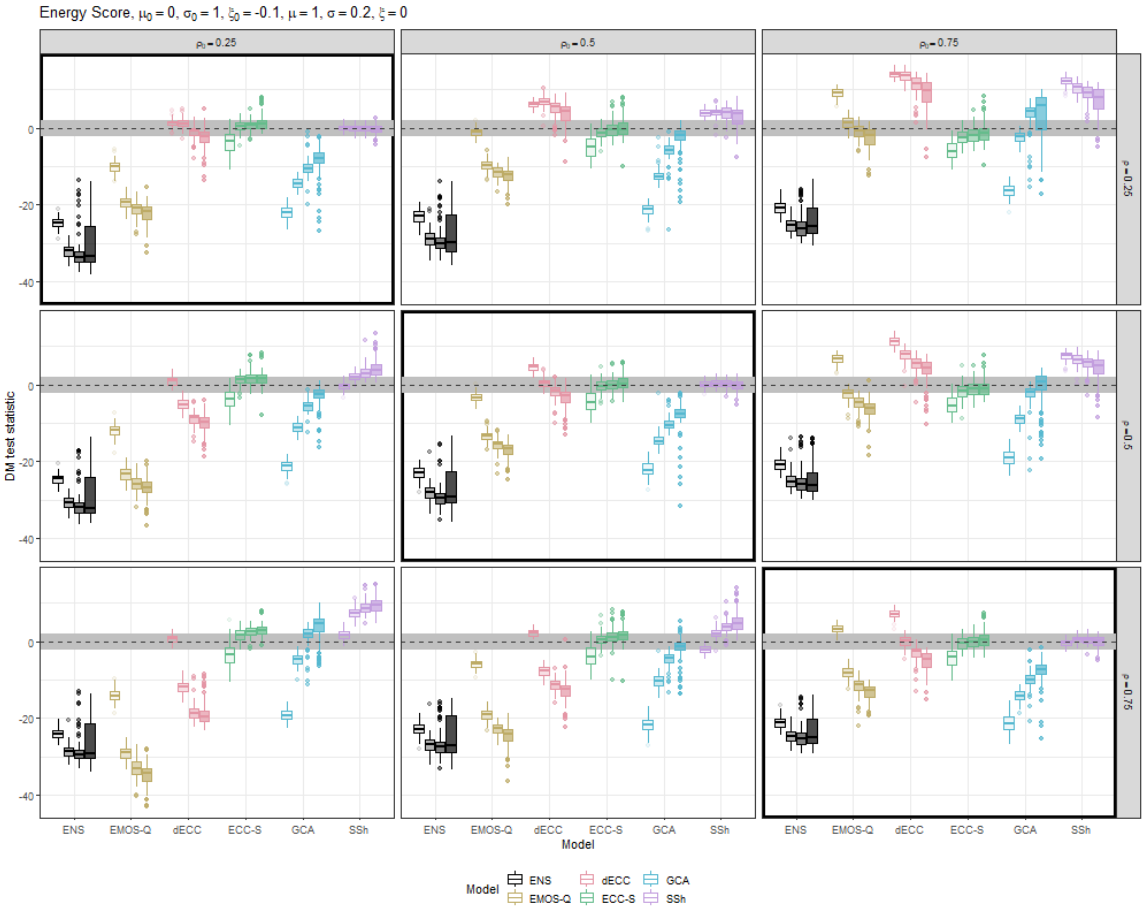


Figure 20: Summaries of DM test statistic values based on ES for Setting 2, scenario A from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

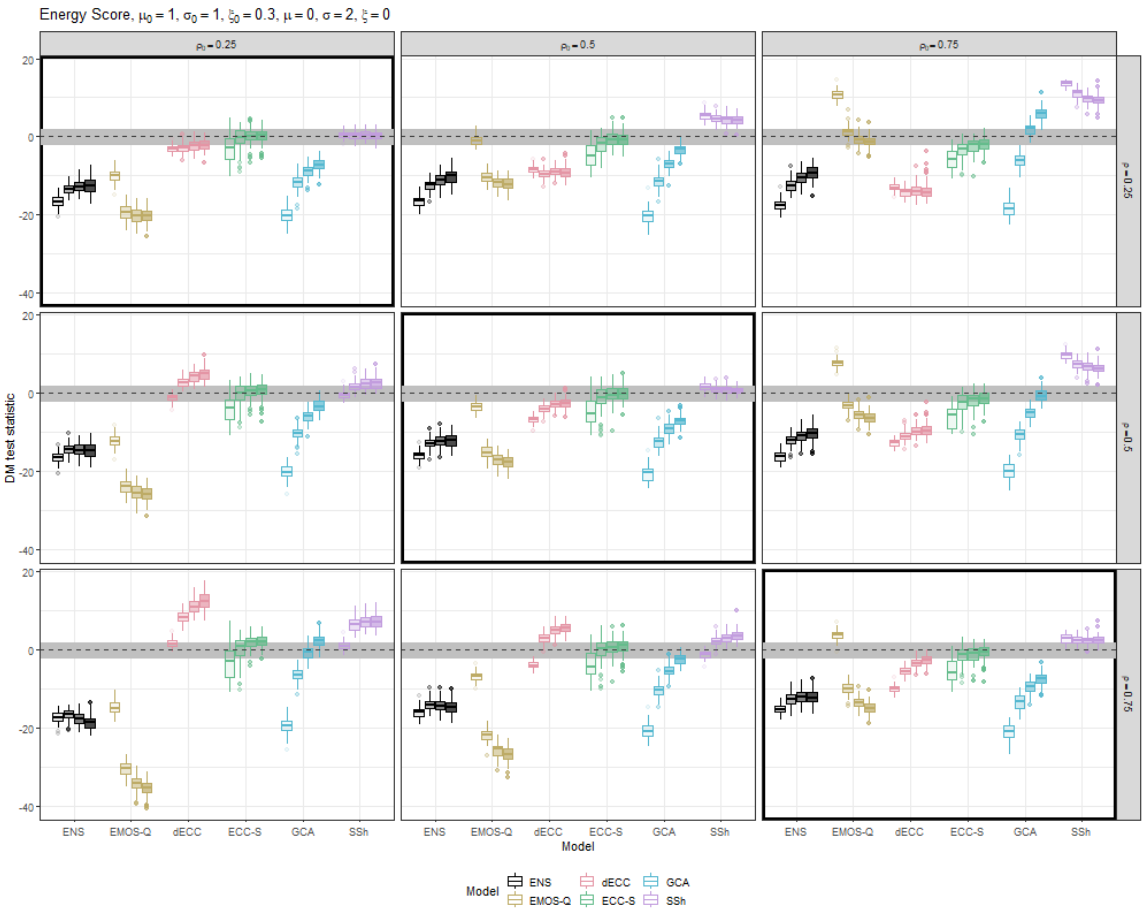


Figure 21: Summaries of DM test statistic values based on ES for Setting 2, scenario B from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

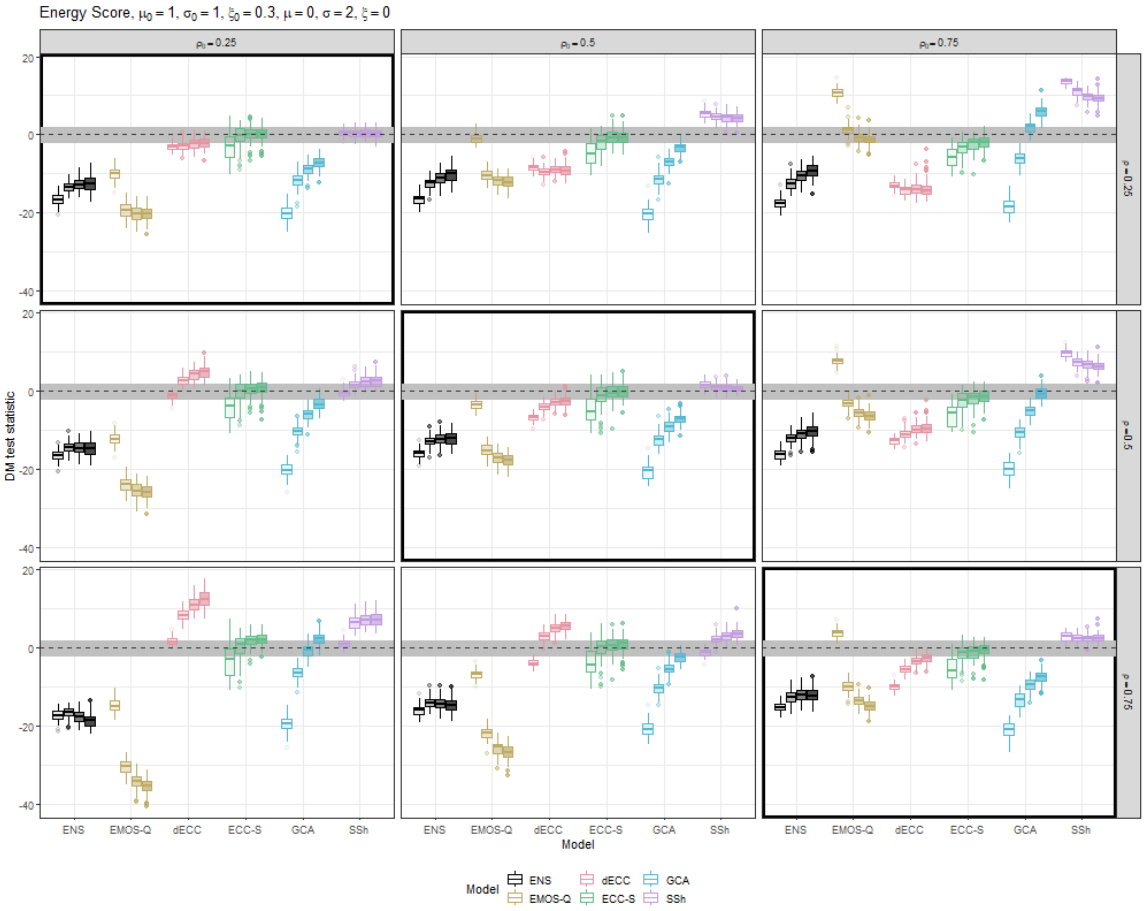


Figure 22: Summaries of DM test statistic values based on ES for Setting 2, scenario C from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

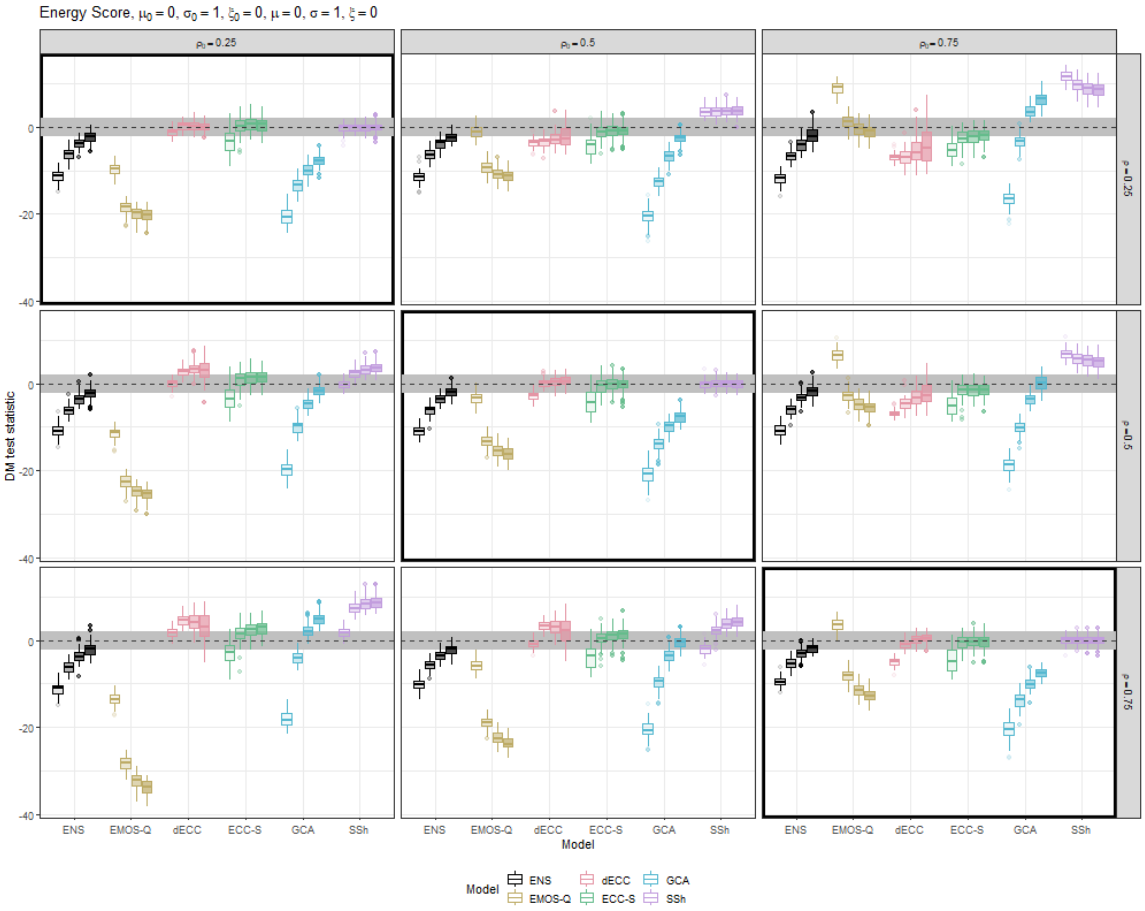


Figure 23: Summaries of DM test statistic values based on ES for Setting 2, scenario D from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

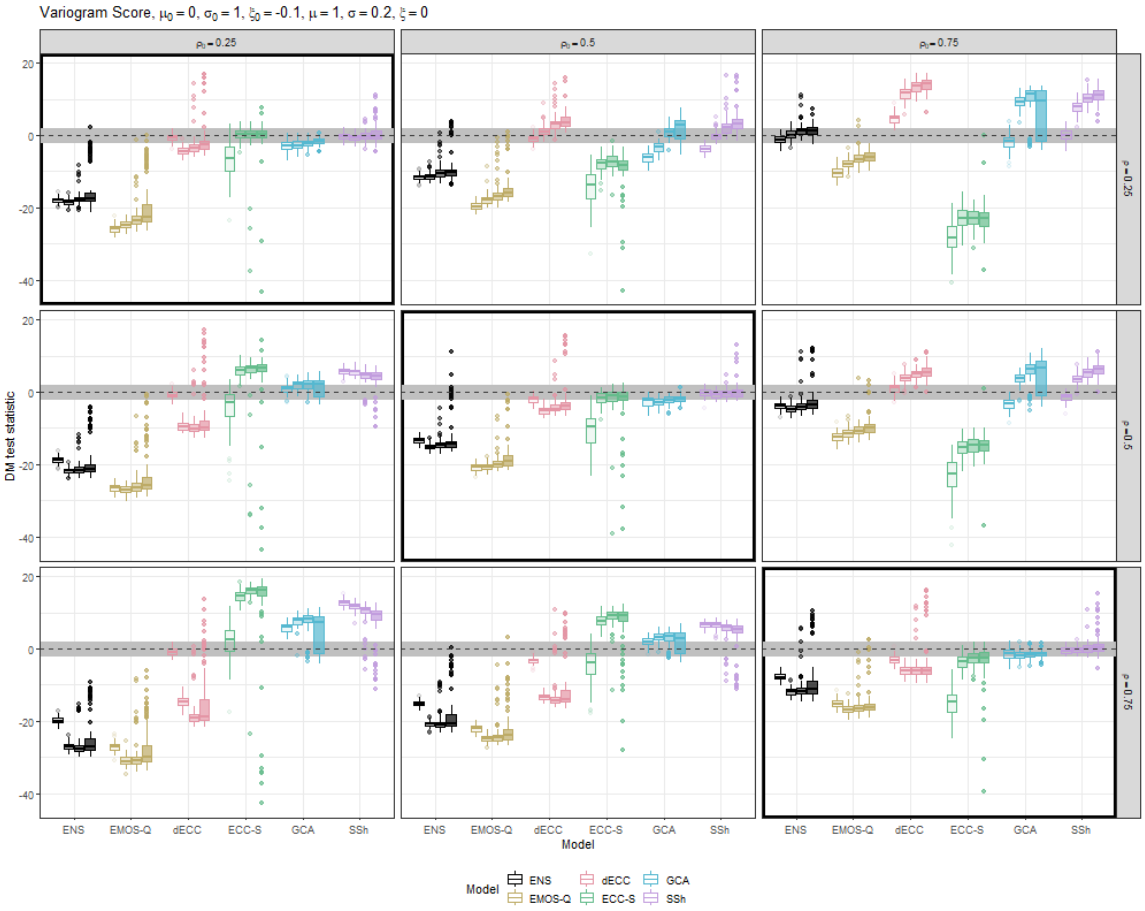


Figure 24: Summaries of DM test statistic values based on VS for Setting 2, scenario A from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

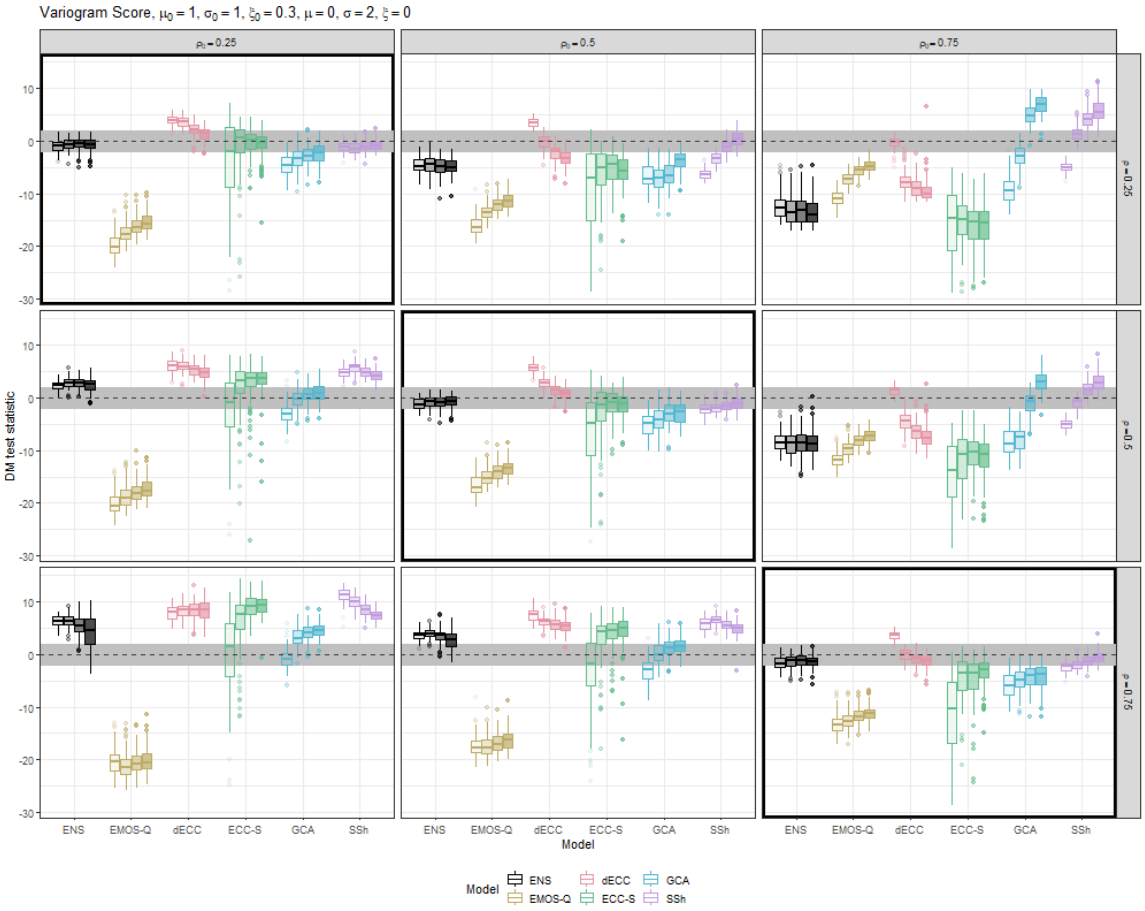


Figure 25: Summaries of DM test statistic values based on VS for Setting 2, scenario B from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

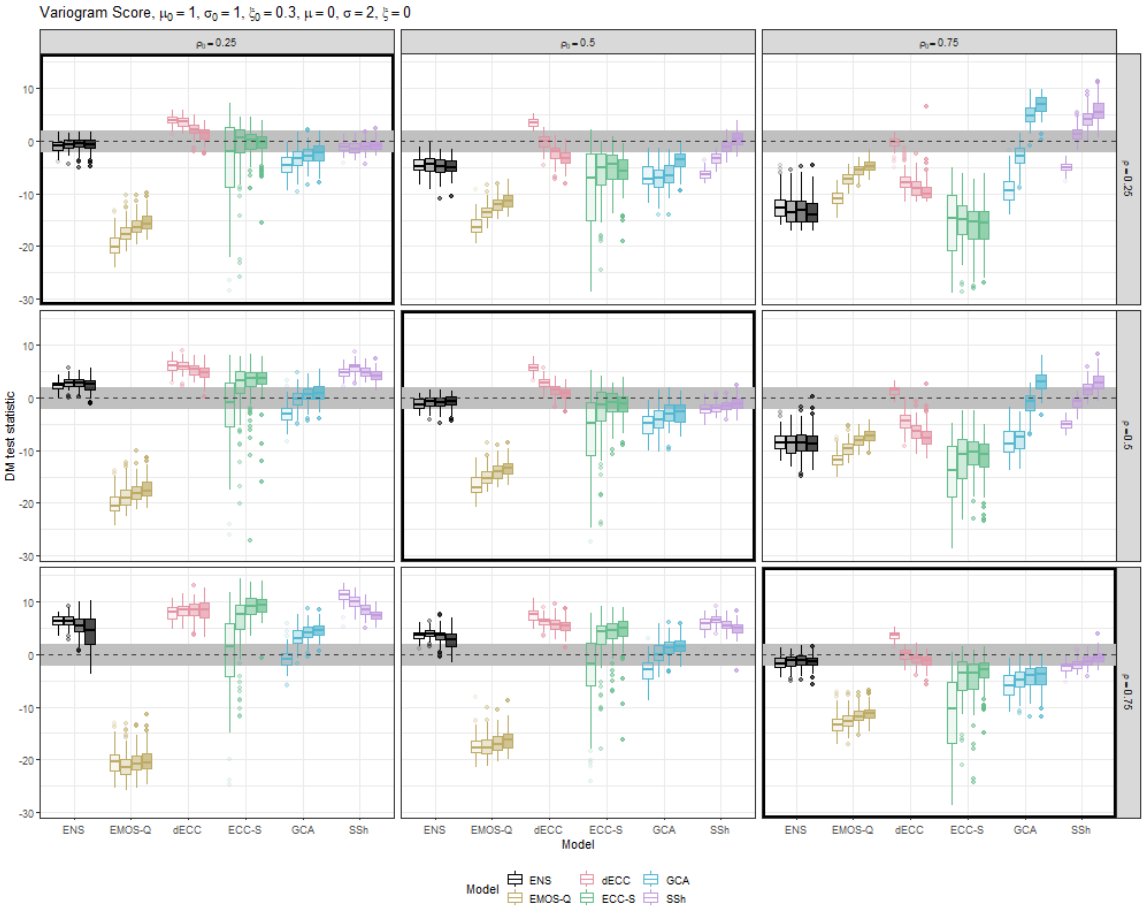


Figure 26: Summaries of DM test statistic values based on VS for Setting 2, scenario C from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

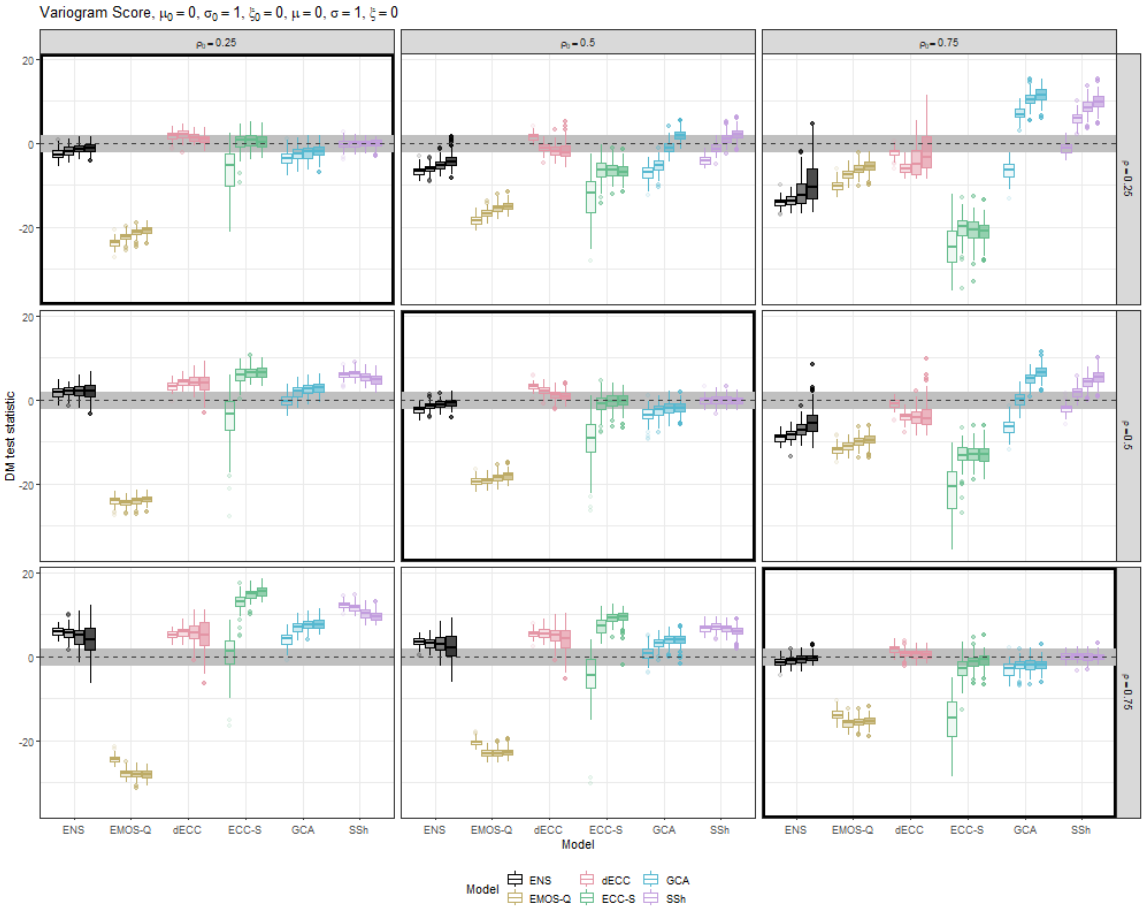


Figure 27: Summaries of DM test statistic values based on VS for Setting 2, scenario D from Table 2, with grouped boxplots of ensemble sizes $m = 5, 20, 50, 100$ (5 corresponds to lightest shade, 100 to darkest shade) for each model. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

3 Additional results for Setting 3A

Results for all values of ρ and ρ_0 are shown in Figure 28.

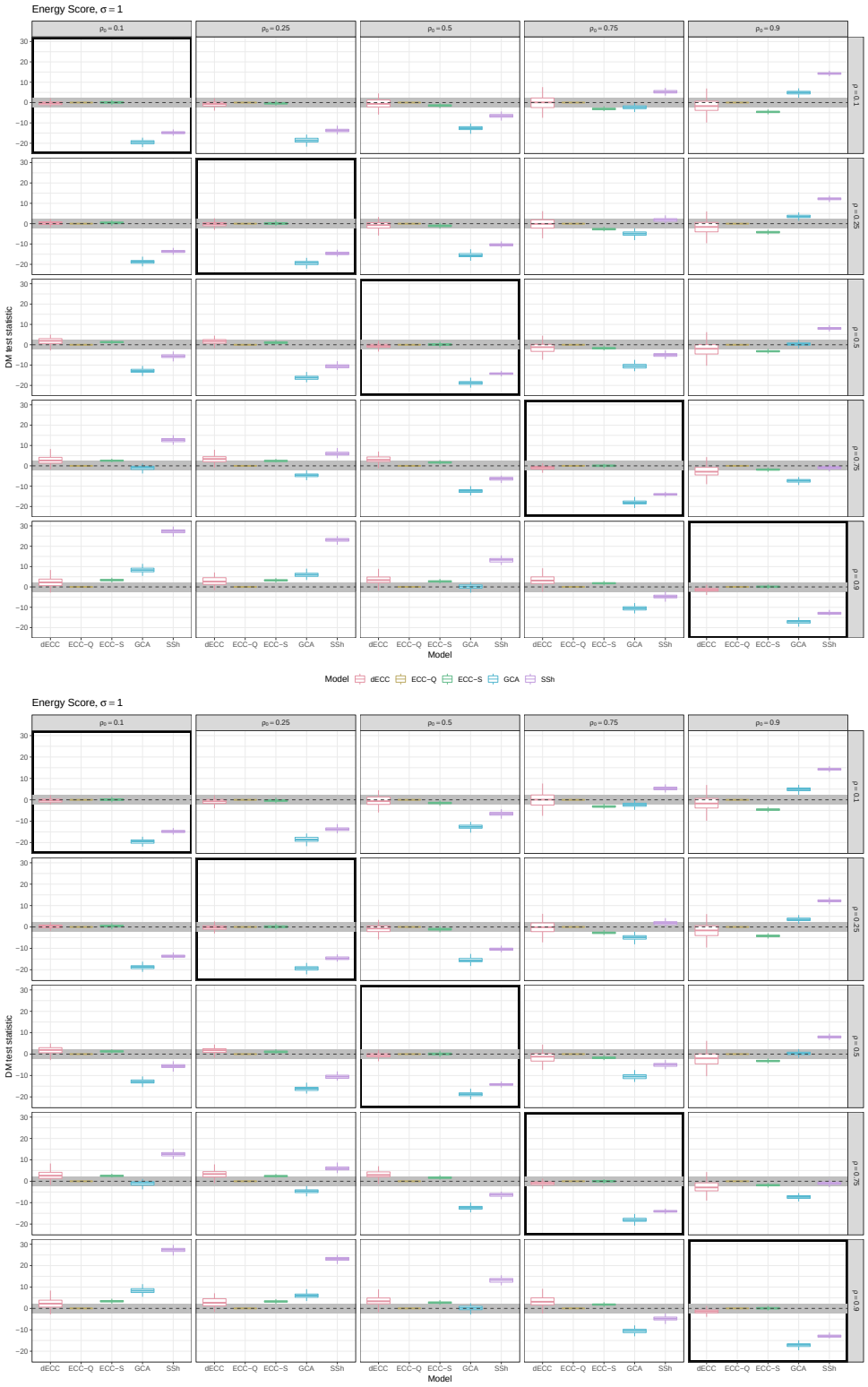


Figure 28: Summaries of DM test statistic values based on ES (top) and VS (bottom) for Setting 4. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

4 Additional results for Setting 3B

An overview the additional results for Setting 3B is provided in Table 3.

Table 3: Overview of parameter combinations for Setting 3B with references to corresponding figures in this supplement.

σ	correlations	Figure
0.5	low	29
1	low	30
5	low	31
0.5	medium	32
1	medium	33
5	medium	34
0.5	high	35
1	high	36
5	high	37

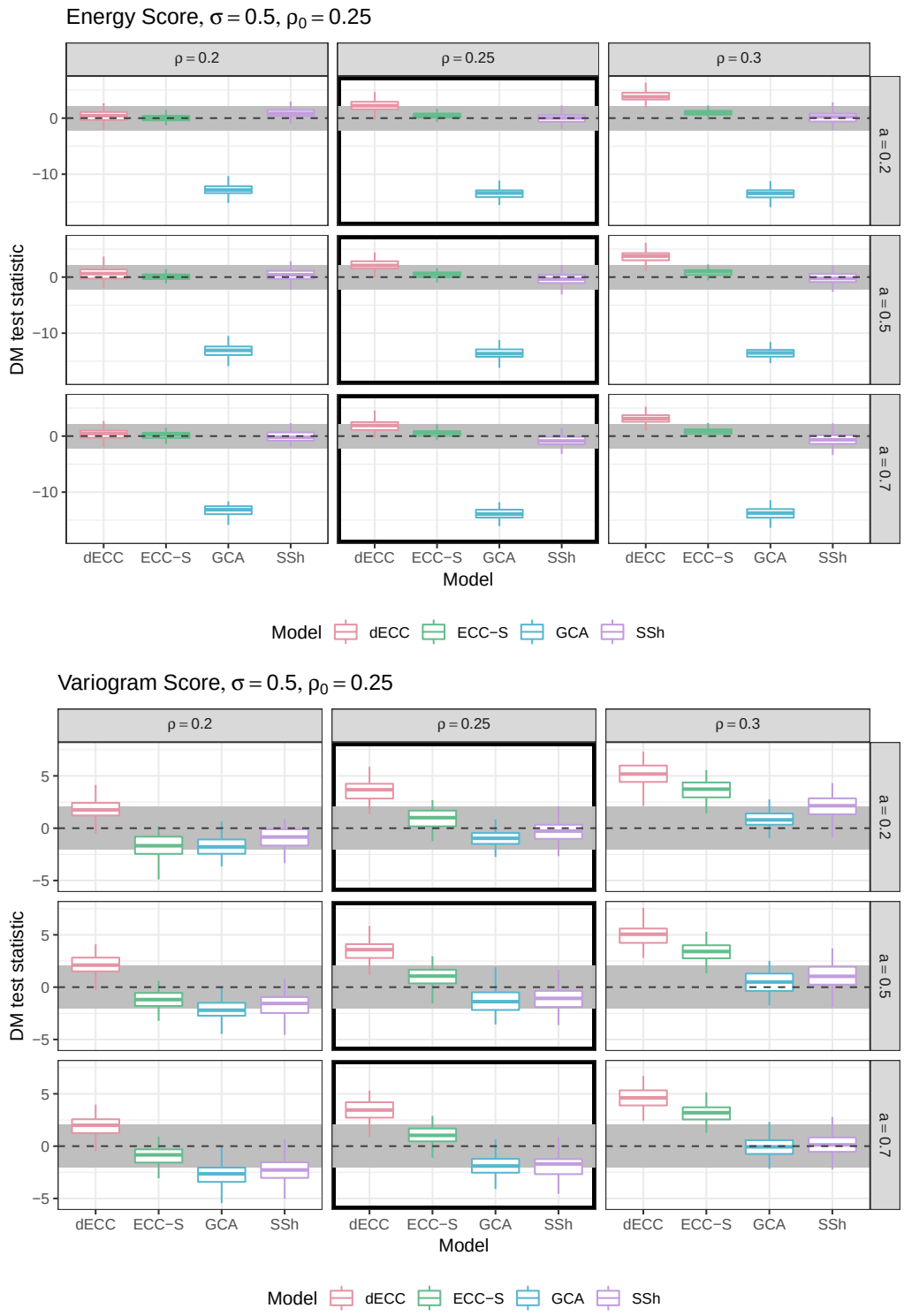


Figure 29: Summaries of DM test statistic values based on ES (top) and VS (bottom) for Setting 3B with $\sigma = 0.5$ and low correlation values. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

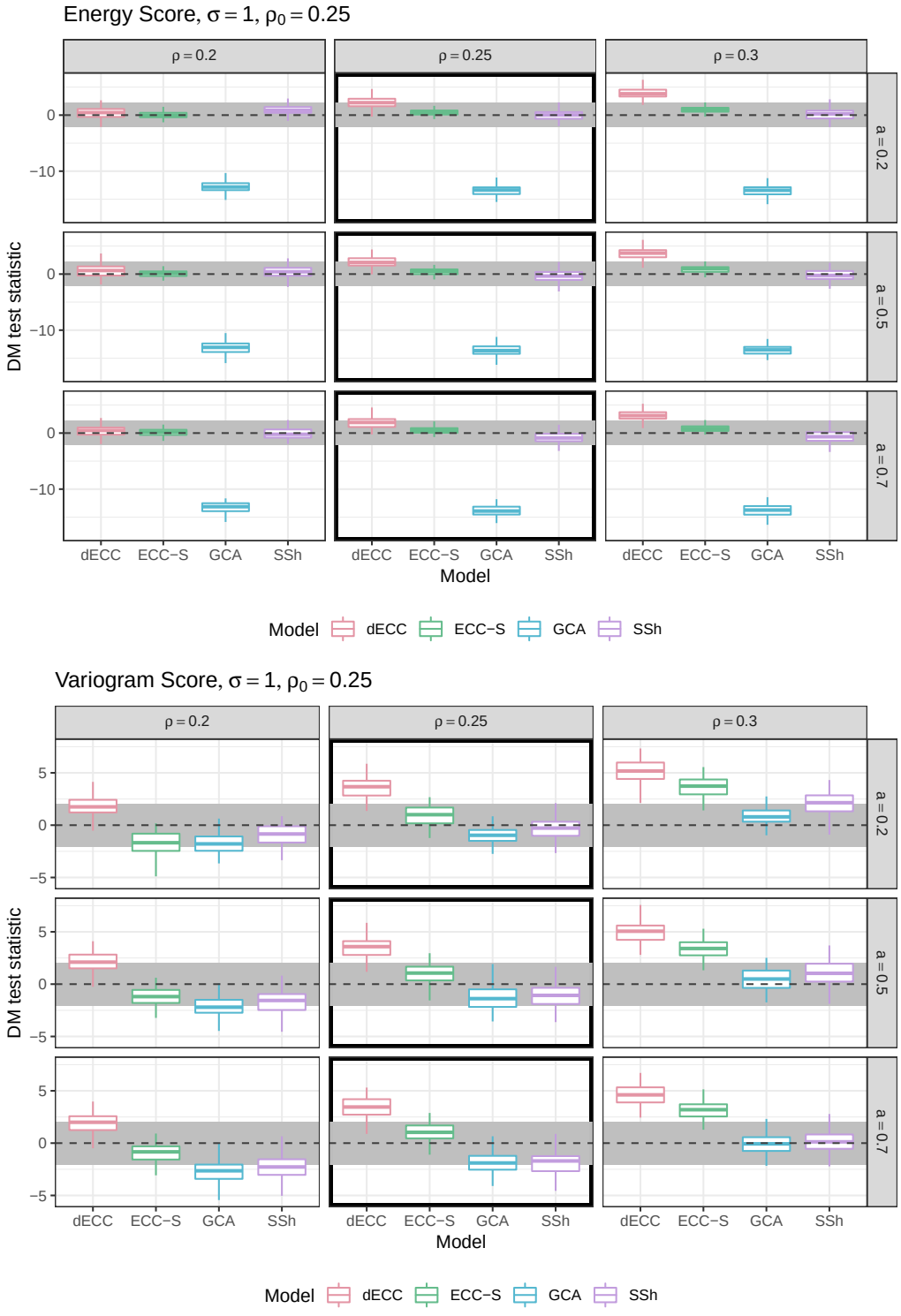


Figure 30: As Figure 29, but for $\sigma = 1$ and low correlation values.

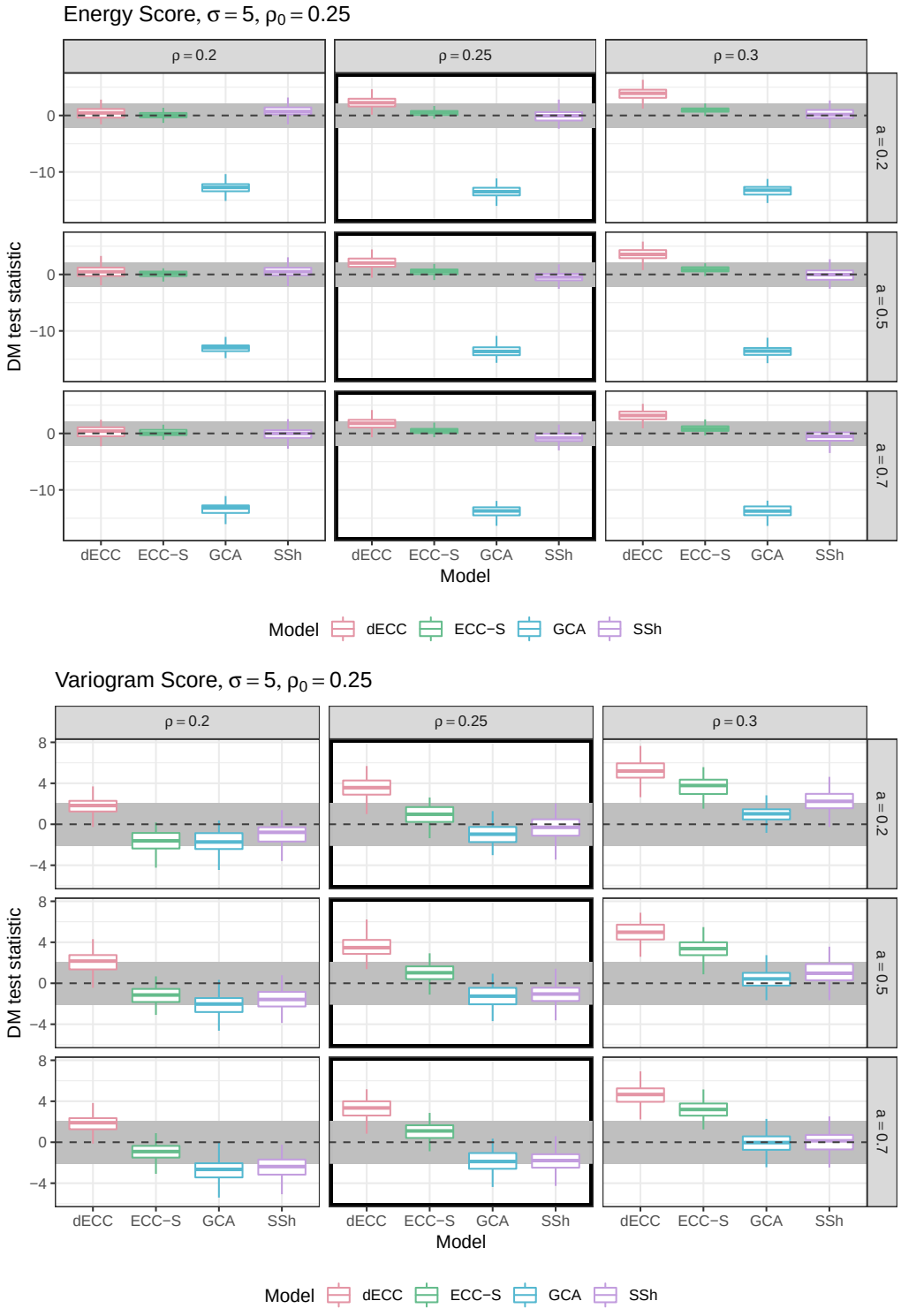


Figure 31: As Figure 29, but for $\sigma = 5$ and low correlation values.

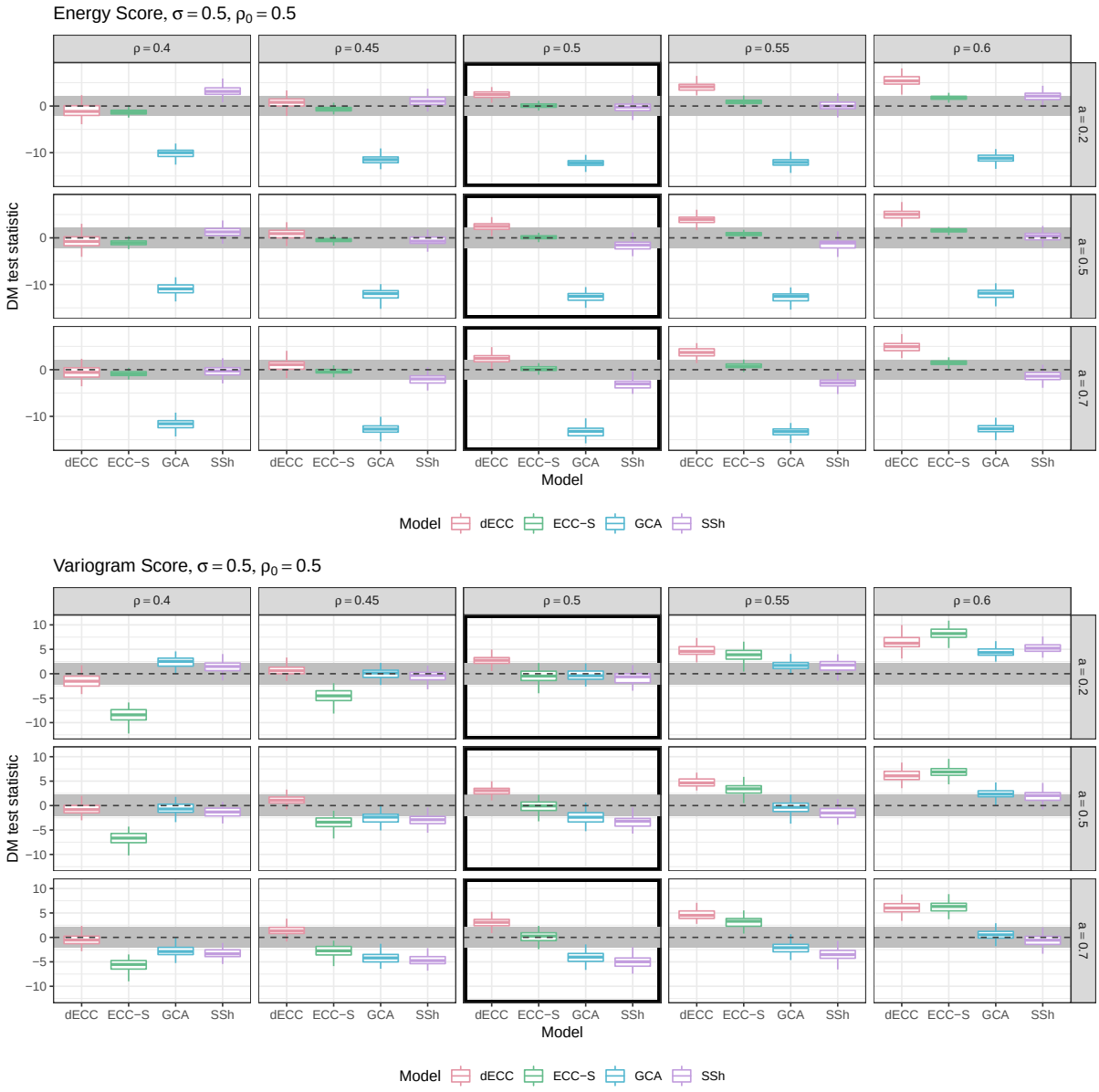


Figure 32: As Figure 29, but for $\sigma = 0.5$ and medium correlation values.

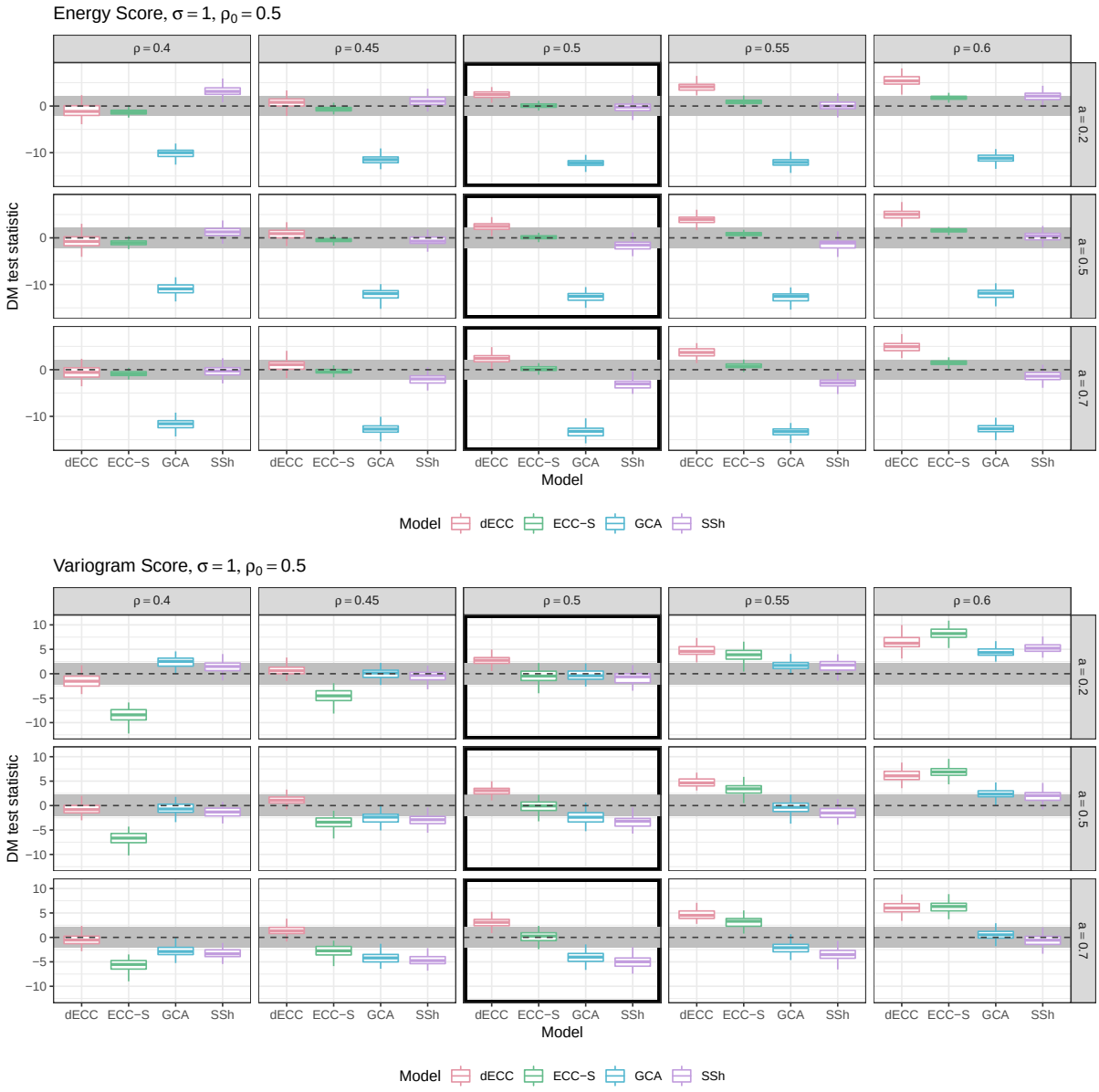


Figure 33: As Figure 29, but for $\sigma = 1$ and medium correlation values.

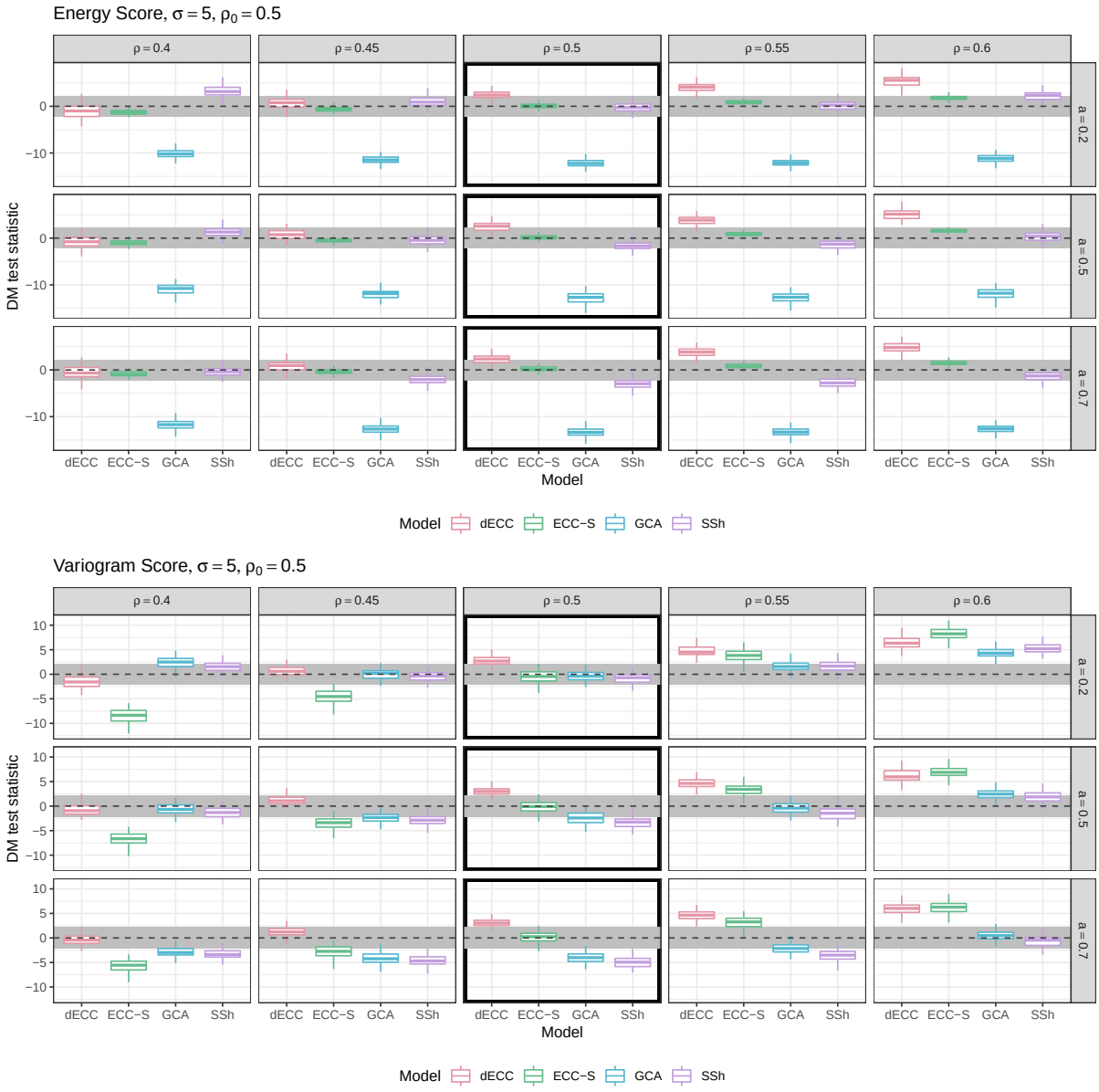


Figure 34: As Figure 29, but for $\sigma = 5$ and medium correlation values.

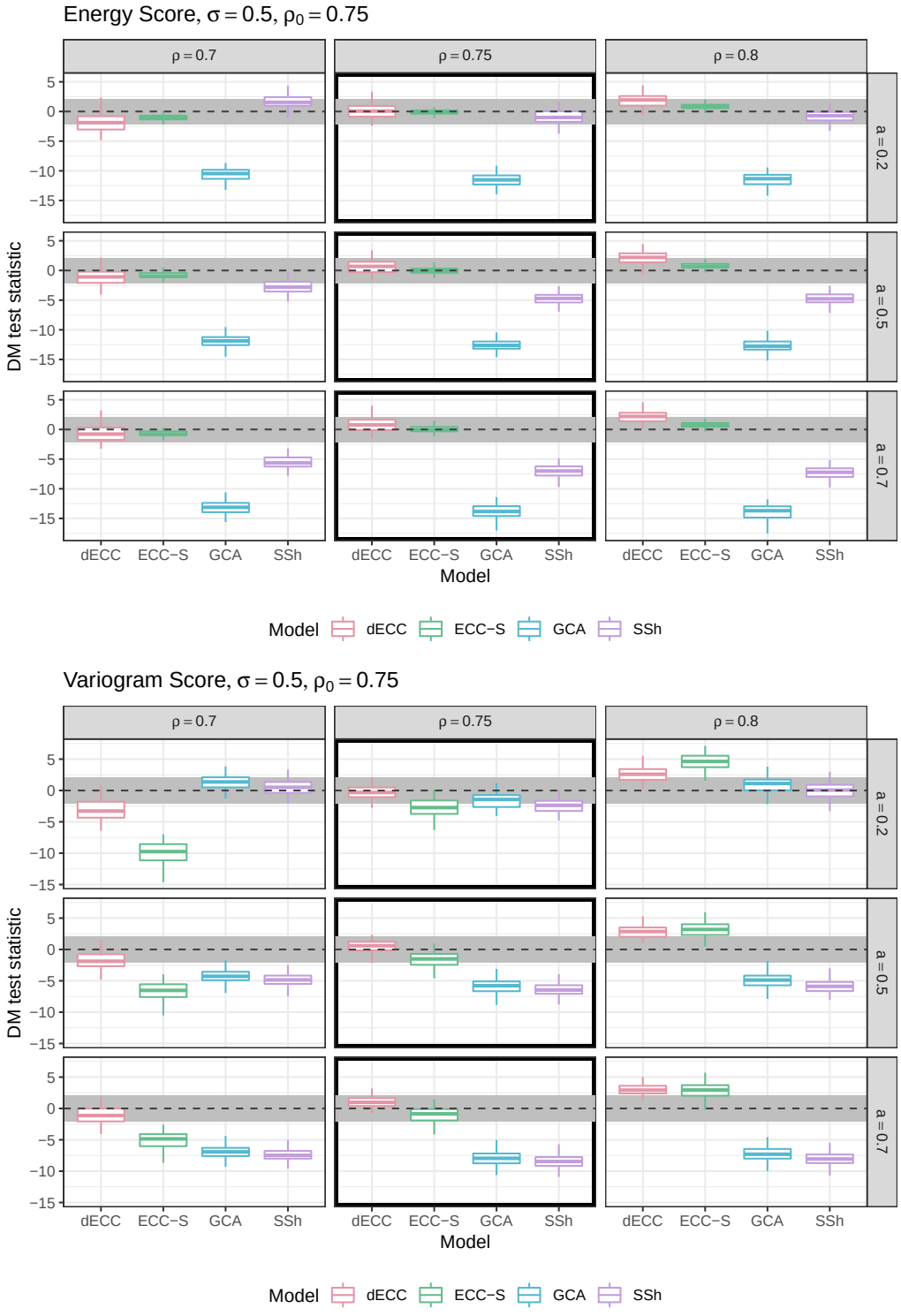


Figure 35: As Figure 29, but for $\sigma = 0.5$ and high correlation values.

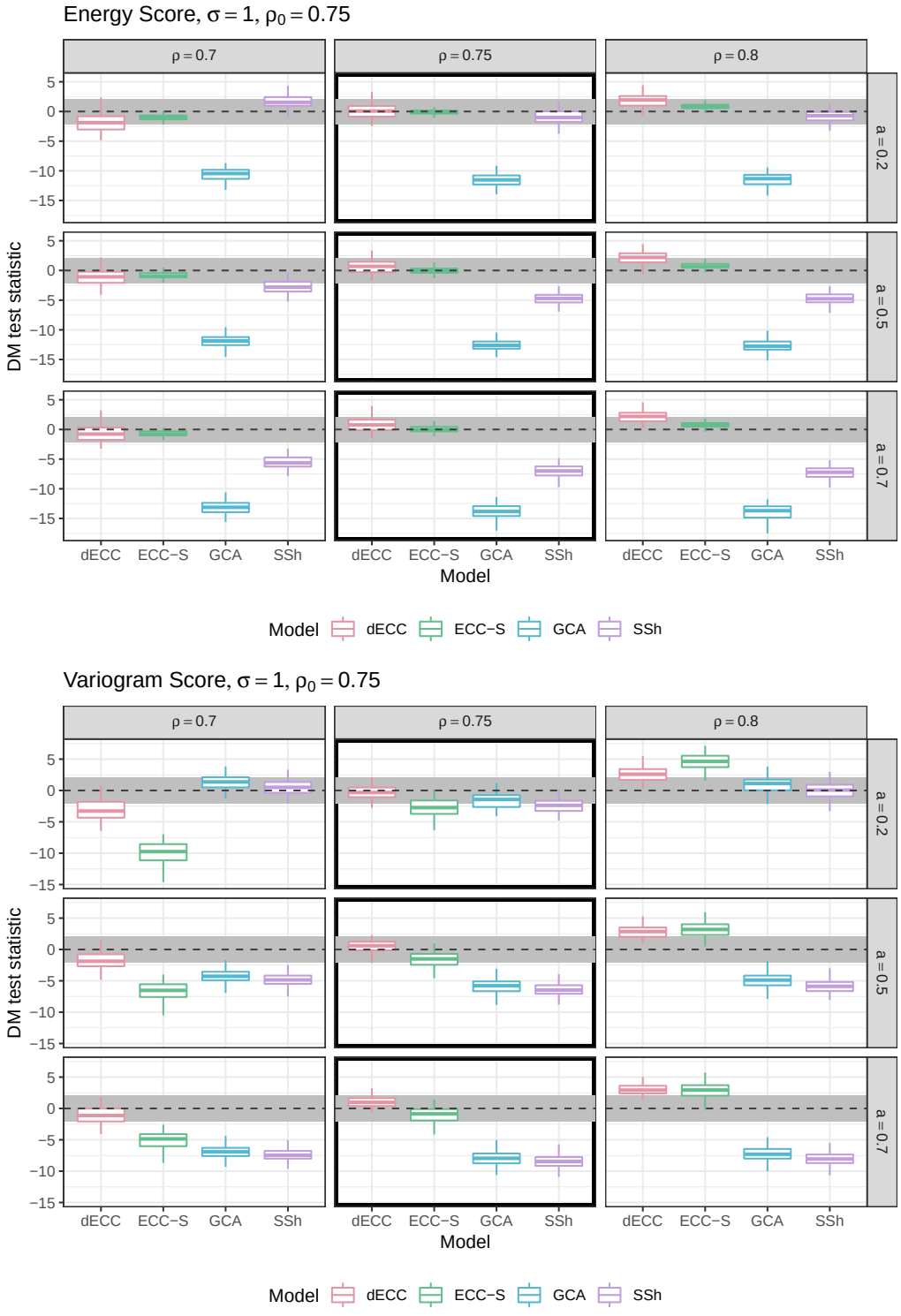


Figure 36: As Figure 29, but for $\sigma = 1$ and high correlation values.

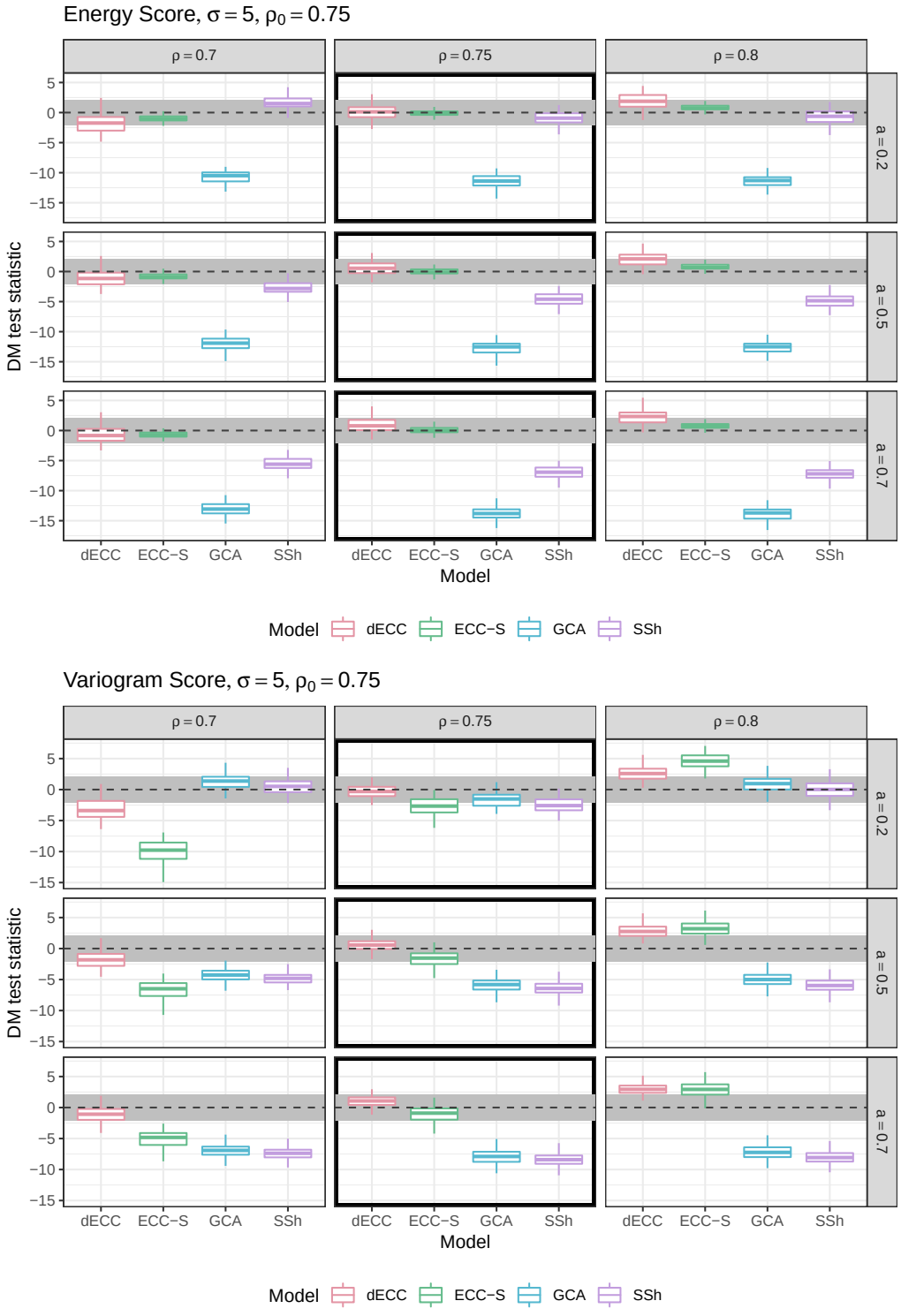


Figure 37: As Figure 29, but for $\sigma = 5$ and high correlation values.

5 Additional results for a multivariate truncated normal distribution-based setting (Setting S1)

A preliminary version of the manuscript included a simulation setting based on a multivariate truncated normal distribution. A description of the setting and simulation results are provided below.

5.1 Simulation setup of Setting S1

Setting 1 can be generalized by replacing the multivariate Gaussian distribution by a multivariate truncated Gaussian distribution $\mathcal{N}^d_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are the vectors of lower and upper truncation points, respectively. In univariate settings this distribution plays important role in wind speed modelling (Thorarinsdottir and Gneiting, 2010) or in post-processing of hydrological forecasts (Hemri and Klein, 2017). Compared to Setting 1, here misspecifications in location vector $\boldsymbol{\mu}$ and/or scale matrix $\boldsymbol{\Sigma}$ result in more complex deviations in mean vectors and covariance matrices.

(S1) For iterations $t = 1, \dots, n$, independent and identically distributed samples of observations $\mathbf{y}$ and ensemble forecasts $\mathbf{X}_1, \dots, \mathbf{X}_m$ are generated from $\mathcal{N}^d_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{\mu_0}, \boldsymbol{\Sigma^0})$ and $\mathcal{N}^d_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, respectively, where $\boldsymbol{\Sigma^0}$ and $\boldsymbol{\Sigma}$ are defined as in Setting 1, $\boldsymbol{\mu_0} = (\mu_0, \dots, \mu_0) \in \mathbb{R}^d$ and $\boldsymbol{\mu} = (\mu, \dots, \mu) \in \mathbb{R}^d$.

(S2) Univariate post-processing is based on the truncated normal EMOS model of Hemri and Klein (2017), where the EMOS coefficients are calculated by optimizing the mean CRPS over the training set consisting of the n_{init} initial iterations. Similar to Setting 1, the obtained EMOS models are used to produce out of sample forecasts for the n_{test} iterations in the test set.

(S3) Identical to (S3) of Setting 1.

For simplicity, we consider a lower truncation at 0 only, i.e., $\boldsymbol{a} = (0, \dots, 0)$ and $\boldsymbol{b} = (\infty, \dots, \infty)$. The truncated Gaussian setting is implemented for $d = 5$ and all combinations of

$$\mu_0 \in \{2, 3\},\ \mu \in \{2, 3, 5\},\ \rho_0 \in \{0.25, 0.5, 0.75\},\ \rho \in \{0.1, 0.25, 0.5, 0.75, 0.9\},\ \sigma \in \{0.25, 0.5, 1, 3, 5\}\ ,$$

resulting in 450 experiments which are repeated 100 times each.

5.2 Results for Setting S1

Figure 38 summarizes results for Setting S1 in terms of the ES (top) and the VS (bottom). Following a similar structure as in the main paper, we only show results for $\sigma = 1$ and $\rho, \rho_0 \in \{0.25, 0.5, 0.75\}$ here, but discuss the effect of misspecifications of the variance below. Corresponding plots are provided in Section 5.3.

Overall, SSh consistently provides significant improvements over ECC-Q except for settings in which the correlation structure of the ensemble is correctly specified ($\rho = \rho_0$), where no significant differences can be detected. The relative differences in favor of SSh are increased for larger absolute differences of ρ and ρ_0. While these findings are in line with the results from Setting 1 and hold for both ES and VS, the relative performance of GCA is now somewhat different from before. In terms of the ES, GCA here performs worse for many parameter settings. In particular, GCA is significantly worse than ECC-Q if $\rho = \rho_0$, and does not offer any improvements over ECC-Q if ρ and ρ_0 are not too different. By contrast, in terms of the VS GCA shows much better performance and outperforms ECC-Q in almost all cases. Even for cases where $\rho = \rho_0$, no significant differences in terms of VS can be detected.

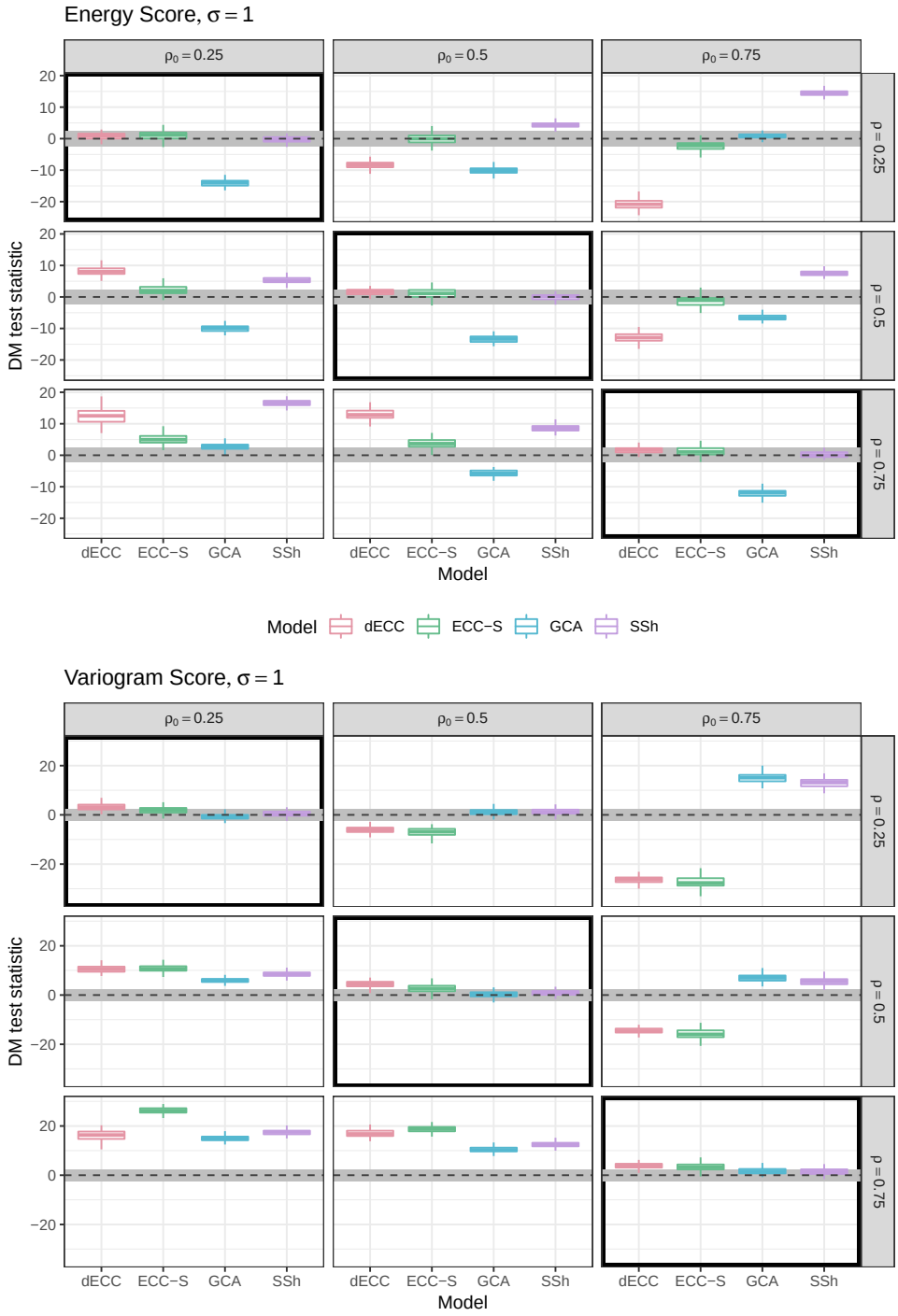


Figure 38: Summaries of DM test statistic values based on the ES (top) and the VS (bottom) for Setting S1 with $\epsilon = 3, \mu_0 = 2$, and $\sigma = 1$. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

As for Setting 1, the results for ECC-S and dECC strongly depend on the misspecification of the correlation structure. Further, the results now additionally vary with the variance misspecification parameter σ also for ECC-S. In terms of the ES, ECC-S provides improvements over ECC-Q for $\rho > \rho_0$ and similar forecast quality for $\rho \leq \rho_0$. However, the forecast quality deteriorates for larger values of σ and the ECC-S forecasts are significantly worse than those of ECC-Q, even for cases where $\rho = \rho_0$. In terms of the VS, ECC-S also provides significant improvements over ECC-Q for $\rho > \rho_0$ and even provides the best forecasts among all models, but shows significantly worse performance for $\rho < \rho_0$ across all values of σ.

The results for dECC are similar for ES and VS. For both scoring rules and $\rho < \rho_0$, the values of the DM test statistics move from positive (improvement over ECC-Q) to negative (deterioration compared to ECC-Q) values with increasing σ. For $\rho > \rho_0$, they instead change from negative to positive values. While improvements over ECC-Q can be observed for cases with a correctly specified correlation structure of the ensemble ($\rho = \rho_0$) and $\sigma = 1$, dECC performs worse than ECC-Q for those cases if $\sigma \neq 1$.

5.3 Additional results for Setting S1

An overview the additional results for Setting S1 is provided in Table 4.

Table 4: Overview of parameter combinations for Setting S1 with references to corresponding figures in this supplement.

μ_0	ϵ	σ	Figure
2	2	0.25	39
2	3	0.5	40
2	3	1	41
2	3	3	42
2	3	5	43
2	3	0.25	44
2	3	0.5	45
2	3	1	46
2	3	3	47
2	3	5	48
2	5	0.25	49
2	5	0.5	50
2	5	1	51
2	5	3	52
2	5	5	53
3	2	0.25	54
3	3	0.5	55
3	3	1	56
3	3	3	57
3	3	5	58
3	3	0.25	59
3	3	0.5	60
3	3	1	61
3	3	3	62
3	3	5	63
3	5	0.25	64
3	5	0.5	65
3	5	1	66
3	5	3	67
3	5	5	68

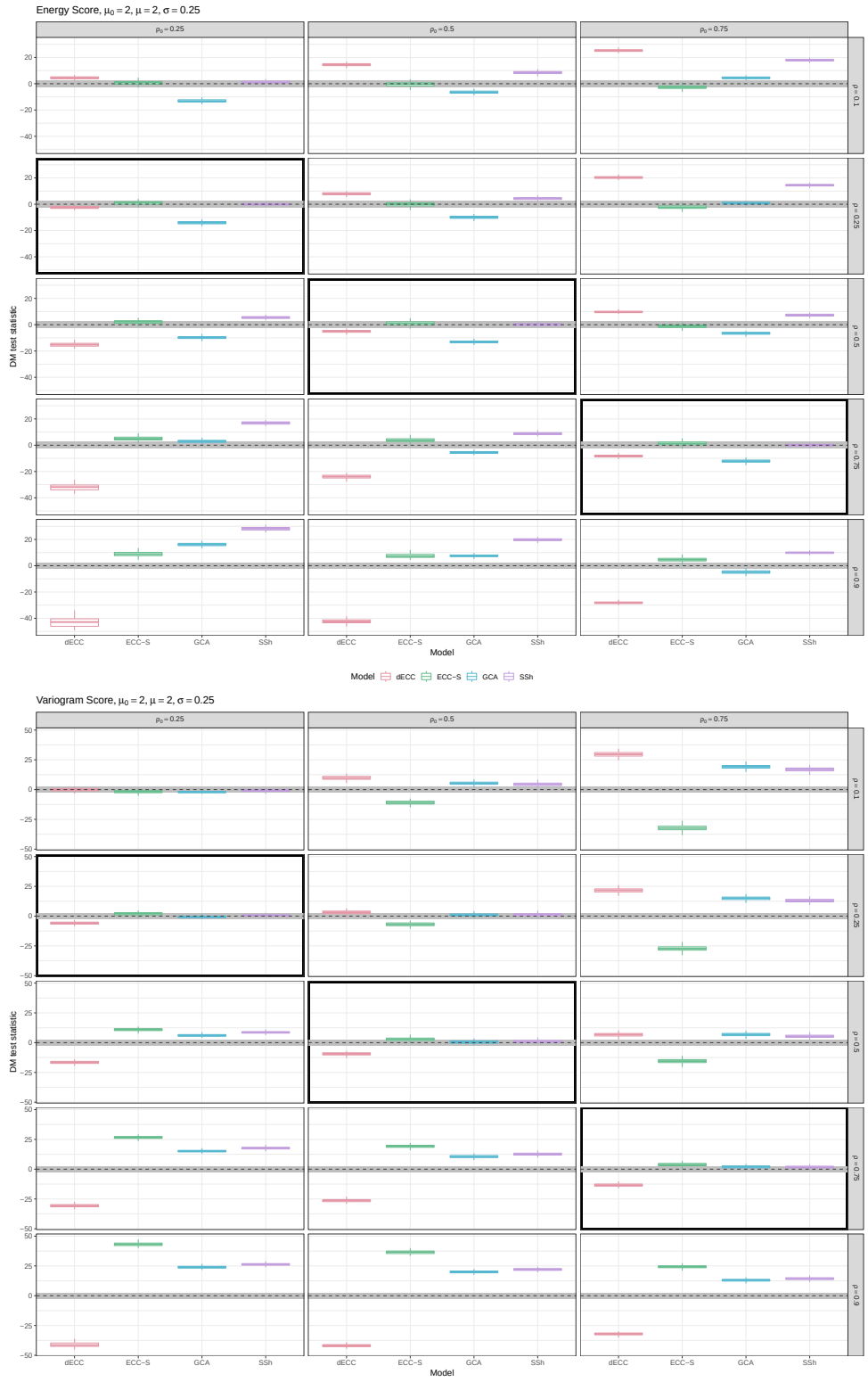


Figure 39: Summaries of DM test statistic values based on ES (top) and VS (bottom) for Setting S1 with $\mu_0 = 2, \epsilon = 2$, and $\sigma = 0.25$. ECC-Q forecasts are used as reference model such that positive values of the test statistic indicate improvements over ECC-Q and negative values indicate deterioration of forecast skill. Boxplots summarize results of the 100 Monte Carlo repetitions of each individual experiment. The horizontal gray stripe indicates the acceptance region of the two-sided DM test under the null hypothesis of equal predictive performance at a level of 0.05. Simulation parameter choices where the correlation structure of the raw ensemble is correctly specified ($\rho = \rho_0$) are surrounded by black boxes.

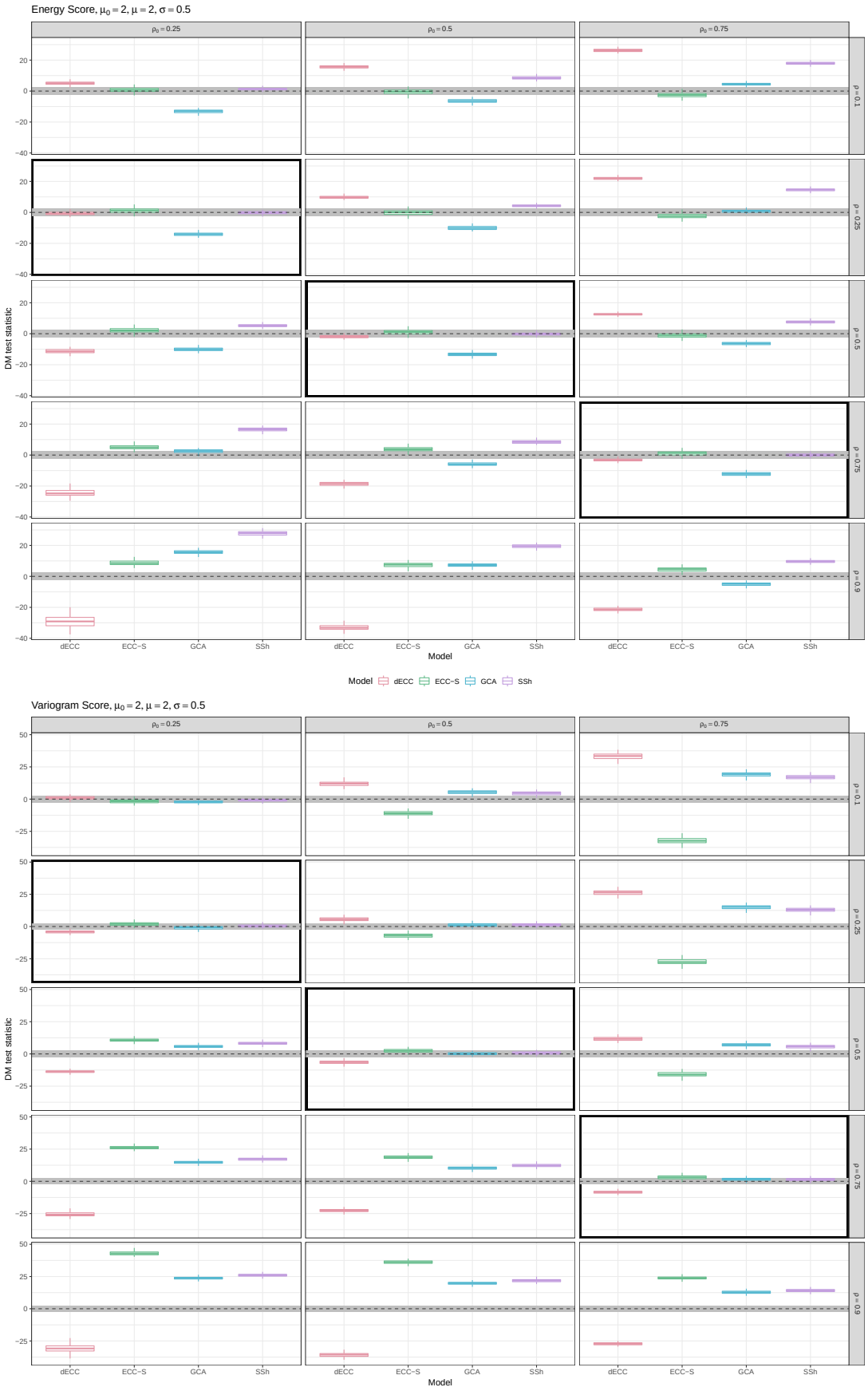


Figure 40: As Figure 39, but for $\mu_0 = 2, \epsilon = 2$, and $\sigma = 0.5$.

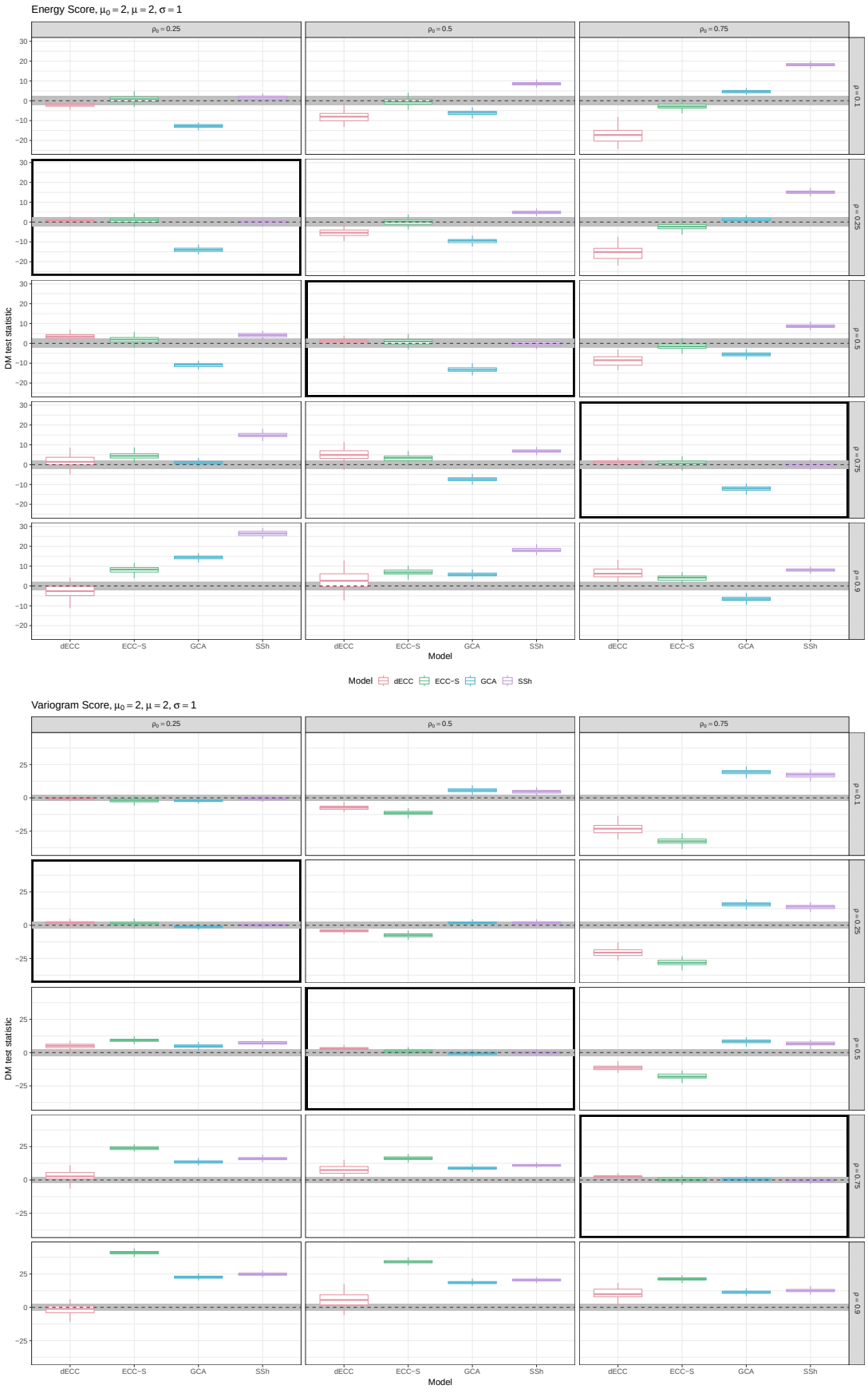


Figure 41: As Figure 39, but for $\mu_0 = 2, \epsilon = 2$, and $\sigma = 1$.

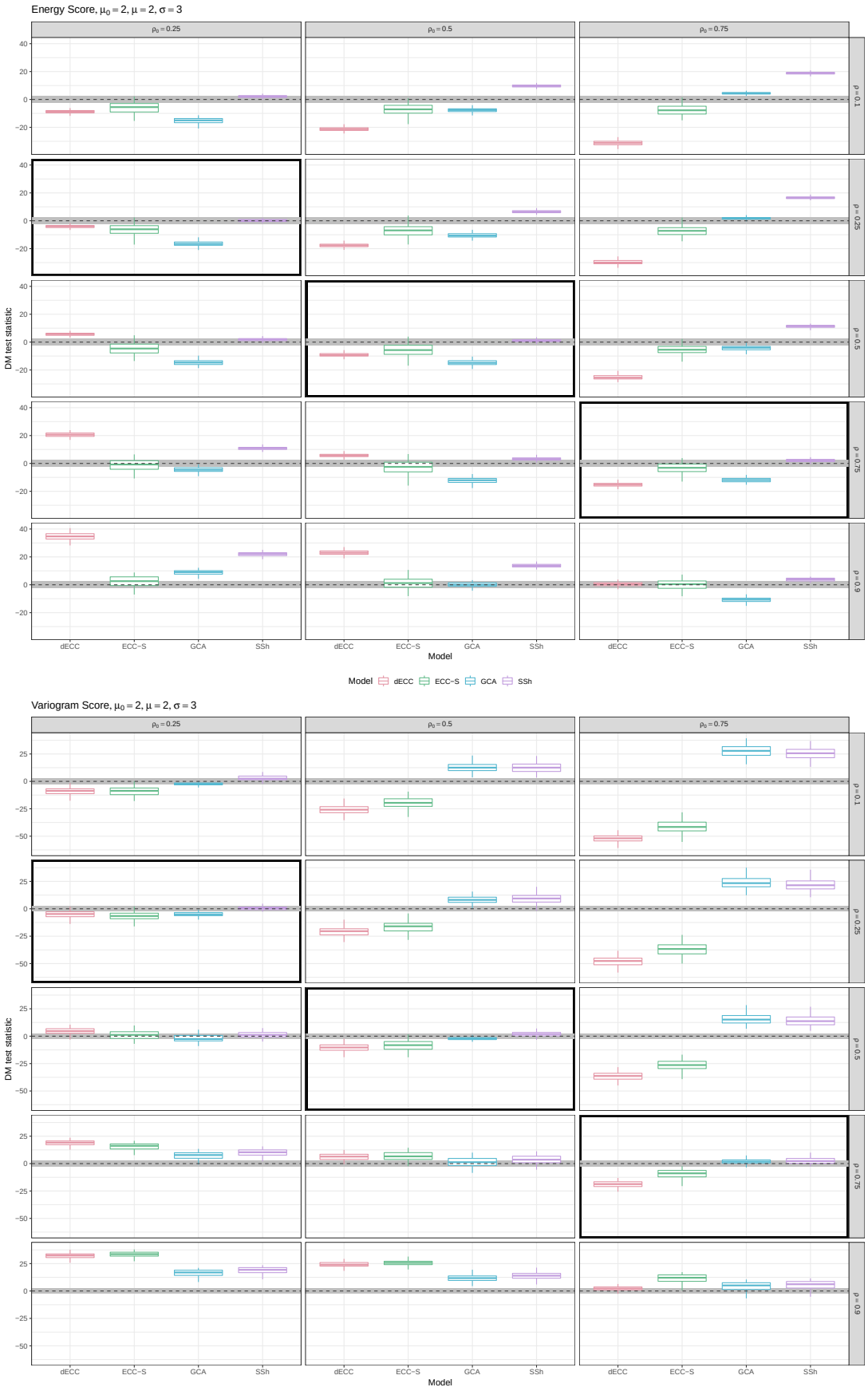


Figure 42: As Figure 39, but for $\mu_0 = 2, \epsilon = 2$, and $\sigma = 3$.

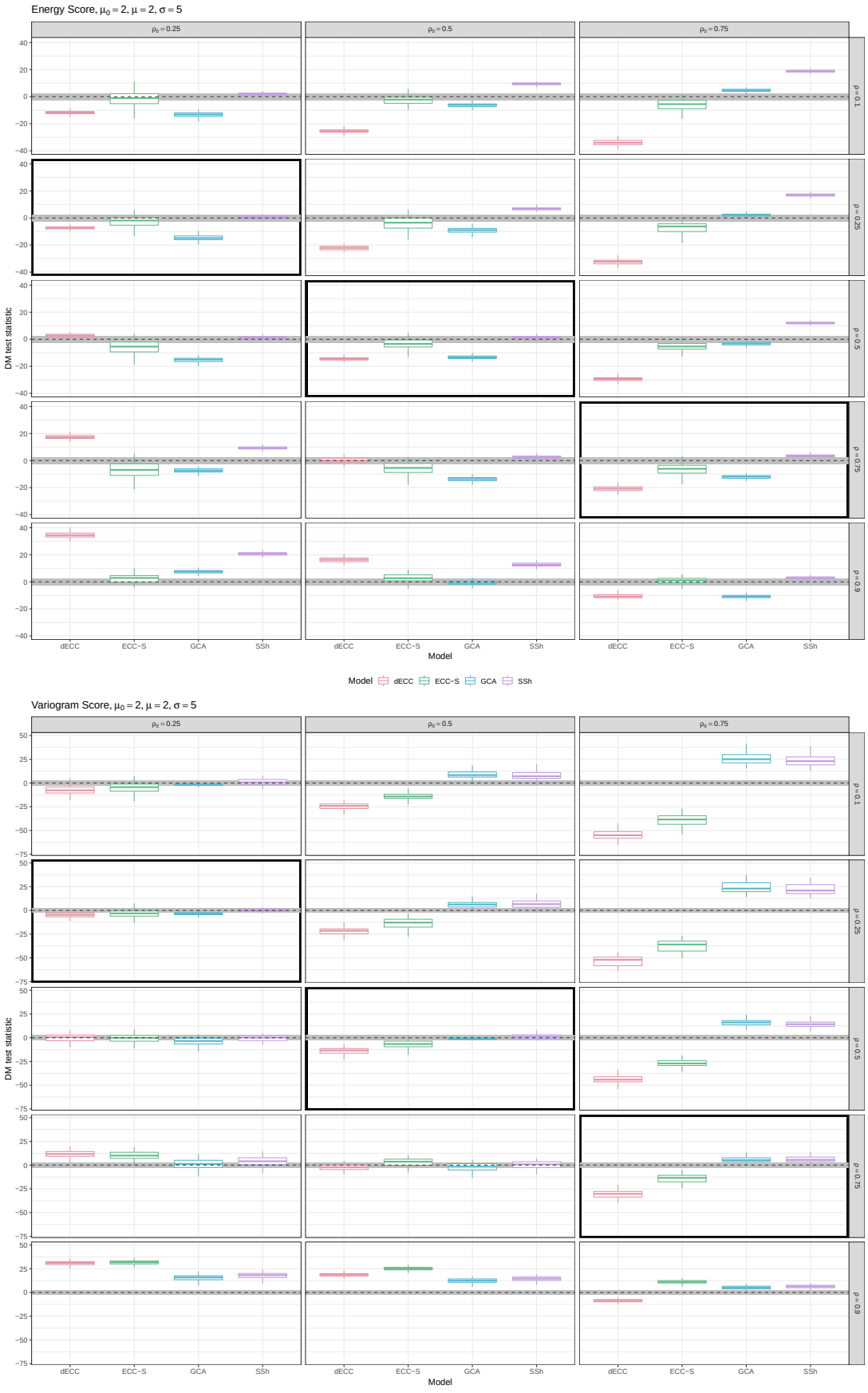


Figure 43: As Figure 39, but for $\mu_0 = 2, \epsilon = 2$, and $\sigma = 5$.

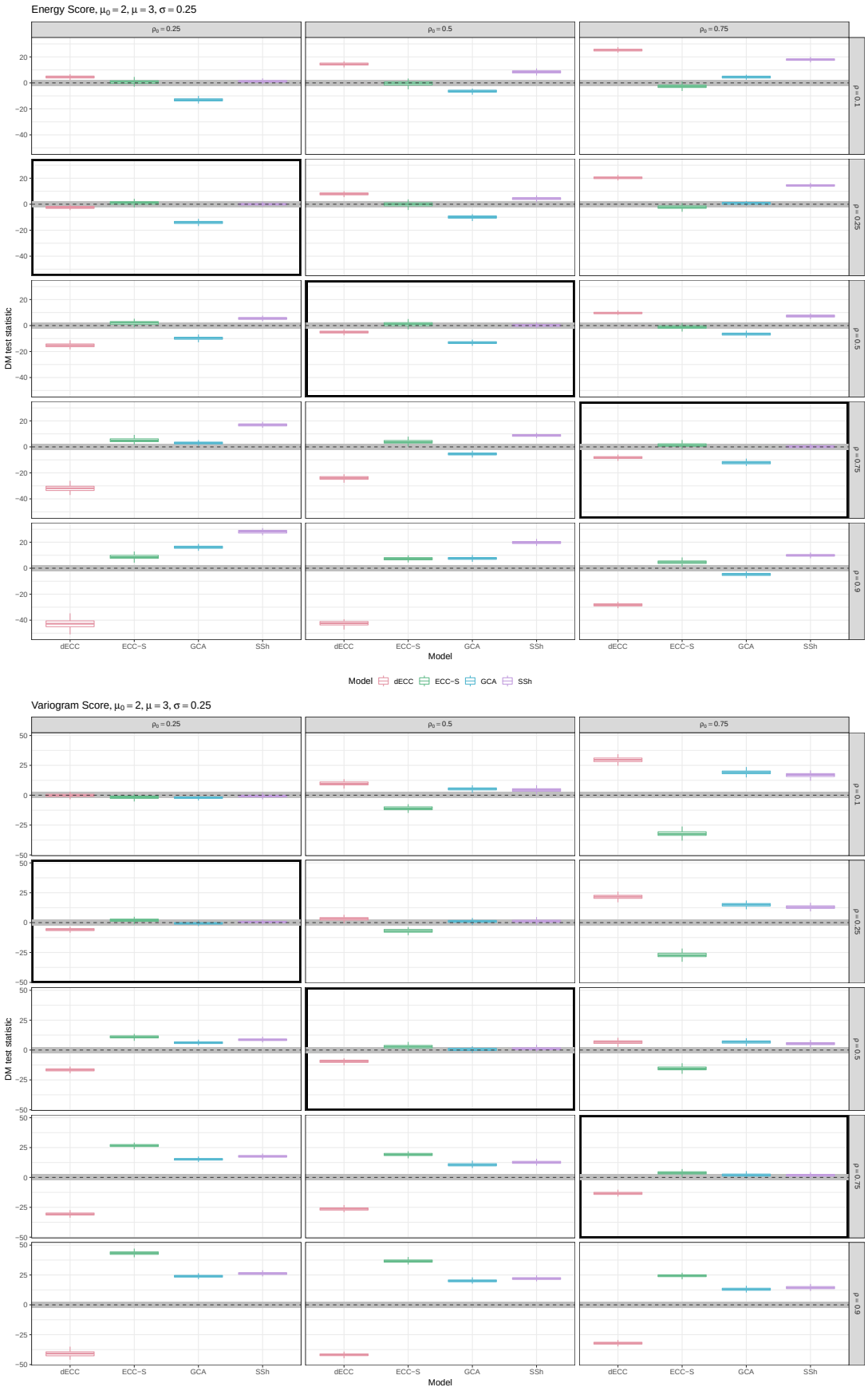


Figure 44: As Figure 39, but for $\mu_0 = 2, \epsilon = 3$, and $\sigma = 0.5$.

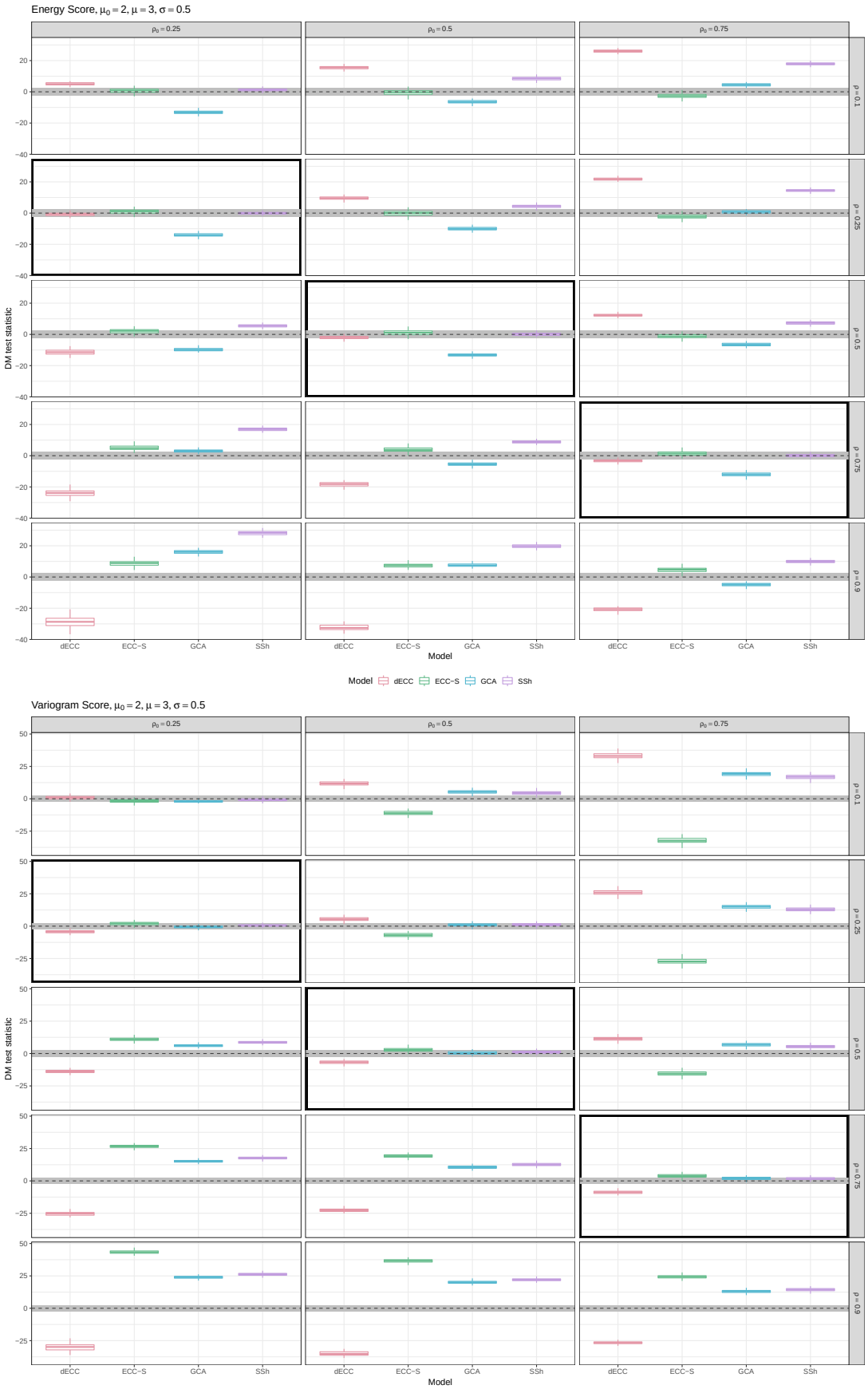


Figure 45: As Figure 39, but for $\mu_0 = 2, \epsilon = 3$, and $\sigma = 0.5$.

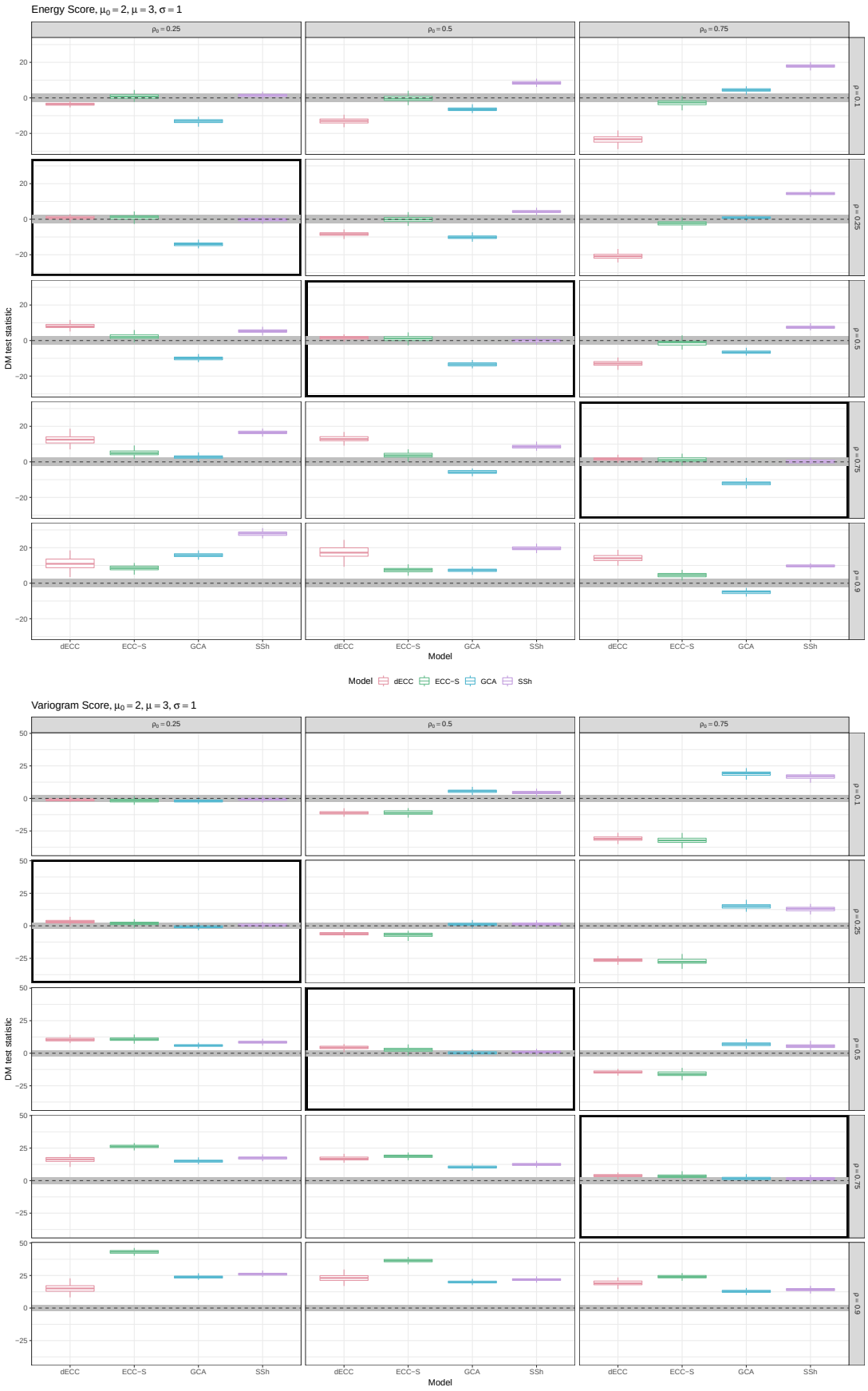


Figure 46: As Figure 39, but for $\mu_0 = 2, \epsilon = 3$, and $\sigma = 1$.

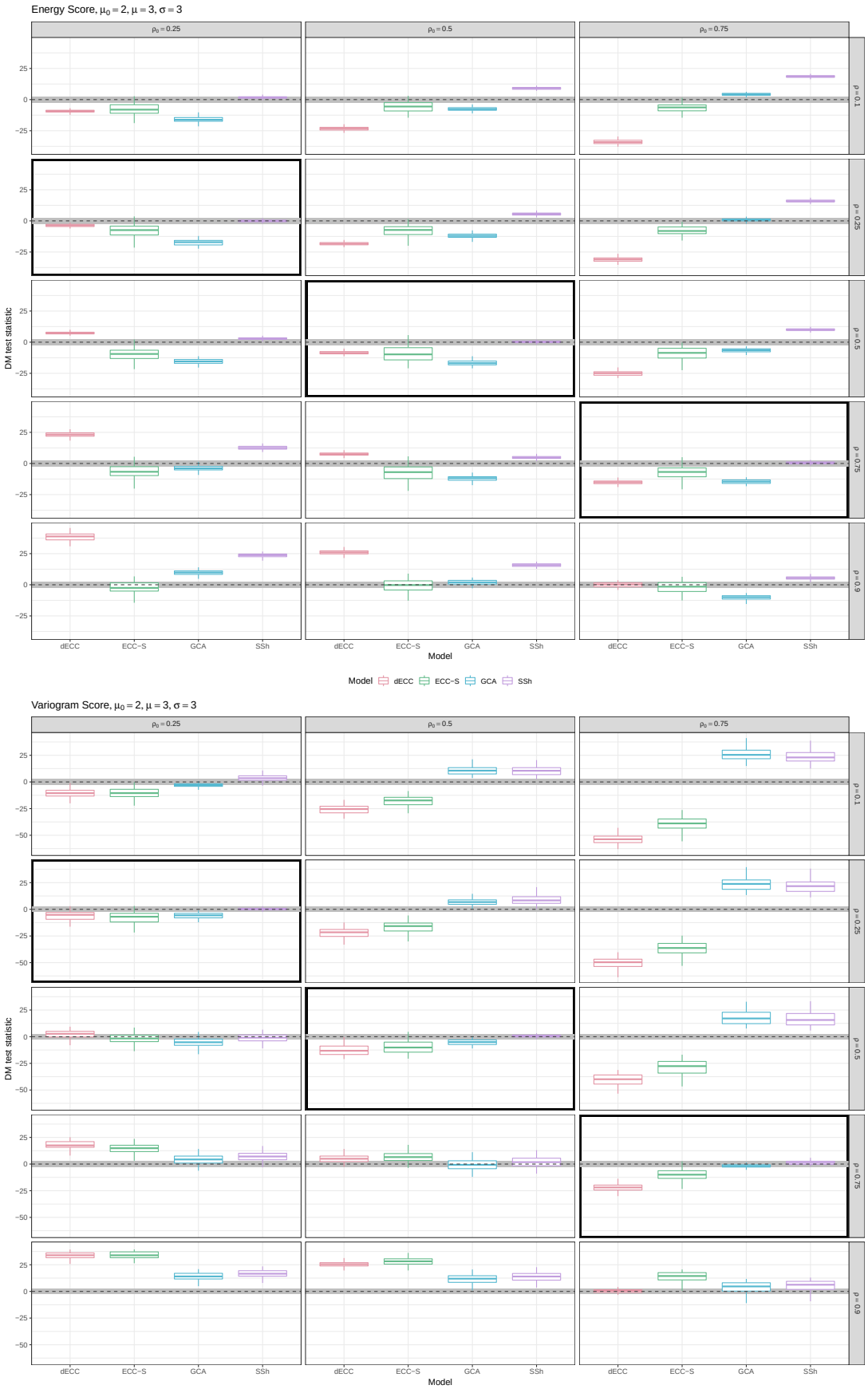


Figure 47: As Figure 39, but for $\mu_0 = 2, \epsilon = 3$, and $\sigma = 3$.

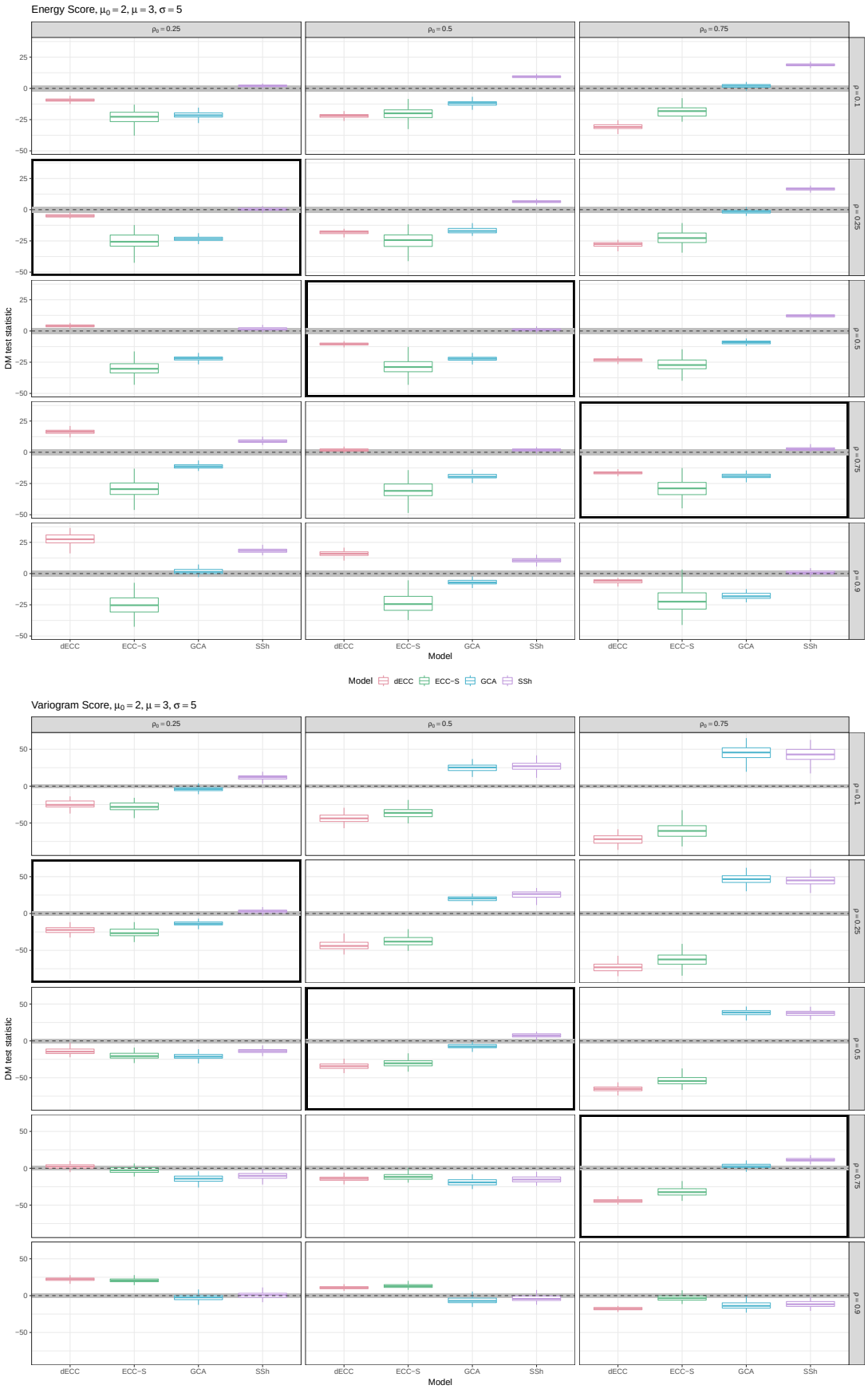


Figure 48: As Figure 39, but for $\mu_0 = 2, \epsilon = 3$, and $\sigma = 5$.

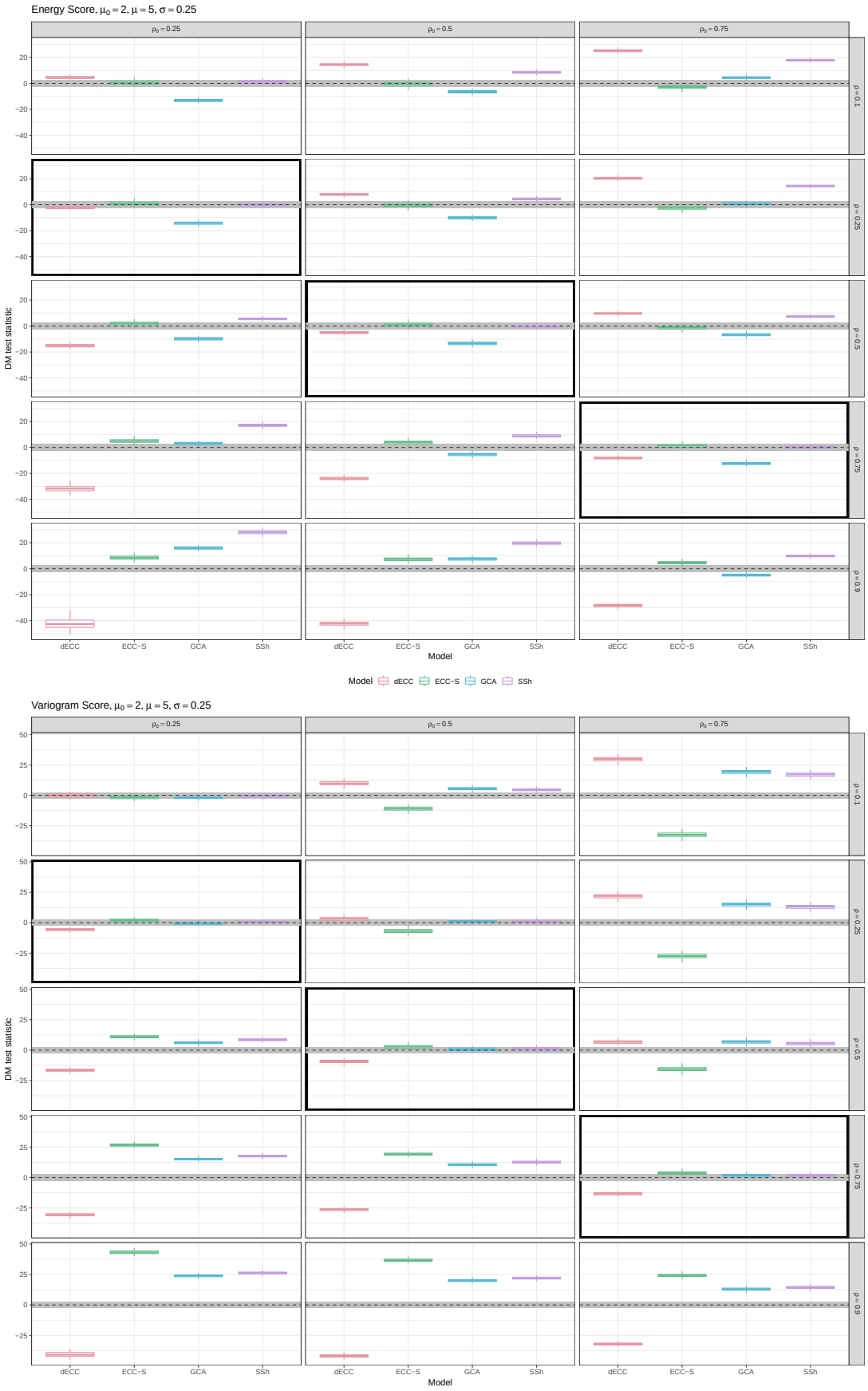


Figure 49: As Figure 39, but for $\mu_0 = 2, \epsilon = 5$, and $\sigma = 0.5$.

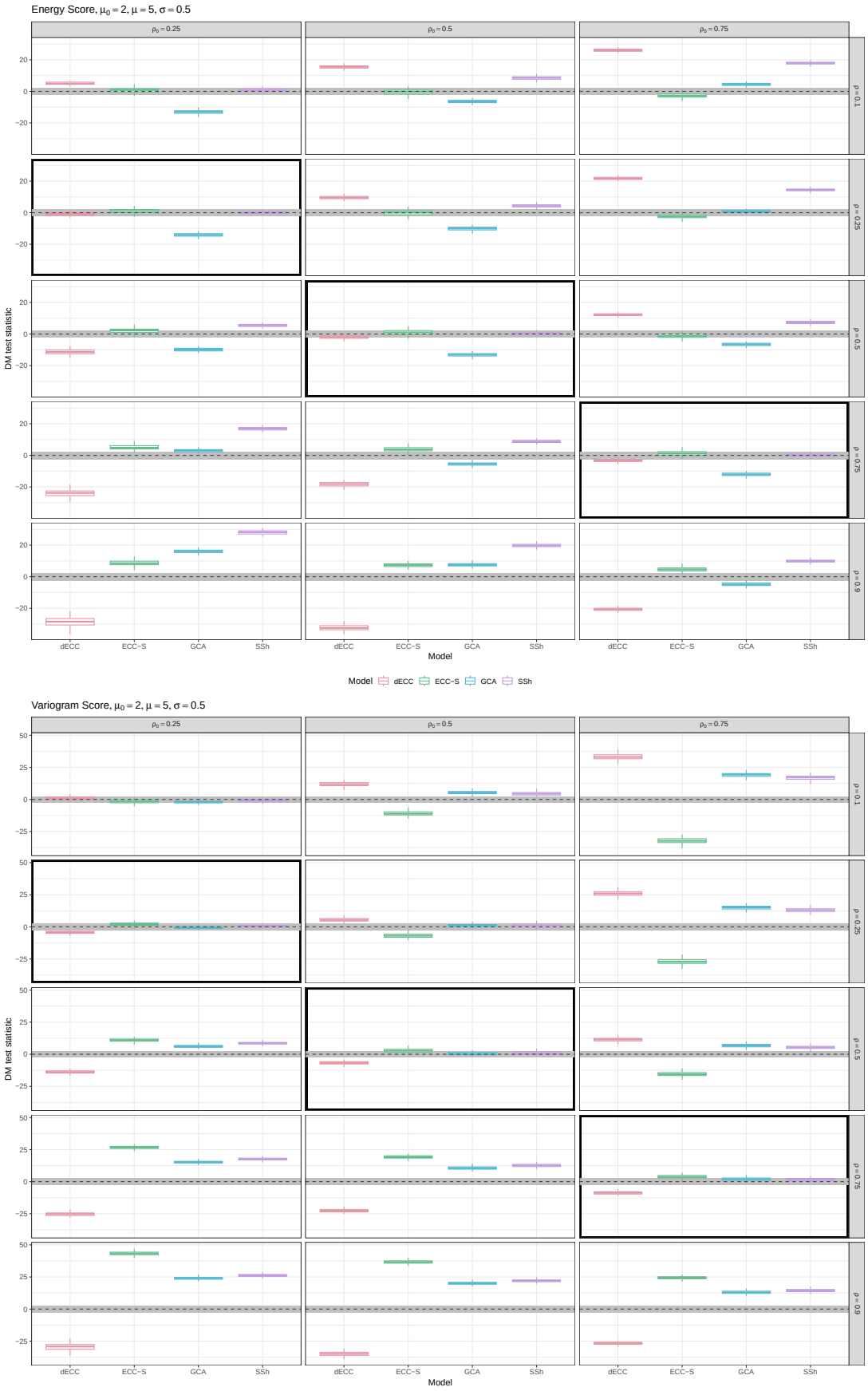


Figure 50: As Figure 39, but for $\mu_0 = 2, \epsilon = 5$, and $\sigma = 0.5$.

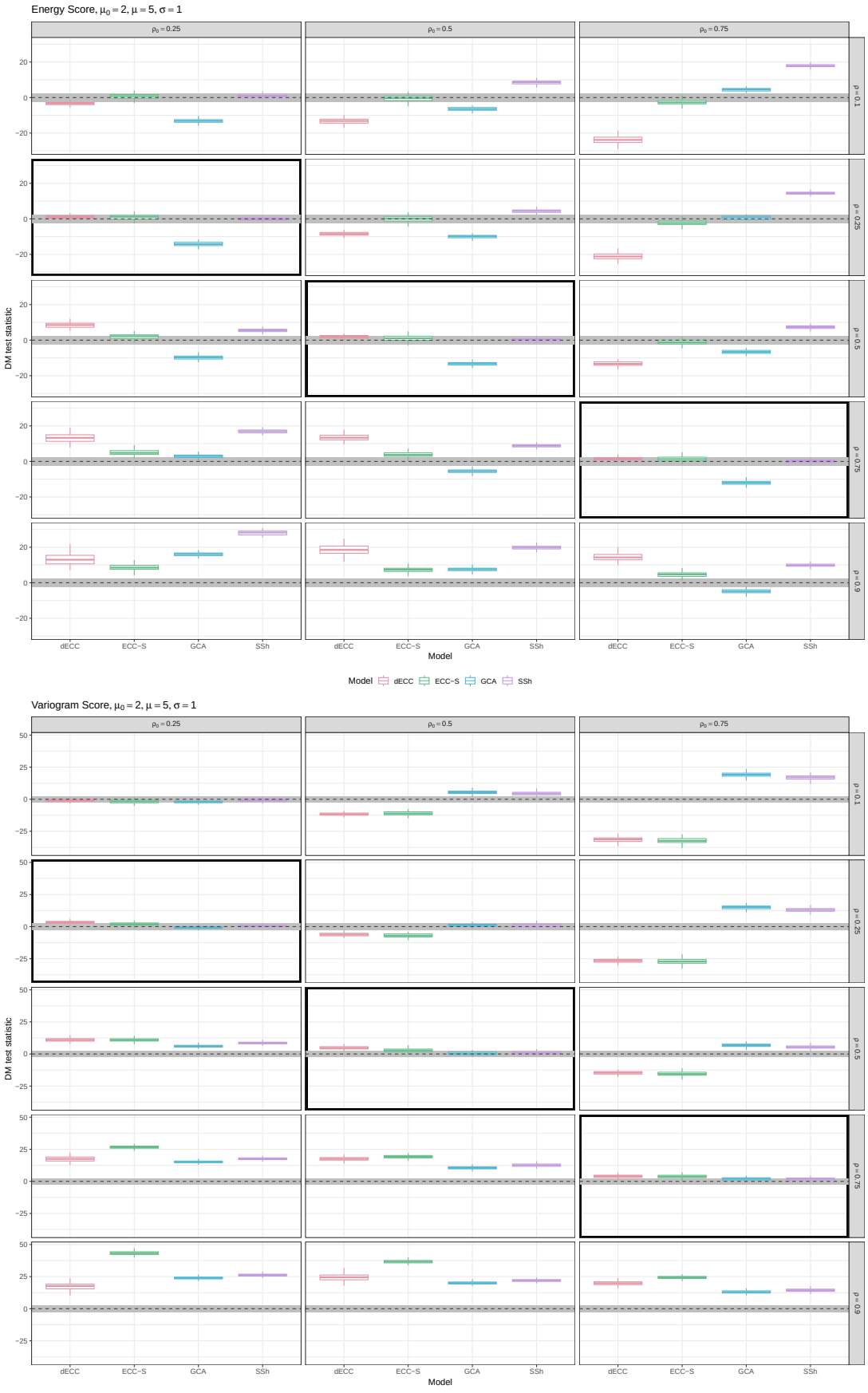


Figure 51: As Figure 39, but for $\mu_0 = 2, \epsilon = 5$, and $\sigma = 1$.

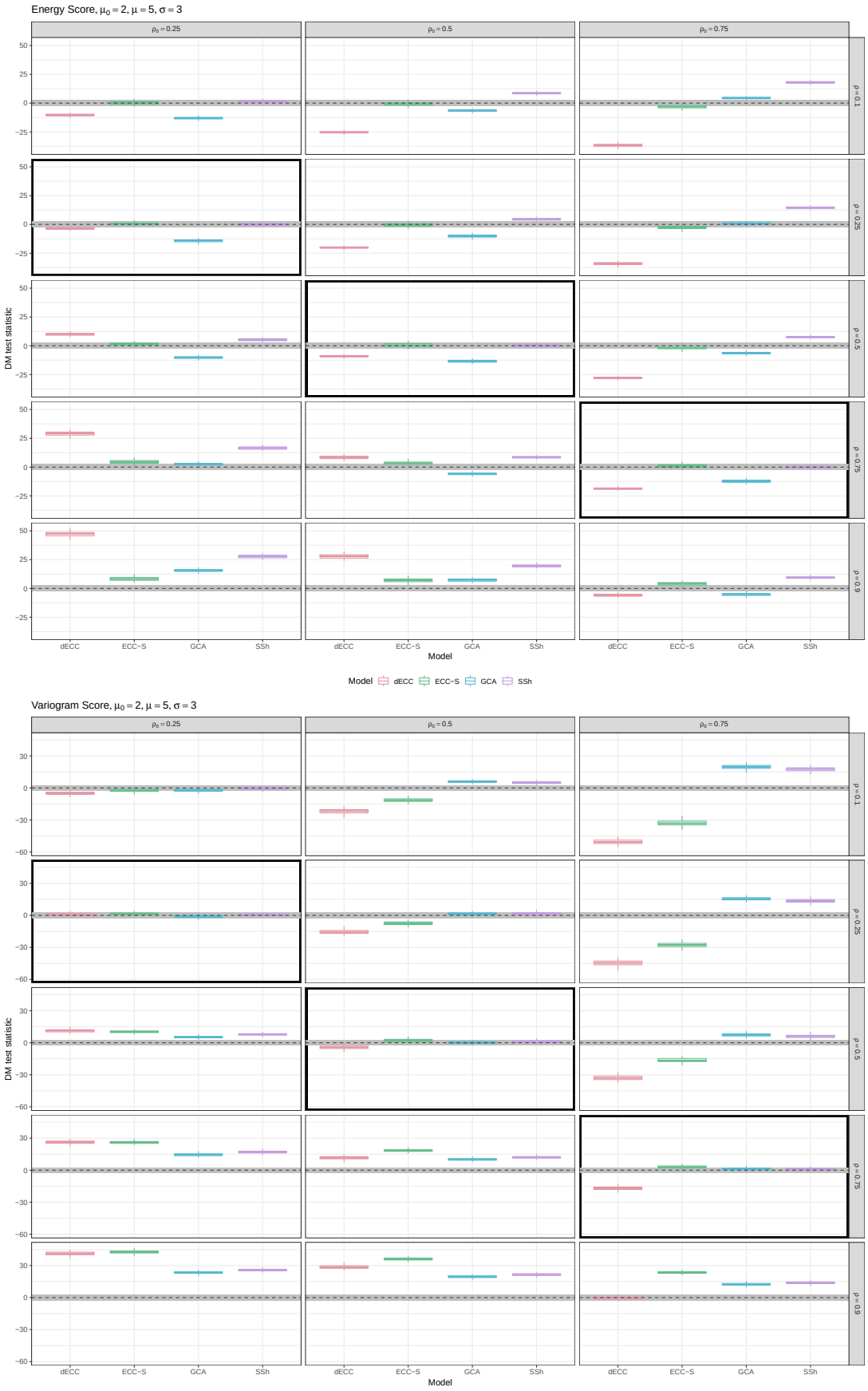


Figure 52: As Figure 39, but for $\mu_0 = 2, \epsilon = 5$, and $\sigma = 3$.

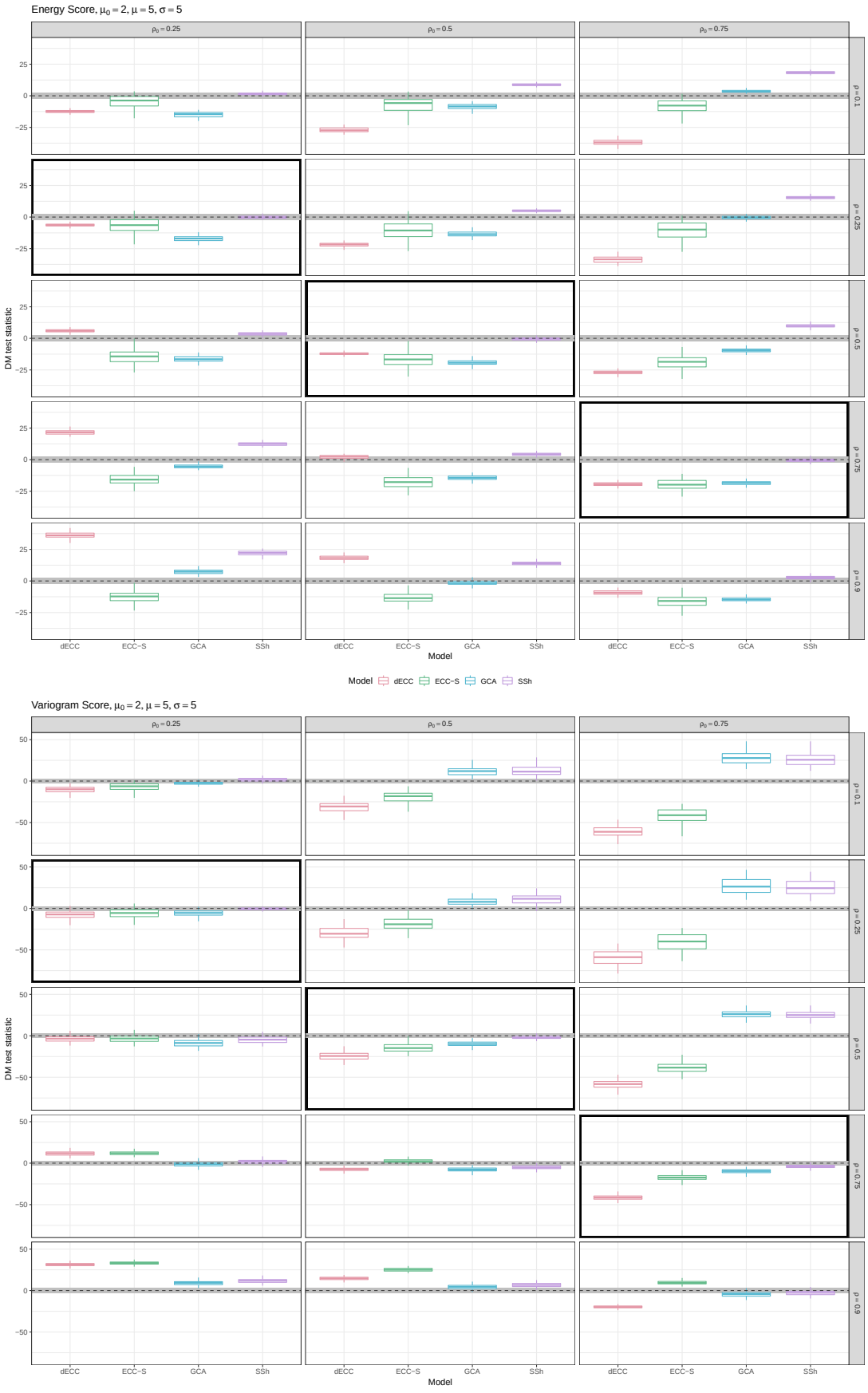


Figure 53: As Figure 39, but for $\mu_0 = 2, \epsilon = 5$, and $\sigma = 5$.

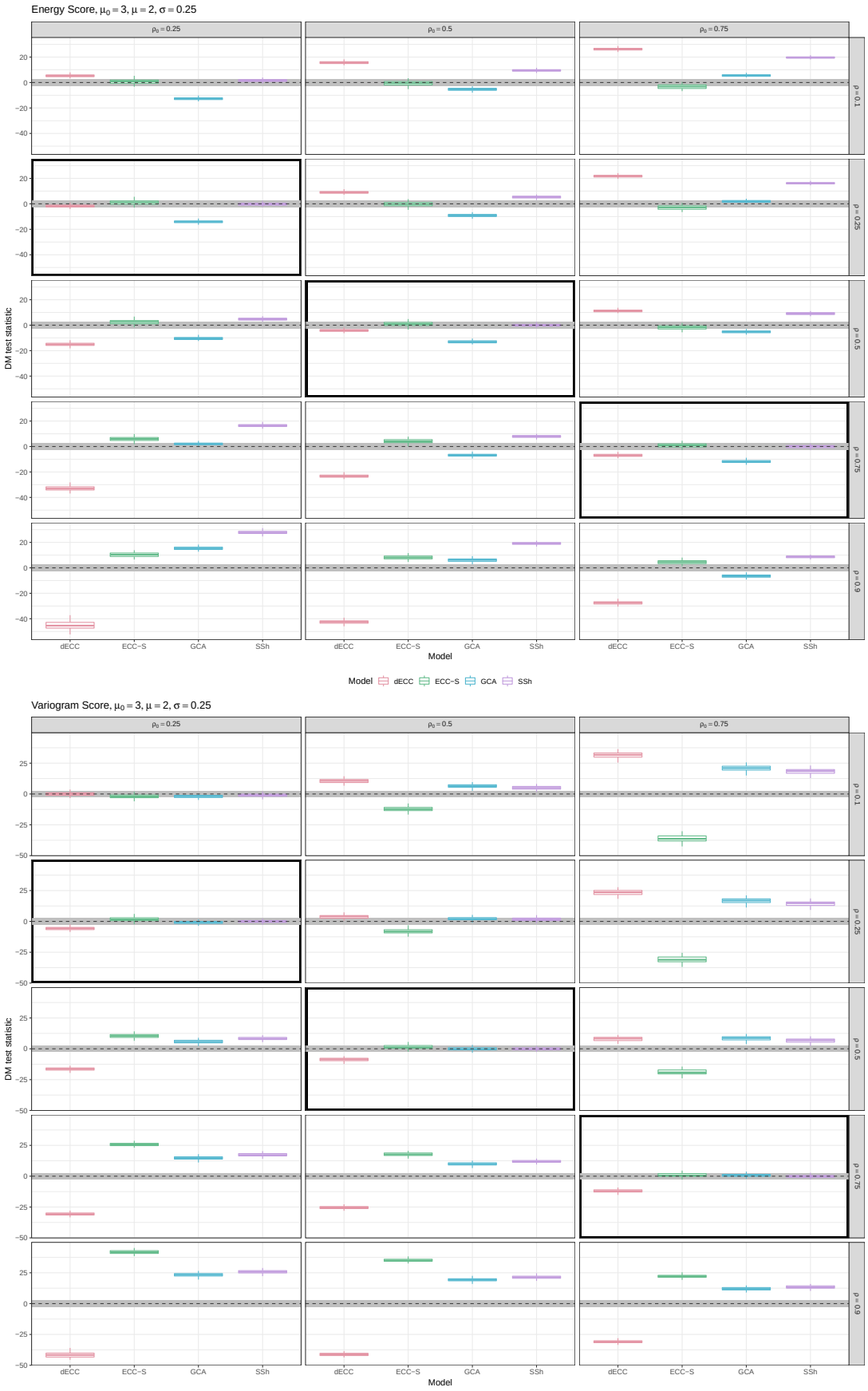
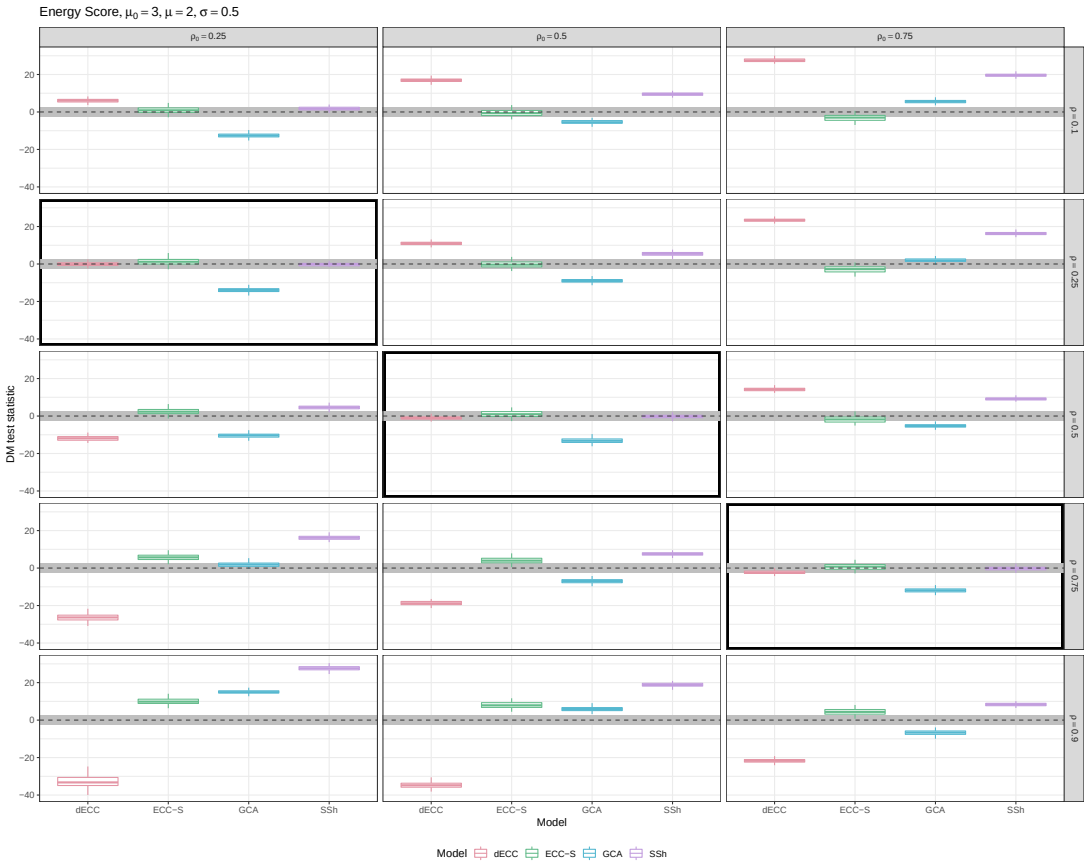


Figure 54: As Figure 39, but for $\mu_0 = 3, \epsilon = 2$, and $\sigma = 0.5$.



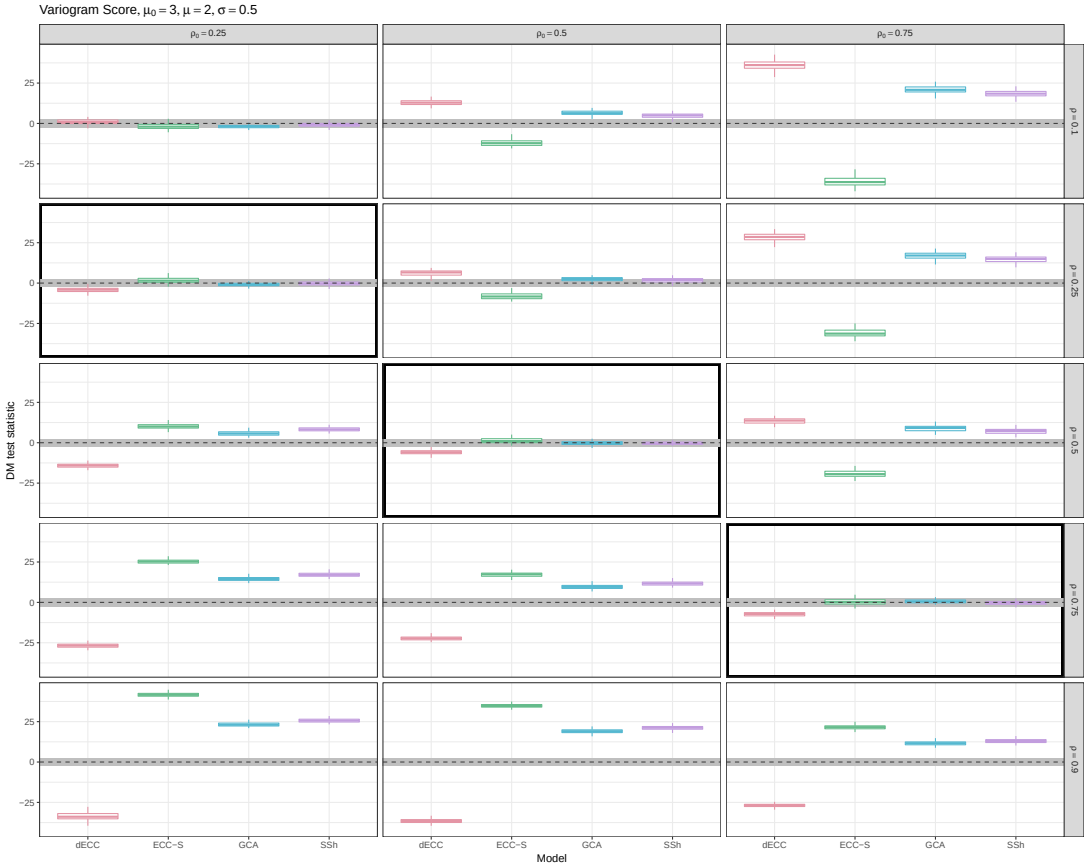


Figure 55: As Figure 39, but for $\mu_0 = 3, \epsilon = 2$, and $\sigma = 0.5$.

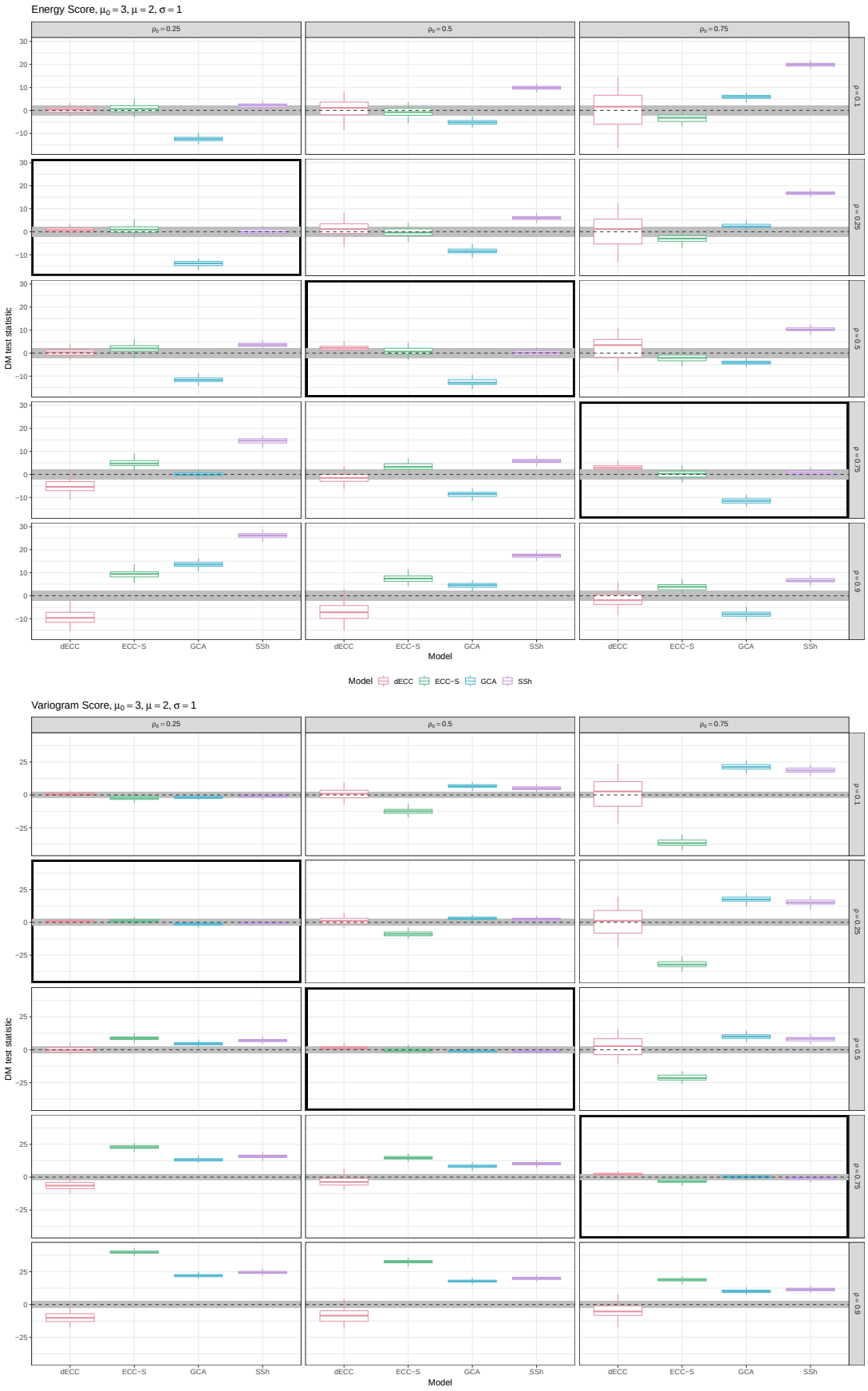


Figure 56: As Figure 39, but for $\mu_0 = 3, \epsilon = 2$, and $\sigma = 1$.

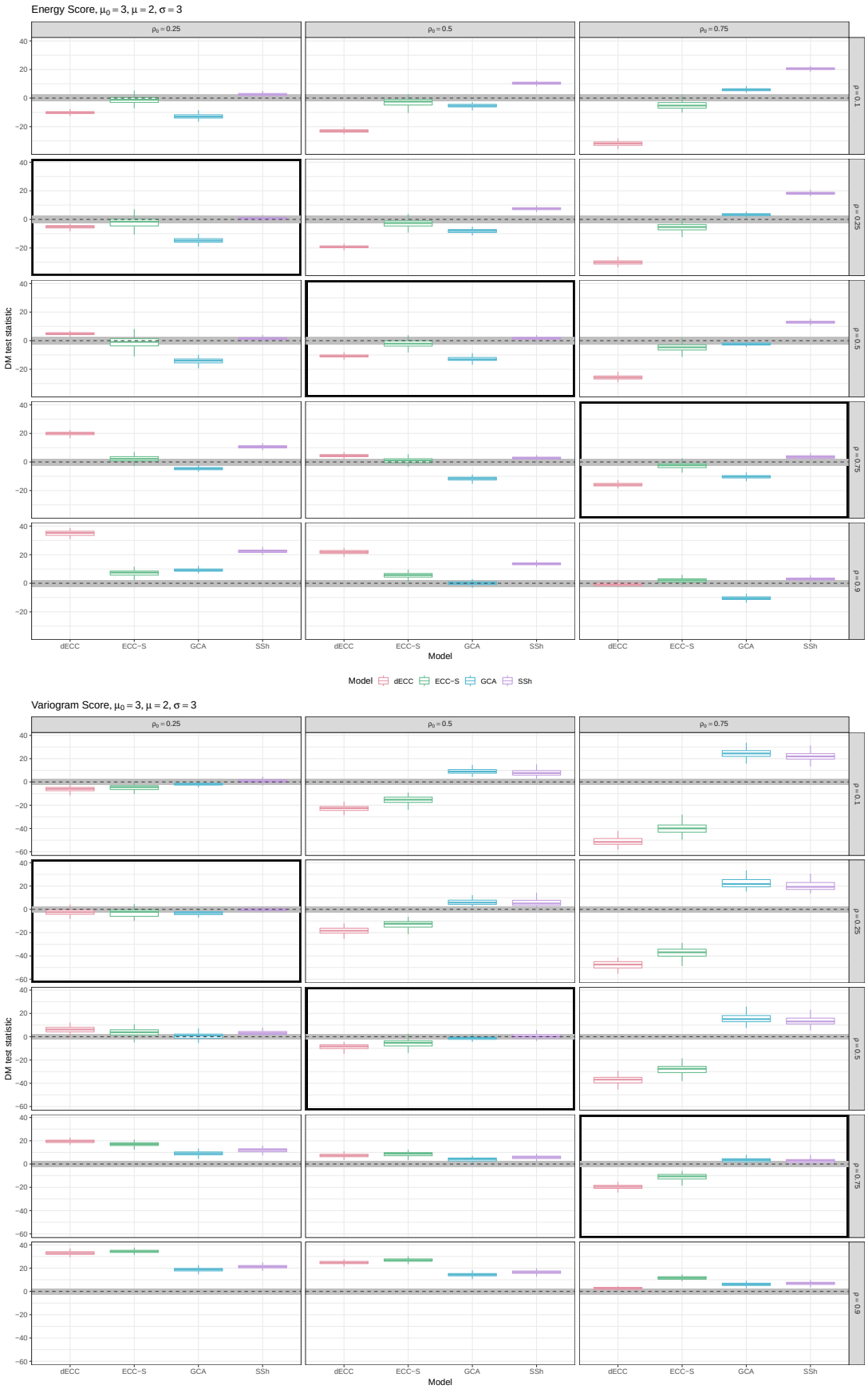


Figure 57: As Figure 39, but for $\mu_0 = 3, \epsilon = 2$, and $\sigma = 3$.

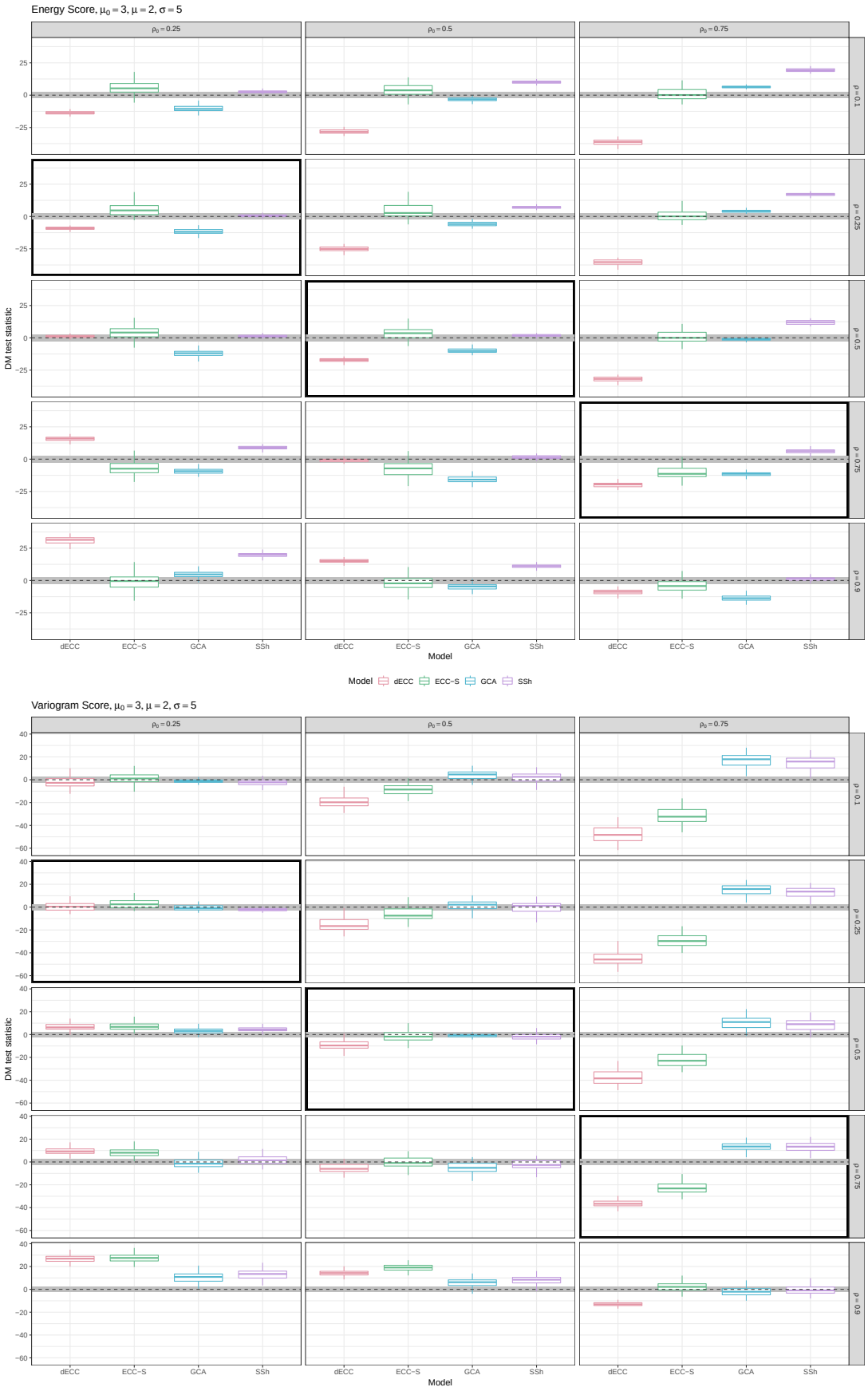


Figure 58: As Figure 39, but for $\mu_0 = 3, \epsilon = 2$, and $\sigma = 5$.

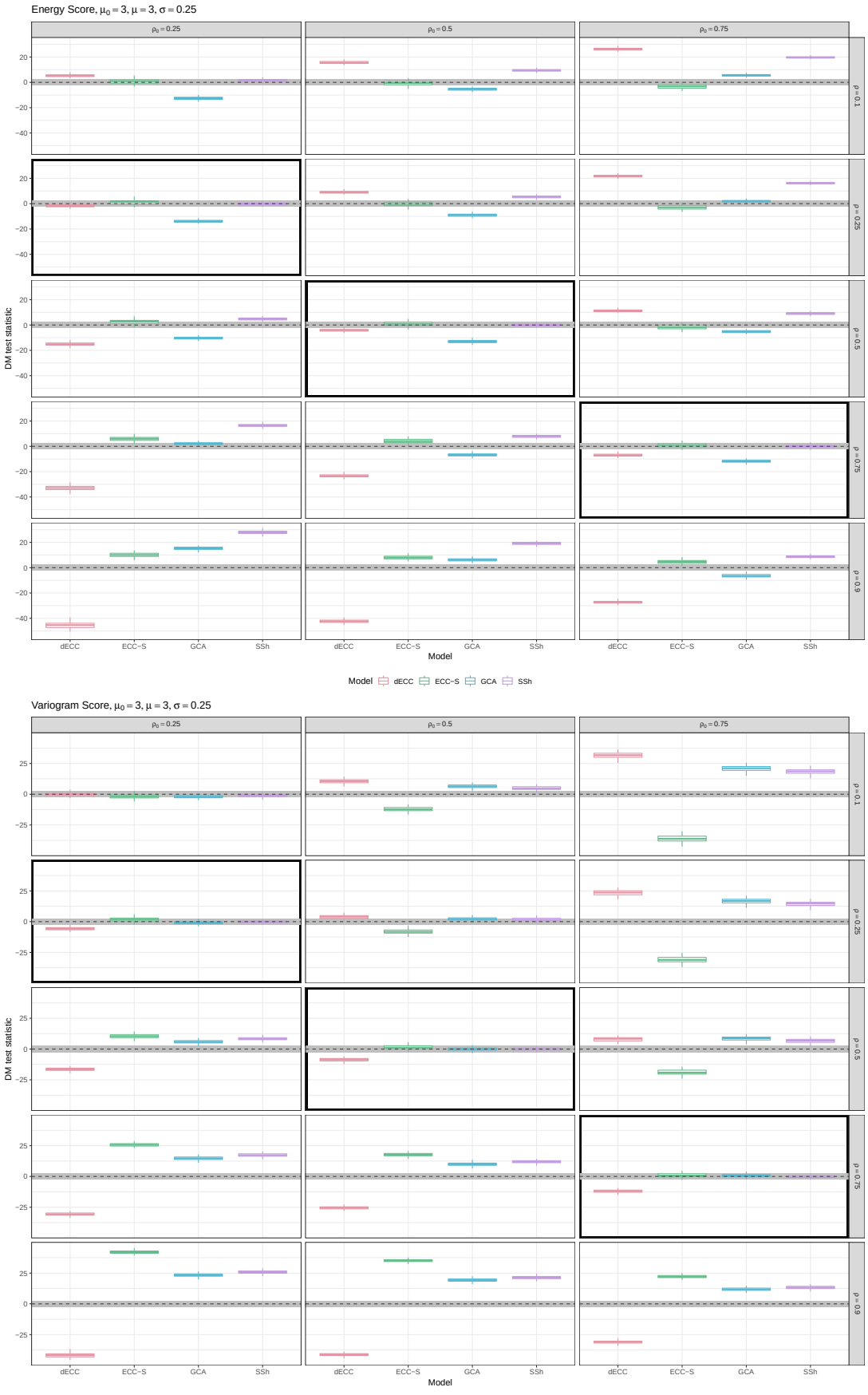


Figure 59: As Figure 39, but for $\mu_0 = 3, \epsilon = 3$, and $\sigma = 0.5$.

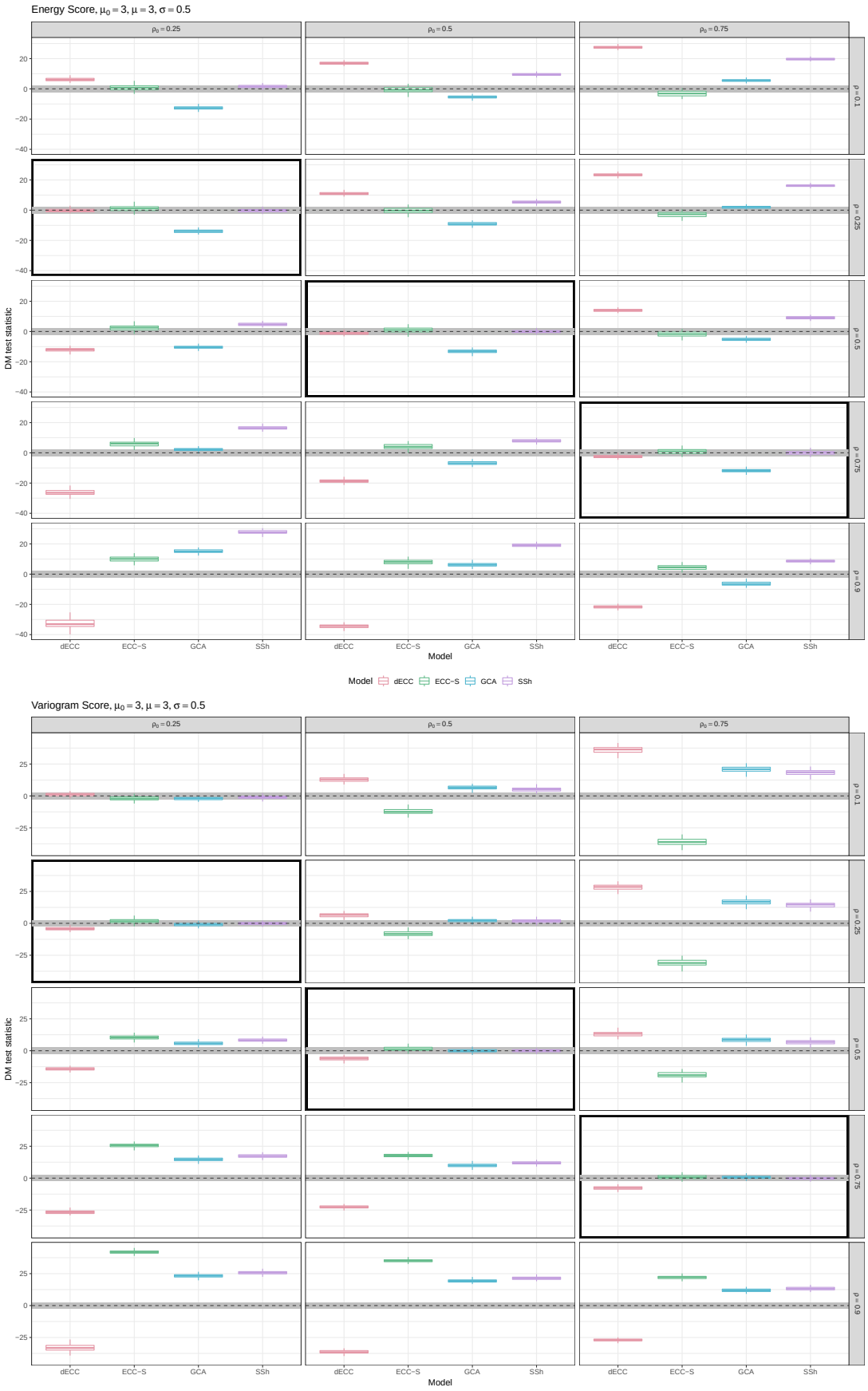


Figure 60: As Figure 39, but for $\mu_0 = 3, \epsilon = 3$, and $\sigma = 0.5$.

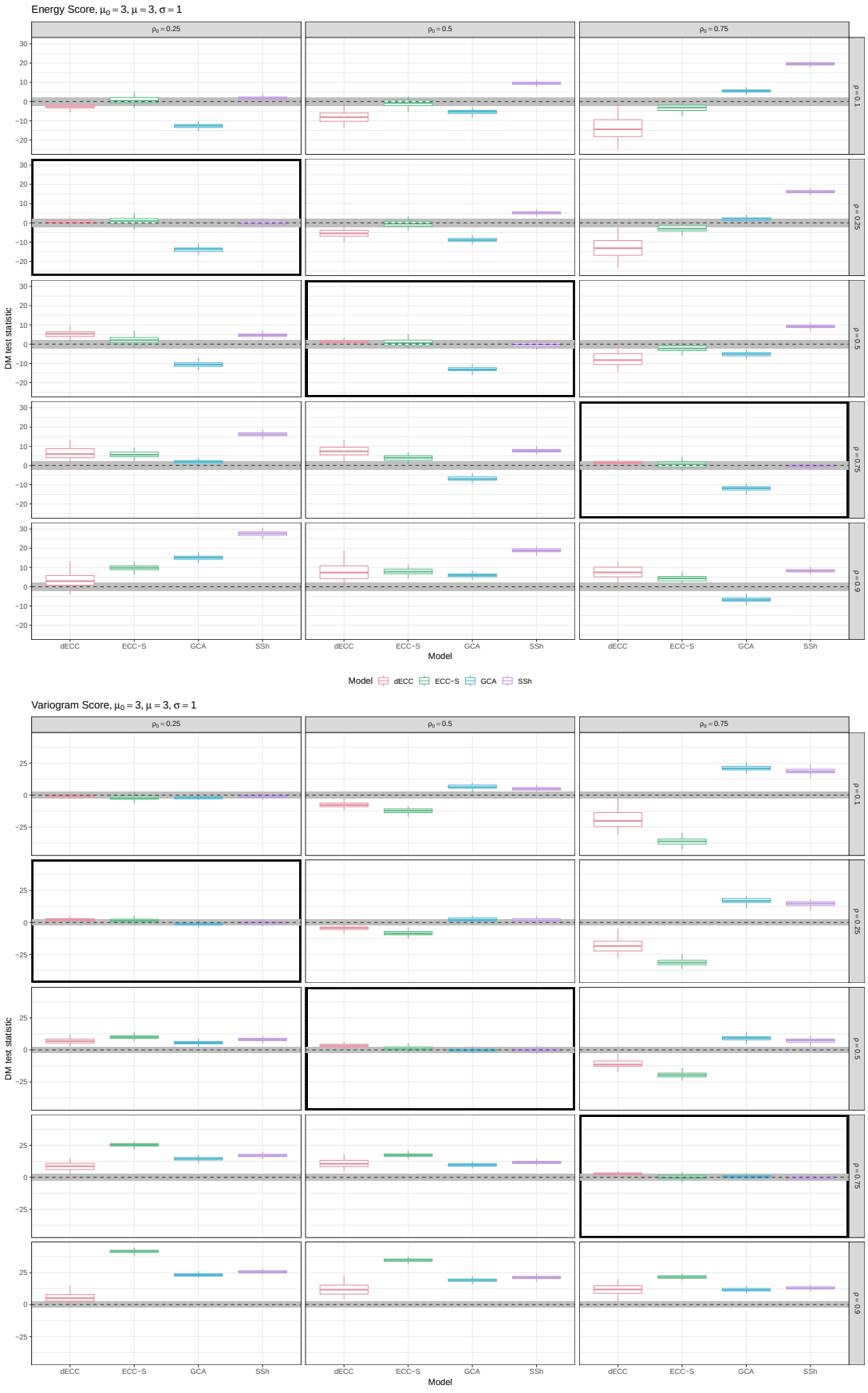
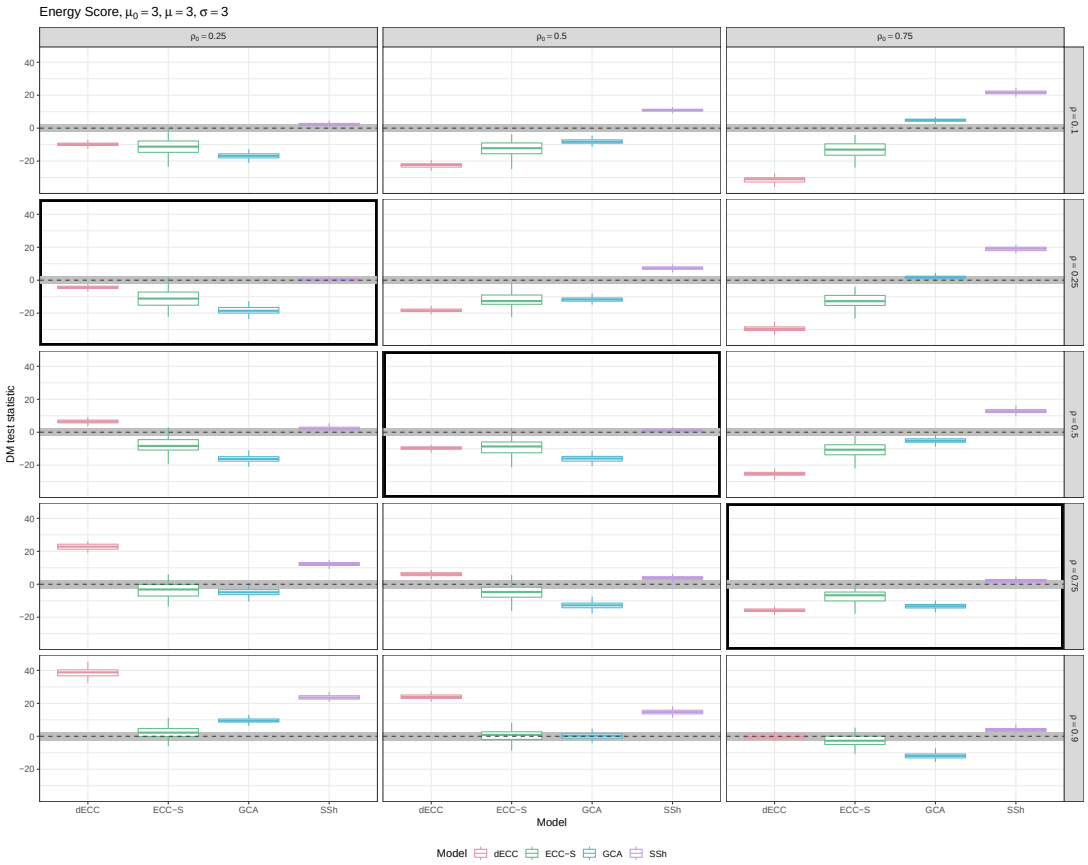


Figure 61: As Figure 39, but for $\mu_0 = 3, \epsilon = 3$, and $\sigma = 1$.



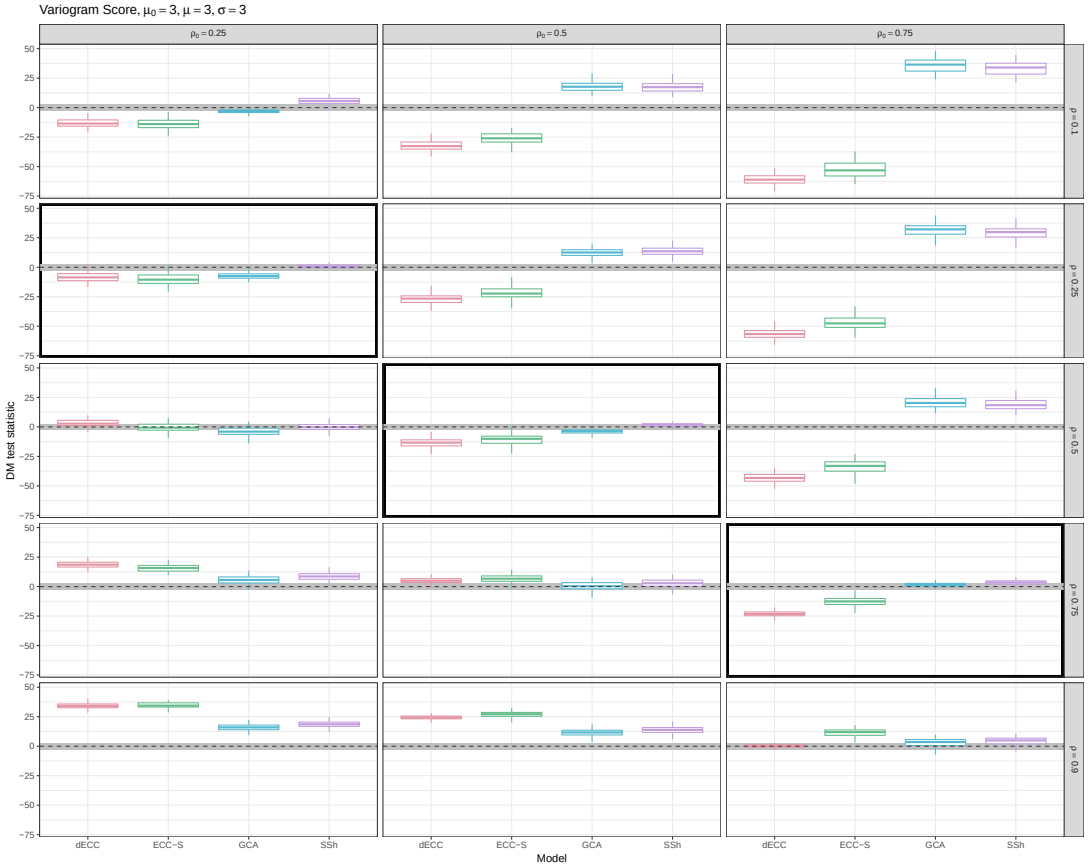


Figure 62: As Figure 39, but for $\mu_0 = 3, \epsilon = 3$, and $\sigma = 3$.

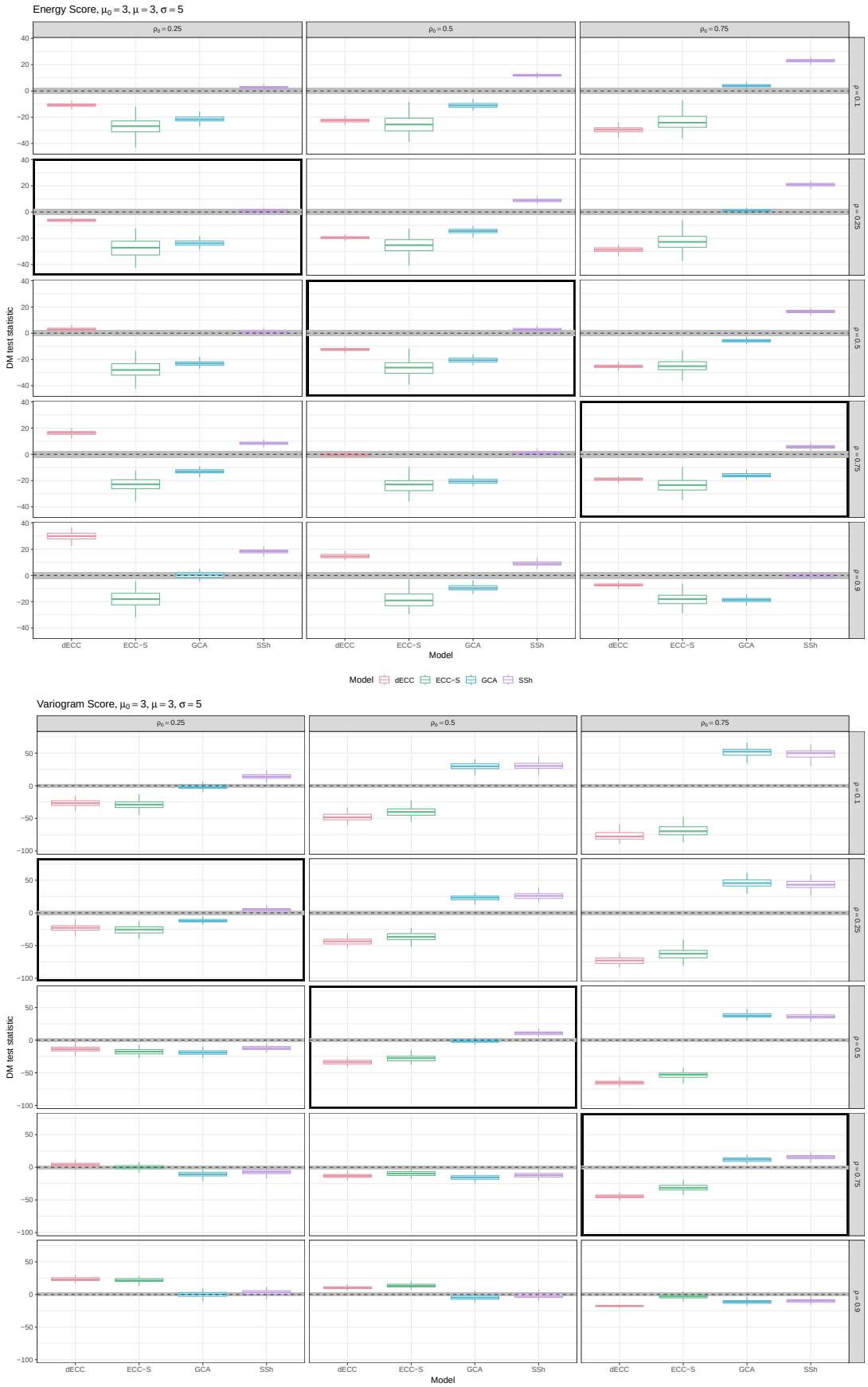


Figure 63: As Figure 39, but for $\mu_0 = 3, \epsilon = 3$, and $\sigma = 5$.

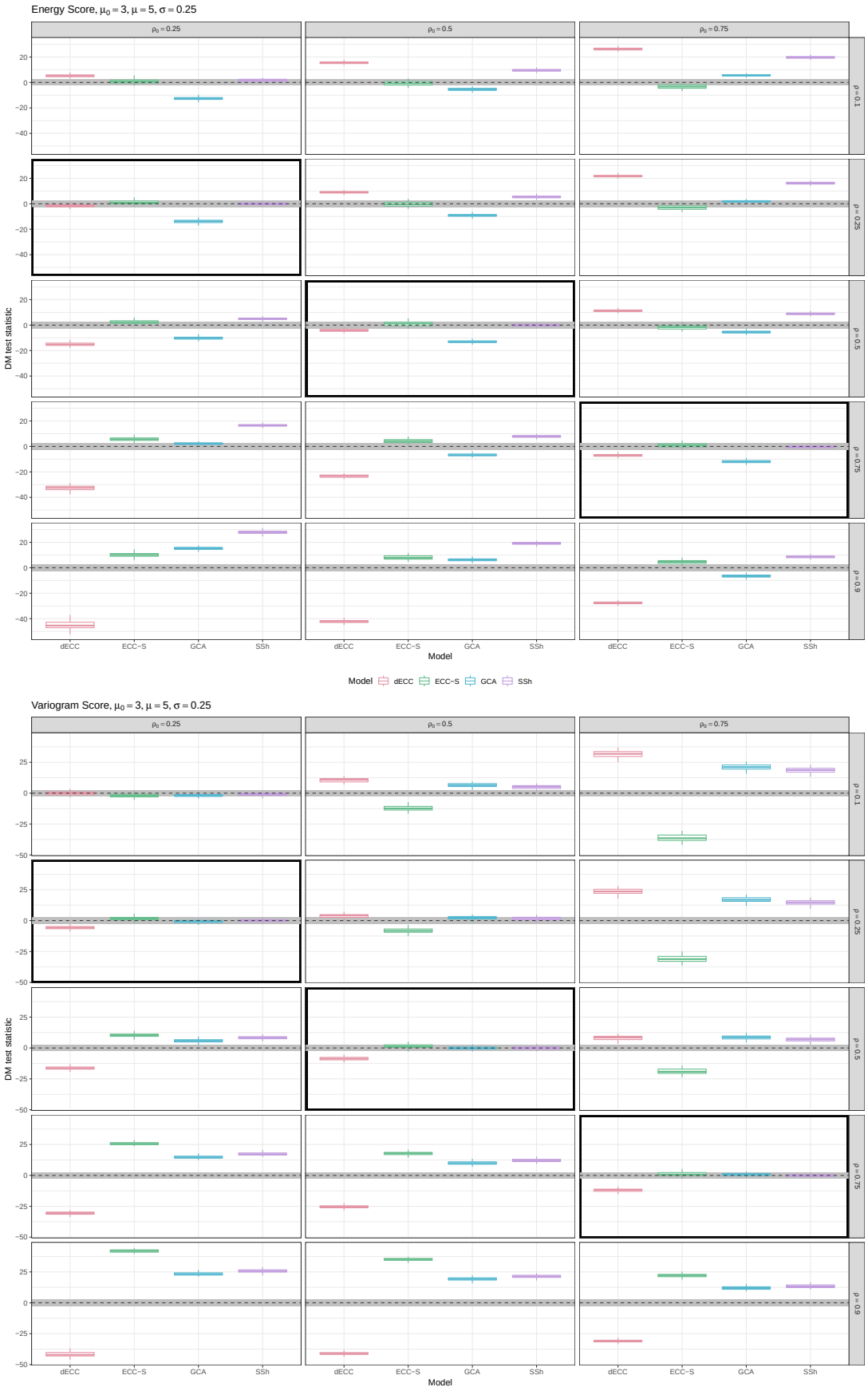


Figure 64: As Figure 39, but for $\mu_0 = 3, \epsilon = 5$, and $\sigma = 0.5$.

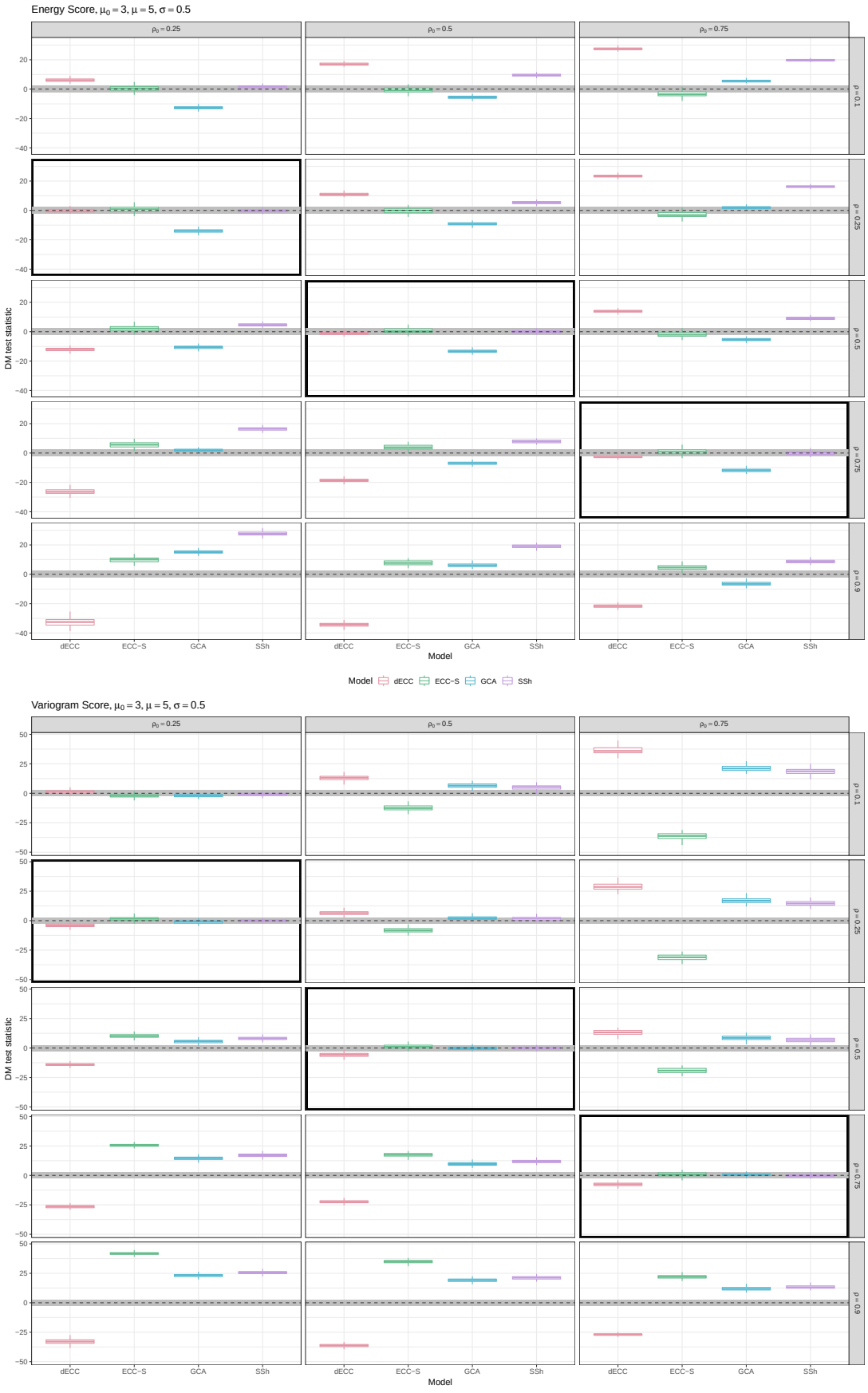


Figure 65: As Figure 39, but for $\mu_0 = 3, \epsilon = 5$, and $\sigma = 0.5$.

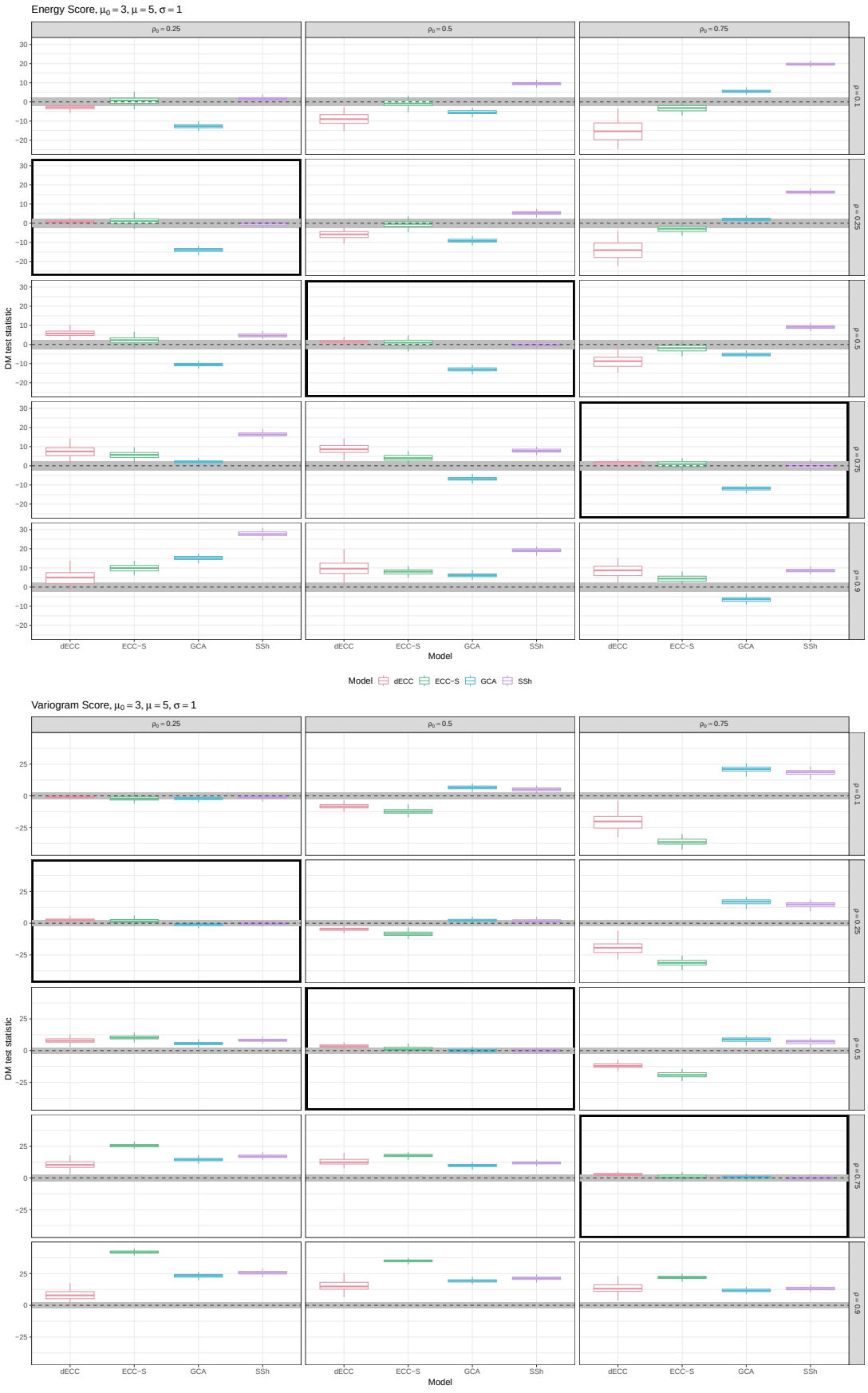


Figure 66: As Figure 39, but for $\mu_0 = 3, \epsilon = 5$, and $\sigma = 1$.

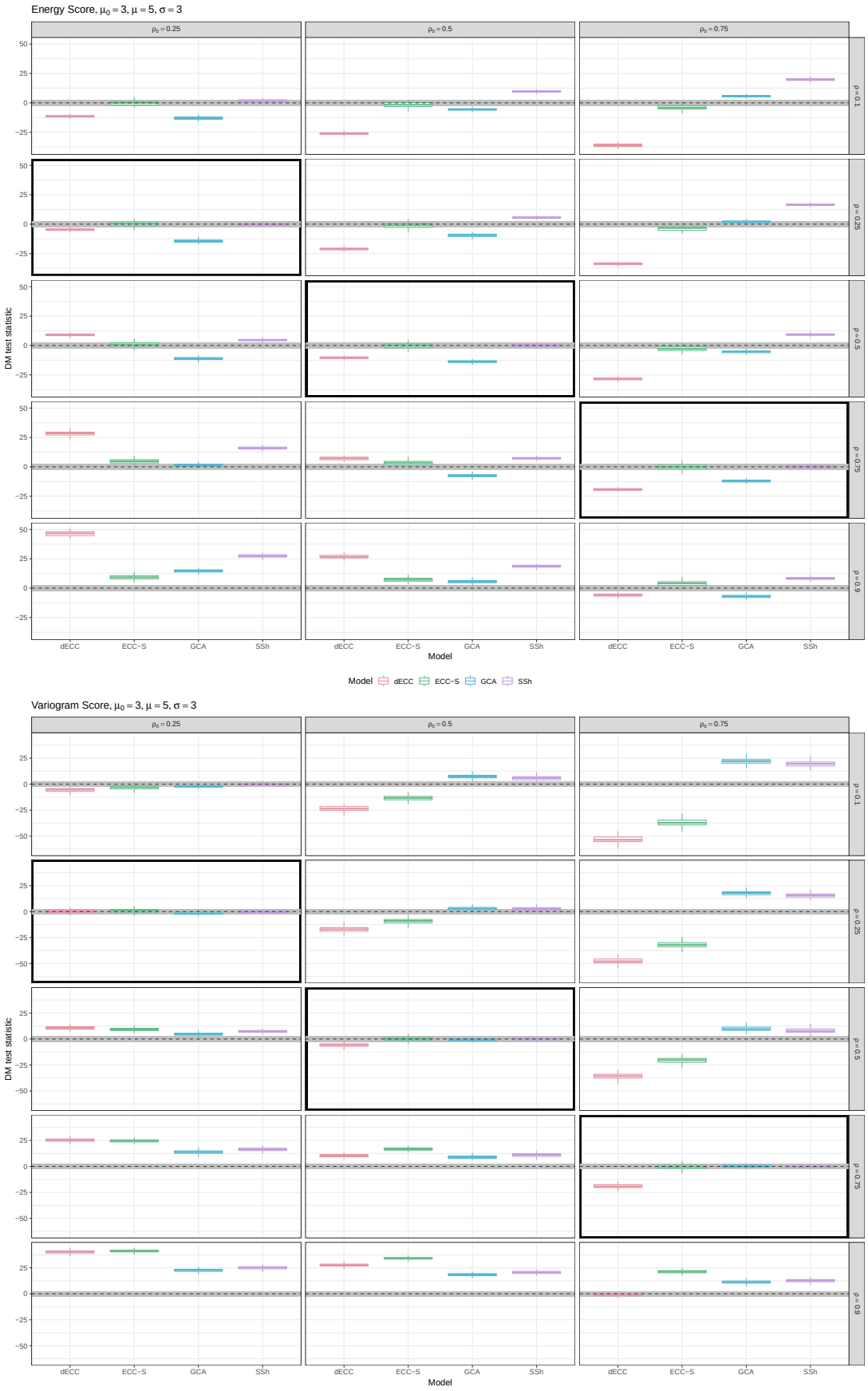


Figure 67: As Figure 39, but for $\mu_0 = 3, \epsilon = 5$, and $\sigma = 3$.

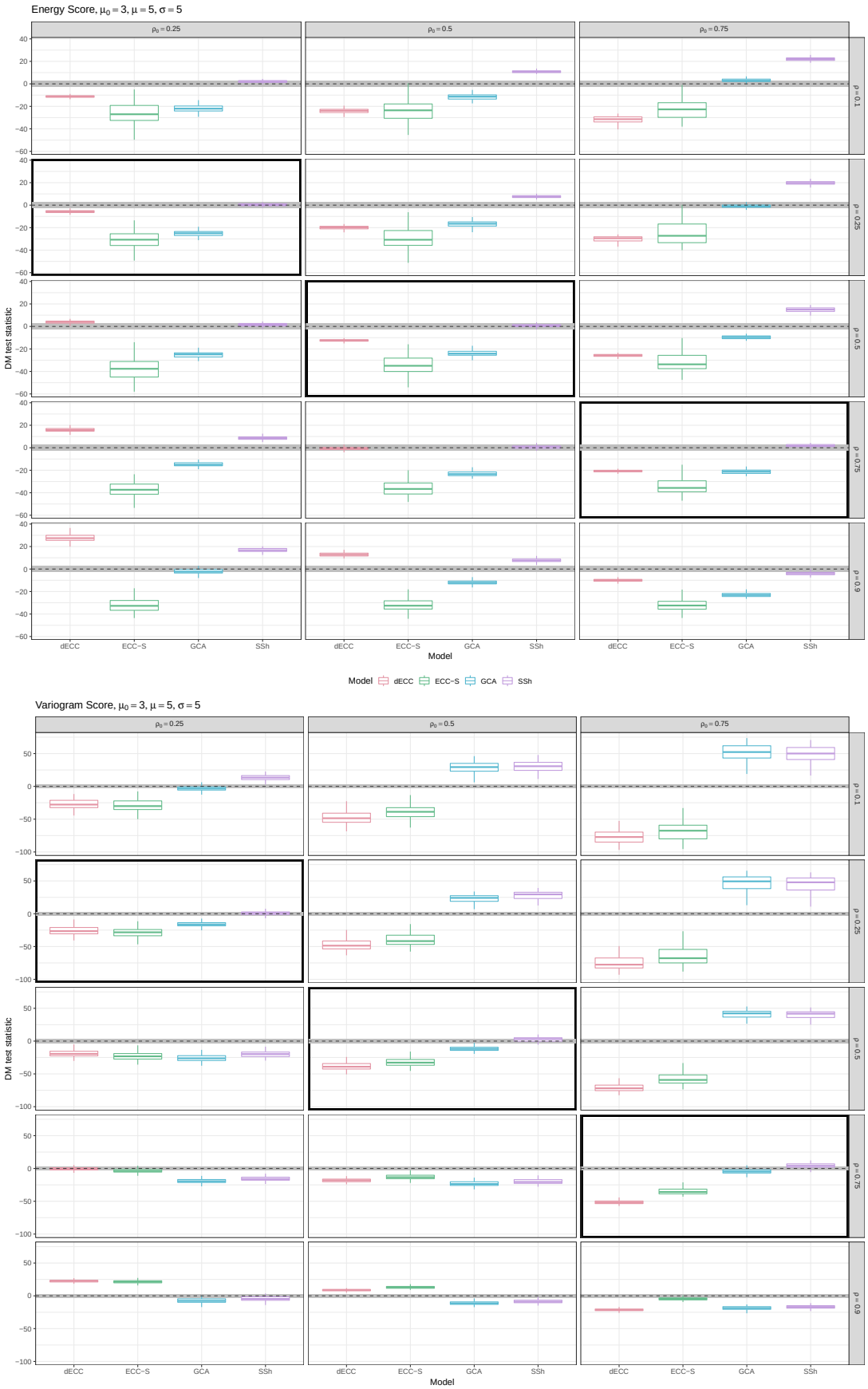


Figure 68: As Figure 39, but for $\mu_0 = 3, \epsilon = 5$, and $\sigma = 5$.

References

Hemri, S. and Klein, B.: Analog-Based Postprocessing of Navigation-Related Hydrological Ensemble Forecasts, Water Resources Research, 53, 9059–9077, https://doi.org/10.1002/2017WR020684, 2017.

Thorarinsdottir, T. L. and Gneiting, T.: Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression, Journal of the Royal Statistical Society: Series A (Statistics in Society), 173, 371–388, https://doi.org/10.1111/j.1467-985X.2009.00616.x, 2010.