# Factors affecting inferences on natural mortality and associated environmental effects in state-space age-structured assessment models

Timothy J. Miller[1,2] Gregory L. Britten[3] Elizabeth N. Brooks[2]

Gavin Fay[4] Alexander C. Hansell[2] Christopher M. Legault[2]

Brandon Muffley[5] John Wiedenmann[6]

18 December, 2025

[1]corresponding author: timothy.j.miller@noaa.gov

[2]Northeast Fisheries Science Center, Woods Hole Laboratory, 166 Water Street, Woods Hole, MA 02543 USA

[3]Biology Department, Woods Hole Oceanographic Institution, 266 Woods Hole Rd. Woods Hole, MA, USA

[4]Department of Fisheries Oceanography, School for Marine Science and Technology, University of Massachusetts Dartmouth, 836 S Rodney French Boulevard, New Bedford, MA 02744, USA

[5]Mid-Atlantic Fishery Management Council, 800 North State Street, Suite 201, Dover, DE 19901 USA

[6]Department of Ecology, Evolution, and Natural Resources. Rutgers University

# Abstract

Treatment of natural mortality is a major consideration in assessment models and state-space approaches allow estimation of temporal variation in this mortality rate as well as effects of specific covariates. However, there has been no investigation of the reliability of inferences made regarding natural mortality, associated covariate effects, and important assessment output from state-space assessment models. We conducted a large-scale simulation study that considers models fit to data simulated from operating models with alternative assumptions defined by several factors, but we focus on scenarios where there is temporal contrast in fishing pressure and lower uncertainty in population observations (age composition and indices of abundance). We fit estimating models to simulated observations with alternative assumptions on inclusion of environmental effects, estimation of the median natural mortality rate, and the source of temporal variability in the population demography. Our results suggest that estimation of environmental effects on natural mortality is possible and reliable even when the population process error source was misspecified in some scenarios with lower uncertainty in covariate observations, and higher covariate temporal variability.



**keywords:** state-space assessment models, time-varying natural mortality, bias, AIC

# Introduction

State-space population models are now used widely for fisheries stock assessment in Europe, the United States, and Canada (Nielsen and Berg, 2014; Cadigan, 2016; Pedersen and Berg, 2017; Stock and Miller, 2021). Because application of these methods are considered best practice and recommended for the next generation of stock assessment models (Hoyle et al., 2022; Punt, 2023), it is expected their use will only grow globally. An appeal of state-space models lies in their separation of sources of biological and measurement variability by treating latent population characteristics as statistical time series with periodic observations measured with error. Through advances in computational capacity, we can use sophisticated numerical approaches to estimate model parameters as mixed effects (Thorson and Minto, 2015; Kristensen et al., 2016).

State-space stock assessment models, with non-linear functions of latent processes and numerous observation types with different probability distribution assumptions represent one of most complex classes of state-space models. The literature on the effects of various factors on reliability of inferences from state-space assessment models is growing (Li et al., 2024; Miller et al., In reviewa). The importance of contrast in population size and fishing mortality ($F$) and quality of data used to fit assessment models including the state-space variety is known (Magnusson and Hilborn, 2007; Miller et al., In reviewa). Furthermore, estimation of natural mortality ($M$), and even temporal variability in $M$ is possible in many scenarios (Lee et al., 2011; Cadigan, 2016; Miller and Hyun, 2018; Miller et al., In reviewa).

The effects of temporal variation in recruitment via undefined or explicit environmental factors have been extensively investigated in both traditional assessment models and state space models (Myers, 1998; Haltuch and Punt, 2011; Johnson et al., 2016; Miller et al., 2016). Reliability of estimating environmental and spawning biomass effects on recruitment in state-space assessment models requires a combination of strong effects, good age composition data quality, contrast in the environmental covariate and lower recruitment variability (Britten et al., In review; Miller et al., In reviewa).

A critical aspect of fisheries assessment models and their use in management is short-term projections that are used to determine catch advice. While understanding drivers of recruitment is important particularly for subsequent effects on reference points, recruitment in short-term projections typically has little impact on the exploitable biomass in the first few projection years. However, assumptions for $M$ have immediate and larger effects on projected biomass because they affect the abundances at older age classes at the end of the data time series that constitute spawning biomass and catch (Brodziak et al., 2008; Stock et al., 2021).

Because of the effects of $M$ on both biological reference points and short term projections, better understanding sources of variation in $M$ would provide more accurate estimation of abundance and productivity and therefore improved management. Temporal variation in $M$ is less studied than recruitment, but its importance for explaining variability in observations has been demonstrated in state-space assessment models for Atlantic cod and yellowtail flounder (Cadigan, 2016; Stock et al., 2021). Deriso et al. (2008) also demonstrated the importance of several factors affecting $M$ for Pacific herring.

Assessment models could include temporal variation in many aspects of population dynamics or how observations are related to the population. For example the Woods Hole Assessment Model (WHAM) can include process errors treated as random effects for transition in cohorts over time (hereafter referred to as apparent survival), catchability for indices of abundance, selectivity of fishing fleets or indices, movement between regions, or in $M$ (Stock and Miller, 2021; Miller et al., In reviewb). However, misspecified temporal population process errors could lead to biased population and stock status estimation, and, therefore, poor fisheries management decisions (Legault and Palmer, 2016; Szuwalski et al., 2018). Studies of the reliability of inferences regarding the presence of temporal variability in $M$ are limited. Miller et al. (In reviewa) found AIC could accurately distinguish process errors in apparent survival, but not for those specifically due to $M$ except when uncertainty in population observations (indices, catch and age composition) was low and there was greater temporal variation in $M$. In their simulation studies looking at models with multiple sources of process error, Li et al. (2024) found including more sources of process error than existed in the operating model was a better model-building approach than excluding them a priori.

Here we conduct a simulation study with operating models (OMs) varying by degree of observation error uncertainty, sources of process error, fishing history, temporal variation in environmental covariates, and magnitude of the effect of the covariate on $M$. The simulated observations from these OMs are fitted with estimating models that make alternative assumptions for sources of process error, and whether median $M$ and covariate effects are estimated. We evaluate the effects of these factors on convergence of fitted models, whether Akaike's information criterion (AIC) can determine the correct source of process error and correct assumption about covariate effects on $M$, and the accuracy for estimators of relevant parameters and stock size and harvest rates derived from the assessment model.

# Methods

Our analyses used the Woods Hole Assessment Model (WHAM) to construct both OMs and estimation models (EMs) (Miller and Stock, 2020; Stock and Miller, 2021; Miller et al., In reviewb). The WHAM package has been used extensively to configure OMs and EMs for several other simulation studies (Stock et al., 2021; Legault et al., 2023; Li et al., 2024; Britten et al., In review; Li et al., In reviewa) and is used to assess many commercially important stocks in the Northeast U.S. (e.g., NEFSC, 2022a,b, 2024). We used version 1.0.6.9000, commit 77bbd94 to generate all results.

We completed a simulation study with 288 operating models. The factors defining the configuration of each operating model, described in detail in subsequent sections, include source of population process error (3 levels), index and catch observation uncertainty (2 levels), environmental covariate uncertainty (2 levels), temporal variation in the latent environmental covariate (4 levels), and fishing history (2 levels). We simulated 100 data sets for each operating model that included simulations of process errors.

For each simulated data set we fit a set of 12 EMs. The factors that distinguish the estimating models, also described in detail below, include source of population process error type (3 levels) whether (median) $M$ was estimated or assumed known (2 levels), and whether the effect of the environmental covariate on $M$ was estimated or not (2 levels).

The sources of population process error that were used in the OMs or assumed in the EMs were on recruitment only (R), recruitment and apparent survival (R+S), or recruitment and $M$ (R+M). We did not use the log-normal bias-correction feature for process errors or observations described by Stock and Miller (2021) for OMs and EMs (Li et al., In reviewb). Simulations were all carried out on the University of Massachusetts Green High-Performance Computing Cluster. Code for completing the simulations and summarizing results can be found at https://github.com/timjmiller/SSRTWG/ecov_study/mortality.

# Operating models

## Environmental covariate

In the WHAM model, environmental covariates are assumed to be described as state-space processes with annual observations of the true latent covariate (Miller et al., 2016; Stock and Miller, 2021). In our simulations, the latent covariate is assumed to be a stationary first order autoregressive (AR1) process

$$X_y|X_{y-1} \sim \mathrm{N}\left(\mu_E\left(1-\rho_E\right)+\rho_E X_{y-1}, \left(1-\rho_E^2\right)\sigma_E^2\right)$$

with marginal mean $\mu_E = 0$ and variance $\sigma_E^2$. The four configurations of the latent environmental covariate in the operating models assume one of two values for the marginal standard deviation $\sigma_E \in \{0.1, 0.5\}$ and for the autocorrelation parameter $\rho_E \in \{0, 0.5\}$.

The observations of the latent environmental covariate are assumed to be unbiased and Gaussian

$$x_y|X_y \sim \mathrm{N}\left(X_y, \sigma_e^2\right)$$

The standard deviation of the environmental observations in the operating models is one of two values $\sigma_e \in \{0.1, 0.5\}$. Figure S2 provides example simulations of the latent covariate and observations under the alternative configurations.

## Population

Many of the characteristics of the population biology and structure including the age classes (10 age classes (ages 1 to 10+)), time span (40 years), maturity (Figure S1, top left), growth (Figure S1, top right), time of spawning (1/4 of the year), and recruitment (Figure S1, bottom right) are identical to Miller et al. (In reviewa). The maturity at age is a logistic function with age at 50% maturity ($a_{50}$) = 2.89 and slope = 0.88 and weight at age is derived from a von Beralanffy growth function where $t_0 = 0$, $L_\infty = 85$, and $k = 0.3$, and a length-weight relationship

$$W_a = \theta_1 L_a^{\theta_2}$$

where $\theta_1 = e^{-12.1}$ and $\theta_2 = 3.2$.

The general model for $M$ in year $y$ is a log-linear function of both covariate effects $X_y$ and process errors $\varepsilon_{M,y}$ and a parameter $\beta_M$ that defines median $M$

$$\log M_y = \beta_M + \beta_E X_y + \varepsilon_{M,y}$$

where the process errors are modeled as random effects that may, in general, be autocorrelated normal random variables

$$\varepsilon_{M,y}|\varepsilon_{M,y-1} \sim \mathrm{N}\left(\varepsilon_{M,y-1}, (1-\rho_M^2)\sigma_M^2\right)$$

(Stock and Miller, 2021), but we assume $\rho_M = 0$ in our R+M OMs. We assume the median $M$ rate $e^{\beta_M} = 0.2$ is constant across ages. For R and R+S OMs and EMs, $\varepsilon_{M,y} = 0$. For all R+M OMs, we assume the same standard deviation $\sigma_M = 0.3$, which is estimated in the R+M EMs. The covariate effect is one of 3 alternative values in the operating models, $\beta_{\mathrm{E}} \in \{0, 0.25, 0.5\}$. The parameters defining the simulated covariate time series, size of the covariate effect, and any $M$ random effects result in a range of different levels of variation in annual values (Figure S3).

We assumed expected recruitment each year is from a Beverton-Holt stock-recruit relationship (SRR)

$$R_y = \frac{a\mathrm{SSB}_{y-1}}{1 + b\mathrm{SSB}_{y-1}}.$$

All biological inputs to calculations of spawning biomass per recruit (i.e., weight, maturity, and $M$ at age) are constant in the R and R+S OMs without covariate effects on $M$. Therefore, steepness and equilibrium unfished recruitment are also constant over the time period for those OMs (Miller and Brooks, 2021). As in Miller et al. (In reviewa), our assumed biological inputs and selectivity (defined below) with constant $M$ result in equilibrium $F$ that reduces spawning biomass per recruit to 40% of the unfished level is $F_{40\%} = 0.348$. With an assumed unfished recruitment of $R_0 = e^{10}$, setting $F_{\mathrm{MSY}} = F_{40\%}$ results in a steepness of 0.69 and $a = 0.60$ and $b = 2.4 \times 10^{-5}$. For R+M OMs and all OMs with covariate effects on $M$, steepness is not constant, but we used the same $a$ and $b$ parameters as other operating models which equates to a steepness and $R_0$ at the median of the time series models for $M$ and the covariate.

We also used the same two fishing scenarios as Miller et al. (In reviewa) for OMs. In the first scenario, the stock experiences overfishing at $2.5F_{\mathrm{MSY}}$ for the first 20 years followed by fishing at $F_{\mathrm{MSY}}$ for the last 20 years (denoted $2.5F_{\mathrm{MSY}} \rightarrow F_{\mathrm{MSY}}$). In the second scenario, the stock is fished at $F_{\mathrm{MSY}}$ for the entire time period (40 years). The magnitude of the overfishing assumptions is intended to reflect estimates of overfishing for Northeast US groundfish stocks from Wiedenmann et al. (2019).

We configured all R, R+S, and R+M OMs with uncorrelated random effects on recruitment with standard deviation on log(recruitment) $\sigma_R = 0.5$. This same assumption was used by Miller et al. (In reviewa) for R+M OMs and other OMs with fishery selectivity and index catchability process errors. For R+S OMs, apparent survival process errors were uncorrelated with $\sigma_{2+} = 0.3$.

## Catch and index observations

We define the generation of observations of aggregate (total combined across ages) catch and indices, and corresponding age composition identical to Miller et al. (In reviewa). There is a single fleet operating year round for catch observations with logistic selectivity for the fleet with $a_{50} = 5$ and slope = 1 (Figure S1, bottom left). Observations are generated for all 40 years of the model. There are two index time series intended to represent fishery-independent surveys occurring in the spring (0.25 way through the year) and the fall (0.75 way through the year). Catchability of both surveys are assumed to be 0.1. We assumed catch and index age composition observations are generated from a logistic-normal distribution where errors on the multivariate normal scale are independent. The standard deviation parameter is also constant across ages.

Standard deviation for log-aggregate catch was 0.1. There were two levels of observation error variance for aggregate indices and age composition for both indices and fleet catch. A low uncertainty specification assumed standard deviation of both series of log-aggregate index observations was 0.1 and the standard deviation of the logistic-normal for age composition observations was 0.3. In the high uncertainty specification the standard deviation for log-aggregate indices was 0.4 and that for the age composition observations was 1.5. For all estimating models, the standard deviation for log-aggregate observations was assumed known at the true value whereas that for the logistic-normal age composition observations was estimated.

## Estimating models

Estimating models were fit to each of 100 simulated data sets from each operating model. There were three factors defining the configuration of each estimating model: 1) whether $\beta_M$ was estimated or assumed known, 2) whether an environmental effect $\beta_E$ was estimated or not, and 3) whether the process errors were assumed on recruitment only (R), recruitment and survival (R+S), or recruitment and $M$ (R+M).

The configuration of the process errors in the estimating models generally matched the corresponding options in the operating models. For example, uncorrelated R+S was assumed for both the estimating and operating model. However, R+M EMs did not assume $M$ random effects were uncorrelated ($\rho_M$ was estimated). The environmental covariate observations were included in all estimation models, whether effects on $M$ were estimated or not, to ensure comparability of AIC. All fixed effects parameters for selectivity, catchability, fully-selected $F$, mean recruitment, initial abundance at age, and variances for logistic-normal age composition distributions were estimated. Any process error variance parameters for recruitment, survival, and $M$ were also estimated. The observation error variance of the environmental observations and aggregate catch and indices were all assumed known at the true values.

# Performance measures

## EM convergence

We measured the frequency of convergence when fitting each EM to the simulated data sets for each OM. There are various ways to assess convergence of the fit (e.g., Carvalho et al., 2021), but we defined successful convergence as the Hessian of the marginal log-likelihood being invertible and providing variance estimates for the fixed effects parameters as recommended by Miller et al. (In reviewa). We calculated 95% confidence intervals (CIs) for probability of convergence using the Clopper-Pearson exact method (Clopper and Pearson, 1934; Thulin, 2014).

## AIC for model selection

We measured the frequency of correct model selection using marginal AIC. For a given operating model the set of models that were considered all made the same assumptions on whether or not to estimate $\beta_M$ or it is assumed at the true value. For model $m$, the marginal AIC is a function of the marginal log-likelihood maximized with respect to the fixed effects in the model $\boldsymbol{\theta}$ and the number of fixed effects $n(\boldsymbol{\theta})$ estimated,

$$\mathrm{AIC}_m = -2\left[\mathrm{argmax}_{\boldsymbol{\theta}} \log L_m(\boldsymbol{\theta}) - n(\boldsymbol{\theta})\right].$$

All model fits that successfully completed the optimization were used for this set of analyses. We used all of these fits because some lack of convergence would be expected for the correct behavior of more complicated models that include process errors that did not exist in the operating model. For example R+M EMs fit to R OMs would be expected to estimate no variance in the $M$ random effects and the estimated variance parameter going to zero would cause poor convergence.

## Parameter estimation

All results here use OM simulations with fits that satisfied the convergence criterion described above. We used this conditioning to reflect how practitioners would proceed in analyses of model fits with real assessment data. That is, practitioners would ensure models converged such that Hessian-based standard errors were available for all model parameter estimates.

We focused on statistical behavior of estimators of the covariate effect on $M$ ($\widehat{\beta}_E$), the estimator of the median $M$ parameter ($\widehat{\beta}_M$), and estimators of terminal year $M$ ($\widehat{M}$), spawning stock biomass ($\widehat{\mathrm{SSB}}$), and fully-selected $F$ ($\widehat{F}$). In preliminary analyses we examined results for the estimators of all the annual values for $M$, SSB and $F$ over the whole time series, but we found no appreciable differences in patterns across the various factors defining the OMs and EMs. Furthermore, results for terminal year $F$ were generally inversely related to those for spawning stock biomass, and are provided in the Supplementary Materials and not discussed further.

We calculated median errors (ME) of $\widehat{\beta}_E$ and $\widehat{\beta}_M$, and the Hessian-based standard error estimators of these parameters ($\widehat{SE}\left(\widehat{\beta}_E\right)$ and $\widehat{SE}\left(\widehat{\beta}_M\right)$). We calculated the median relative errors (MRE) of terminal year $\widehat{M}$, $\widehat{\mathrm{SSB}}$, and $\widehat{F}$. We also calculated the root mean square error (RMSE) and estimated probability of coverage of constructed 95% CIs for $\widehat{\beta}_E$ and $\widehat{\beta}_M$ for EMs that estimated these parameters. We constructed the CIs for probabilities of CI coverage using the same methods as those for probabilities of convergence.

The true values for terminal year SSB and $F$ vary among simulations. For the $i$th simulated data the relative error for terminal year value $\theta_i$ provided from the fitted estimation model is

$$\mathrm{RE}_i\left(\theta\right) = \frac{\widehat{\theta}_i - \theta_i}{\theta_i}$$

We calculated RMSE as

$$\mathrm{RMSE}\left(\widehat{\theta}\right) = \sqrt{\frac{1}{n}\sum^{n}\left(\widehat{\theta}_i - \theta_i\right)^2}$$

For ME and MRE of estimators, we constructed 95% CIs using the binomial distribution approach as in Stock and Miller (2021) and Miller et al. (In reviewa). Because of the inverse relationship of the CI and the null hypothesis significance test, when the CI contains zero we do not have evidence of bias from the simulation study.

# Results

Miller et al. (In reviewa) found inferences are most reliable for distinguishing process error sources and stock-recruit relationships in scenarios with lower observation error for indices and age composition data and with temporal contrast in fishing pressure. Our expectation is that inferences for covariate effects on $M$ would require data to be at least that informative and inspection of the comprehensive set of results across all OMs generally confirms this (See Supplemental Materials). Therefore, we restrict our attention to these OM scenarios with more informative data (temporal contrast in fishing and lower population observation error).

## EM Convergence

In OMs with contrast in fishing pressure and lower uncertainty in population observations (catch, indices and age composition), R EMs generally converged with high frequency except when EMs estimated covariate effects on $M$ and the covariate uncertainty ($\sigma_e$) was high (Figure S4). R+S EMs generally converged with low frequency when the OM process error did not match the EM configuration. R+M EMs generally did not converge with high frequency except for R+S OMs with low covariate uncertainty and where the EMs assumed median $M$ as known. There was generally little effect of the treatment of median $M$ (known or estimated) on convergence for any of the other EMs. The largest impact of the size of the true covariate effect ($\beta_E$) on convergence was observed for R and R+M EMs that estimated covariate effects in R and R+M OMs with high uncertainty in covariate observations ($\sigma_e$) and high temporal variability in the latent covariate ($\sigma_E$). There was an increase in convergence frequency with increased temporal variability in the covariate for many EMs in OMs with large covariate uncertainty.

## AIC performance

We only present results for EMs where the median $M$ rate parameter was estimated because results differed little whether this parameter was estimated or assumed known (see Figures S15 to S17). When there was temporal contrast in fishing pressure and lower uncertainty in population observations, the only OM scenarios where AIC was consistently accurate for both covariate effect and process error configurations were R OMs with high variability in the covariate and low uncertainty in observations of the covariate (Figure S5). R+M OMs with those same covariate attributes were only accurate for the covariate effect (low Type I and II errors) and the incorrect process error almost always chosen was R rather than R+S. We also observed high accuracy for both covariate effect and process error configurations for R+S OMs where either no effect of the covariate or the strongest effect of the covariate was simulated. All R and R+S OMs were accurate for process error configuration and across all OMs, AIC accurately selects models without covariate effects when there are none (low Type I error). See Figures S15 to S17 for full results.

## Covariate effect estimation

When there was temporal contrast in fishing pressure, lower uncertainty in population observations, low uncertainty in environmental observations, and larger temporal contrast in the simulated true environmental covariate, we observed little or no bias in estimation of covariate effects ($\widehat{\beta}_E$) across all EM and OM process error assumptions, both EM configurations of median $M$, and all true covariate effect sizes (Figure S6, rows 3 and 4). The only exception to this was when R EMs with median $M$ estimated were fit to R+S OMs with those configurations. Decreasing trends in bias of $\widehat{\beta}_E$ for EMs fit to many configurations of R and R+M OMs indicate that $\widehat{\beta}_E$ values were closer to zero on average even when the true effect increased. For R+S OMs, EMs with the matching process error configuration showed little evidence of bias in estimating $\beta_E$ across a range of covariate and covariate observation configurations. R+S and R+M EMs generally performed similarly when fitted to either R+S or R+M OMs except R+S OMs with high covariate uncertainty and low covariate variability (Figure S6, rows 5 and 6 of middle columns).

Bias of Hessian-based standard error estimation for $\widehat{\beta}_E$ was also close to zero in the same OM scenarios where bias in estimation of $\beta_E$ itself was close to zero (Figure S21, rows 3 and 4). However, in these high information OMs, bias of standard error estimation was observed for several EMs that did not have the matching R+S and R+M process error configuration. Estimation of standard errors was essentially unreliable for most EMs when OMs had higher uncertainty in covariate observations (Figure S21, rows 5 to 8).

Despite the reliable estimation of covariate effects and standard errors with low uncertainty in environmental observations, and larger temporal contrast in the simulated true environmental covariate, we found poor CI coverage for many EMs fitted to R+S and R+M OMs (Figure S7, rows 3 and 4). For example, R+M EMs fit to R+M OMs showed little bias in estimation of the covariate effect and standard errors, but CI coverage was negatively biased (rows 3 and 4 in right two columns of Figures S6, S7, and S21). However, there appears to be a positive correlation of the covariate and standard error estimates such that estimates less than the true value had negatively biased standard error estimates which would make CIs too small (Figure S32). CI coverage was most reliable for R+S OMs when the EMs had the matching process error configuration (Figure S7, middle 2 columns).

## Median natural mortality rate estimation

We observed negligible bias in estimating the median $M$ parameter ($\beta_M$) over a much wider set of operating models than that for estimating the covariate effect on $M$. When there is temporal contrast in fishing pressure, lower uncertainty in population observations, all EMs fit to R and R+M OMs showed little evidence of bias for $\widehat{\beta}_M$ (Figure S8, left and right sets of columns). For R+S OMs with those conditions, only EMs with the matching process error assumption showed little evidence of bias.

We also found little or no bias of Hessian-based standard error estimation for $\beta_M$ in many of the same OM scenarios where bias of $\widehat{\beta}_M$ was close to zero except for some EMs with the process error source mis-specified or where the OM simulated the strongest covariate effect (Figure S36). Like $\widehat{\beta}_M$, reliability of standard error estimation in R+S OMs required the EMs to have the matching process error assumption.

we found CI coverage to be unreliable for many EM-OM combinations even for OM and EM combinations where bias in $\widehat{\beta}_M$ and its standard error estimators was negligible, much like that for $\widehat{\beta}_E$ (Figure S9). In the same EM OM combination we investigated above for $\widehat{\beta}_E$, we observed an opposite negative correlation of $\widehat{\beta}_M$ and its standard error estimates (Figure S43), which would result in CIs being too narrow when $\widehat{\beta}_M$ values are larger than average. The exceptions to the general poor CI coverage were for all EMs fit to several R OMs with low covariate uncertainty and covariate variability (Figure S9, first two columns) and for R+S EMs in many R+S OMs (Figure S9, middle two columns) .

## Terminal year natural morality rate

The EMs with either R or R+S process error configurations with median $M$ assumed known and covariate effects not estimated will have terminal year $M$ correctly specified when fitted to either R or R+S OMs with no covariate effect because $M$ is constant in both the EM and OM at the same value. We exclude those OM-EM combinations from our results here.

In many of the OMs with temporal contrast in fishing pressure and lower uncertainty in population observations, there was little difference in bias for terminal year $M$ among EMs whether the median $M$ parameter ($\beta_M$) was estimated (Figure S10). However, when there were differences, bias in terminal year $M$ was closer to zero when the EM assumed $\beta_M$ known, as would be expected. For R OMs, all estimating models exhibited little evidence of bias except when OMs simulated the largest covariate effect ($\beta_E = 0.5$). However, bias estimates were at worst around 10% (Figure S10, left columns). For R+S OMs, the R+S EMs generally exhibited the least evidence for bias and R+M EMs generally estimated terminal $M$ lower than the true value, particularly when observation uncertainty in the covariate was lower. For R+M OMs, many of the EM configurations provided little bias in terminal $\widehat{M}$, but the biases were generally further from zero than the results for the R OMs (Figure S10, right columns).

Accuracy (lower RMSE) of terminal year $\widehat{M}$ was similar for all of the EMs when fitted to R+M OMs (Figure S51). For R+S OMs, accuracy was generally better for R+S EM than other process error assumptions, particularly when $\beta_E$ was also estimated. Accuracy for R+S EMs was also generally as good as other EM process error assumptions even when the OM process error type was different. As would be expected, the accuracy of terminal year $\widehat{M}$ where median $M$ was assumed known was better than when it was estimated, when there were differences.

## Terminal year spawning stock biomass

In OMs with temporal contrast in fishing pressure and lower uncertainty in population observations, we found little bias in estimates of terminal year SSB for all EM configurations when fitted to R or R+M OMs (Figure S11). We also found little or no evidence of bias for R+S OMs when the EMs assumed the correct process error configuration. We found the worst bias in terminal year SSB estimation for R EMs with $\beta_M$ estimated when fitted to

R+S OMs, and R+M EMs also had more bias than R+S EMs in these OMs. Like the bias results, accuracy, as measured by RMSE, was similar for many of EMs when fit to R and R+M OMs. Accuracy was generally worse (higher RMSE) when EMs were fit to R+S OMs, and the worst accuracy occurred when the EM process error assumption did not match that of the OM (Figure S58).

# Discussion

Our simulation study demonstrated that estimation of environmental effects on $M$ is possible and reliable in certain scenarios even when the process error was misspecified (e.g., R and R+S EMs fit to R+M OM). In many of these same OMs, frequency of convergence of fitted models did not appear to suffer when covariate effects on $M$ were estimated even when there was no effect simulated. However, these scenarios are information rich in that there was contrast in population size (due to changes in fishing pressure) and the covariate affecting the population, and there was low uncertainty in population and covariate observations. Previous research has shown that estimation of a constant $M$ parameter requires contrast in time series and informative data (Lee et al., 2011), so it is no surprise that estimation of these effects also requires relatively good information via more precise observations and higher contrast in the covariate time series.

Even though estimation of covariate effects was unbiased in many scenarios, AIC could only reliably detect covariate effects for R and R+M OMs with contrast in covariates and low covariate uncertainty. In those scenarios where the covariate effect could be detected, CI coverage for the covariate effect was often biased even when there was little or no bias in the estimators of the effect and its standard error. Cadigan et al. (2024) found CI coverage to be biased for SSB and $F$ estimation in a state-space model in some scenarios, but they attributed the poor coverage to bias in Hessian-based standard error estimation, and their simulations held any process error random effects constant. The coverage bias we observed, at least for some OM-EM combinations, may be related to correlation of the estimators of the covariate effect and the corresponding standard error and therefore consideration of other methods of calculating CIs may be warranted (e.g., those based on profile likelihood and/or Monte Carlo sampling of the log-likelihood surface).

Miller et al. (In reviewa) investigated R+M OMs with two levels of $M$ process error variability ($\sigma_M \in \{0.1, 0.5\}$) and only found AIC able to accurately distinguish R+M process errors with the higher level of process error variability ($\sigma_M = 0.5$). We assumed $\sigma_M = 0.3$, intermediate to the values investigated by Miller et al. (In reviewa), and found process error inferences unreliable for the source of process error, indicating that the level of variability required for detecting $M$ process errors must be greater than $\sigma_M = 0.3$, but may still be less than 0.5. In our results and those by Miller et al. (In reviewa), AIC typically chose R EMs which would indicate the fitted R+M EMs would estimate no variability in the $M$ process errors. Future studies like ours where R+M OMs are simulated with greater variability in $M$ process errors would better inform reliability of covariate effect inferences under such scenarios.

Deriso et al. (2008) attempted to estimate process errors as well as covariate effects with $M$ for Pacific herring, but similarly found no variability in $M$, suggesting there was too much uncertainty in the available observations relative to the true temporal variability in $M$. Given that we found covariate effect inferences using R EMs was reliable in R+M OMs with apparently little variability in $M$ process errors, the findings of Deriso et al. (2008) on covariate effects for Pacific herring would presumably also be robust to true low variability in $M$. However, they did not account for uncertainty in covariate observations, some of which would probably have substantial uncertainty (e.g., competition and predation covariates). We found higher covariate observation uncertainty to cause true covariate effects to be less detectable using AIC, but we did not investigate the implications for incorrectly assuming no covariate uncertainty for covariate inferences.

Any bias or poor accuracy for annual SSB estimation was primarily a function of whether or not the median $M$ parameter was estimated or known and the types of process errors, rather than the treatment of the covariate effects on $M$. For example, we found estimation of SSB was better when the EM had the process error correctly specified for R+S OMs. Fortunately, our results and those by Miller et al. (In reviewa) demonstrate that marginal AIC seems to be a good tool for determining whether this source of process error should be included in the model. However, the reliability of the estimation of SSB does break down in the less ideal scenarios when there is higher population observation error, and lack of contrast in fishing pressure (e.g., Figures S55 to S57).

The R+S and R+M EMs both include process error for the survival of cohorts and would be expected to perform similarly, and they did in our simulations when the OM and EM matched. However, we found that the biases using R+M EMs for R+S OMs were generally worse than using R+S EMs for R+M OMs. Additionally, there are implications for biological reference points for R+M EMs (e.g., Legault and Palmer, 2016) that are not present with R+S EMs. So, we recommend following the suggestion from Li et al. (2024) to prefer R+S EMs over R+M OMs unless strong biological evidence is present to support a particular R+M OM. Such support could be found through both biological understanding (e.g., Trijoulet et al., 2020) as well as statistical properties such as a large delta AIC for R+M compared to R+S associated with greater temporal variability in natural mortality (Miller et al., In reviewa).

The ability to accurately infer covariate effects on $M$ in some realistic situations indicates that such investigations may be fruitful. Ability to make inferences could improve further when WHAM is extended to incorporate tagging data (Miller et al., In reviewb). Tagging data can greatly inform natural mortality estimation (Pollock et al., 1991; Hampton, 2000), and this impact on $M$ estimation should also apply to estimation of covariate effects or unexplained temporal variation in $M$. Given our findings and planned future WHAM development, we expect investigations of and accounting for covariate effects on $M$ to become more common within the fisheries stock assessment process. At the same time, it will be equally important to conduct research that will improve our understanding of how best to measure the depletion of stocks and determine catch advice for these stocks with covariate effects on $M$.


# Acknowledgements

This work was funded by NOAA Fisheries Northeast Fisheries Science Center.

# References

Britten, G., Brooks, E.N., and Miller, T.J. In review. Identification and performance of environmentally-driven stock-recruitment relationships in state space assessment models. For Canadian Journal of Fisheries and Aquatic Sciences, State-space assessment models special collection.

Brodziak, J., Cadrin, S.X., Legault, C.M., and Murawski, S.A. 2008. Goals and strategies for rebuilding New England groundfish stocks. Fisheries Research **94**(3): 355–366. doi: 10.1016/j.fishres.2008.03.008.

Cadigan, N.G. 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences **73**(2): 296–308. doi:10.1139/cjfas-2015-0047.

Cadigan, N.G., Albertsen, C.M., Zheng, N., and Nielsen, A. 2024. Are state-space stock assessment model confidence intervals accurate? case studies with sam and barents sea stocks. Fisheries Research **272**: 106950. doi:10.1016/j.fishres.2024.106950.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I., Wetzel, C.R., Doering, K., Johnson, K.F., and Methot, R.D. 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research **240**: 105959. doi:10.1016/j.fishres.2021.105959.

Clopper, C.J. and Pearson, E.S. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika **26**(4): 404–413. doi:10.1093/biomet/26.4.404.

Deriso, R.B., Maunder, M.N., and Pearson, W.H. 2008. Incorporating covariates into fisheries stock assessment models with application to pacific herring. Ecological Applications **18**(5): 1270–1286. doi:10.1890/07-0708.1.

Haltuch, M.A. and Punt, A.E. 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Canadian Journal of Fisheries and Aquatic Sciences **68**(5): 912–926. doi:10.1139/f2011-030.

Hampton, J. 2000. Natural mortality rates in tropical tunas: size really does matter. Canadian Journal of Fisheries and Aquatic Sciences **57**(5): 1002–1010. doi:10.1139/f99-287.

Hoyle, S.D., Maunder, M.N., Punt, A.E., Mace, P.M., Devine, J.A., and A'mar, Z.T. 2022. Preface: Developing the next generation of stock assessment software. Fisheries Research **246**: 106176.

Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., and Punt, A.E. 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? Fisheries Research **183**: 222–232. doi:10.1016/j.fishres.2016.06.004.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software **70**(5): 1–21.

Lee, H.H., Maunder, M.N., Piner, K.R., and Methot, R.D. 2011. Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twelve stock assessments. Fisheries Research **109**(1): 89–94. doi:10.1016/j.fishres.2011.01.021.

Legault, C.M. and Palmer, M.C. 2016. In what direction should the fishing mortality target change when natural mortality increases within an assessment? Canadian Journal of Fisheries and Aquatic Sciences **73**(3): 349–357. doi:10.1139/cjfas-2015-0232.

Legault, C.M., Wiedenmann, J., Deroba, J.J., Fay, G., Miller, T.J., Brooks, E.N., Bell, R.J., Langan, J.A., Cournane, J.M., Jones, A.W., and Muffley, B. 2023. Data-rich but model-resistant: an evaluation of data-limited methods to manage fisheries with failed age-based stock assessments. Canadian Journal of Fisheries and Aquatic Sciences **80**(1): 27–42. doi:10.1139/cjfas-2022-0045.

Li, C., Deroba, J.J., Miller, T.J., Legault, C.M., and Perretti, C.T. 2024. An evaluation of common stock assessment diagnostic tools for choosing among state-space models with multiple random effects processes. Fisheries Research **273**: 106968. doi:10.1016/j.fishres.2024.106968.

Li, C., Deroba, J.J., Berger, A.M., Goethel, D.R., Langseth, B.J., Schueller, A.M., and Miller, T.J. In reviewa. Random effects on numbers-at-age transitions implictly account for movement dynamics and improve performance within a state-space stock assessment. Canadian Journal of Fisheries and Aquatic Sciences .

Li, C., Deroba, J.J., Miller, T.J., Legault, C.M., and Perretti, C. In reviewb. Guidance on bias-correction of log-normal random effects and observations in state-space assessment models. For Canadian Journal of Fisheries and Aquatic Sciences, State-space assessment models special collection.

Magnusson, A. and Hilborn, R. 2007. What makes fisheries data informative? Fish and Fisheries **8**(4): 337–358. doi:https://doi.org/10.1111/j.1467-2979.2007.00258.x.

Miller, T.J. and Brooks, E.N. 2021. Steepness is a slippery slope. Fish and Fisheries **22**(3): 634–645. doi:10.1111/faf.12534.

Miller, T.J. and Hyun, S.Y. 2018. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. Canadian Journal of Fisheries and Aquatic Sciences **75**(5): 691–703. doi:10.1139/cjfas-2017-0035.

Miller, T.J. and Stock, B.C. 2020. The Woods Hole Assessment Model (WHAM).

Miller, T.J., Hare, J.A., and Alade, L. 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to Southern New England yellowtail flounder. Canadian Journal of Fisheries and Aquatic Sciences **73**(8): 1261–1270. doi:10.1139/cjfas-2015-0339.

Miller, T.J., Britten, G., Brooks, E.N., Fay, G., Hansell, A., Legault, C.M., Li, C., Muffley, B., Stock, B.C., and Wiedenmann, J. In reviewa. An investigation of factors affecting inferences from and reliability of state-space age-structured assessment models. For Canadian Journal of Fisheries and Aquatic Sciences, State-space assessment models special collection.

Miller, T.J., Curti, K.L., and Hansell, A.C. In reviewb. Space for WHAM: a multi-region, multi-stock generalization of the Woods Hole Assessment Model with an application to black sea bass. For Canadian Journal of Fisheries and Aquatic Sciences, State-space assessment models special collection.

Myers, R.A. 1998. When do environment–recruitment correlations work? Reviews in Fish Biology and Fisheries **8**(3): 229–249. doi:10.1023/A:1008828730759.

NEFSC 2022a. Final report of the haddock research track assessment working group. Available at https://s3.us-east-1.amazonaws.com/nefmc.org/14b_EGB_Research_Track_Haddock_WG_Report_DRAFT.pdf.

NEFSC 2022b. Report of the american plaice research track working group. Available at https://s3.us-east-1.amazonaws.com/nefmc.org/2_American-Plaice-WG-Report.pdf.

NEFSC 2024. Butterfish research track assessment report. US Dept Commer Northeast Fish Sci Cent Ref Doc. 24-03; 191 p.

Nielsen, A. and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research **158**: 96–101. doi:10.1016/j.fishres.2014.01.014.

Pedersen, M.W. and Berg, C.W. 2017. A stochastic surplus production model in continuous time. Fish and Fisheries **18**(2): 226–243. doi:10.1111/faf.12174.

Pollock, K., Hoenig, J., and Jones, C. 1991. Estimation of fishing and natural mortality when a tagging study is combined with a creel survey or port sampling. American Fisheries Society Symposium **12**: 423–434.

Punt, A.E. 2023. Those who fail to learn from history are condemned to repeat it: A perspective on current stock assessment good practices and the consequences of not following them. Fisheries Research **261**: 106642.

Stock, B.C. and Miller, T.J. 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research **240**: 105967. doi:10.1016/j.fishres.2021.105967.

Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., and Nye, J.A. 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. Fisheries Research **237**: 105873. doi:10.1016/j.fishres.2021.105873.

Szuwalski, C.S., Ianelli, J.N., and Punt, A.E. 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES Journal of Marine Science **75**(2): 596–609. doi:10.1093/icesjms/fsx159.

Thorson, J.T. and Minto, C. 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES Journal of Marine Science **72**(5): 1245–1256. doi:10.1093/icesjms/fsu213.

Thulin, M. 2014. The cost of using exact confidence intervals for a binomial proportion. Electronic Journal of Statistics **8**(1): 817–840. doi:10.1214/14-EJS909.

Trijoulet, V., Fay, G., and Miller, T.J. 2020. Performance of a state-space multispecies model: What are the consequences of ignoring predation and process errors in stock assessments? Journal of Applied Ecology **57**(1): 121–135. doi:10.1111/1365-2664.13515.

Wiedenmann, J., Free, C.M., and Jensen, O.P. 2019. Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks: A case study using northeast U.S. fish stocks. Fisheries Research **209**(1): 129–142. doi:10.1016/j.fishres.2018.09.018.

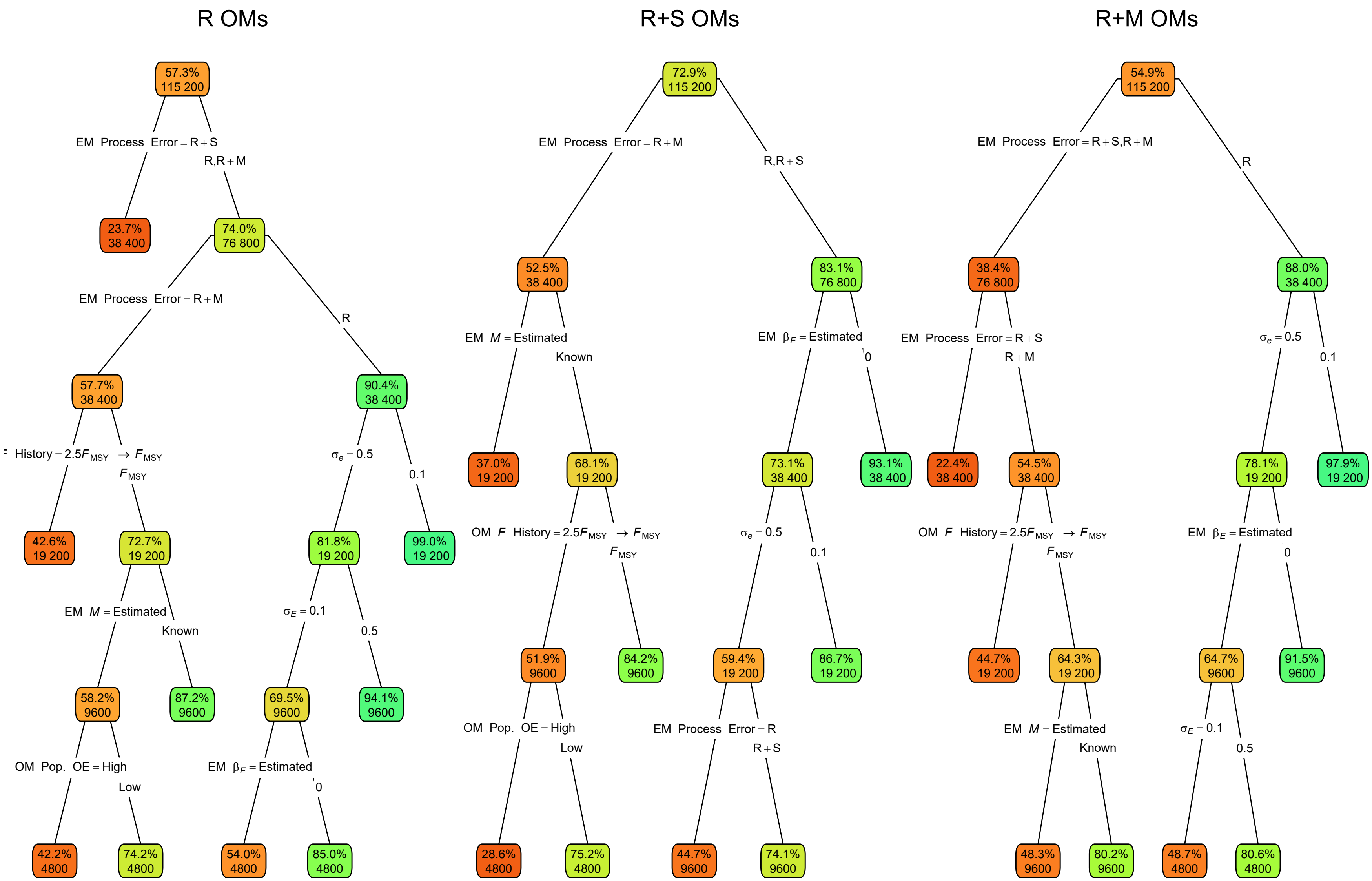



Fig. 1. Classification trees indicating primary factors determining convergence as defined by providing Hessian-based standard errors for R, R+S, and R+M OMs. Nodes denote percent convergence (top) and number of fits (bottom) for the corresponding subset. Lower or higher convergence rates are indicated by more red or green polygons, respectively

# Supplemental Materials

Table 1. For each OM process error source (columns), percent reduction in deviance for logistic regression models fit to indicators of convergence (providing Hessian-based standard errors) with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

| Factor | R | R+S | R+M |
|---|---|---|---|
| OM $F$ history | 0.86 | <0.01 | 0.24 |
| OM Obs. Error | 0.56 | 0.03 | 0.26 |
| OM $\sigma_e$ | 0.66 | 2.99 | 0.93 |
| OM $\sigma_E$ | 0.53 | 2.22 | 0.62 |
| OM $\rho_E$ | <0.01 | 0.02 | <0.01 |
| OM $\beta_E$ | 0.02 | <0.01 | <0.01 |
| EM Process Error | 24.53 | 9.33 | 23.06 |
| EM $\beta_E$ assumption | 0.16 | 2.96 | 0.51 |
| EM $M$ Assumption | 0.18 | 2.83 | 0.40 |
| All factors | 28.92 | 22.67 | 27.34 |
| + All Two Way | 36.15 | 33.54 | 34.03 |
| + All Three Way | 37.95 | 36.58 | 35.57 |

Table 2. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. **??** for the median natural mortality rate parameter with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

| Factor | R | R+S | R+M |
|---|---|---|---|
| OM $F$ history | 4.03 | 16.52 | 6.51 |
| OM Obs. Error | 2.74 | 0.54 | 0.64 |
| OM $\sigma_e$ | 0.01 | $<$0.01 | 0.01 |
| OM $\sigma_E$ | 0.18 | 0.02 | 0.15 |
| OM $\rho_E$ | 0.07 | 0.05 | $<$0.01 |
| OM $\beta_E$ | 0.37 | 0.04 | 0.10 |
| EM Process Error | 2.31 | 11.99 | 2.65 |
| EM $\beta_E$ assumption | 2.83 | 7.20 | 3.25 |
| All factors | 12.14 | 33.71 | 12.81 |
| + All Two Way | 21.74 | 46.46 | 19.84 |
| + All Three Way | 26.38 | 49.41 | 22.24 |

Table 3. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. **??** for the the standard error of the MLE for median natural mortality rate with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

| Factor | R | R+S | R+M |
|---|---|---|---|
| OM $F$ history | 0.55 | 12.91 | 3.65 |
| OM Obs. Error | 0.98 | 0.01 | 0.07 |
| OM $\sigma_e$ | <0.01 | 0.03 | 0.11 |
| OM $\sigma_E$ | 0.13 | 0.03 | 0.19 |
| OM $\rho_E$ | <0.01 | 0.02 | <0.01 |
| OM $\beta_E$ | 0.12 | 0.07 | 0.23 |
| EM Process Error | 3.16 | 8.79 | 6.83 |
| EM $\beta_E$ assumption | 0.18 | 8.96 | 3.81 |
| All factors | 5.32 | 28.39 | 14.71 |
| + All Two Way | 11.80 | 44.67 | 22.70 |
| + All Three Way | 17.83 | 51.54 | 29.46 |

Table 4. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. **??** for the covariate effect on natural mortality ($\beta_E$) with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

| Factor | R | R+S | R+M |
|---|---|---|---|
| OM $F$ history | 0.03 | 0.06 | 0.02 |
| OM Obs. Error | <0.01 | 0.06 | 0.02 |
| OM $\sigma_e$ | 0.04 | 0.08 | 0.06 |
| OM $\sigma_E$ | <0.01 | 0.02 | 0.01 |
| OM $\rho_E$ | <0.01 | 0.01 | <0.01 |
| OM $\beta_E$ | 0.05 | <0.01 | <0.01 |
| EM Process Error | 0.01 | 0.06 | 0.01 |
| EM $M$ assumption | 0.02 | 0.02 | <0.01 |
| All factors | 0.18 | 0.37 | 0.14 |
| + All Two Way | 0.50 | 1.11 | 0.35 |
| + All Three Way | 1.38 | 1.88 | 0.66 |

Table 5. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. **??** for the terminal year SSB with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

| Factor | R | R+S | R+M |
|---|---|---|---|
| Convergence | 0.02 | 0.13 | <0.01 |
| OM $F$ history | 2.08 | 2.74 | 1.75 |
| OM Obs. Error | 1.09 | 0.18 | 1.19 |
| $OM\sigma_e$ | 0.01 | <0.01 | <0.01 |
| $OM\sigma_E$ | <0.01 | <0.01 | <0.01 |
| $OM\rho_E$ | <0.01 | <0.01 | <0.01 |
| OM $\beta_E$ | <0.01 | 0.01 | <0.01 |
| EM Process Error | 0.52 | 1.05 | 0.57 |
| $EM\beta_E$ assumption | <0.01 | 0.05 | 0.01 |
| EM $M$ assumption | 1.76 | 2.12 | 1.51 |
| All factors | 5.97 | 6.58 | 5.21 |
| + All Two Way | 12.81 | 13.00 | 11.97 |
| + All Three Way | 16.41 | 15.57 | 15.73 |

Table 6. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. **??** for the median natural mortality rate parameter with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

| Factor | R | R+S | R+M |
|---|---|---|---|
| OM $F$ history | <0.01 | <0.01 | <0.01 |
| OM Obs. Error | <0.01 | <0.01 | <0.01 |
| OM $\sigma_e$ | <0.01 | <0.01 | <0.01 |
| OM $\sigma_E$ | <0.01 | <0.01 | <0.01 |
| OM $\rho_E$ | <0.01 | <0.01 | <0.01 |
| OM $\beta_E$ | <0.01 | <0.01 | 0.01 |
| EM Convergence | <0.01 | 0.01 | 0.01 |
| EM Process Error | <0.01 | 0.01 | 0.01 |
| EM $\beta_E$ assumption | <0.01 | <0.01 | <0.01 |
| All factors | 0.02 | 0.02 | 0.05 |
| + All Two Way | 0.15 | 0.15 | 0.24 |
| + All Three Way | 0.57 | 0.54 | 0.73 |

Table 7. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. **??** for the covariate effect on natural mortality ($\beta_E$) with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

| Factor | R | R+S | R+M |
|---|---|---|---|
| OM $F$ history | <0.01 | 0.01 | <0.01 |
| OM Obs. Error | <0.01 | <0.01 | <0.01 |
| OM $\sigma_e$ | <0.01 | <0.01 | <0.01 |
| OM $\sigma_E$ | <0.01 | <0.01 | <0.01 |
| OM $\rho_E$ | <0.01 | <0.01 | <0.01 |
| OM $\beta_E$ | <0.01 | 0.01 | <0.01 |
| EM Convergence | <0.01 | <0.01 | <0.01 |
| EM Process Error | <0.01 | 0.01 | 0.01 |
| EM $M$ assumption | <0.01 | <0.01 | <0.01 |
| All factors | 0.02 | 0.03 | 0.03 |
| + All Two Way | 0.16 | 0.20 | 0.18 |
| + All Three Way | 0.58 | 0.73 | 0.59 |

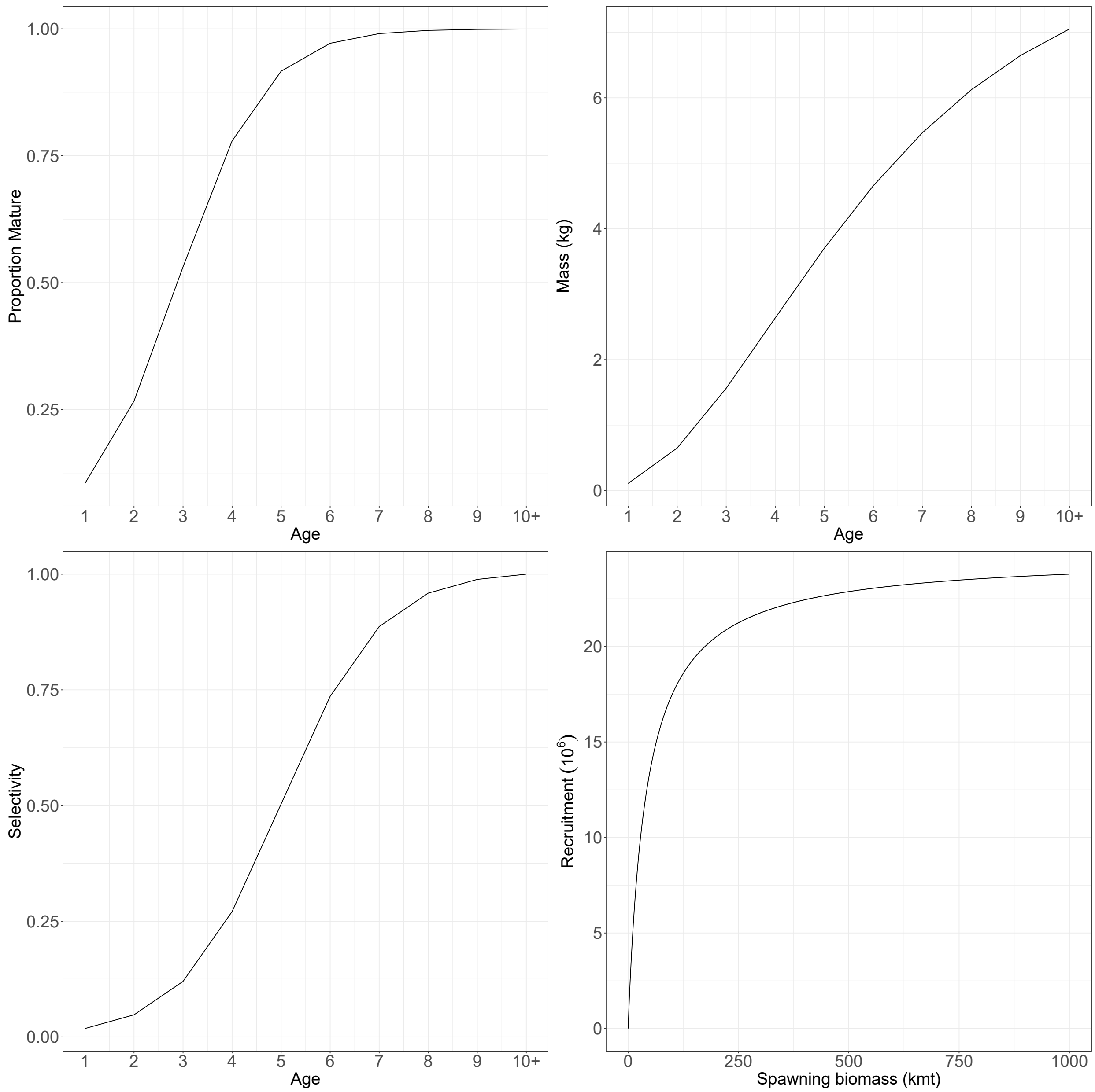



Fig. S1. The proportion mature at age, weight at age, fleet and index selectivity at age, and Beverton-Holt stock-recruit relationship assumed for the population in all operating models. For operating models with random effects on fleet selectivity, this represents the selectivity at the mean of the random effects.

Fig. S2. Example simulations of environmental covariate latent processes and observations with different levels of observation error, and different assumptions about variability of the latent process.

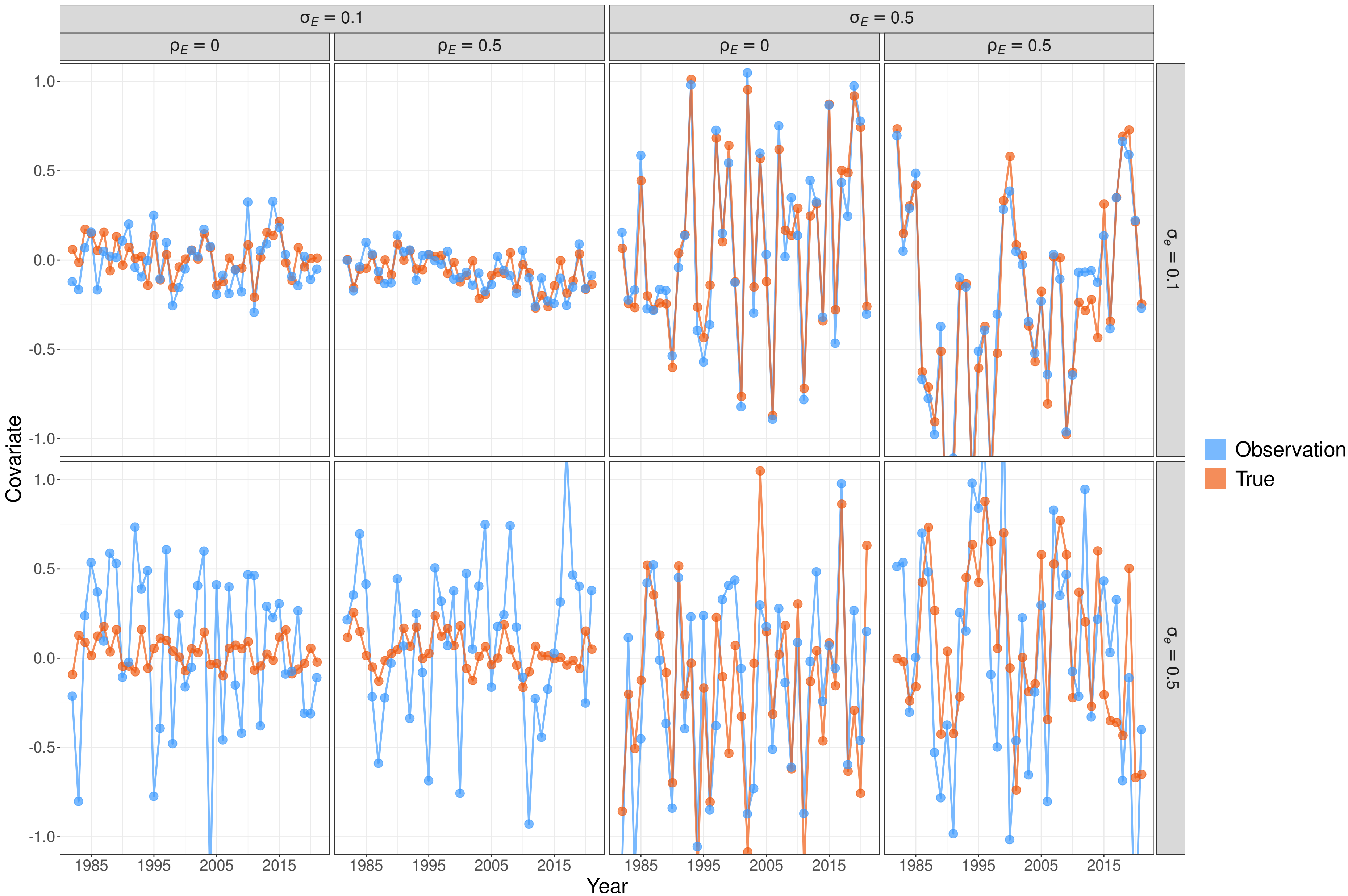
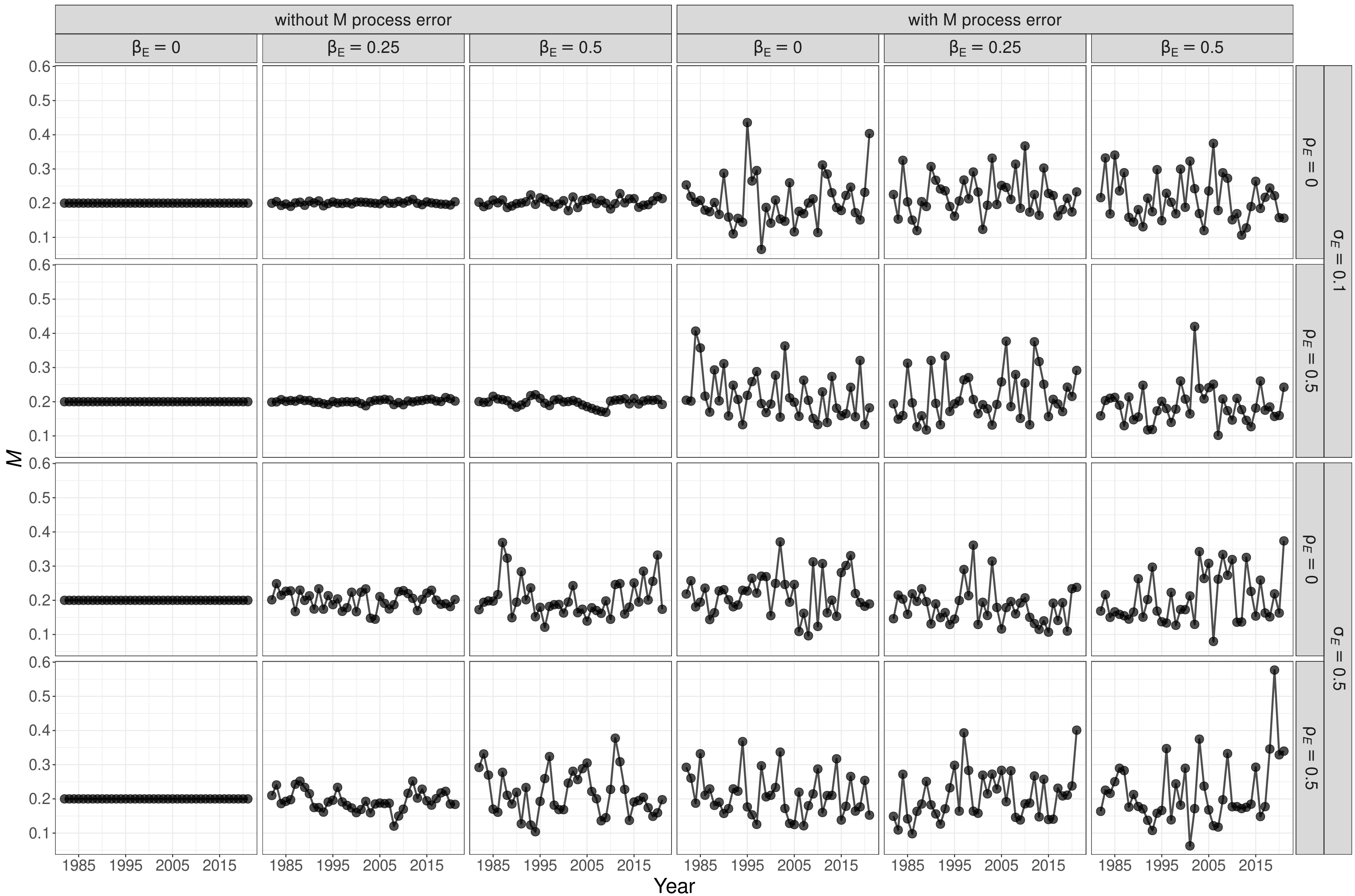
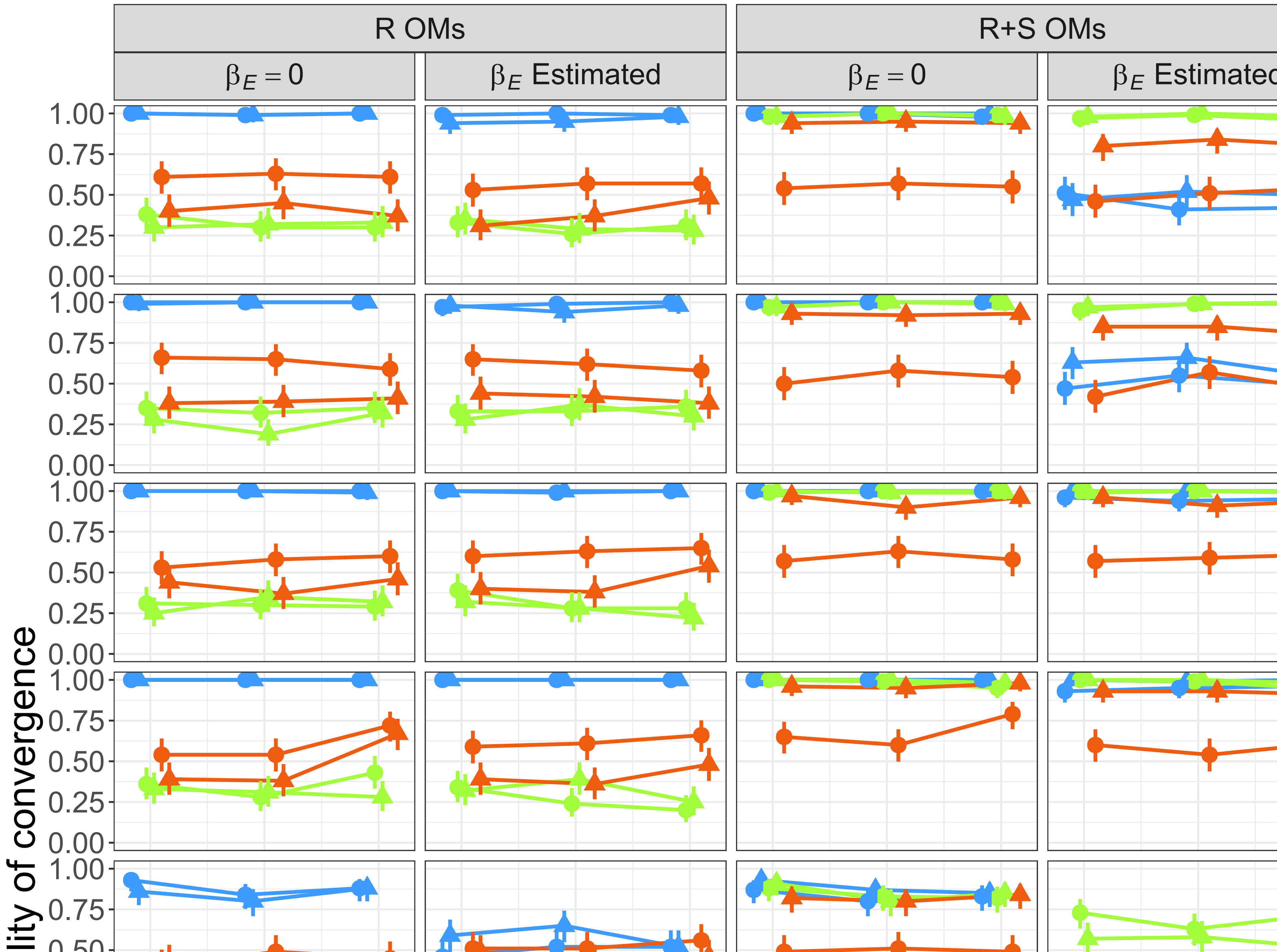


Fig. S3. Example simulations of annual natural mortality rates that may be a function of a temporally varying environmental covariate and autoregressive random effects.




R OMs
R+S OMs
$\beta_E = 0$
$\beta_E$ Estimated
$\beta_E = 0$
$\beta_E$ Estimatec
1.00
0.75
0.50
0.25
0.00
lity of convergence

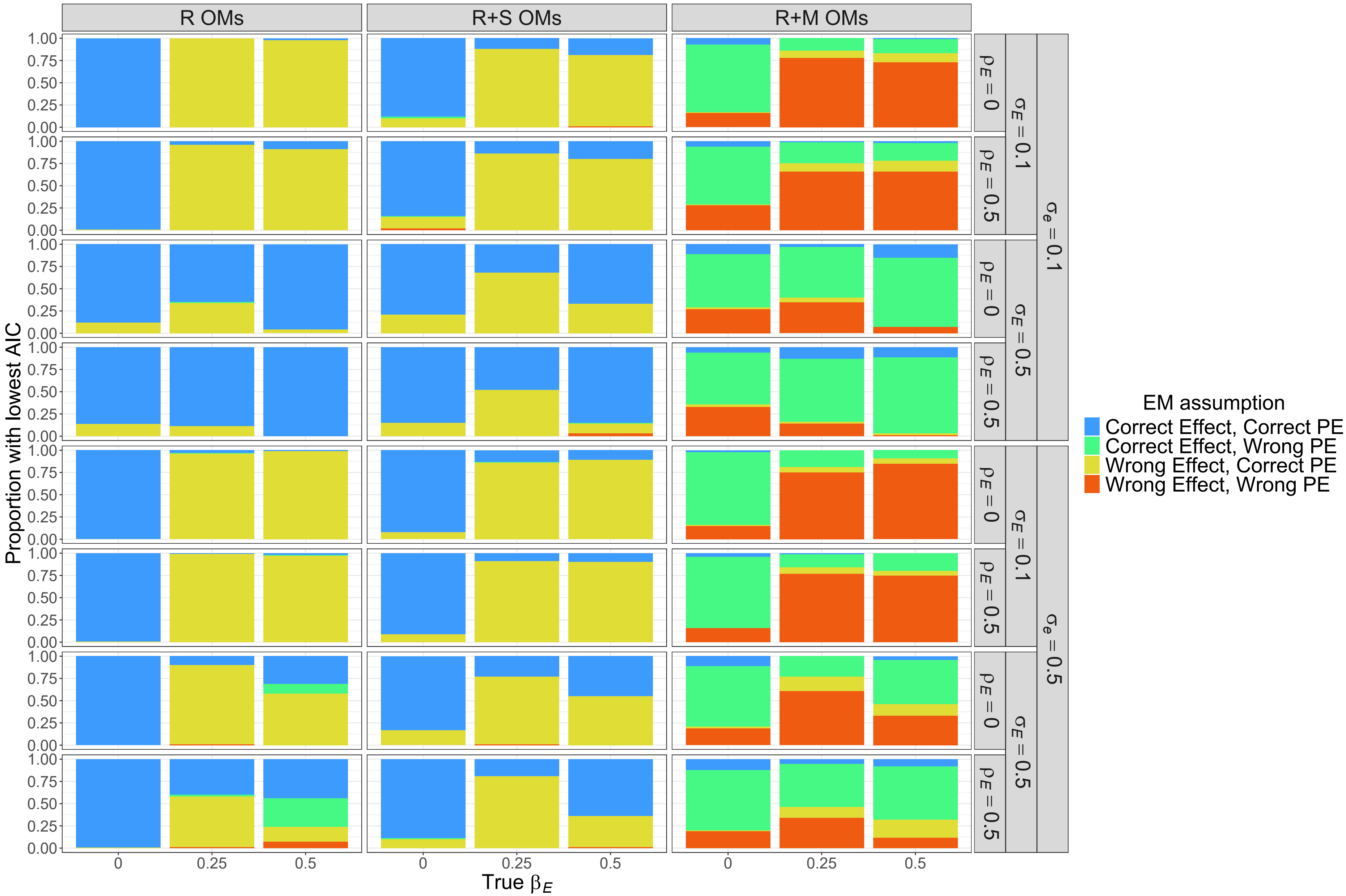



Fig. S5. For each OM, the proportion of simulated data sets where the EM type (treatment of environmental covariate effect and assumed process error type) had the lowest AIC. All OMs had low observation error for fish population observations and temporal contrast in fishing pressure. All EMs estimated median natural mortality rate.

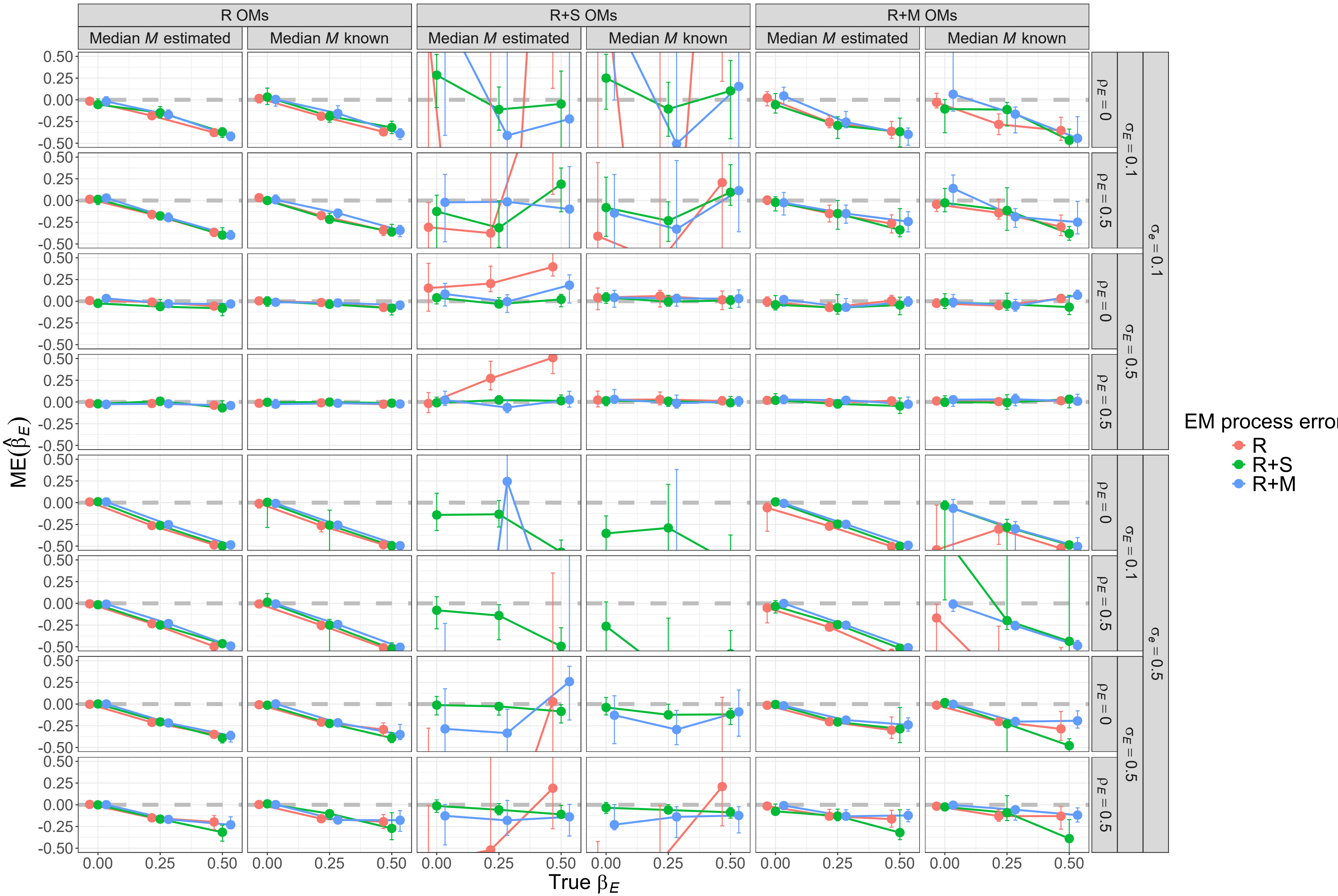



Fig. S6. Median error (ME) of estimates of environmental effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($\beta_M$ known or estimated). All OMs had low observation error and contrast in fishing mortality. Vertical lines represent 95% confidence intervals.

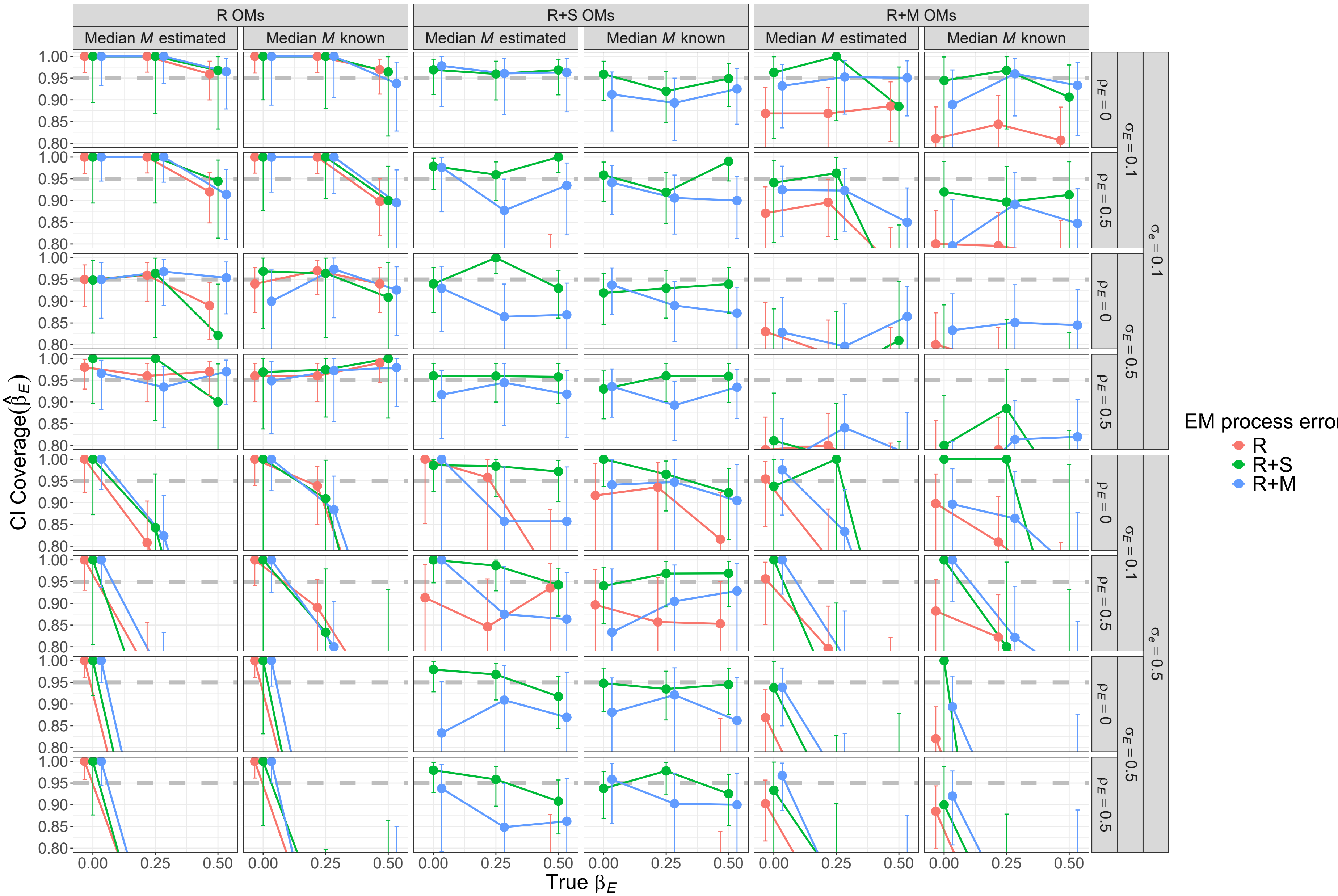



Fig. S7. Probability of 95% confidence interval for $\beta_E$ containing the true value for EMs with alternative process error assumptions and treatment of median natural mortality ($\beta_M$ known or estimated). All OMs had low observation error and contrast in fishing mortality. Vertical lines represent 95% confidence intervals.

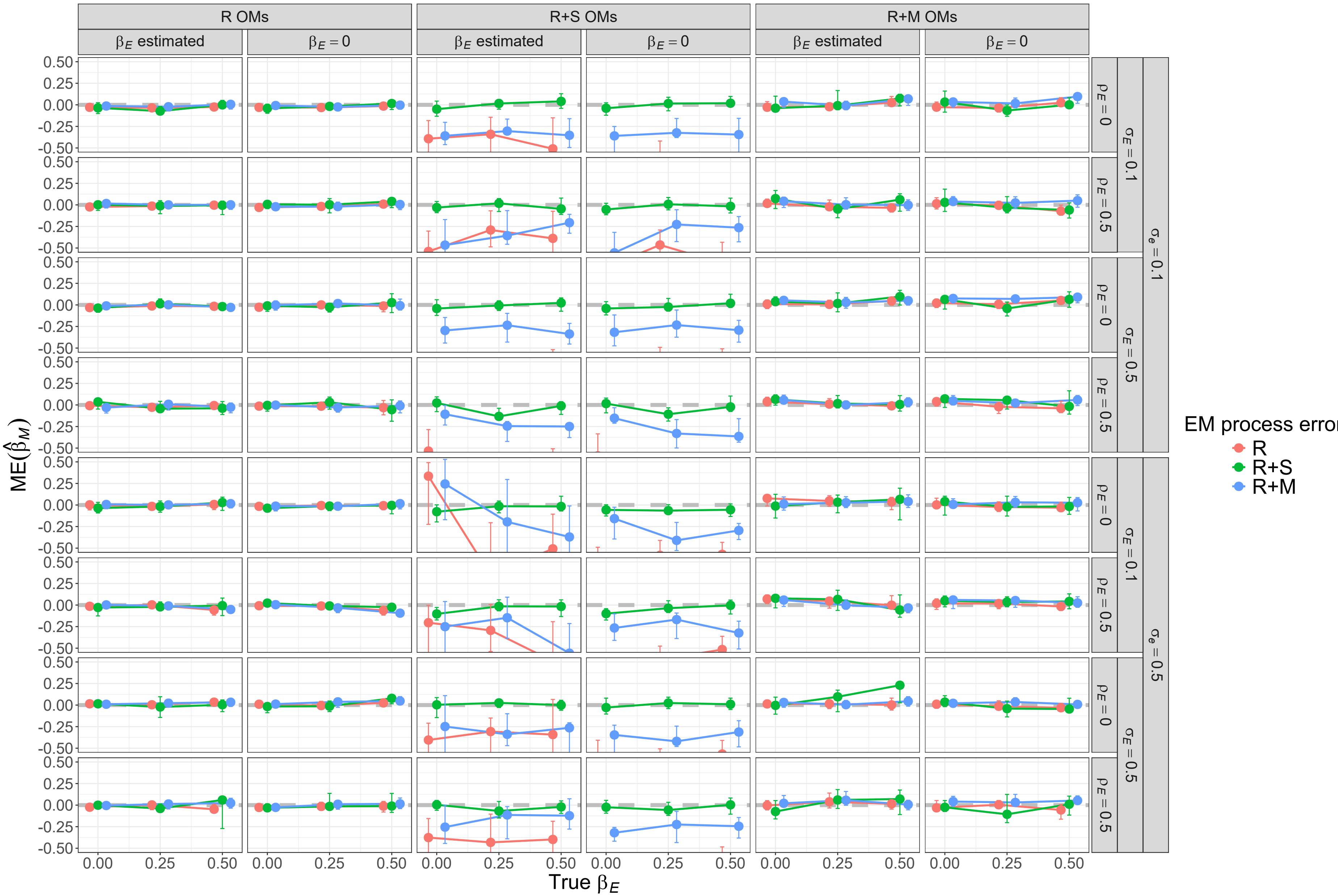



Fig. S8. Median error (ME) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). All OMs had low observation error and contrast in fishing mortality. Vertical lines represent 95% confidence intervals.

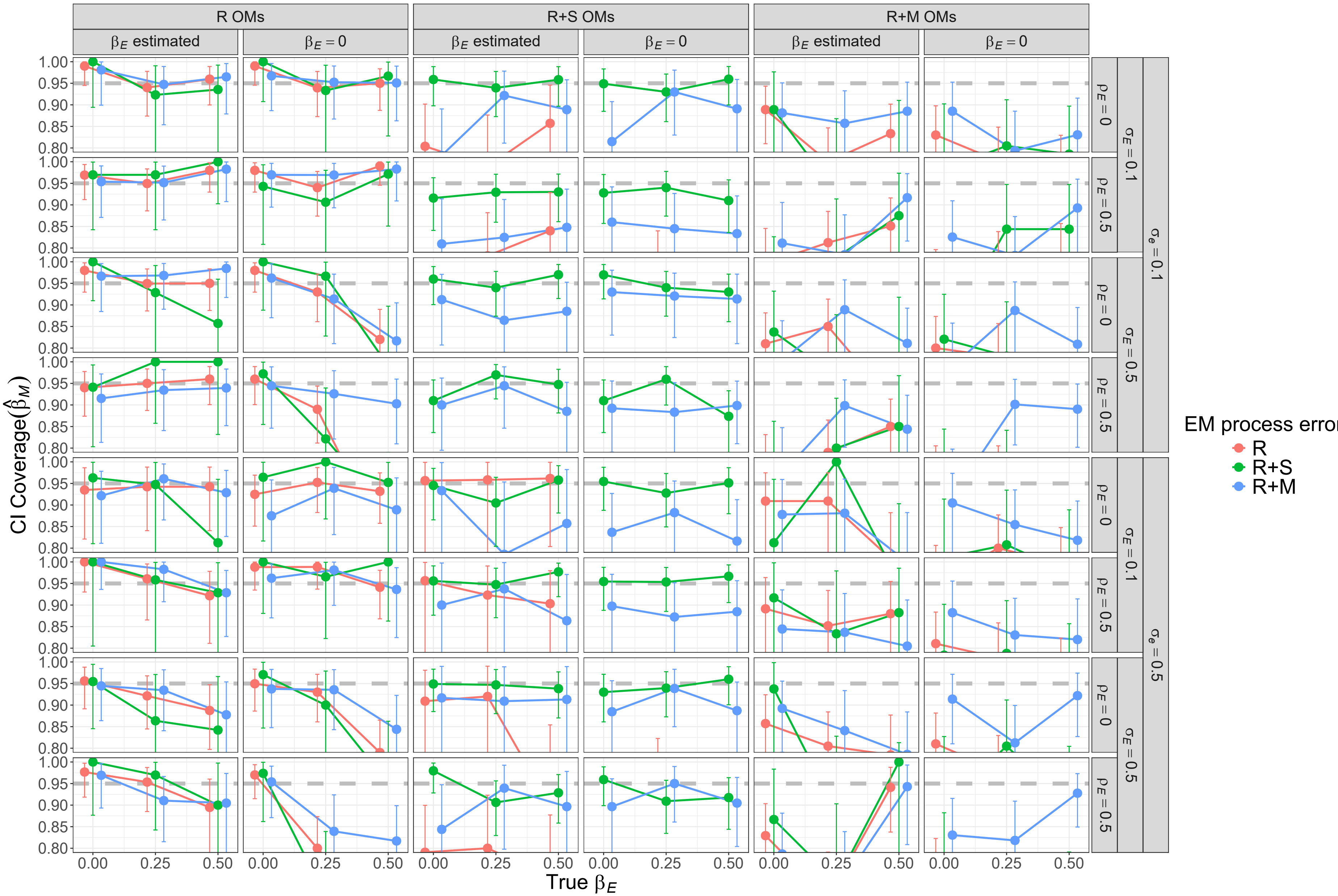



Fig. S9. Probability of 95% confidence interval for $\beta_M$ containing the true value for EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). All OMs had low observation error and contrast in fishing mortality. Vertical lines represent 95% confidence intervals.

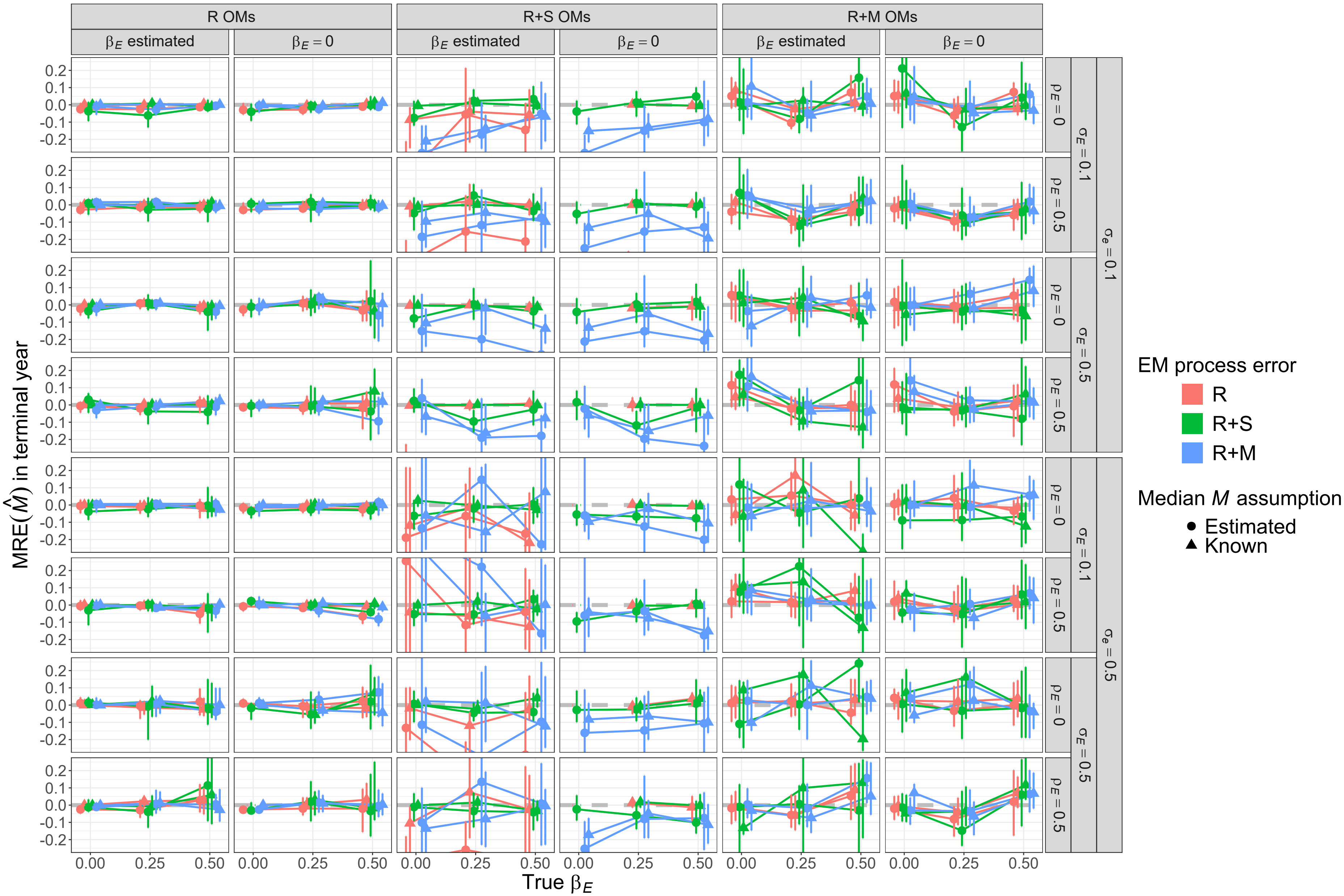



Fig. S10. Median relative error (MRE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low observation error and contrast in fishing mortality.

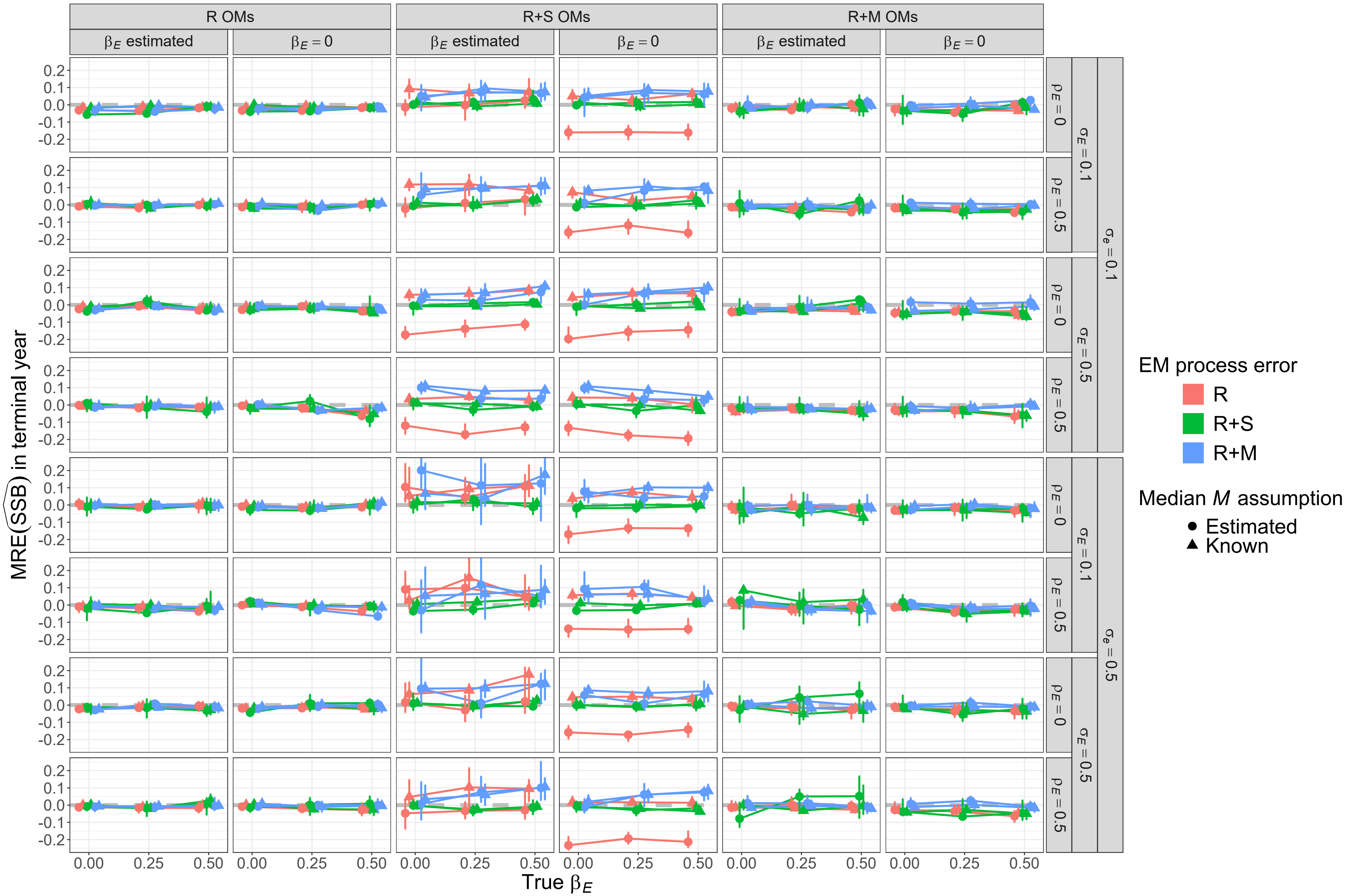
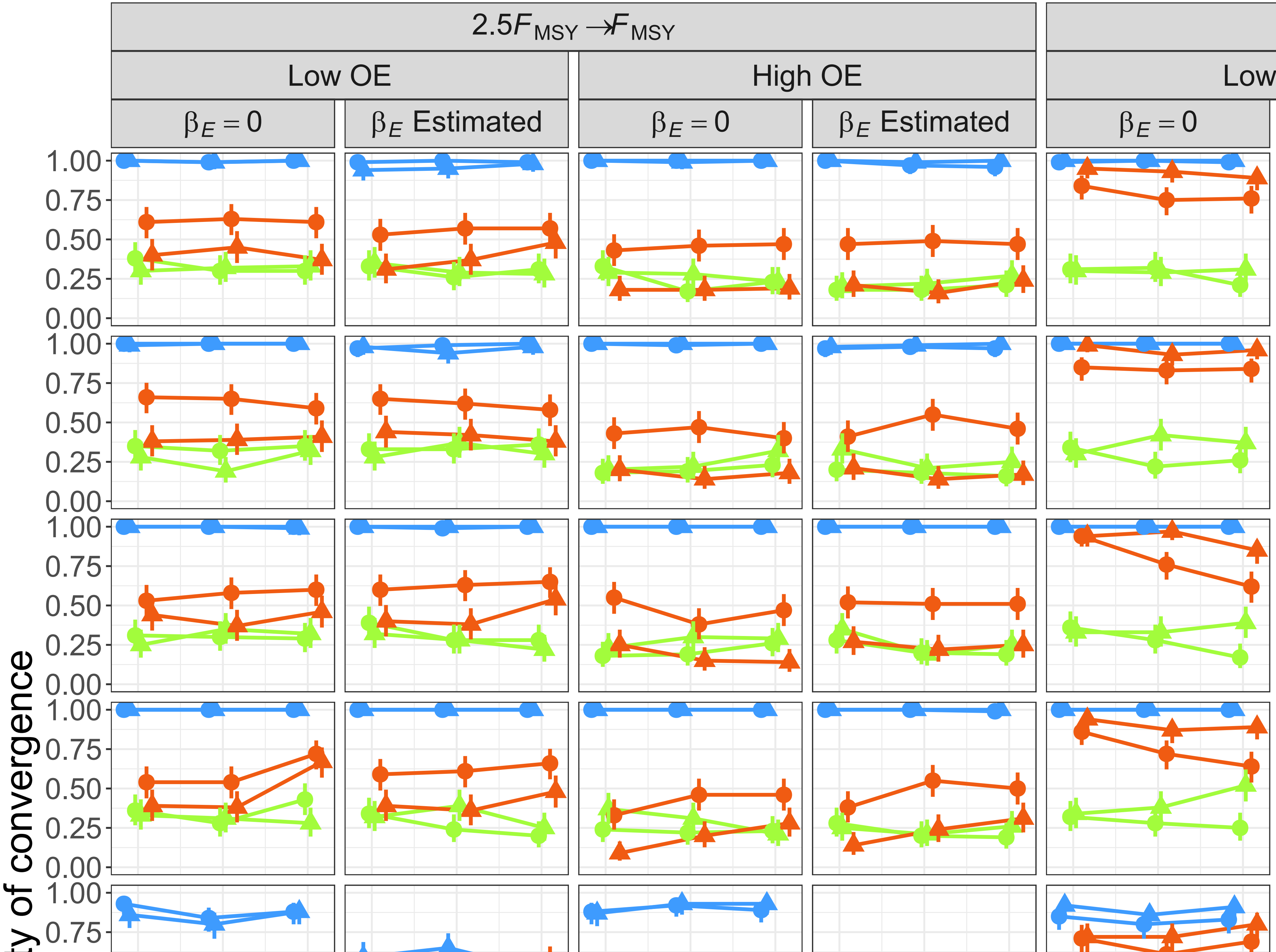
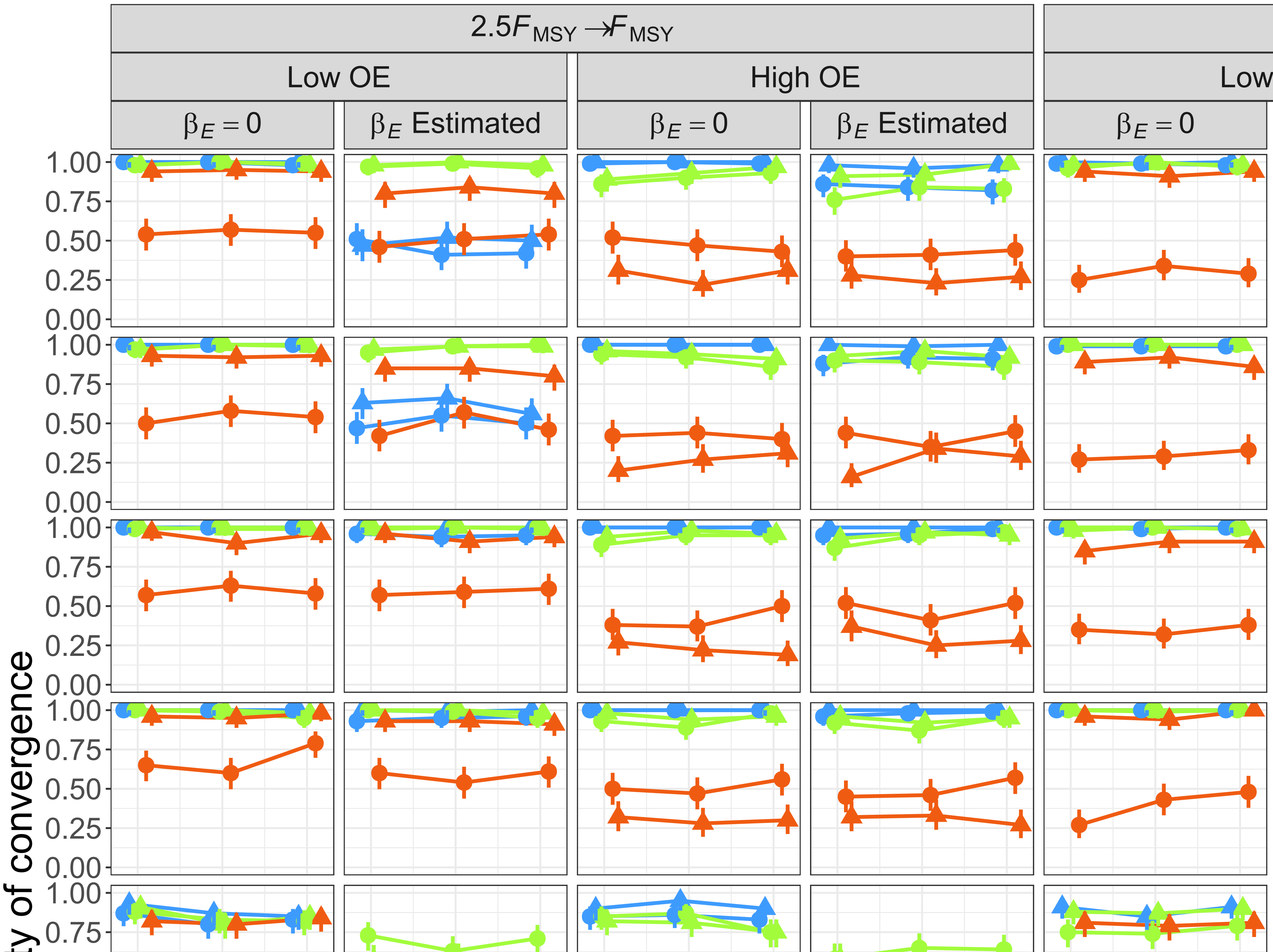



Fig. S11. Median relative error (MRE) of estimates of spawning stock biomass (SSB) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low observation error and contrast in fishing mortality.

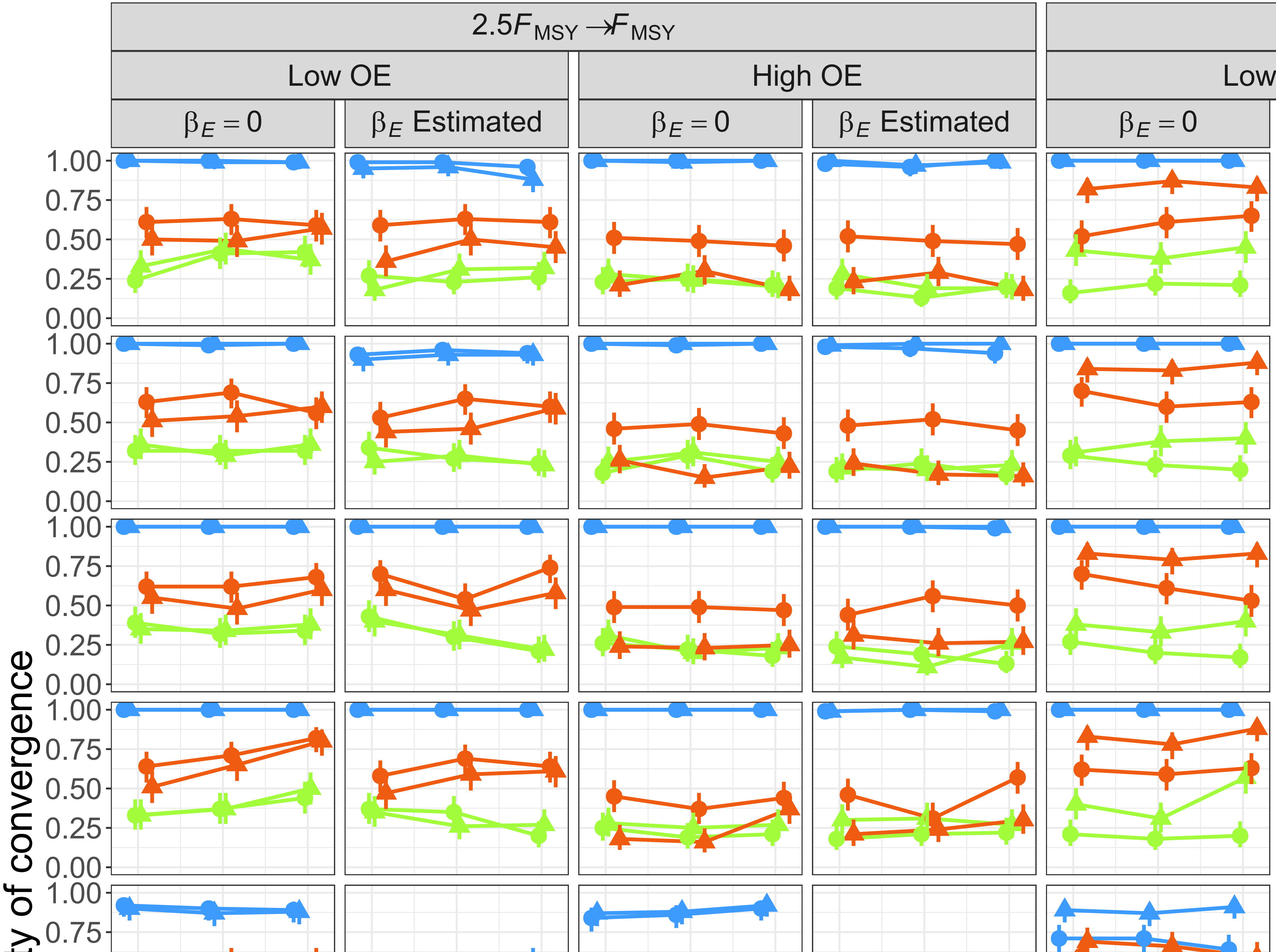
# Convergence results



$2.5F_{\mathrm{MSY}} \rightarrow F_{\mathrm{MSY}}$
Low OE
High OE
Low
$\beta_E = 0$
$\beta_E$ Estimated
$\beta_E = 0$
$\beta_E$ Estimated
$\beta_E = 0$
1.00
0.75
0.50
0.25
0.00
ty of convergence



2.5F_MSY → F_MSY
Low OE
High OE
Low
β_E = 0
β_E Estimated
β_E = 0
β_E Estimated
β_E = 0
1.00
0.75
0.50
0.25
0.00
ty of convergence


2.5$F_{MSY}$ → $F_{MSY}$
Low OE
High OE
Low
$\beta_E = 0$
$\beta_E$ Estimated
1.00
0.75
0.50
0.25
0.00
ty of convergence

## AIC results

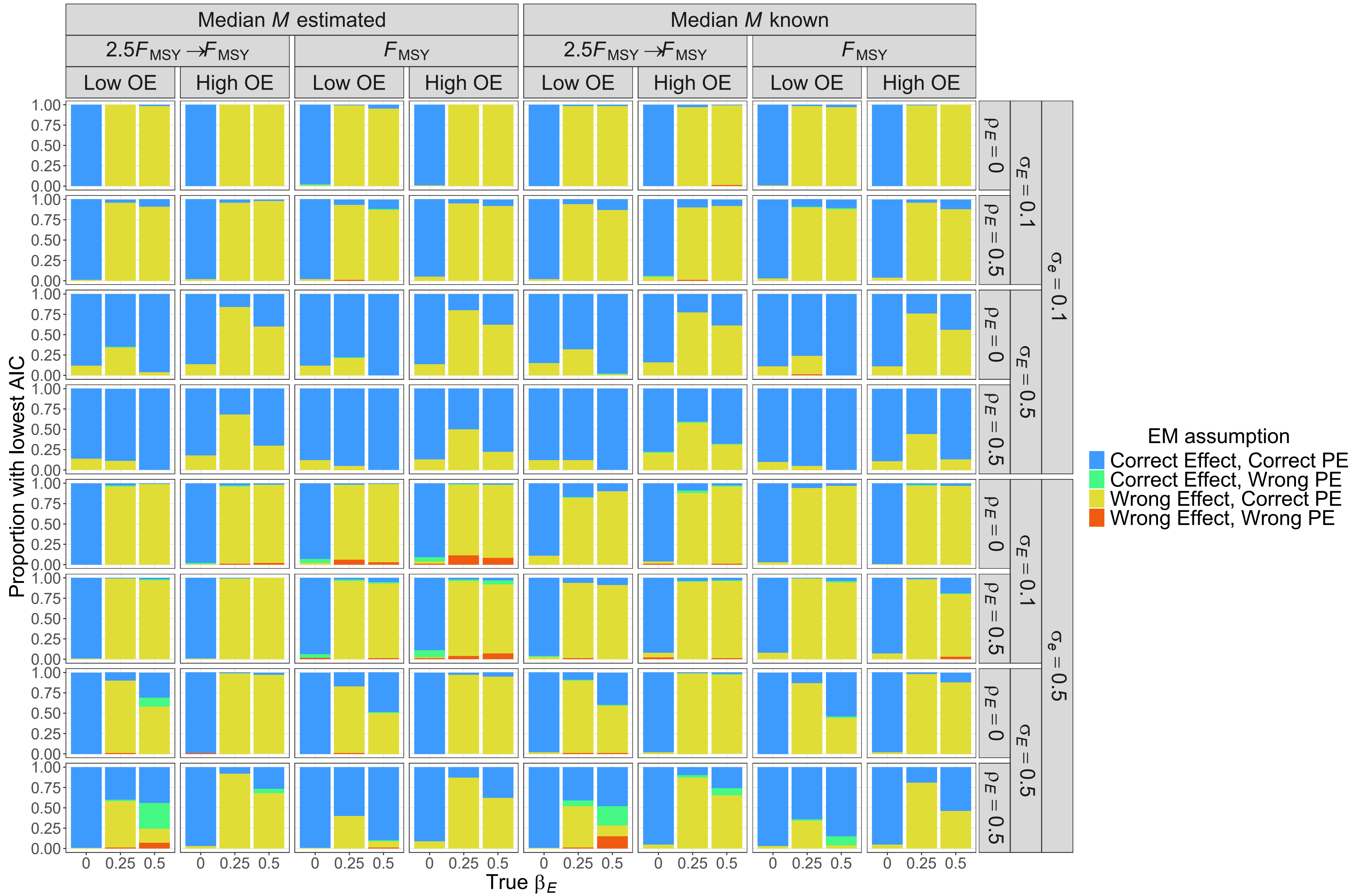



Fig. S15. Proportion of simulated data sets for R OMs where the EM type (treatment of environmental covariate and assumed process error type) had the lowest AIC.

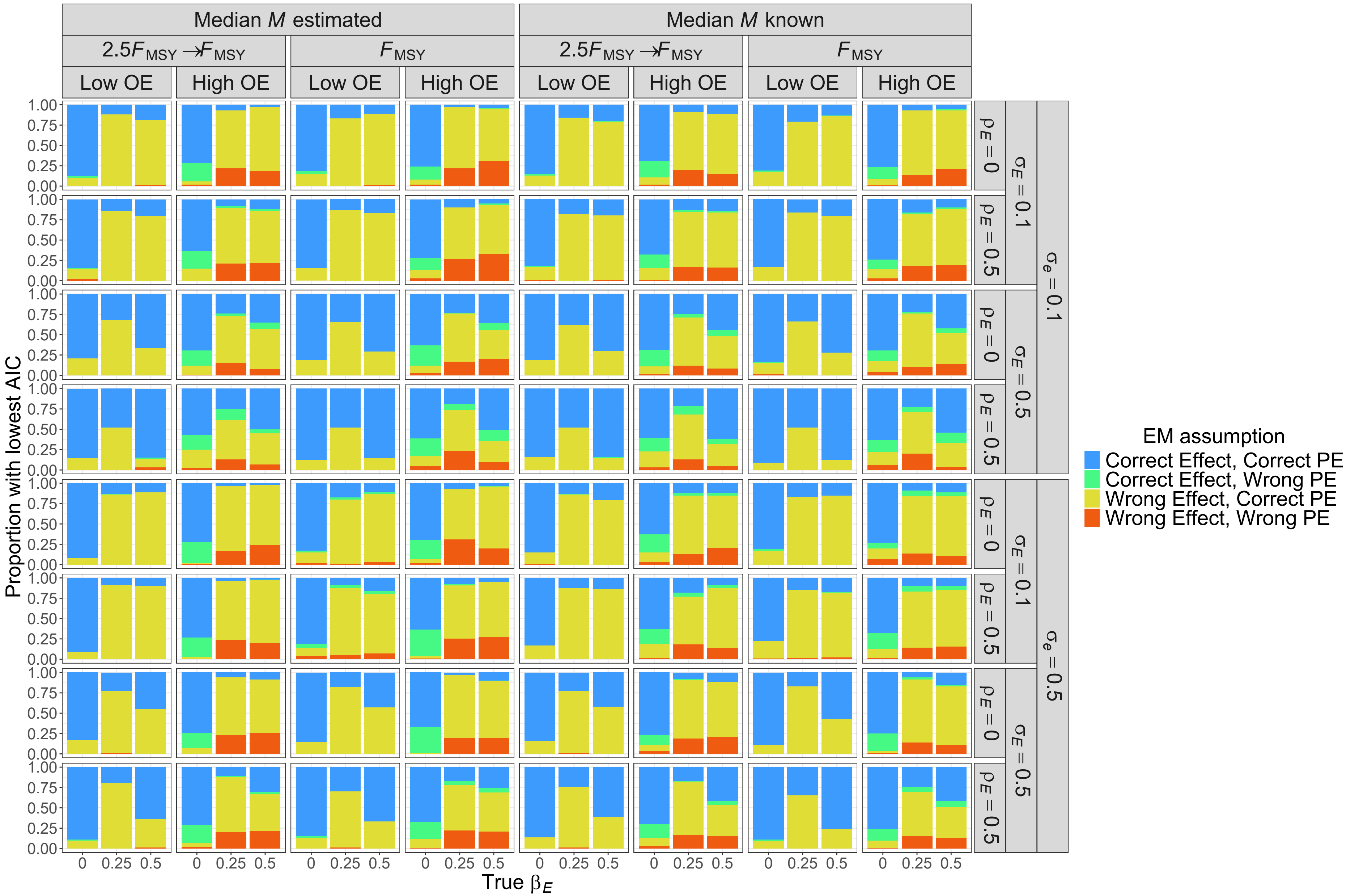



Fig. S16. Proportion of simulated data sets for R+S OMs where the EM type (treatment of environmental covariate and assumed process error type) had the lowest AIC.

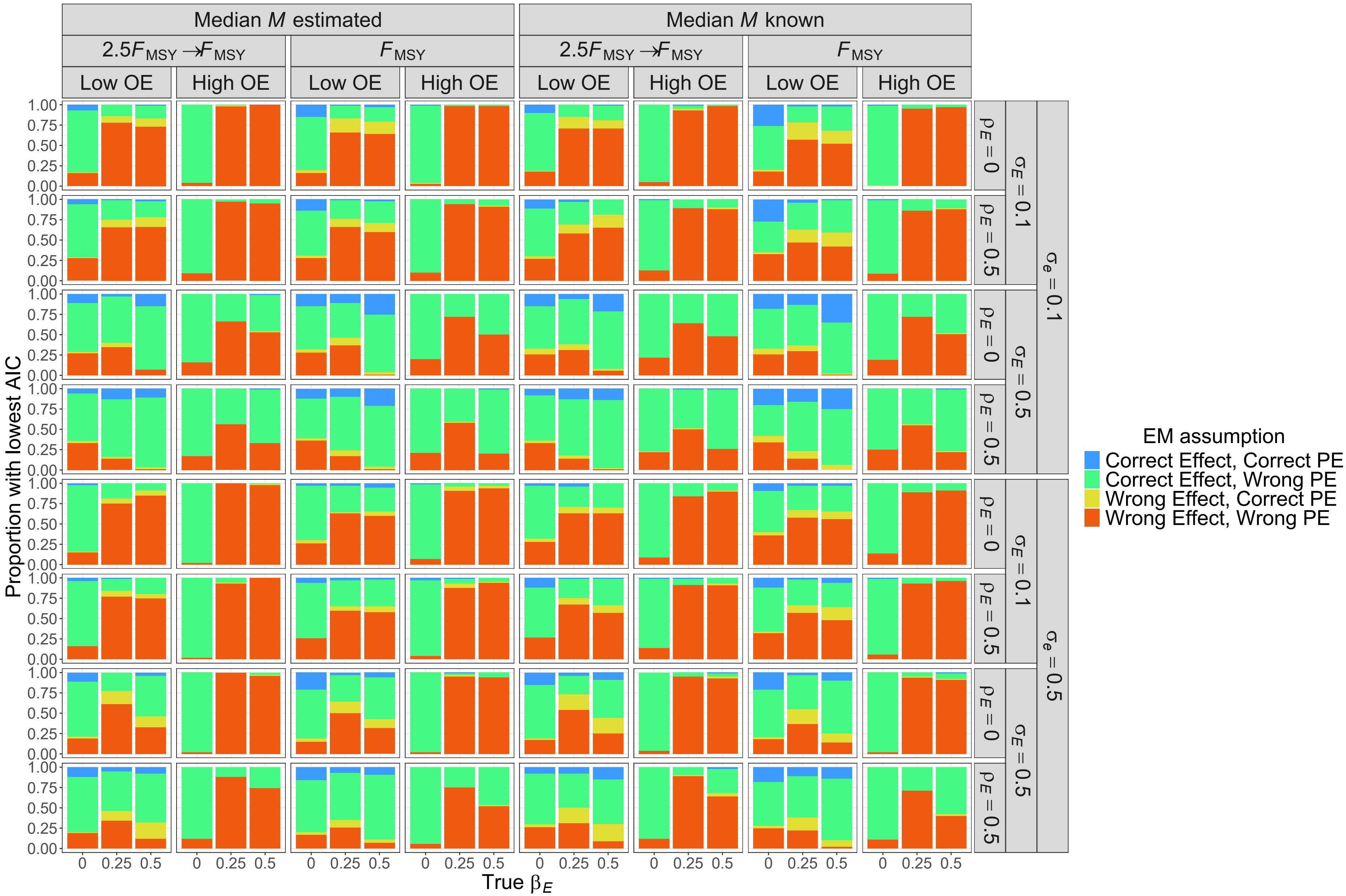



Fig. S17. Proportion of simulated data sets for R+M OMs where the EM type (treatment of environmental covariate and assumed process error type) had the lowest AIC.

## Covariate effect bias

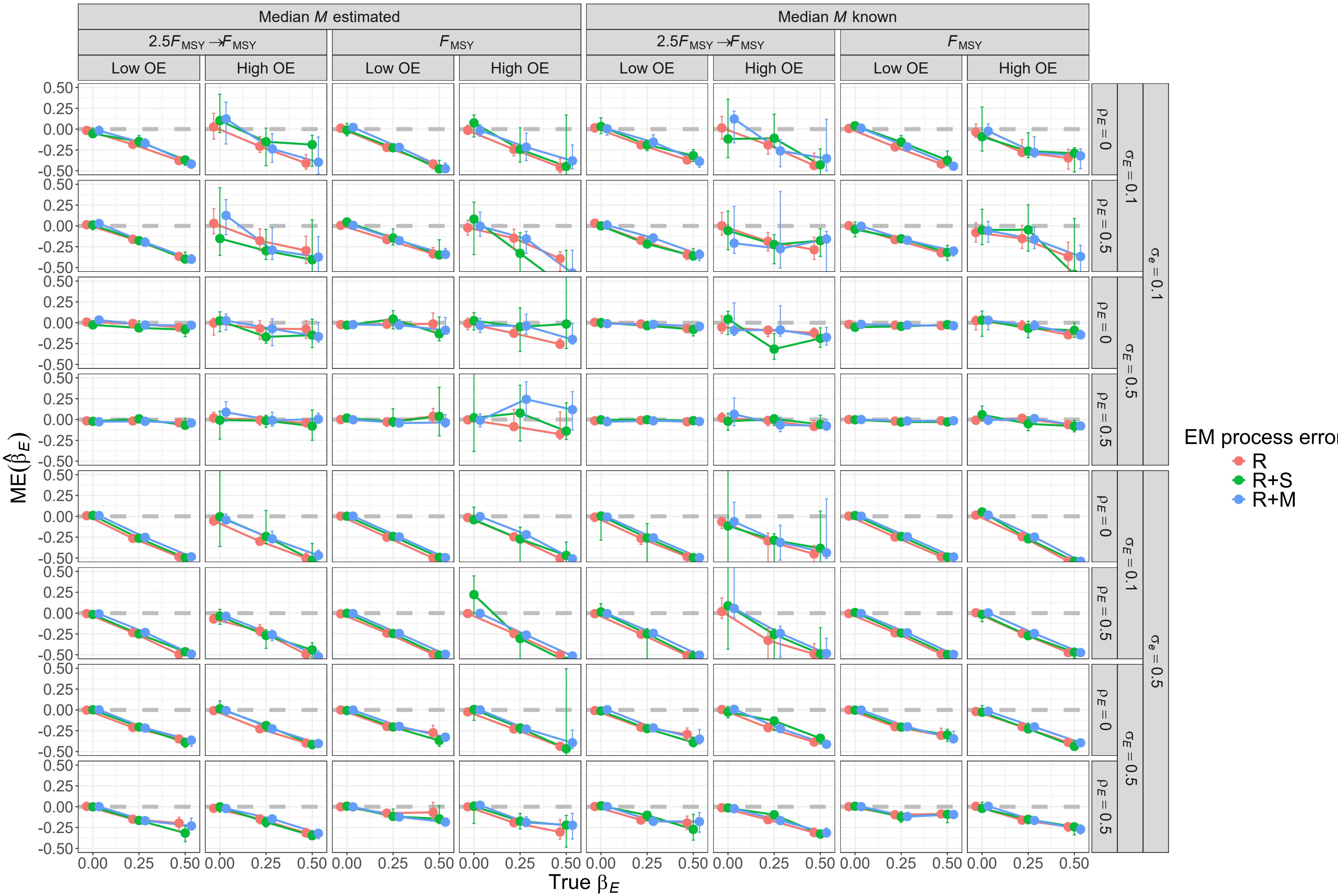



Fig. S18. For R OMs, median error (ME) of estimates of environmental effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). Vertical lines represent 95% confidence intervals.

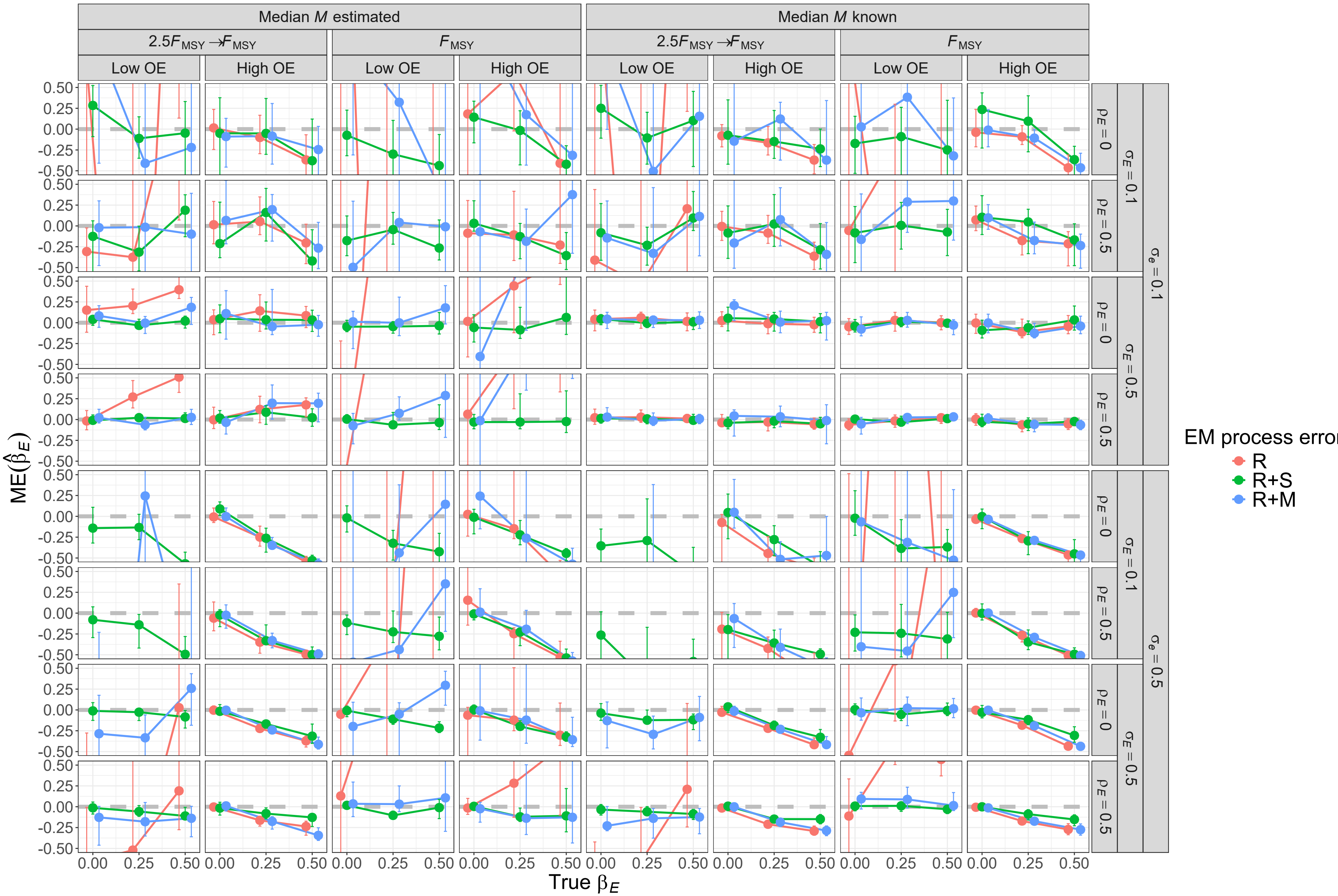



Fig. S19. For R+S OMs, median error (ME) of estimates of environmental effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). Vertical lines represent 95% confidence intervals.

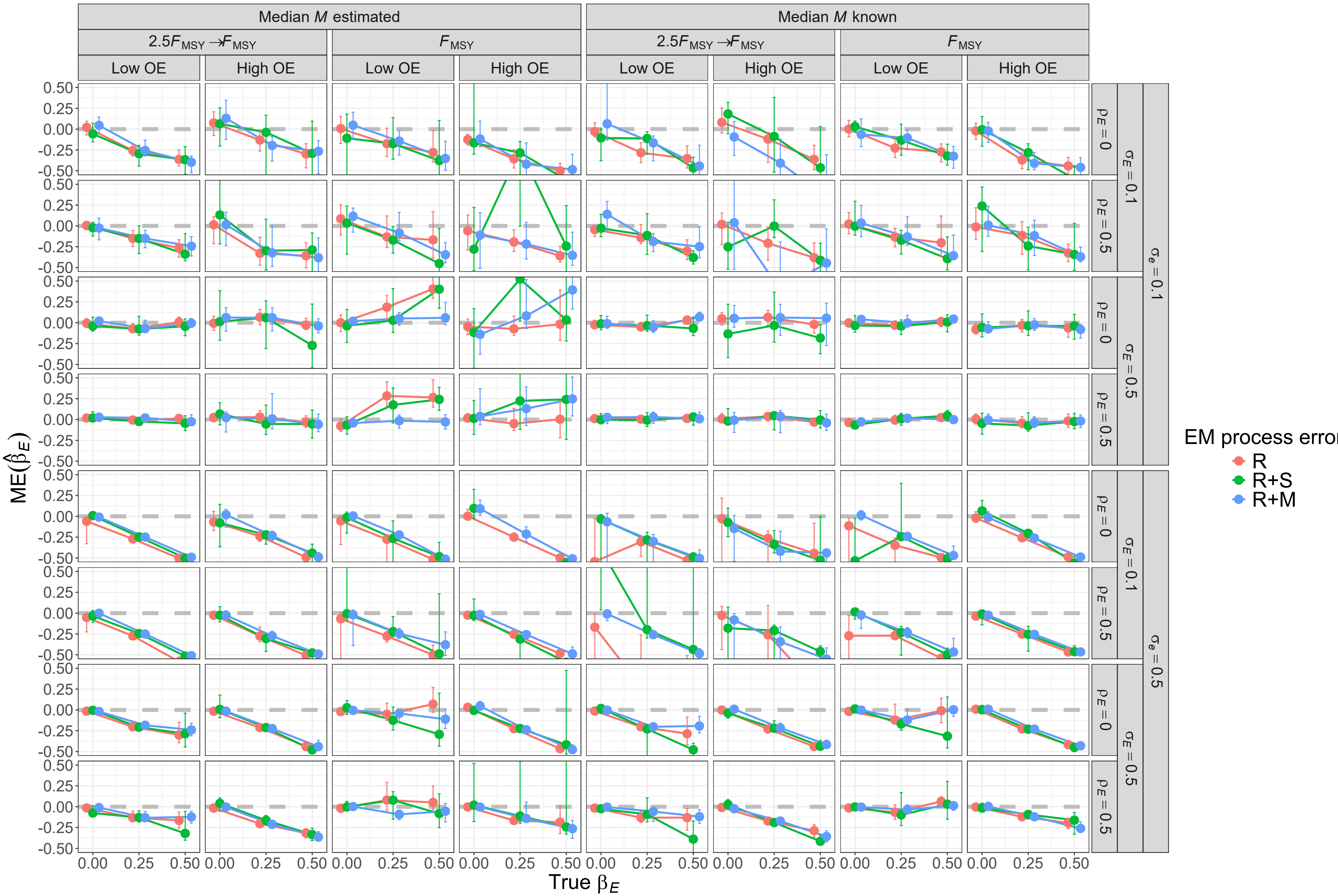



Fig. S20. For R+M OMs, median error (ME) of estimates of environmental effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). Vertical lines represent 95% confidence intervals.

## Covariate effect standard error estimation bias

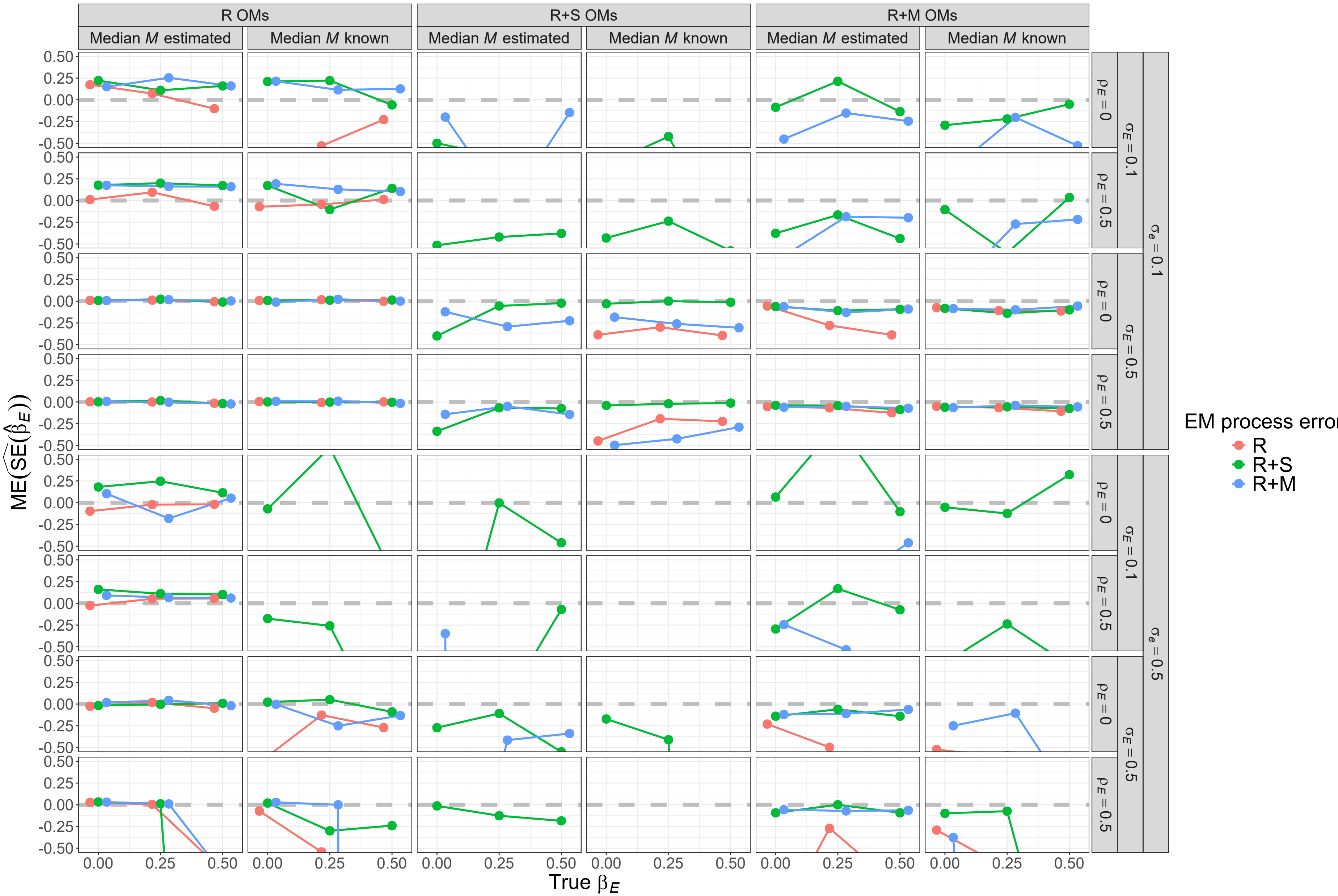



Fig. S21. Median error (ME) of Hessian-based estimates of standard error for covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). All OMs had low observation error and contrast in fishing mortality. True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

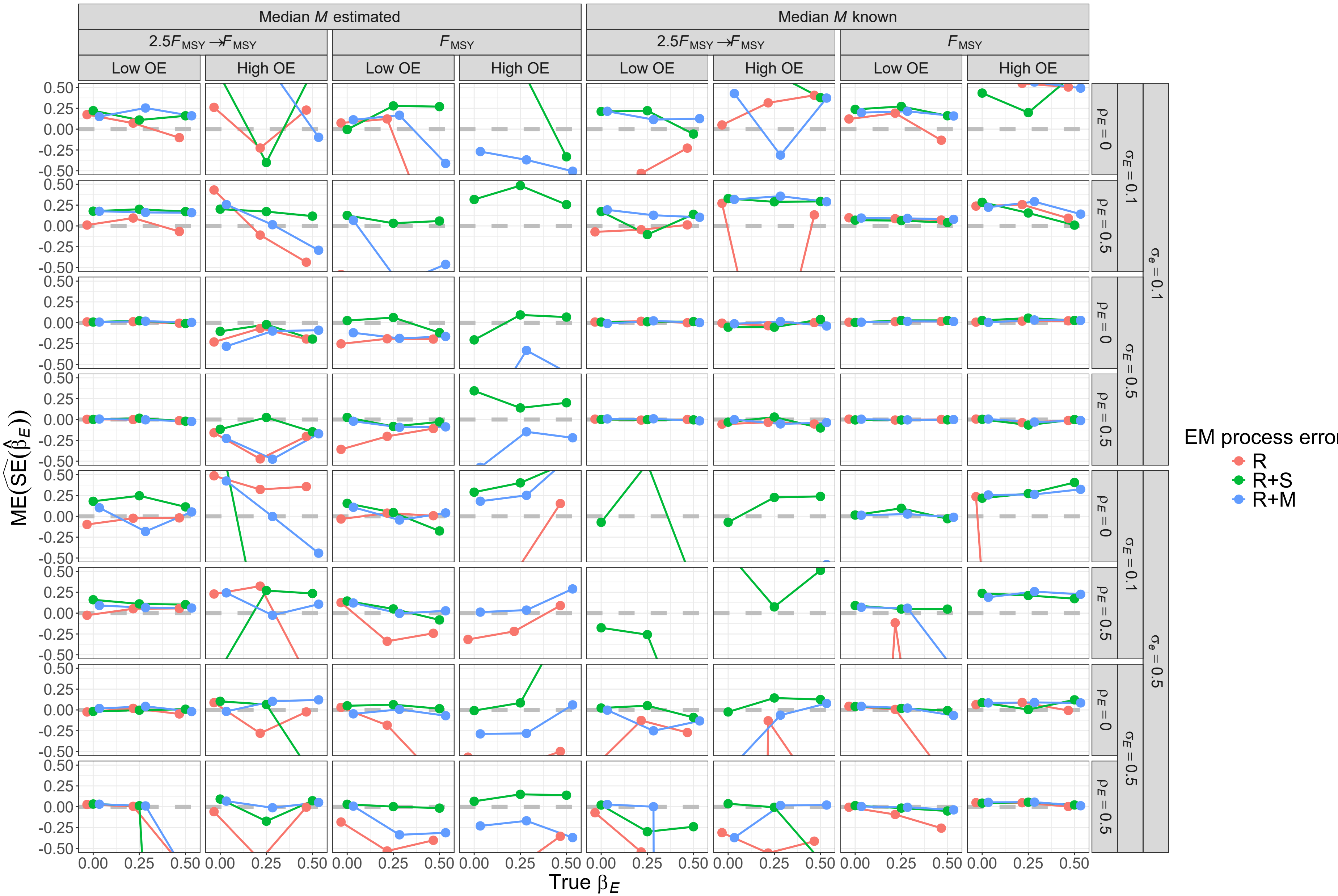



Fig. S22. For R OMs, median error (ME) of Hessian-based estimates of standard error for covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e_M^{\beta}$ known or estimated). True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

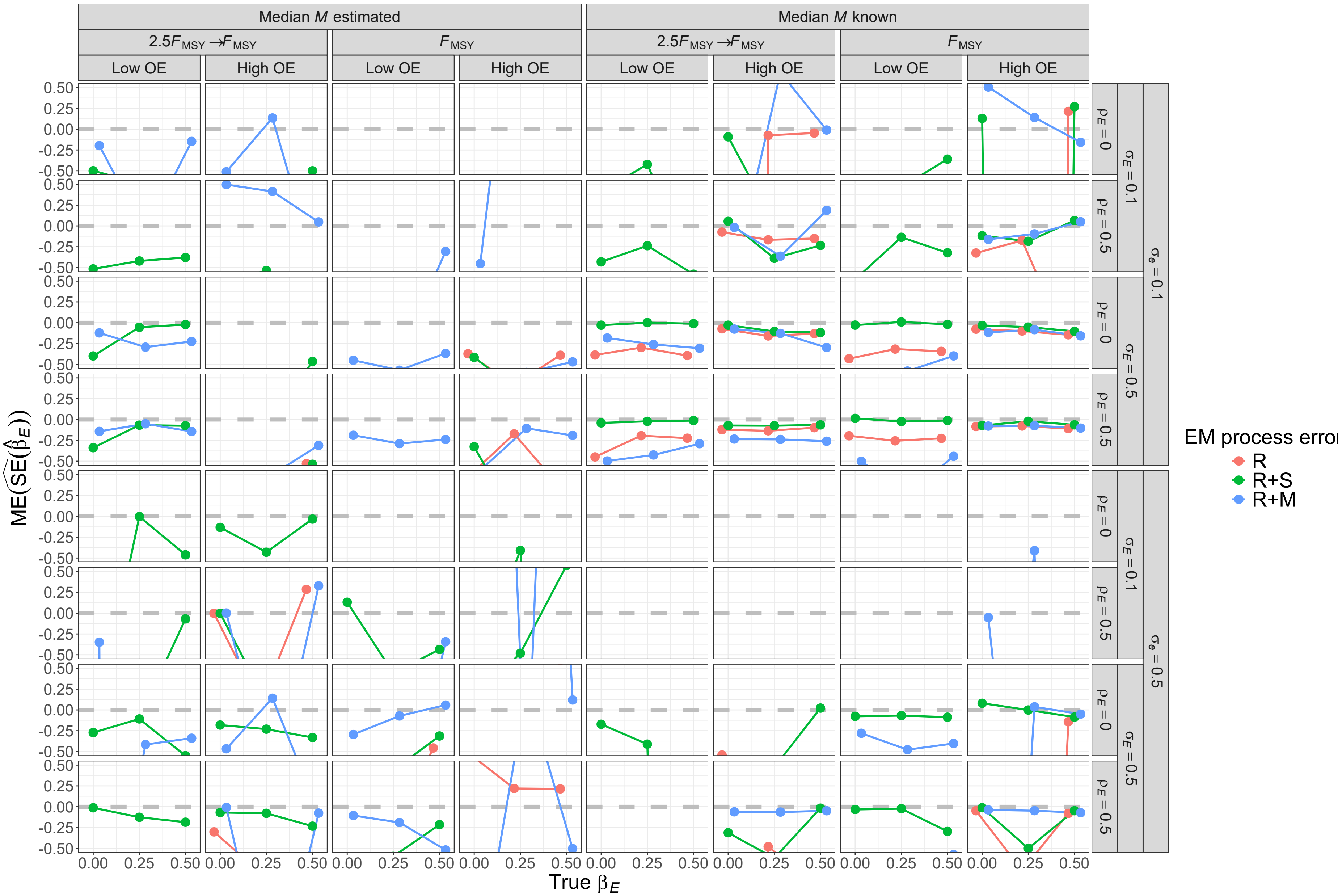



Fig. S23. For R+S OMs, median error (ME) of Hessian-based estimates of standard error for covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

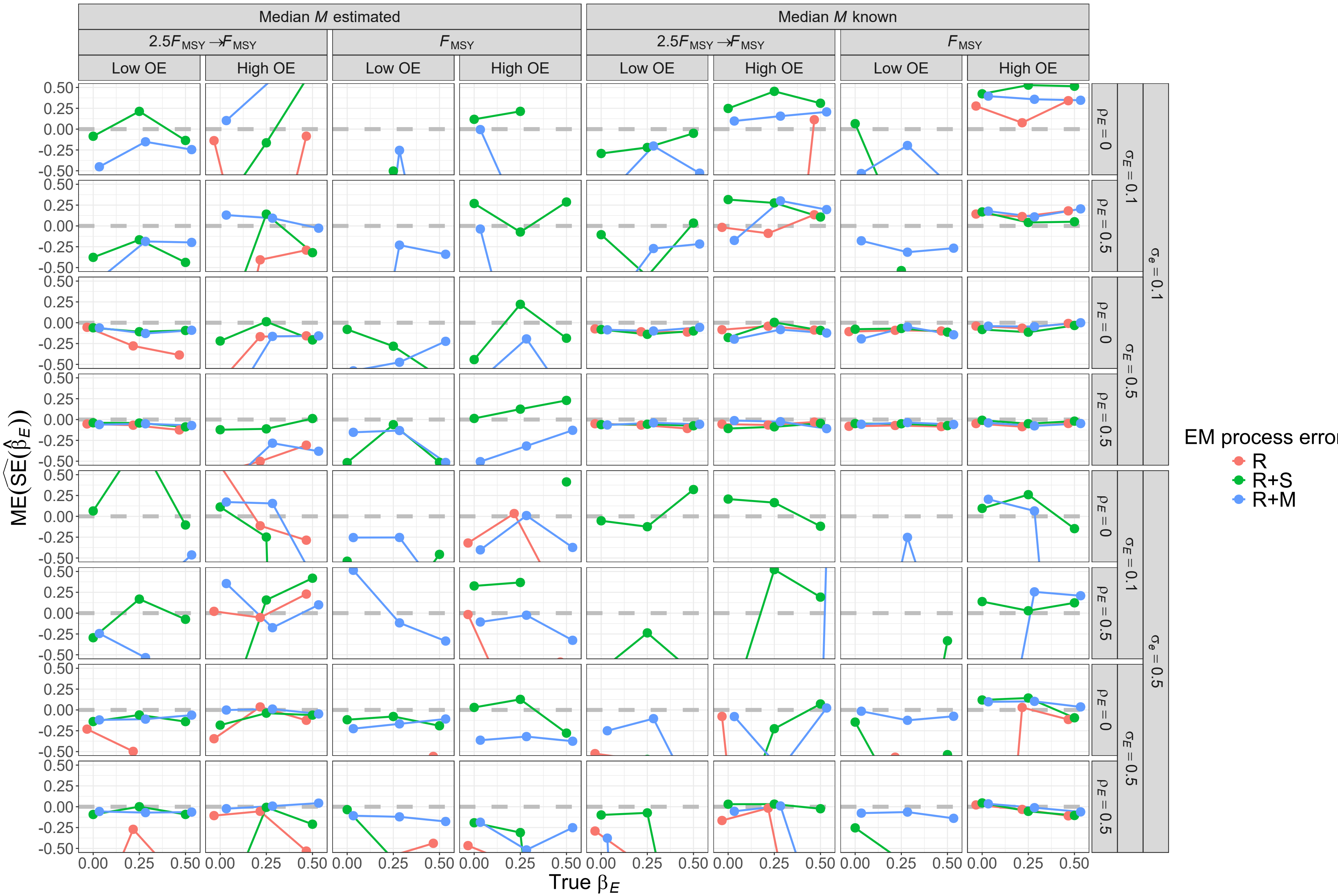



Fig. S24. For R+M OMs, median error (ME) of Hessian-based estimates of standard error for covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e_M^{\beta}$ known or estimated). True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

## Covariate effect confidence interval coverage

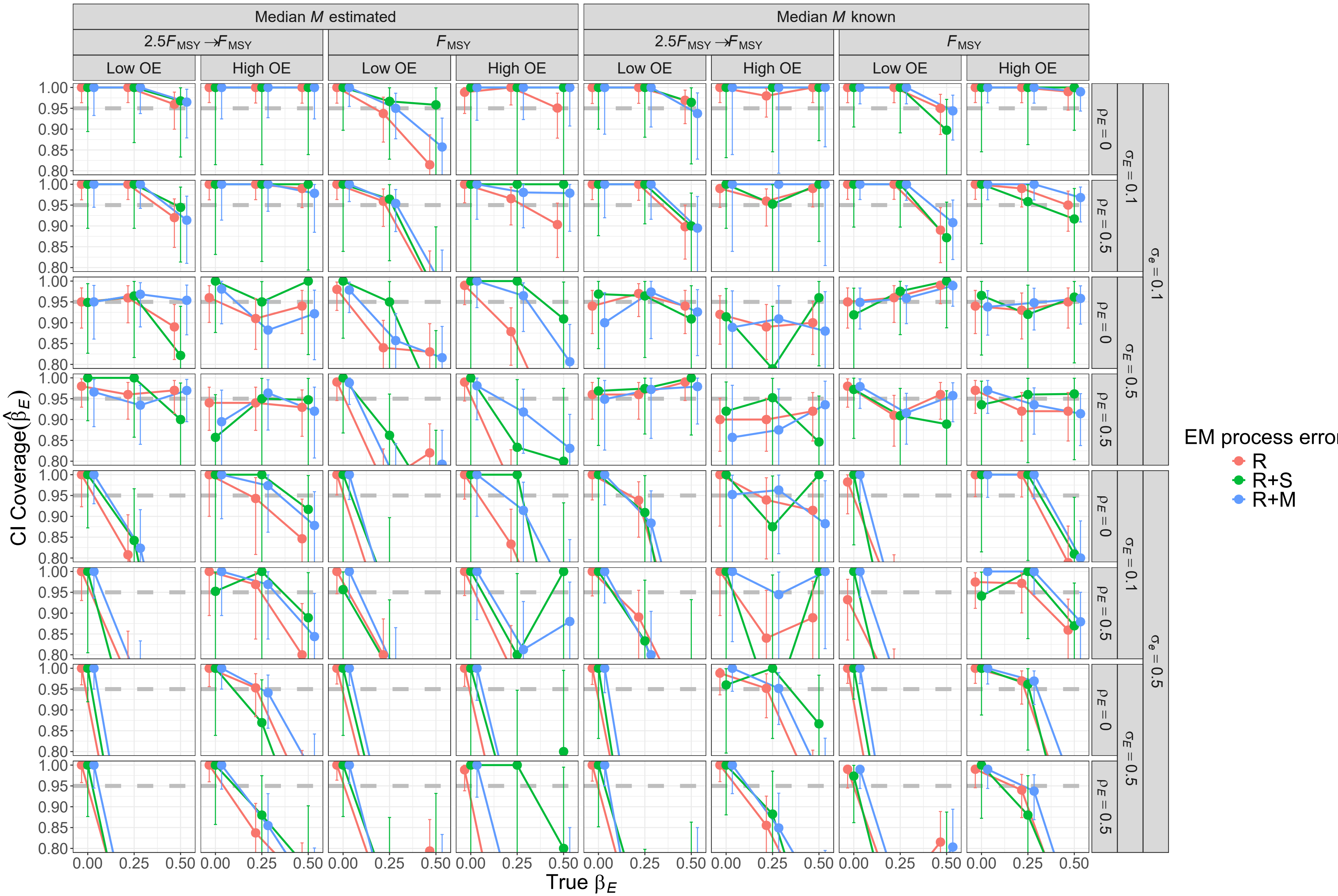



Fig. S25. For R OMs, probability of 95% confidence interval for $\beta_E$ containing the true value for EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). Vertical lines represent 95% confidence intervals.

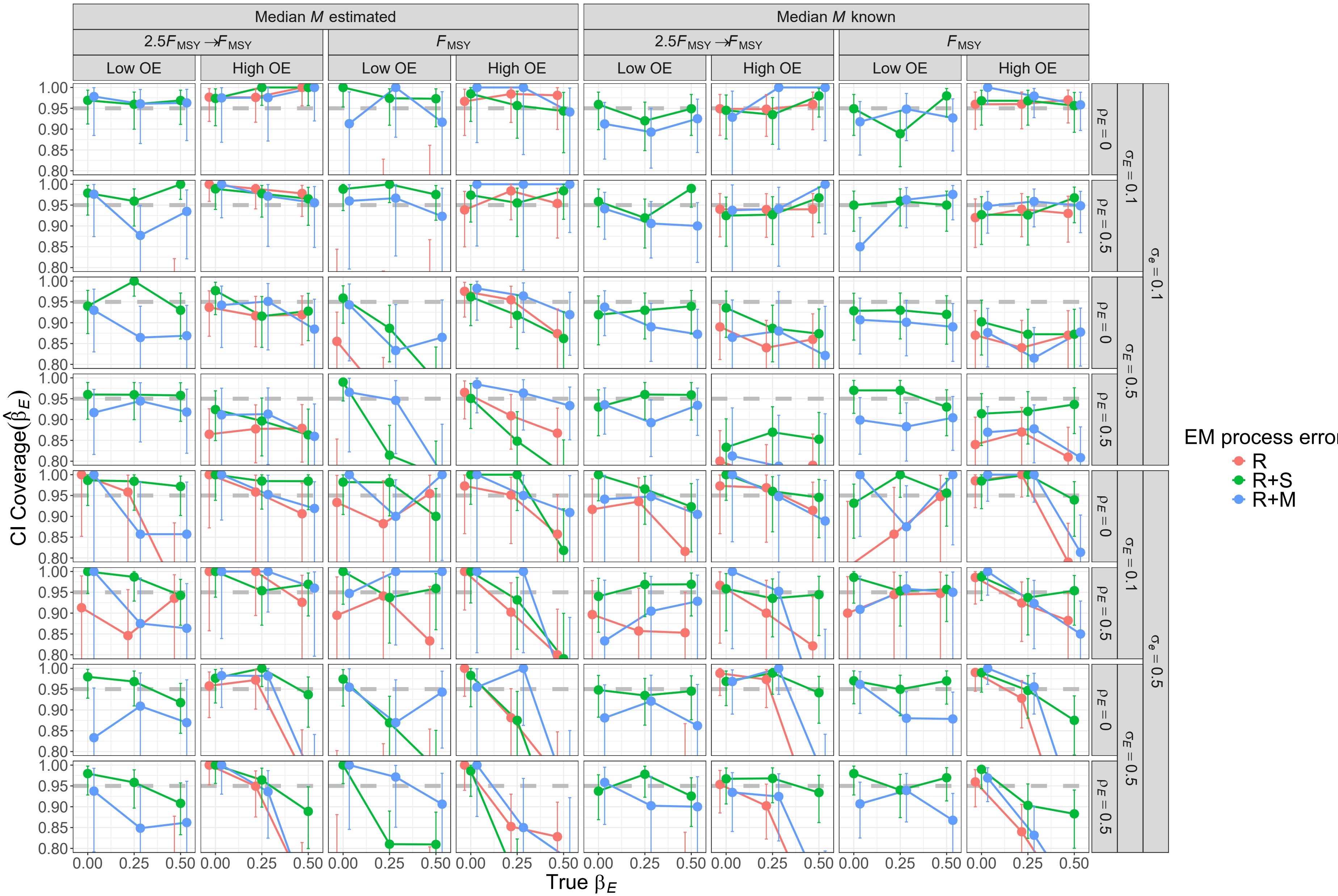



Fig. S26. For R+S OMs, probability of 95% confidence interval for $\beta_E$ containing the true value for EMs with alternative process error assumptions and treatment of median natural mortality ($e_M^\beta$ known or estimated). Vertical lines represent 95% confidence intervals.

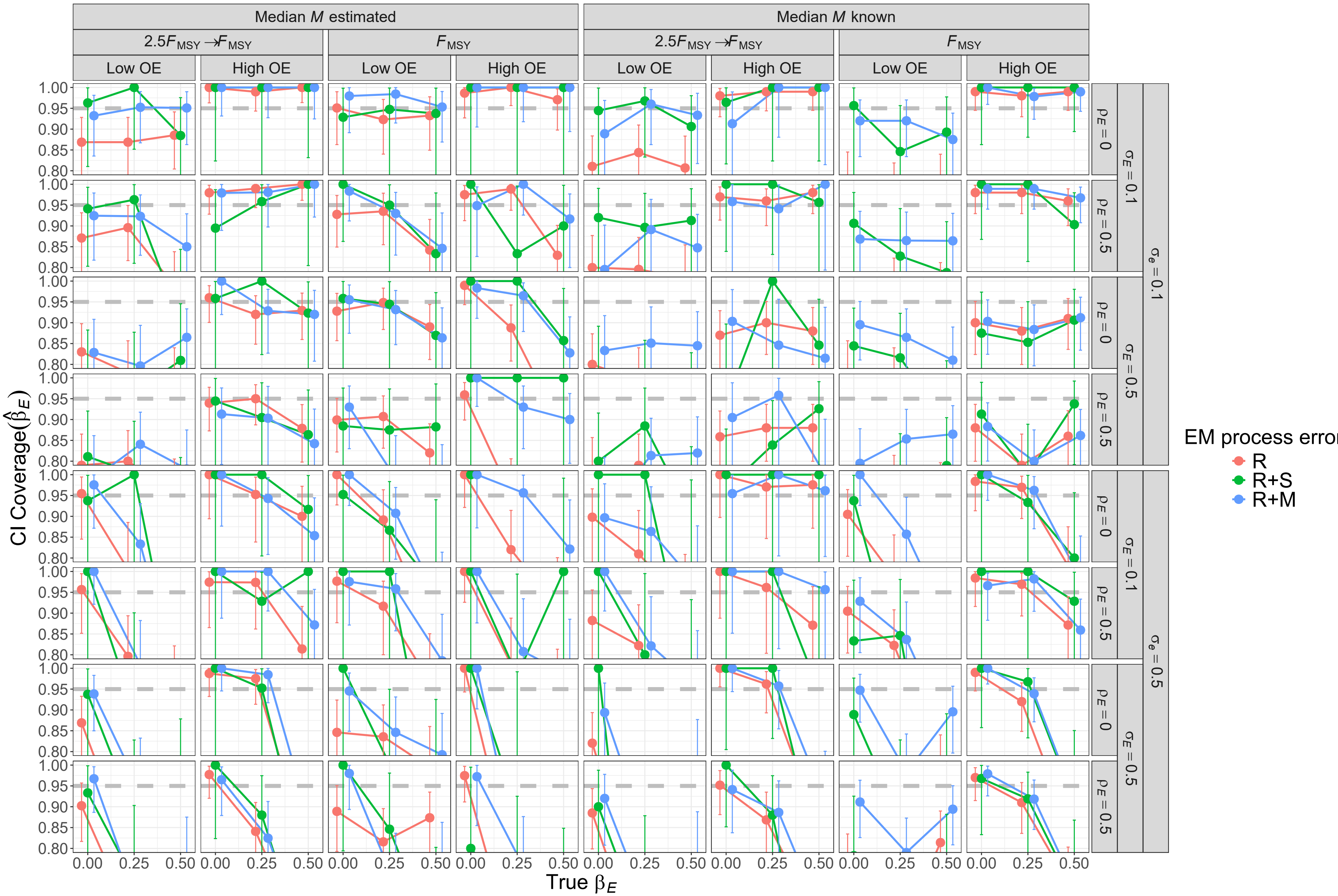



Fig. S27. For R+M OMs, probability of 95% confidence interval for $\beta_E$ containing the true value for EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). Vertical lines represent 95% confidence intervals.

## Covariate effect RMSE

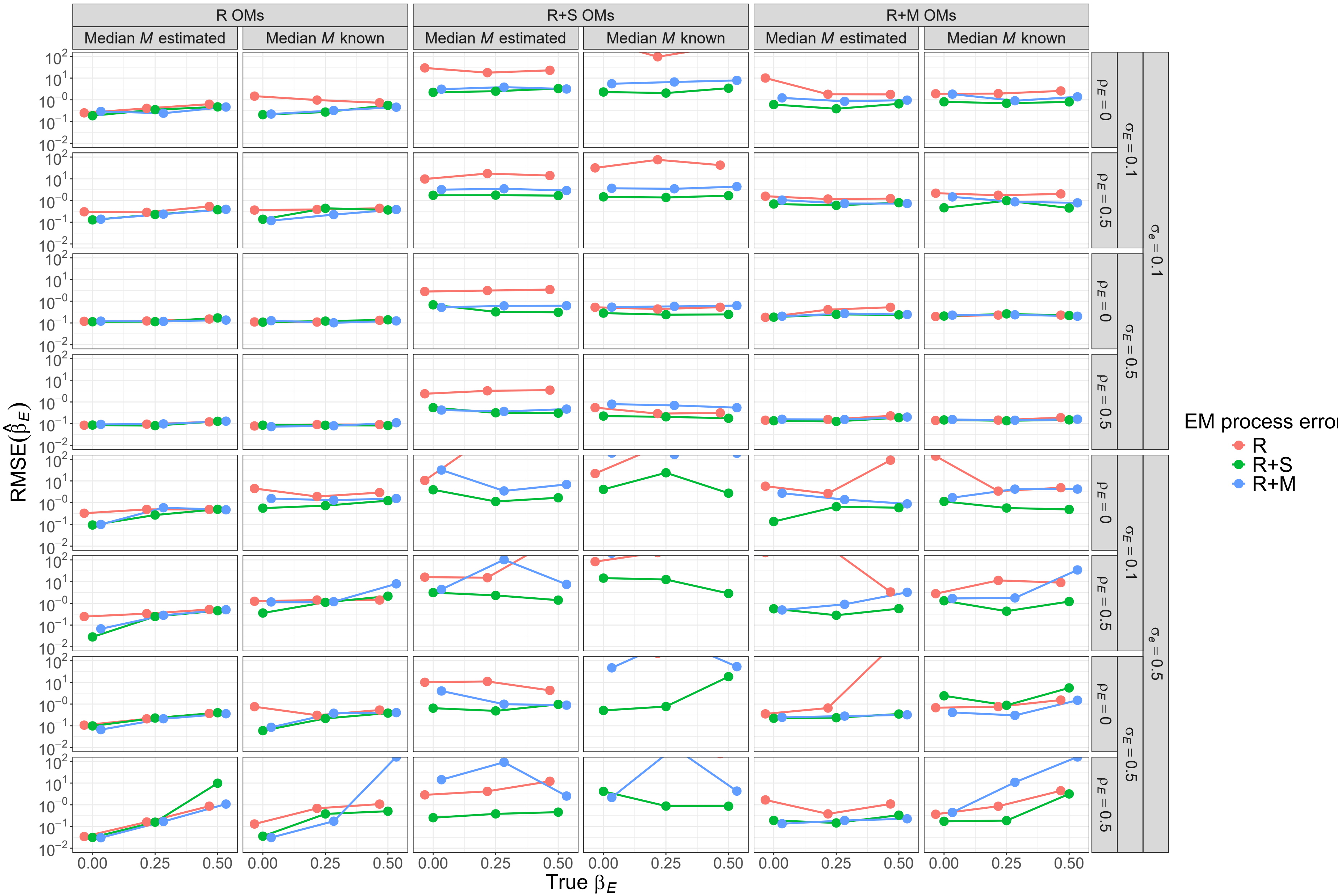



Fig. S28. Root mean square error (RMSE) of estimates of covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated). All OMs had low observation error and contrast in fishing mortality.

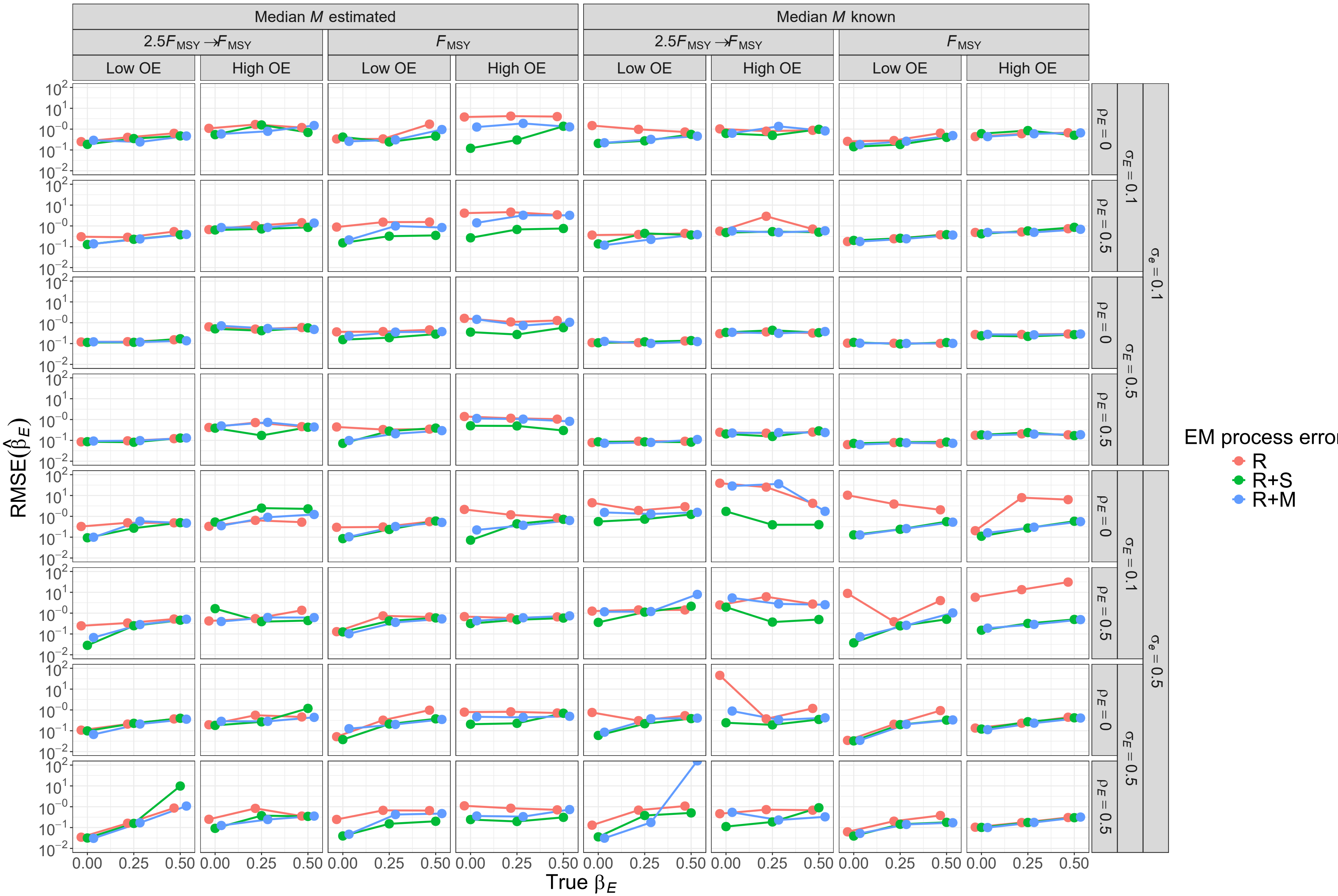



Fig. S29. For R OMs, root mean square error (RMSE) of estimates of covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e_M^\beta$ known or estimated).

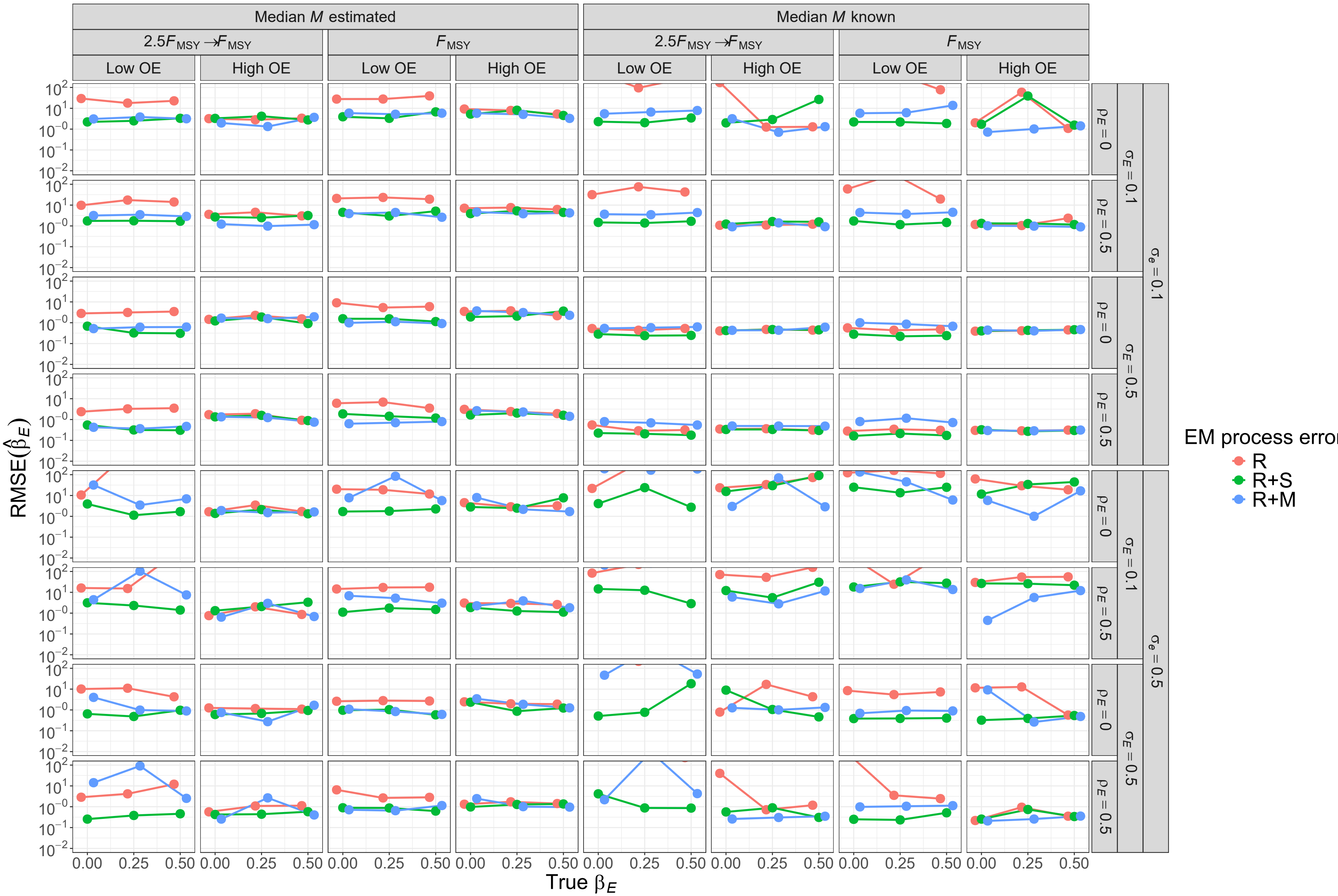



Fig. S30. For R+S OMs, root mean square error (RMSE) of estimates of covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e_M^{\beta}$ known or estimated).

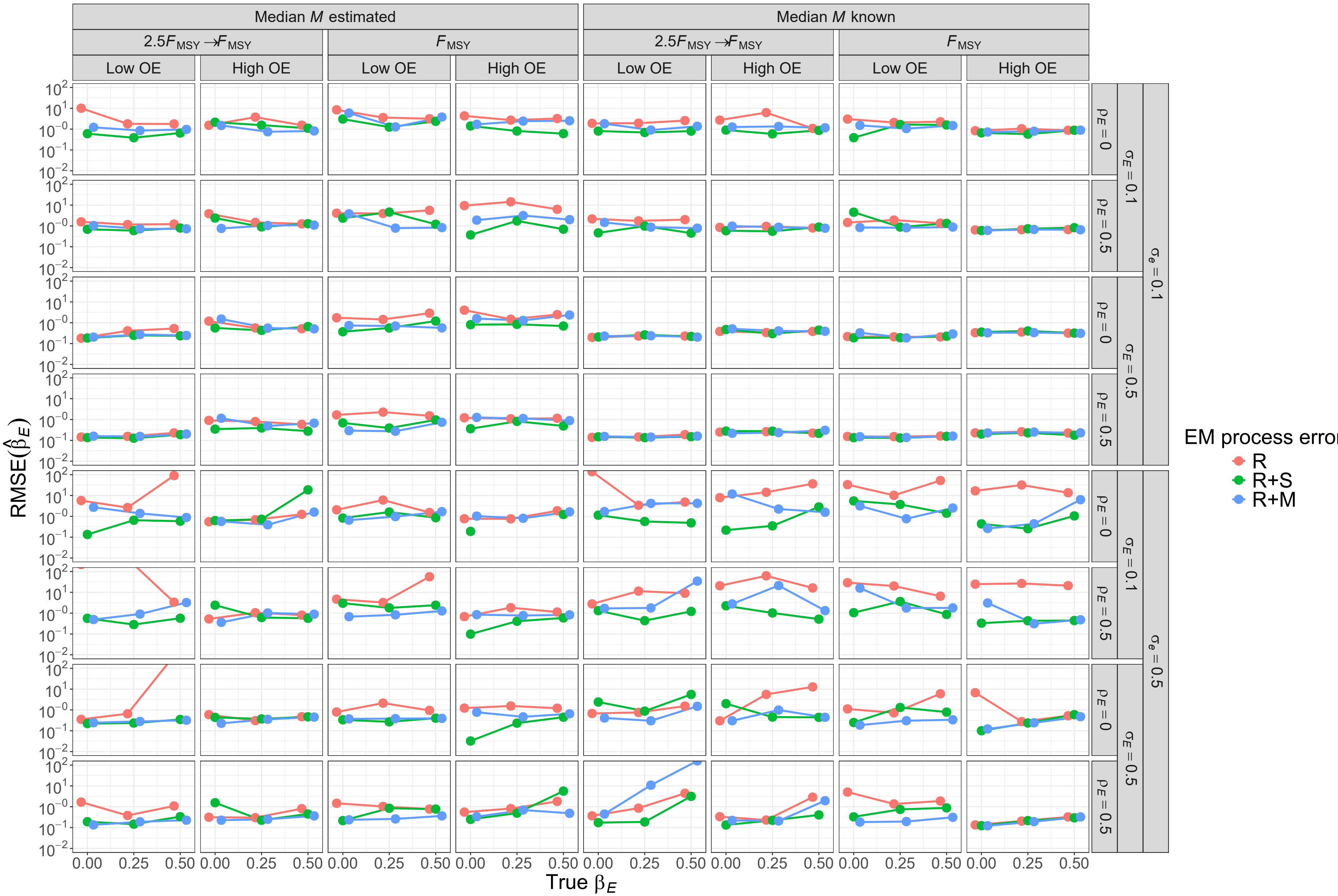



Fig. S31. For R+M OMs, root mean square error (RMSE) of estimates of covariate effect on natural mortality $\beta_E$ from fitting EMs with alternative process error assumptions and treatment of median natural mortality ($e^{\beta}_{M}$ known or estimated).

## Covariate effect estimate and standard error example

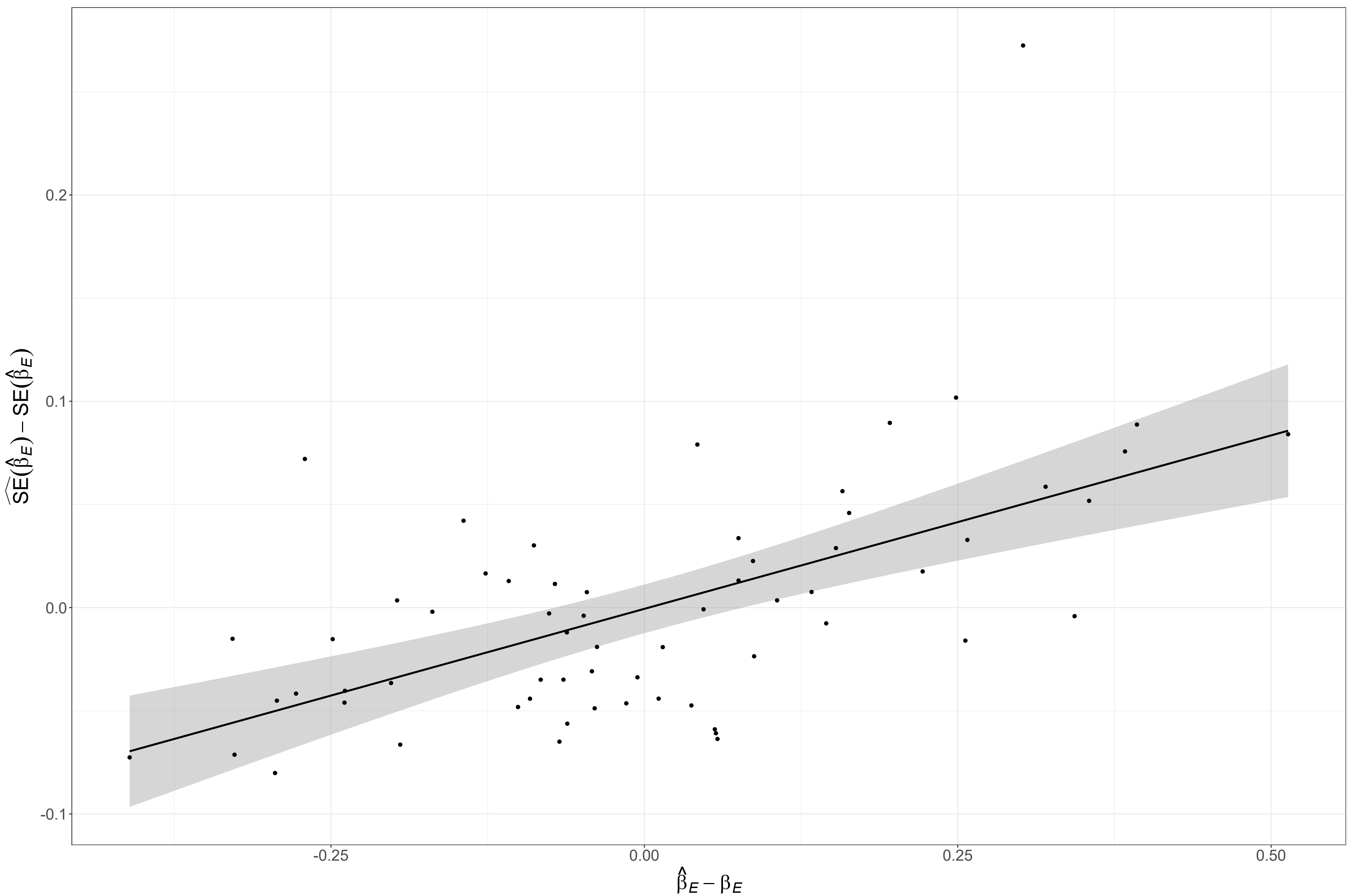



Fig. S32. Positive correlation of covariate effect estimates and Hessian-based standard error estimates for EM that also estimates the median natural mortality parameter and has correct R+M process error assumption fitted to simulated data from the OM with R+M process errors, temporal contrast in fishing pressure, low observation uncertainty for both population ($LowOE$) and covariate observations ($\sigma_e = 0.1$), high and uncorrelated temporal variability in the true covariate ($\sigma_E = 0.5$ and $\rho_E = 0$), and

## Median Natural mortality parameter bias

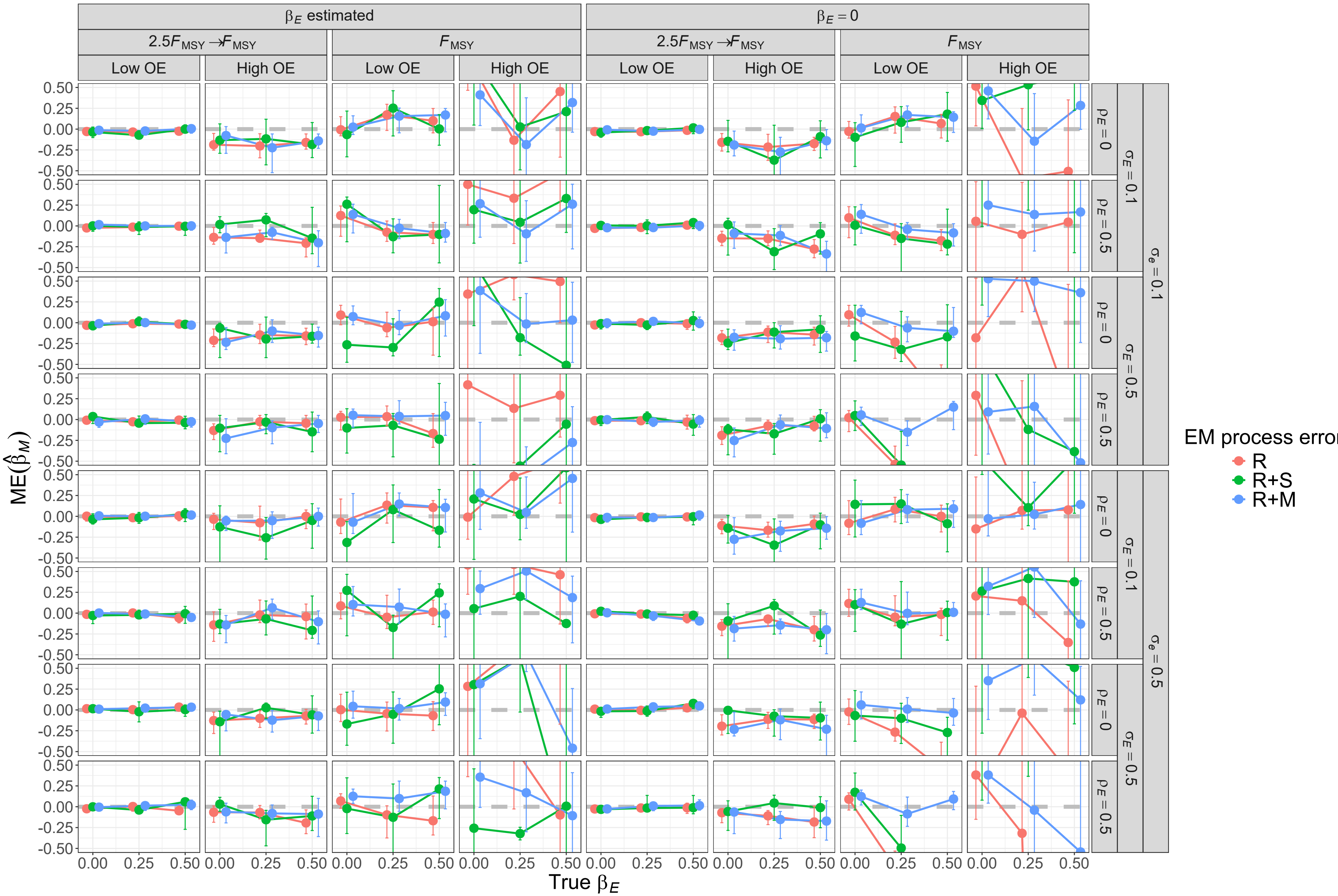



Fig. S33. For R OMs, median error (ME) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). Vertical lines represent 95% confidence intervals.

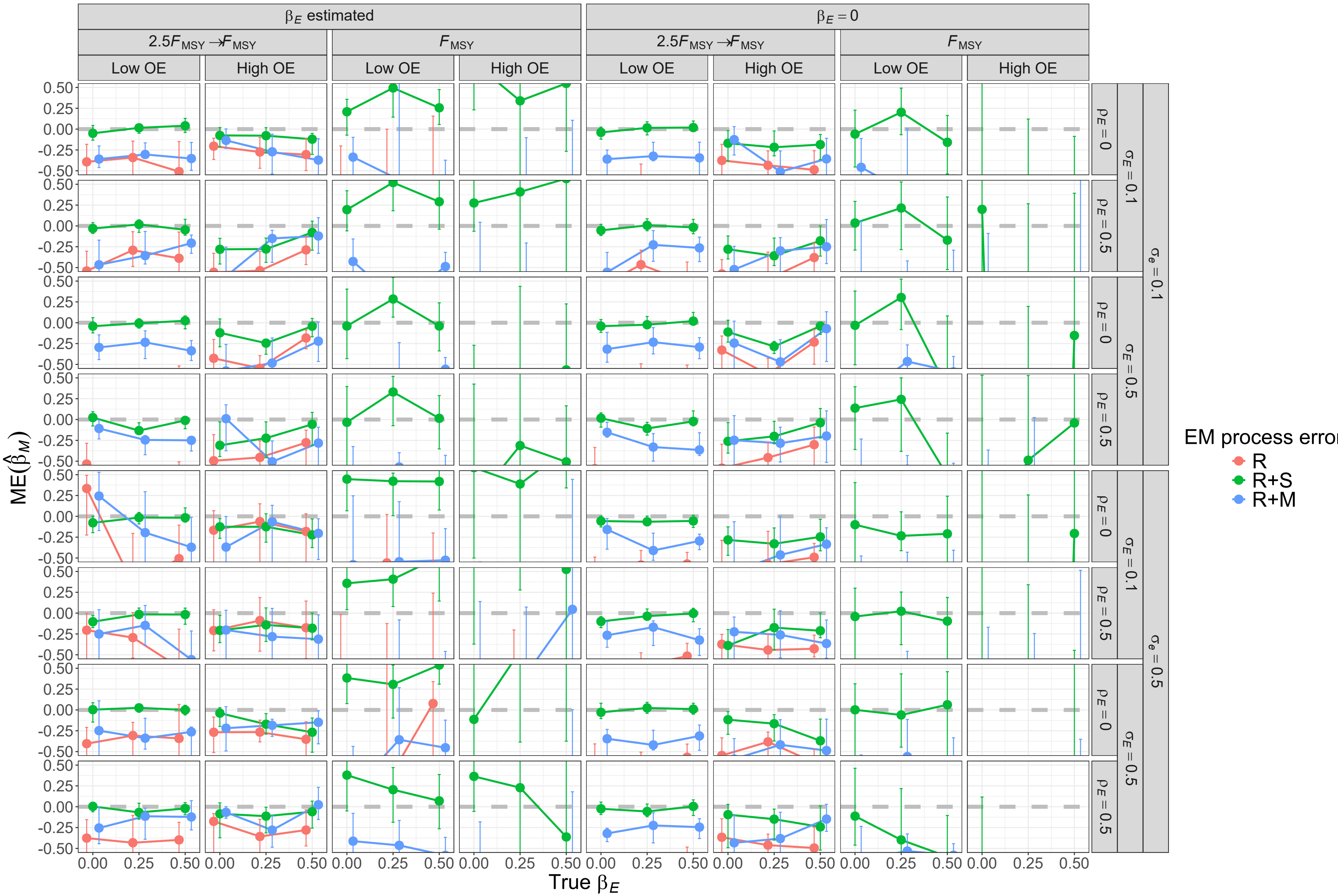



Fig. S34. For R+S OMs, median error (ME) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). Vertical lines represent 95% confidence intervals.

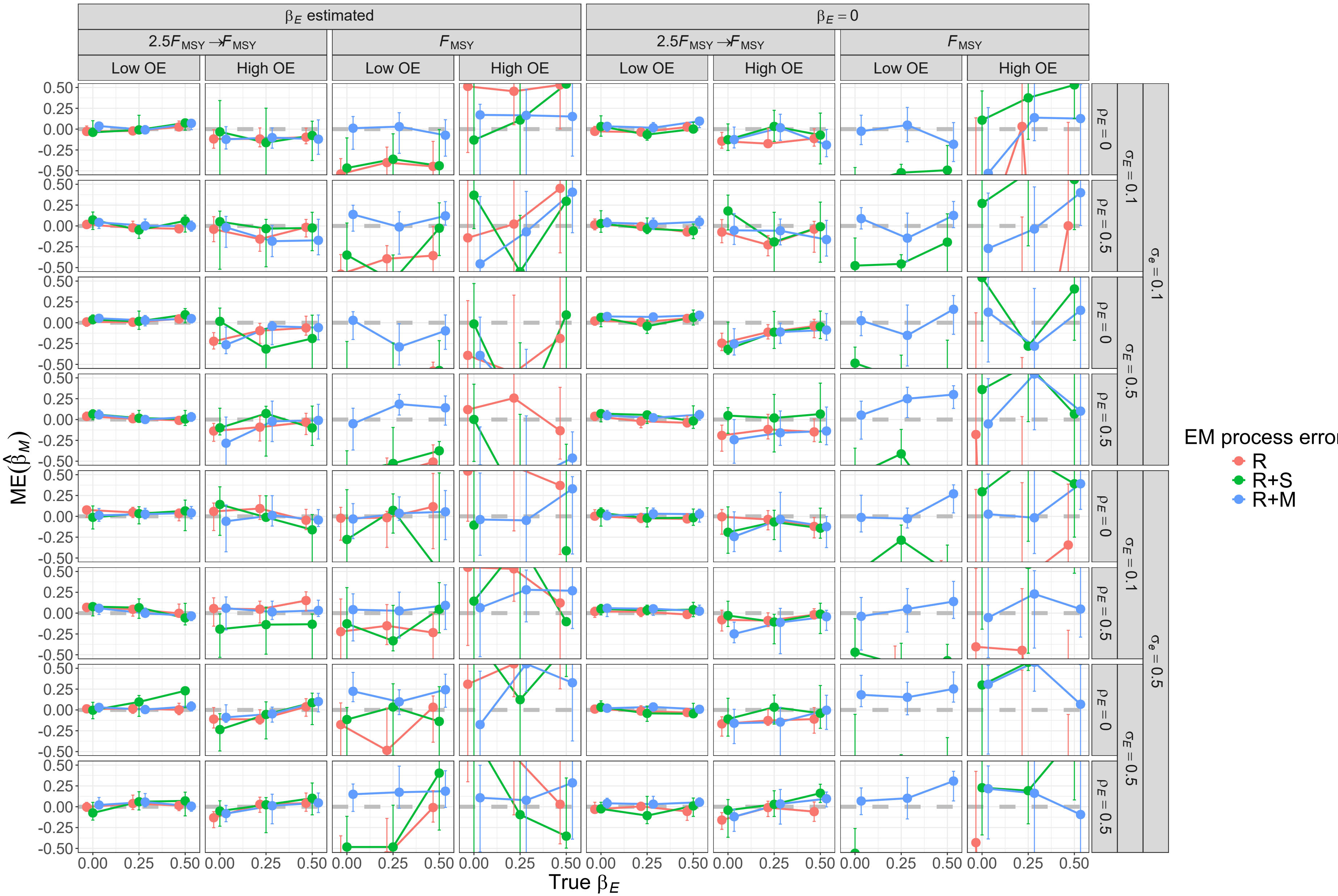



Fig. S35. For R+M OMs, median error (ME) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). Vertical lines represent 95% confidence intervals.

## Median natural mortality parameter standard error estimation bias

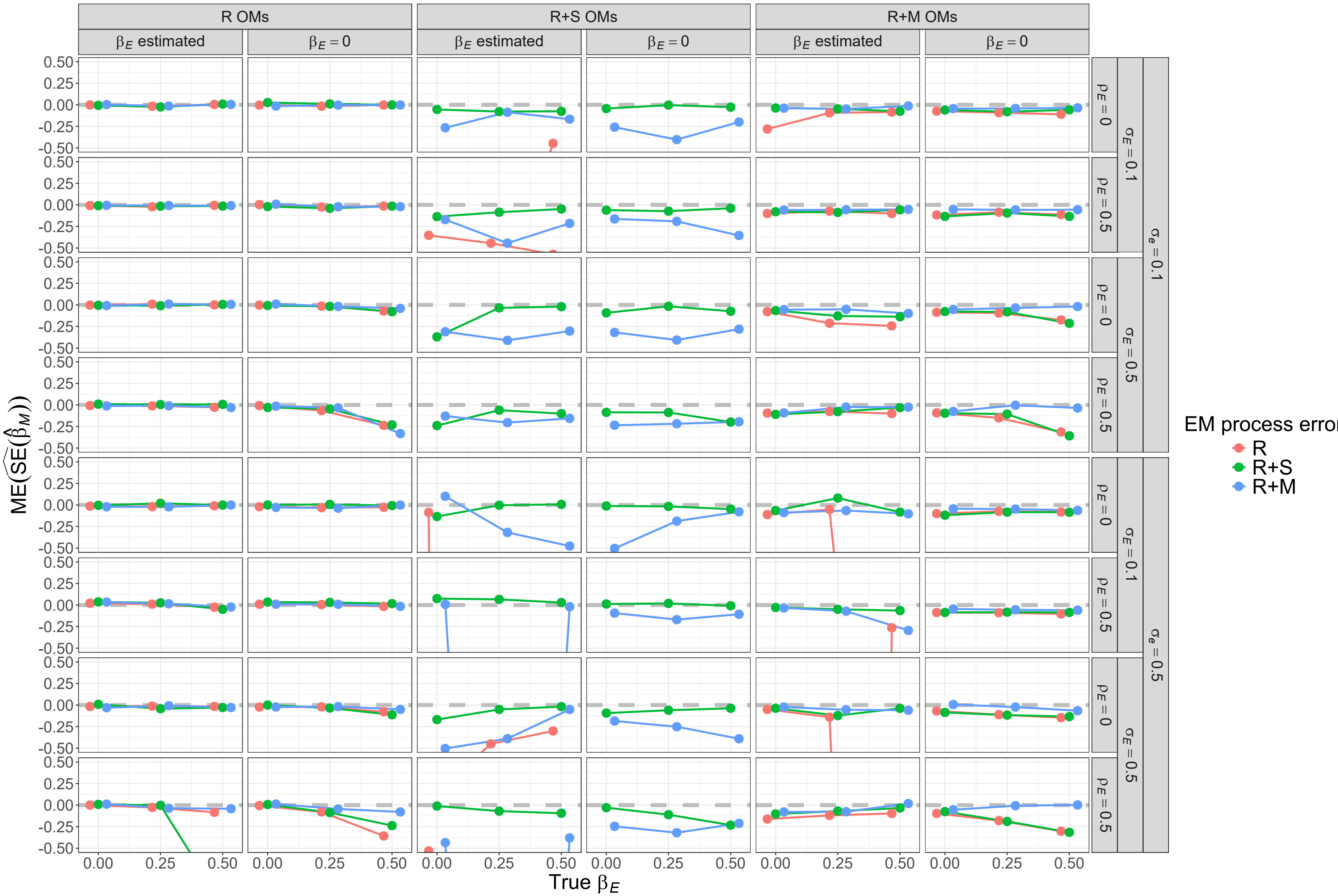



Fig. S36. Median error (ME) of Hessian-based estimates of standard error for median natural mortality parameter $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of the covariate effect ($\beta_E = 0$ or estimated). All OMs had low observation error and contrast in fishing mortality. True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

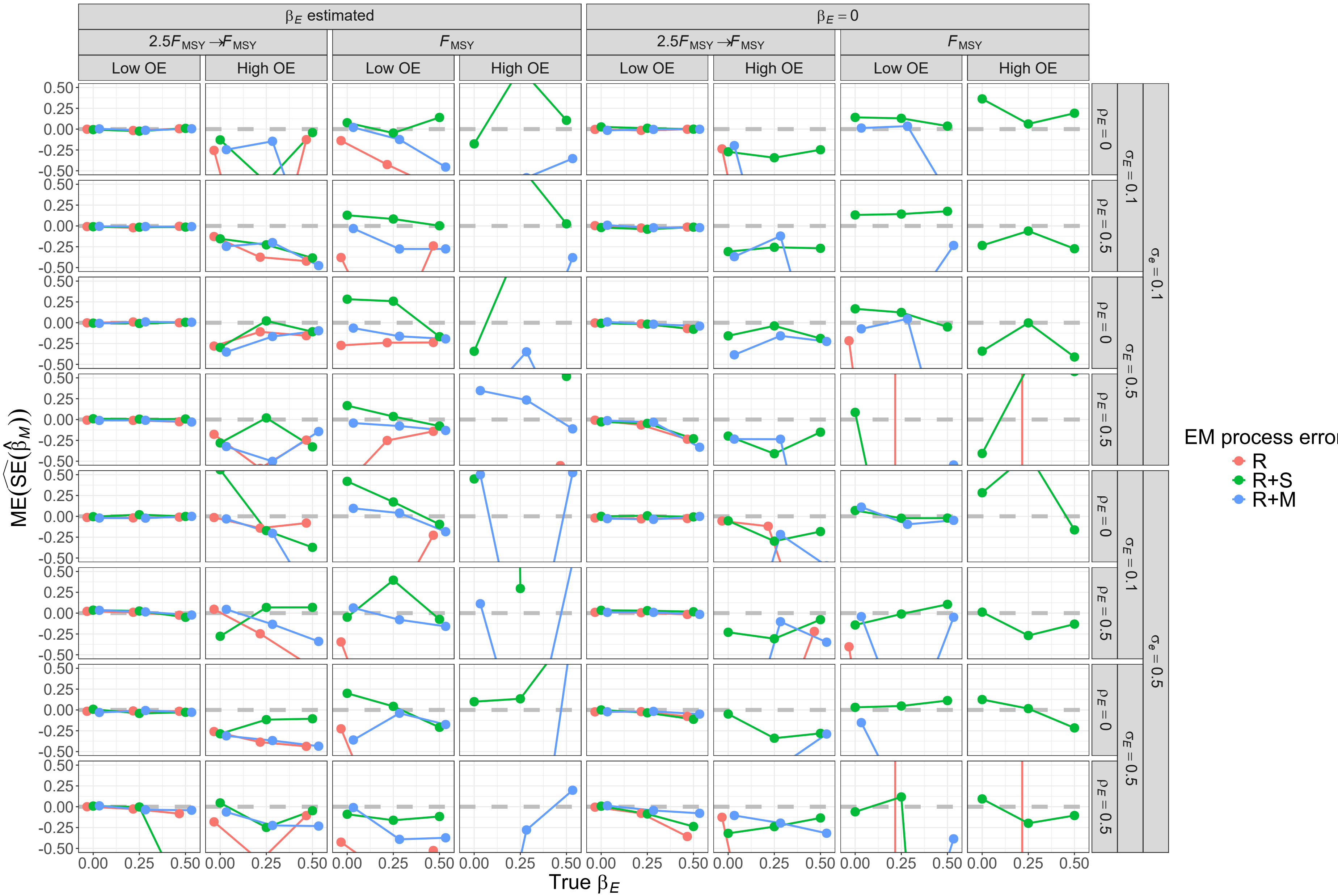



Fig. S37. For R OMs, median error (ME) of Hessian-based estimates of standard error for median natural mortality parameter $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of the covariate effect ($\beta_E = 0$ or estimated). True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

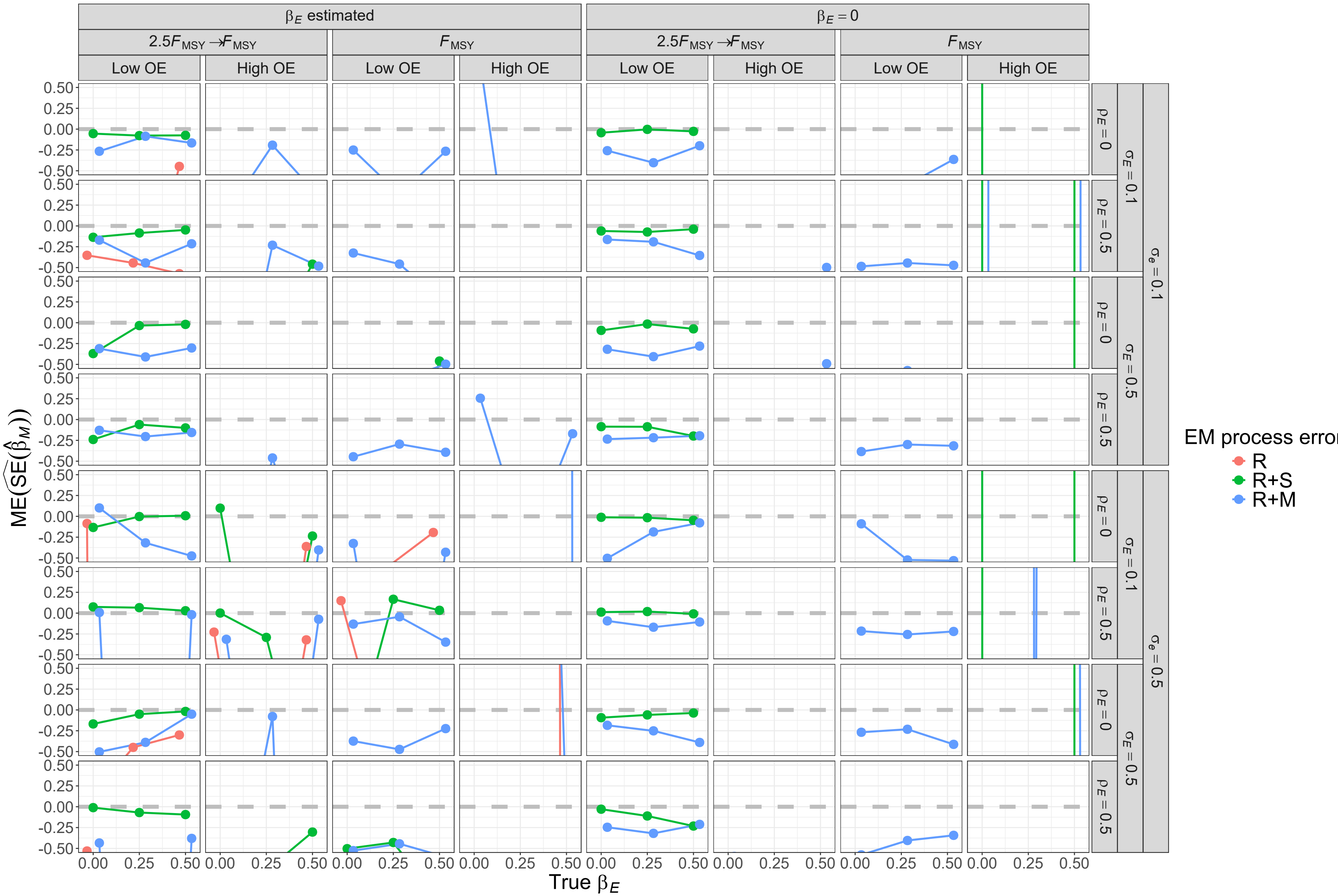



Fig. S38. For R+S OMs, median error (ME) of Hessian-based estimates of standard error for median natural mortality parameter $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of the covariate effect ($\beta_E = 0$ or estimated). True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

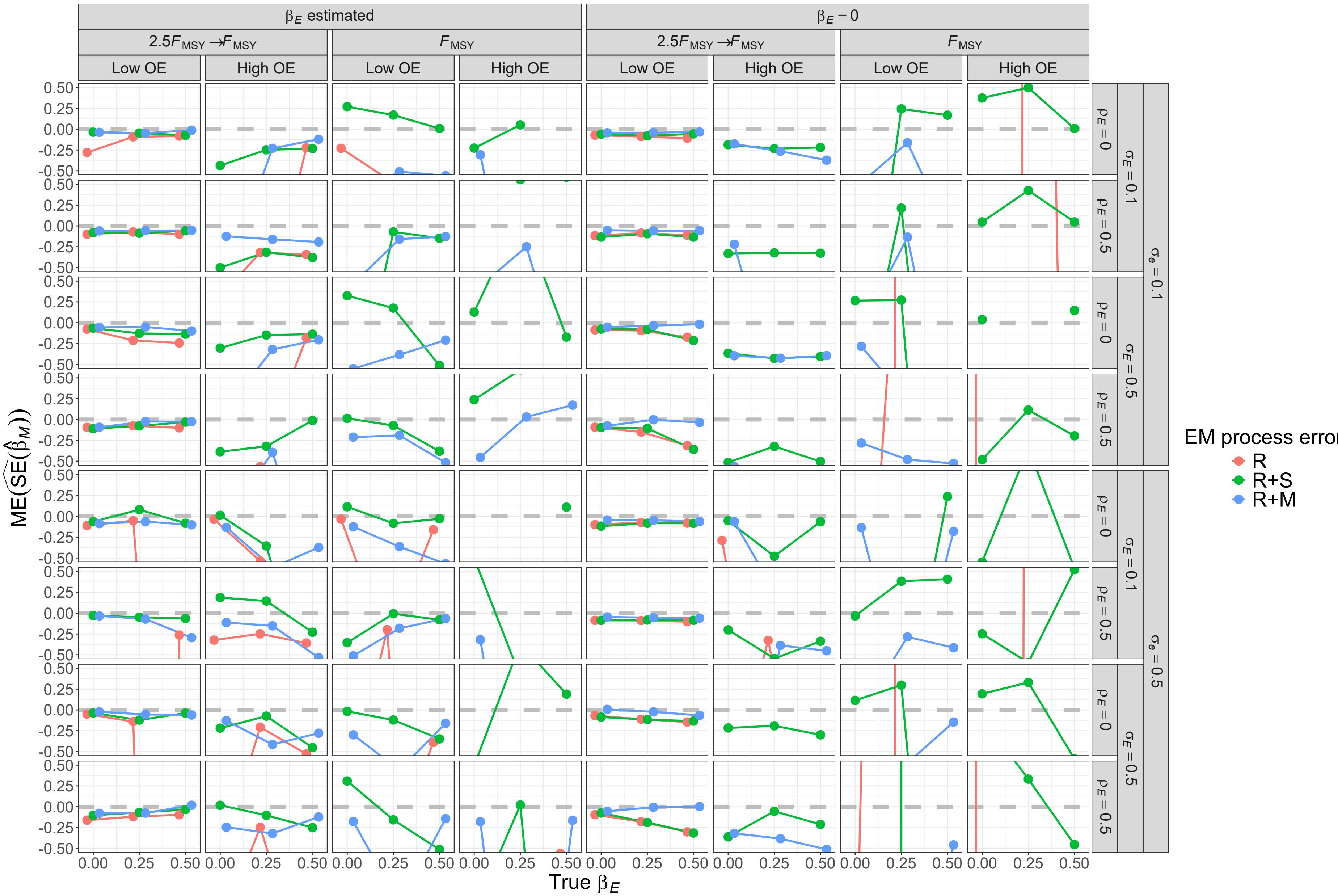



Fig. S39. For R+M OMs, median error (ME) of Hessian-based estimates of standard error for median natural mortality parameter $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of the covariate effect ($\beta_E = 0$ or estimated). True standard error is defined as the mean of the standard error estimates accross converged fits to simulated data sets for a given OM scenario.

## Median Natural mortality parameter confidence interval coverage

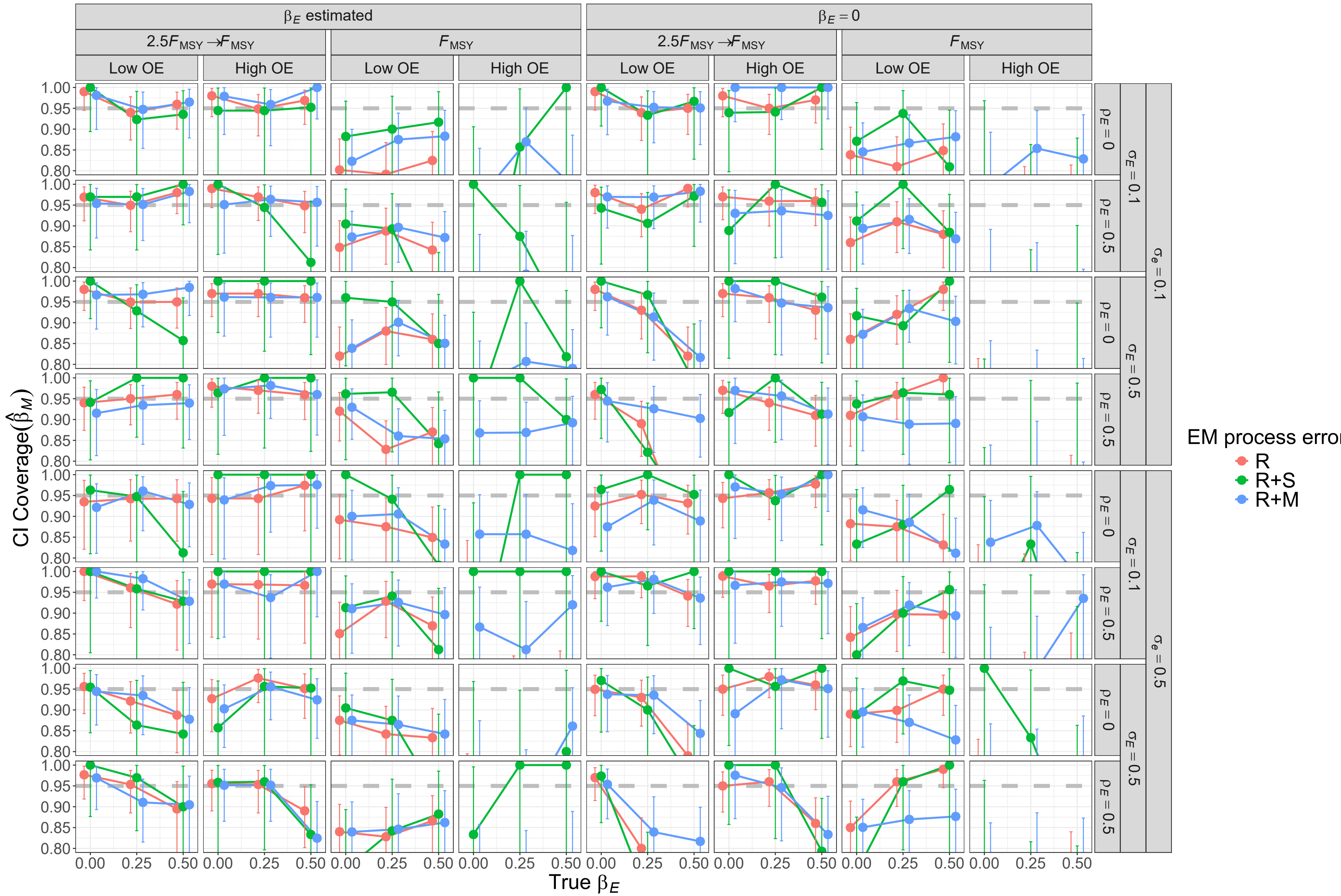



Fig. S40. For R OMs, probability of 95% confidence interval for $\beta_M$ containing the true value for EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). Vertical lines represent 95% confidence intervals.

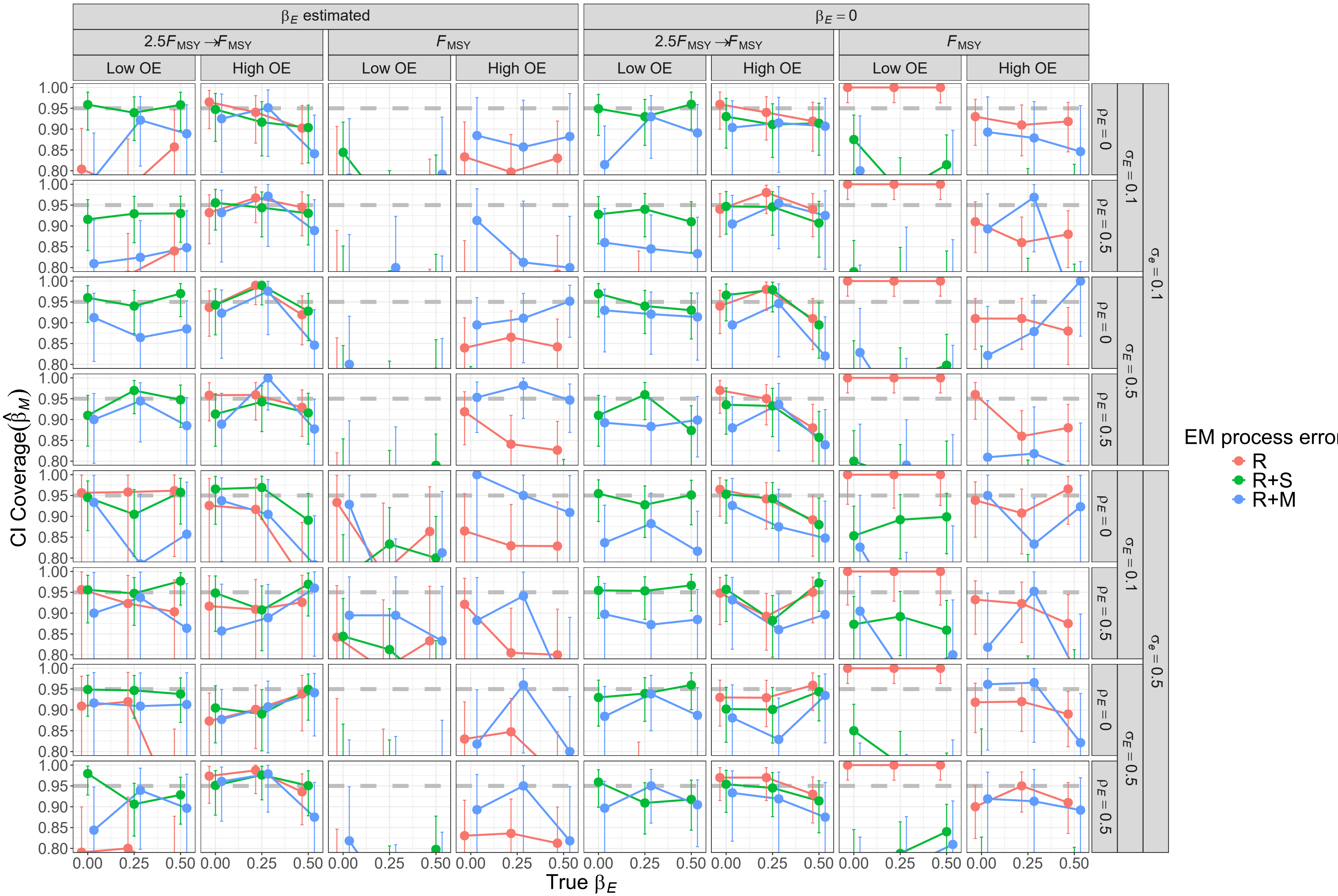



Fig. S41. For R+S OMs, probability of 95% confidence interval for $\beta_M$ containing the true value for EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). Vertical lines represent 95% confidence intervals.

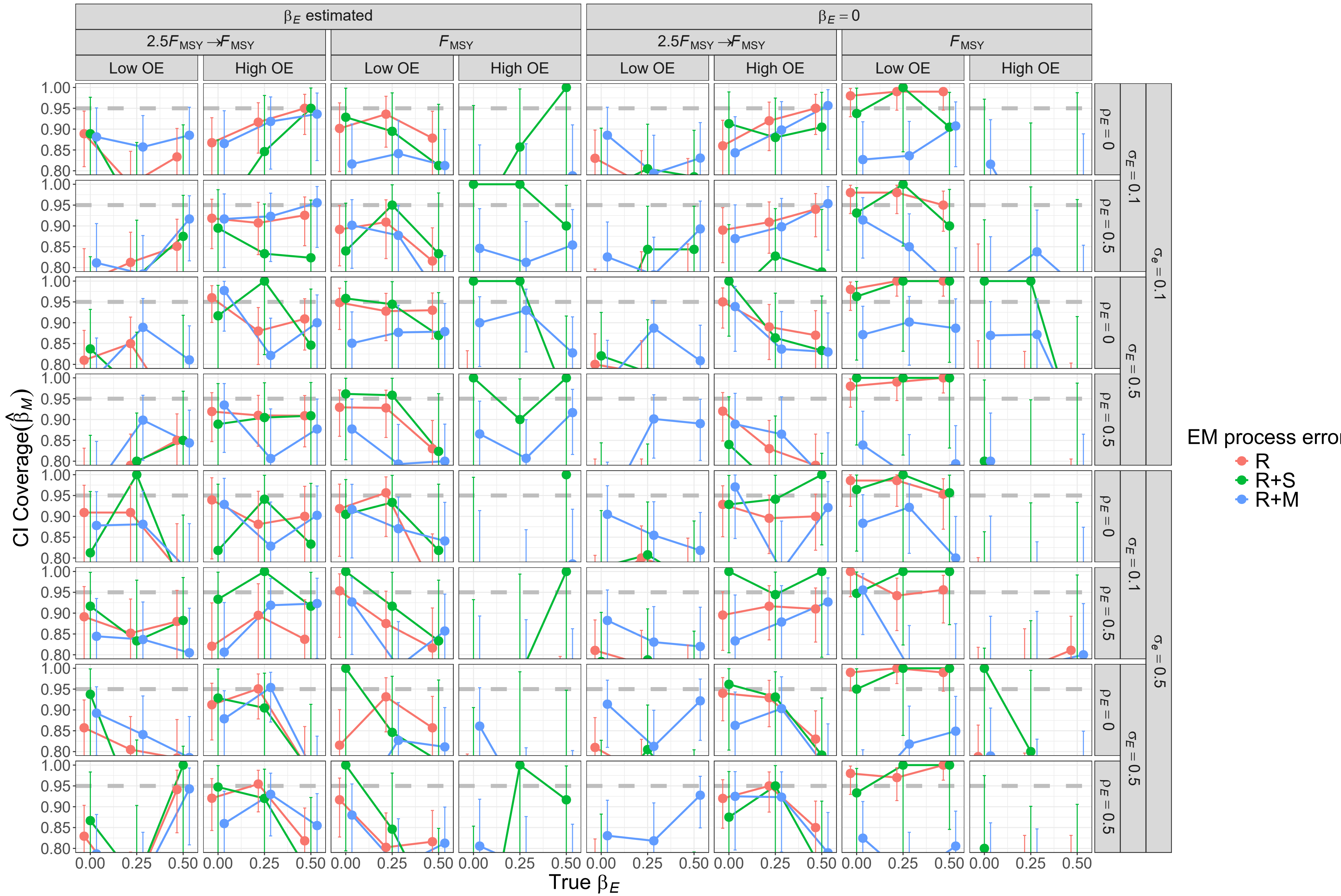



Fig. S42. For R+M OMs, probability of 95% confidence interval for $\beta_M$ containing the true value for EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). Vertical lines represent 95% confidence intervals.

## Median natural mortality parameter estimate and standard error example

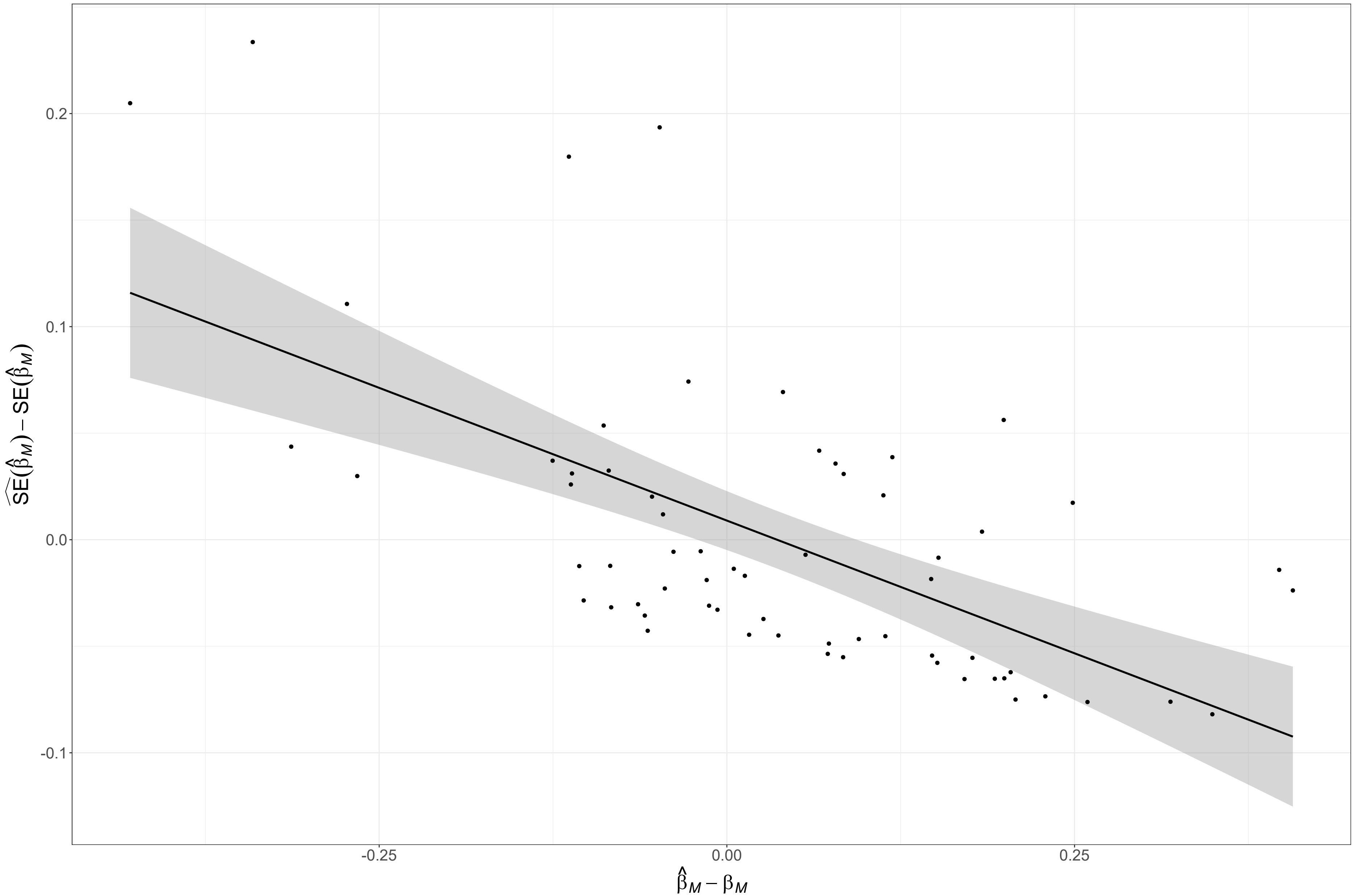



Fig. S43. Negative correlation of $\beta_M$ estimates and Hessian-based standard error estimates for EM that also estimates the covariate effect and has correct R+M process error assumption fitted to simulated data from the OM with R+M process errors, temporal contrast in fishing pressure, low observation uncertainty for both population (*LowOE*) and covariate observations ($\sigma_e = 0.1$), high and uncorrelated temporal variability in the true covariate ($\sigma_E = 0.5$ and $\rho_E = 0$), and the strongest covariate

## Median Natural mortality parameter RMSE

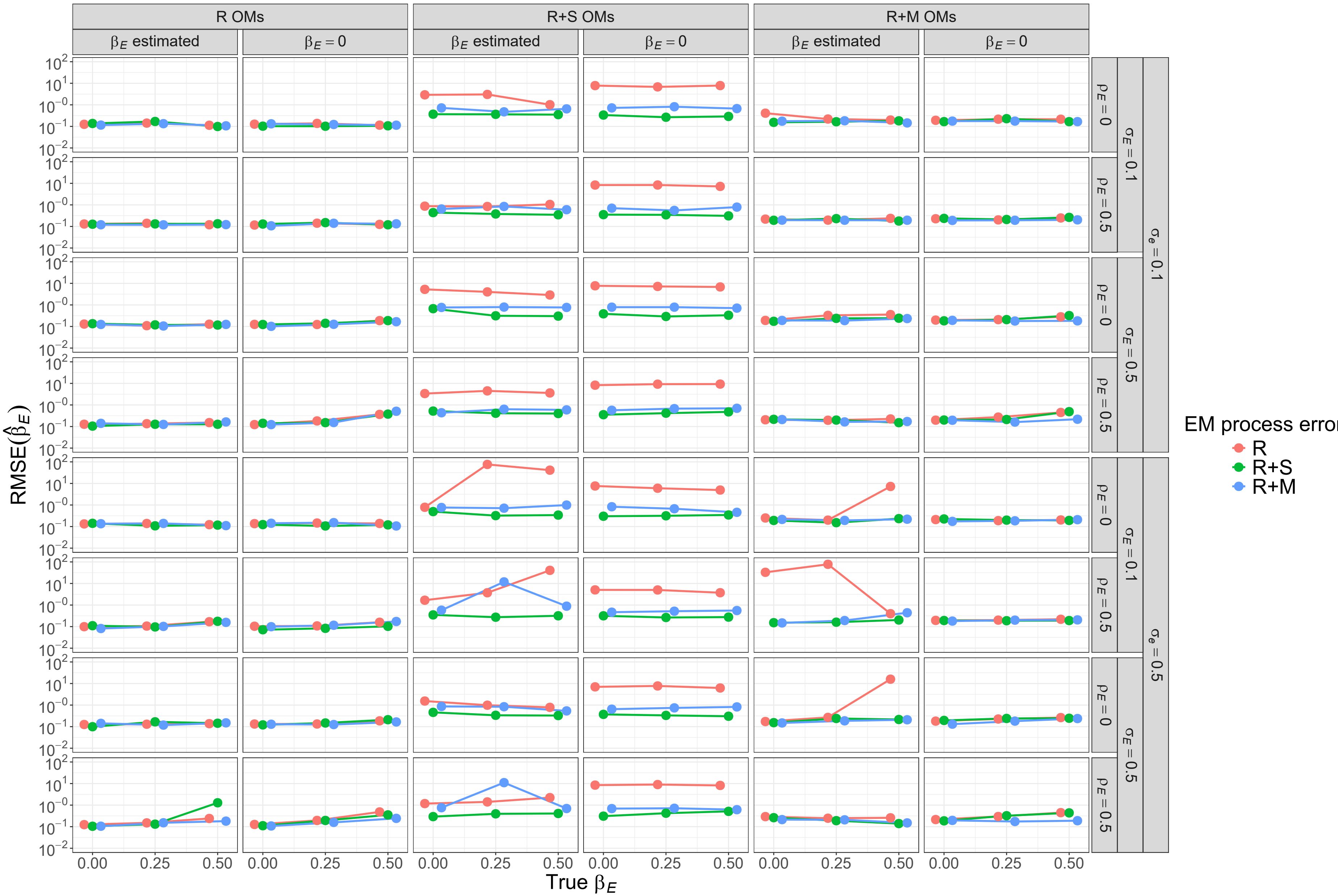



Fig. S44. Root mean square error (RMSE) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated). All OMs had low observation error for population observations and contrast in fishing mortality.

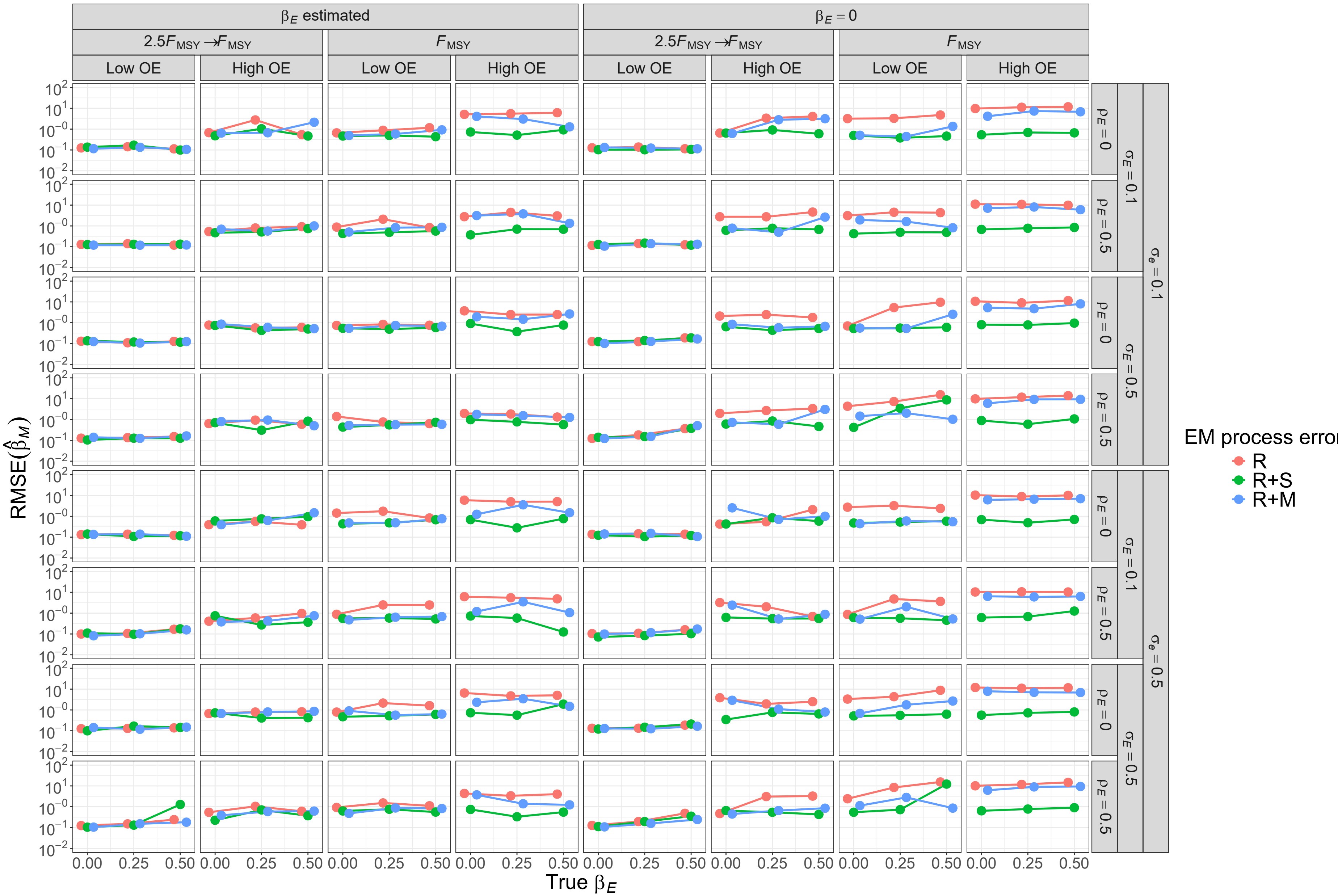



Fig. S45. For R OMs, root mean square error (RMSE) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated).

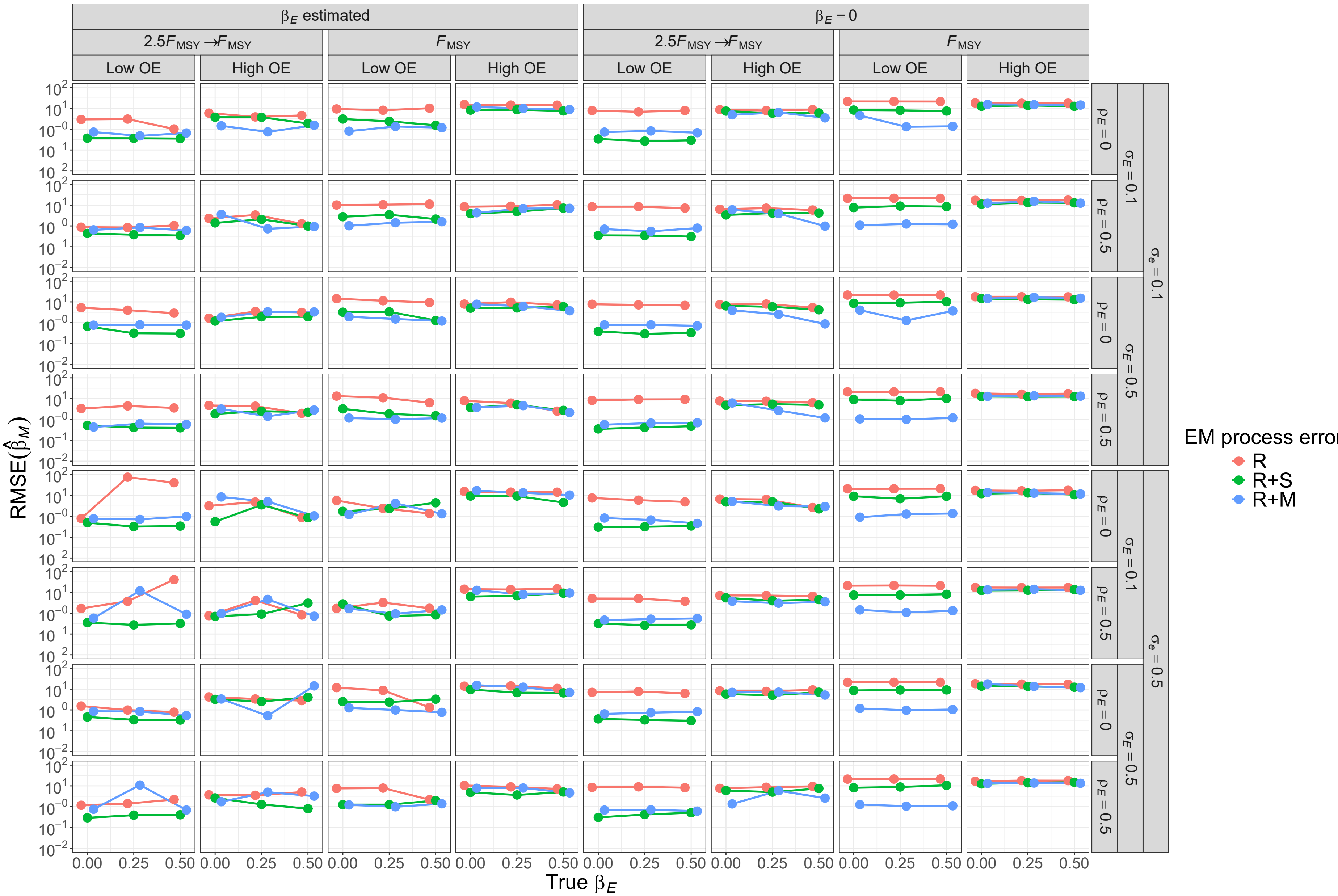



Fig. S46. For R+S OMs, root mean square error (RMSE) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated).

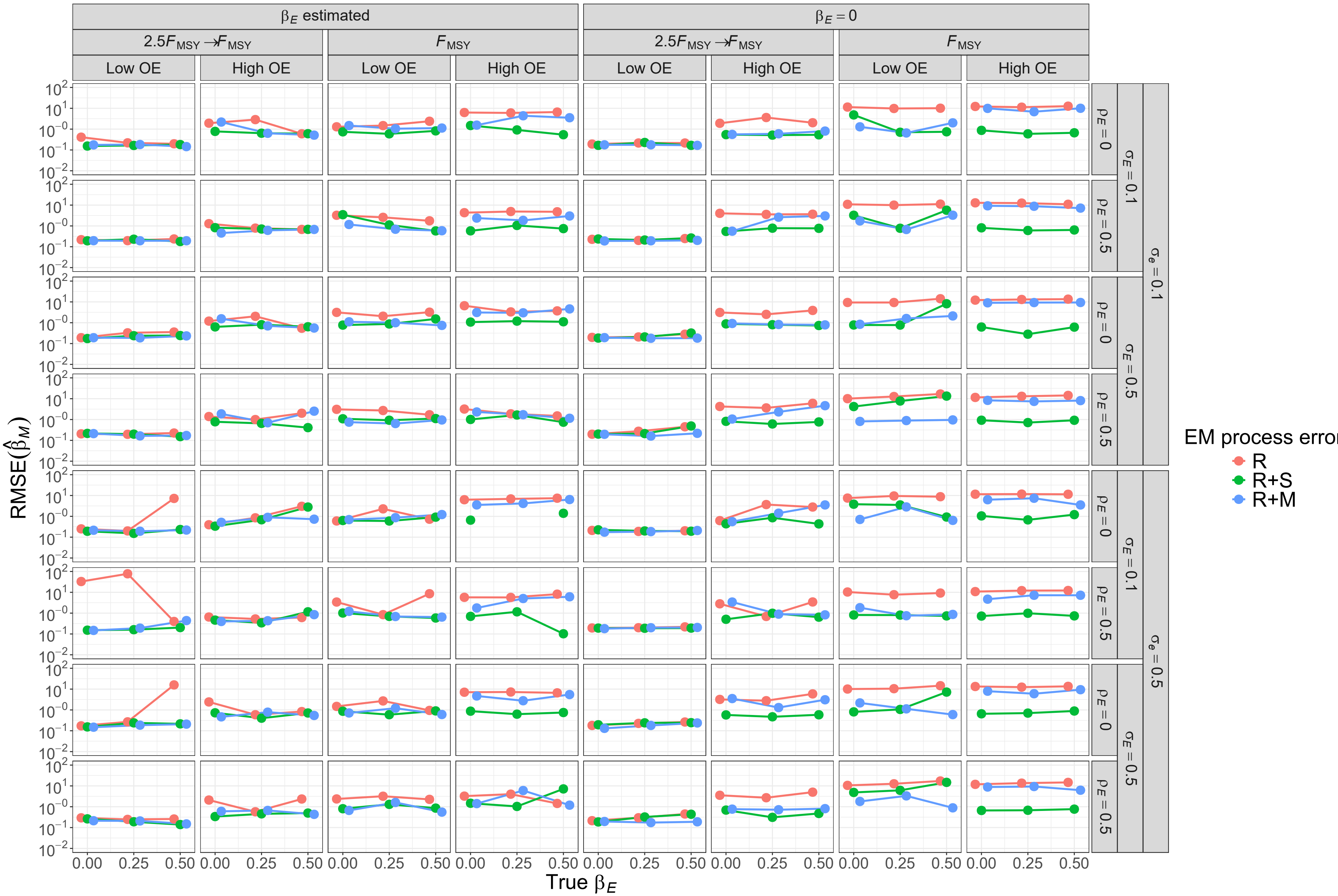



Fig. S47. For R+M OMs, root mean square error (RMSE) of estimates of $\beta_M$ from fitting EMs with alternative process error assumptions and treatment of covariate effect ($\beta_E = 0$ or estimated).

## Terminal year natural mortality bias

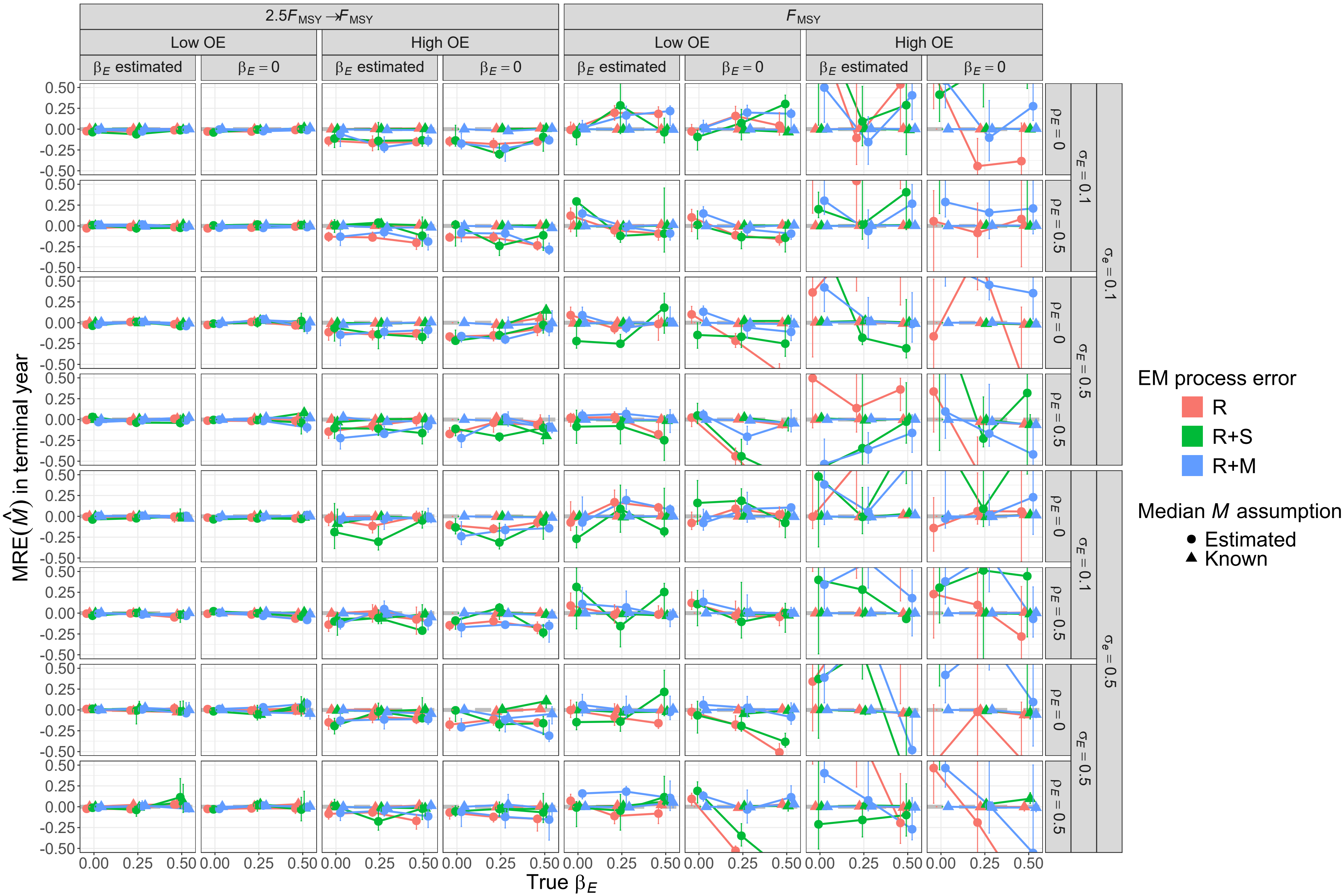



Fig. S48. For R OMs, median relative error (MRE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

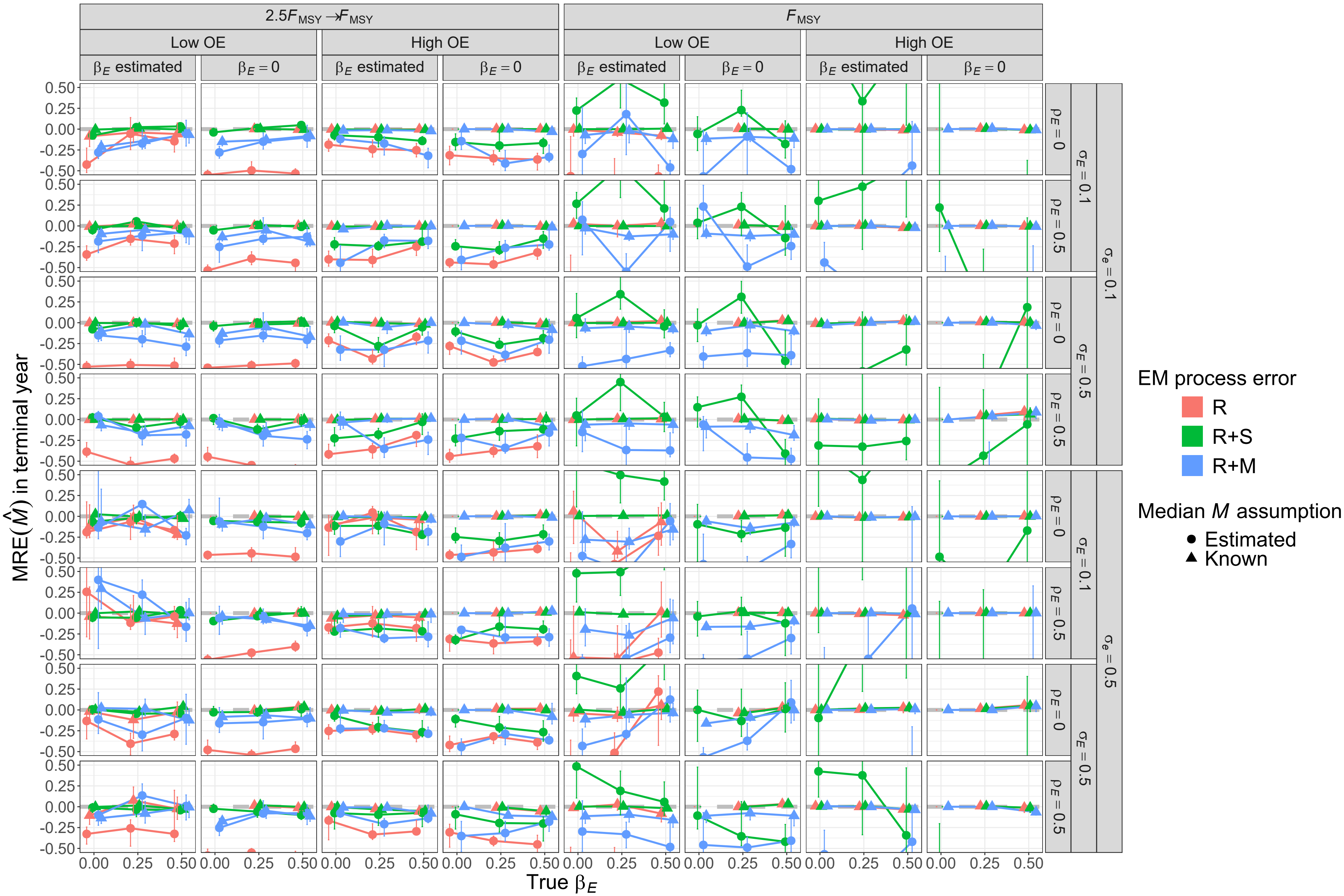



Fig. S49. For R+S OMs, median relative error (MRE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

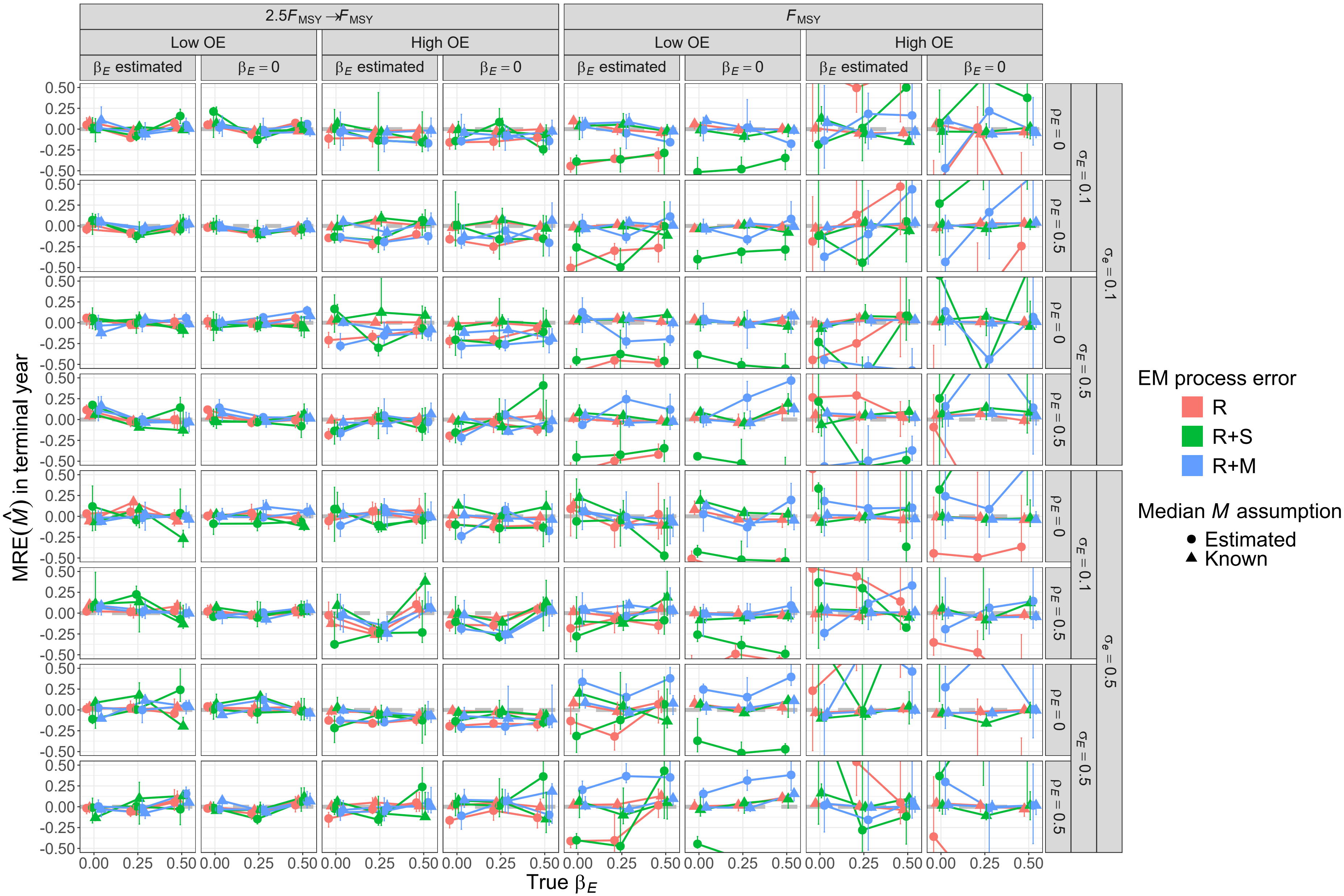



Fig. S50. For R+M OMs, median relative error (MRE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

## Terminal year natural mortality RMSE

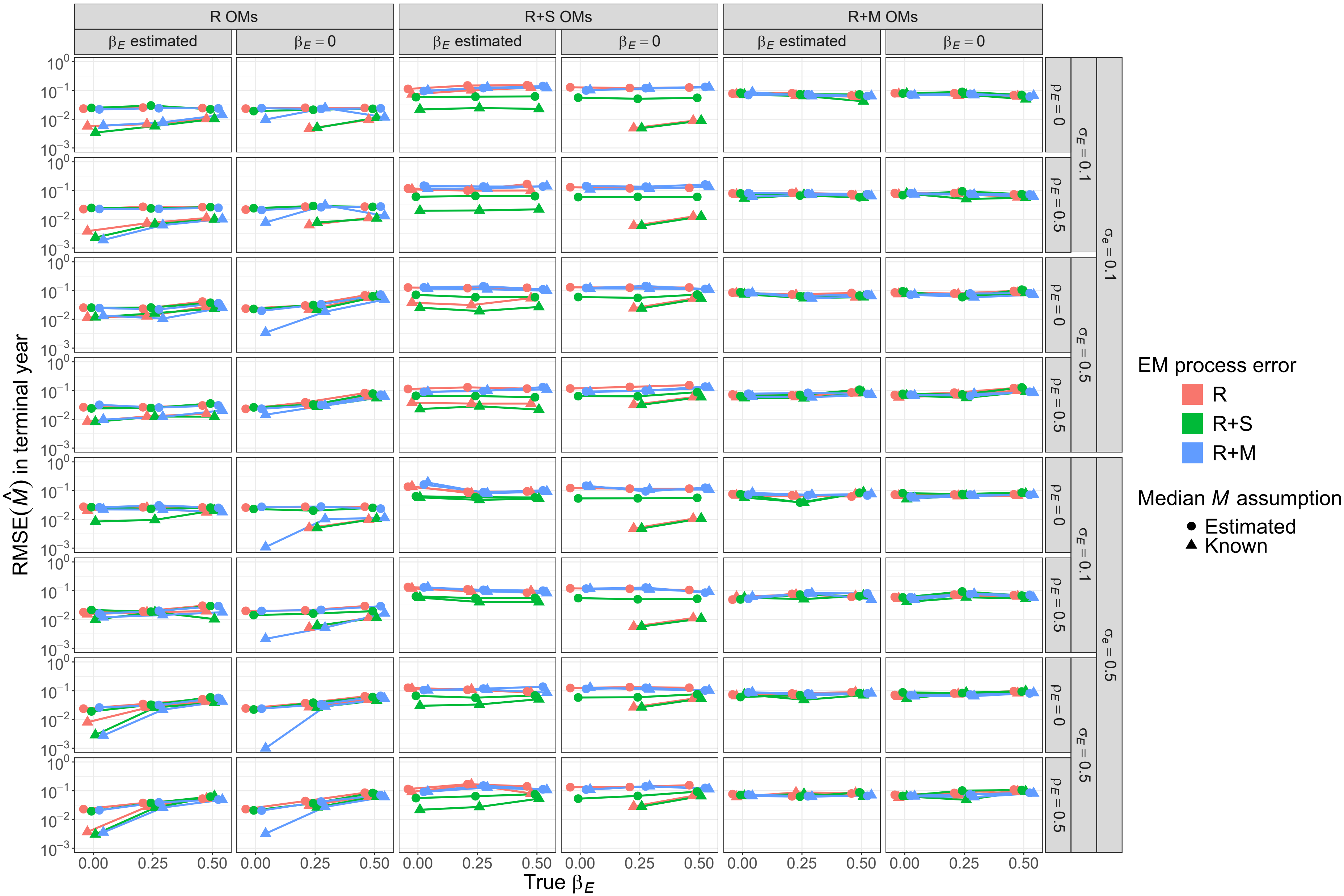



Fig. S51. Root mean square error (RMSE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low population observation error and contrast in fishing mortality.

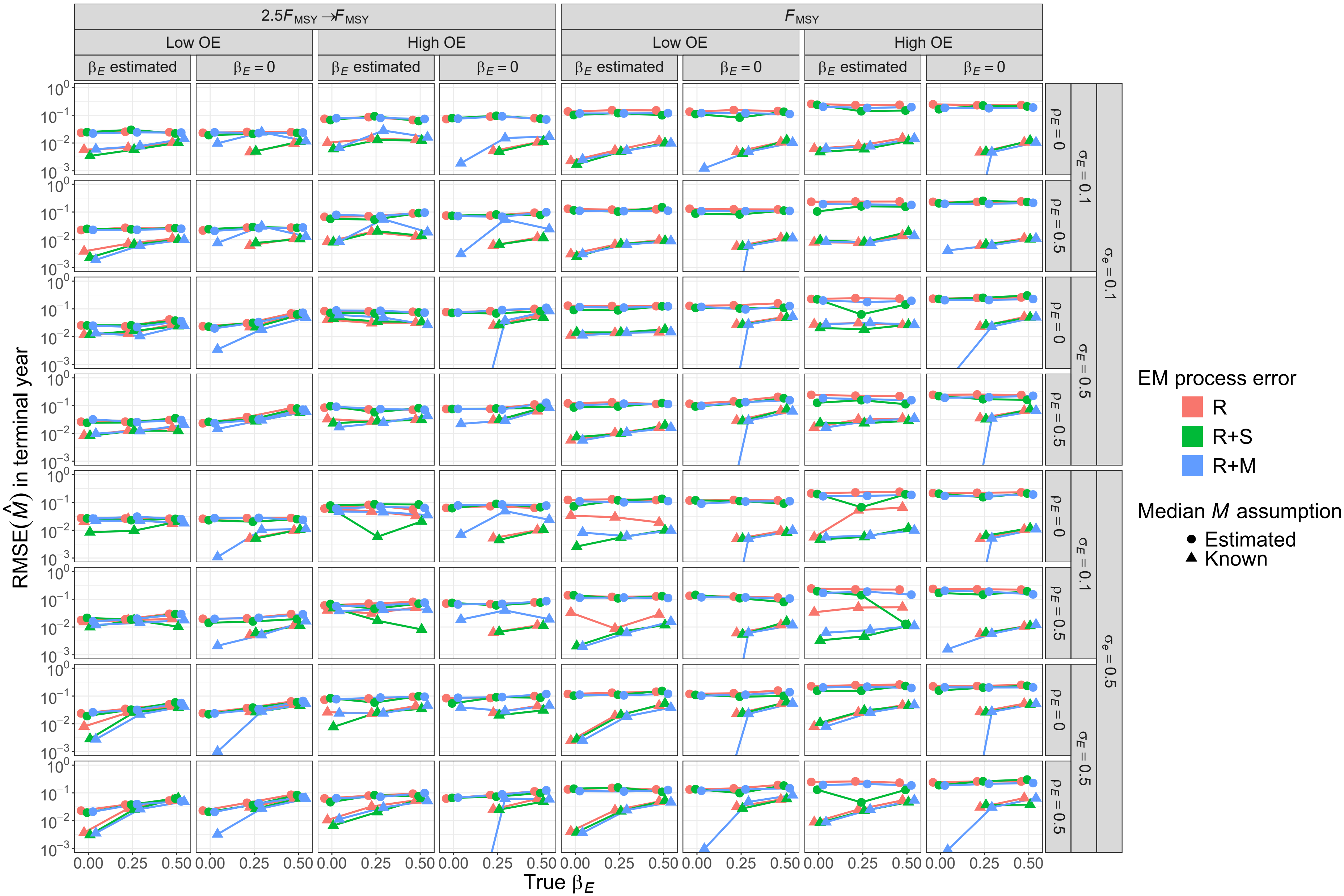



Fig. S52. For R OMs, root mean square error (RMSE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

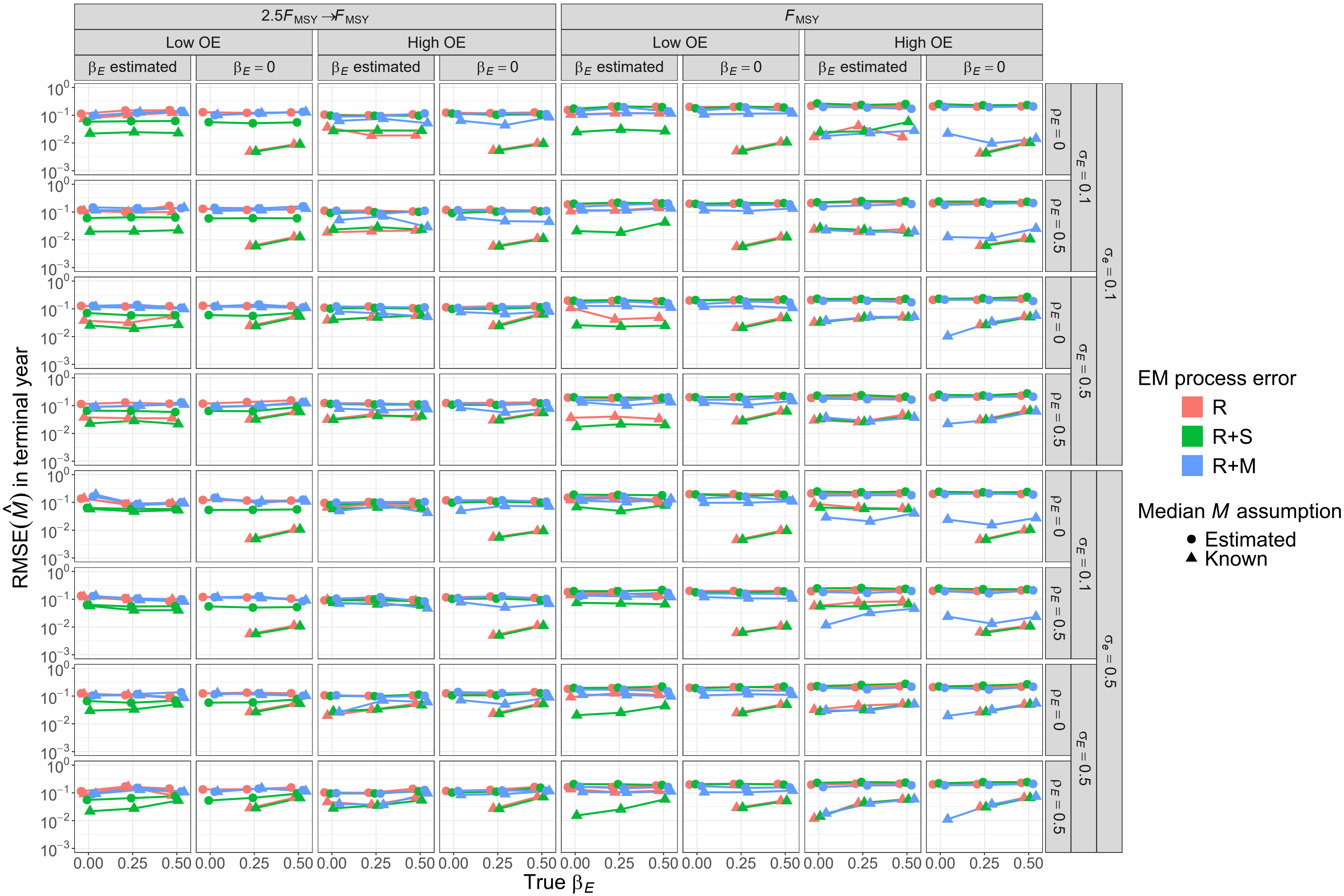



Fig. S53. For R+S OMs, root mean square error (RMSE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

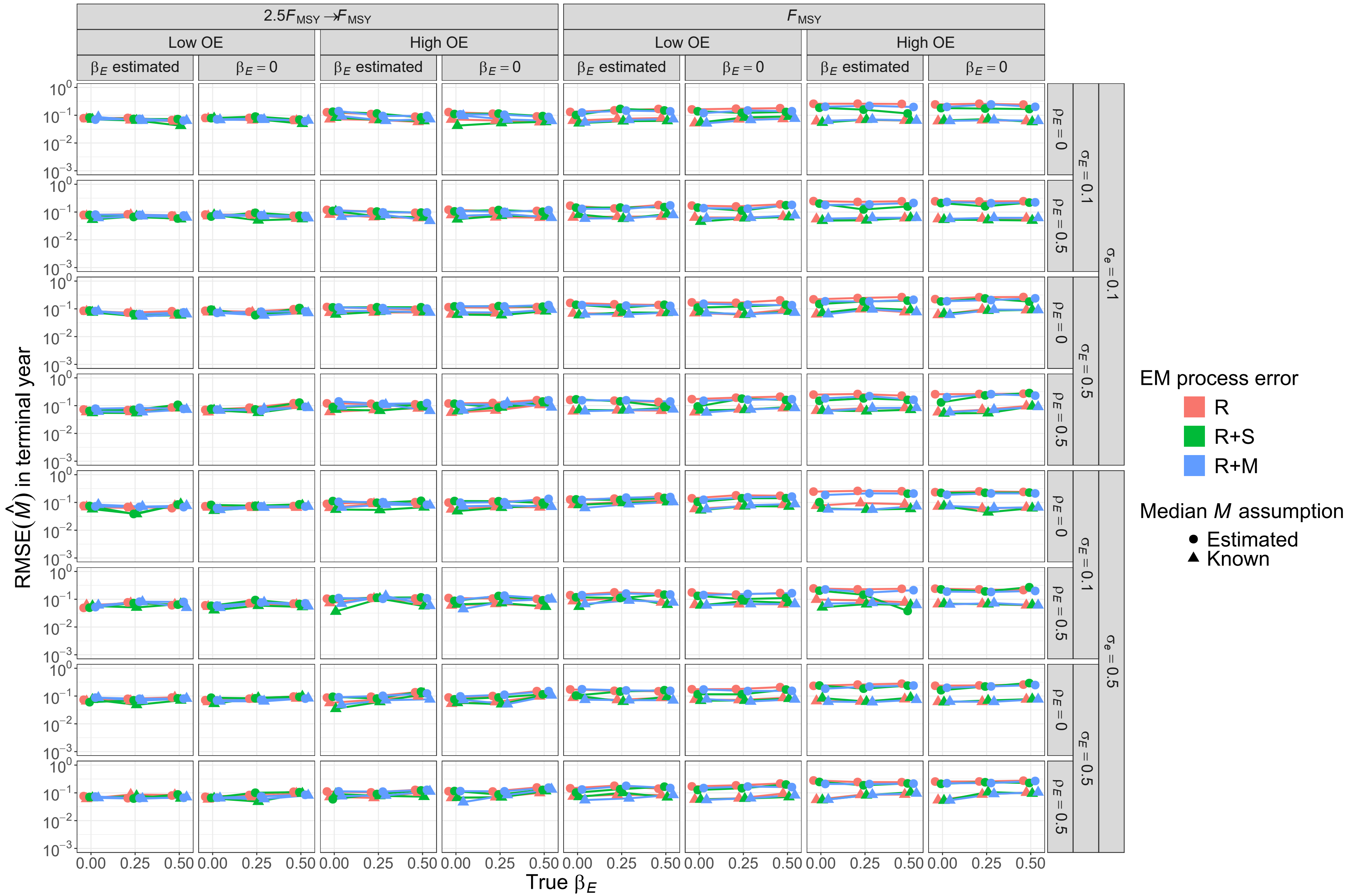



Fig. S54. For R+M OMs, root mean square error (RMSE) of estimates of natural mortality rate in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

## Terminal year spawning stock biomass bias

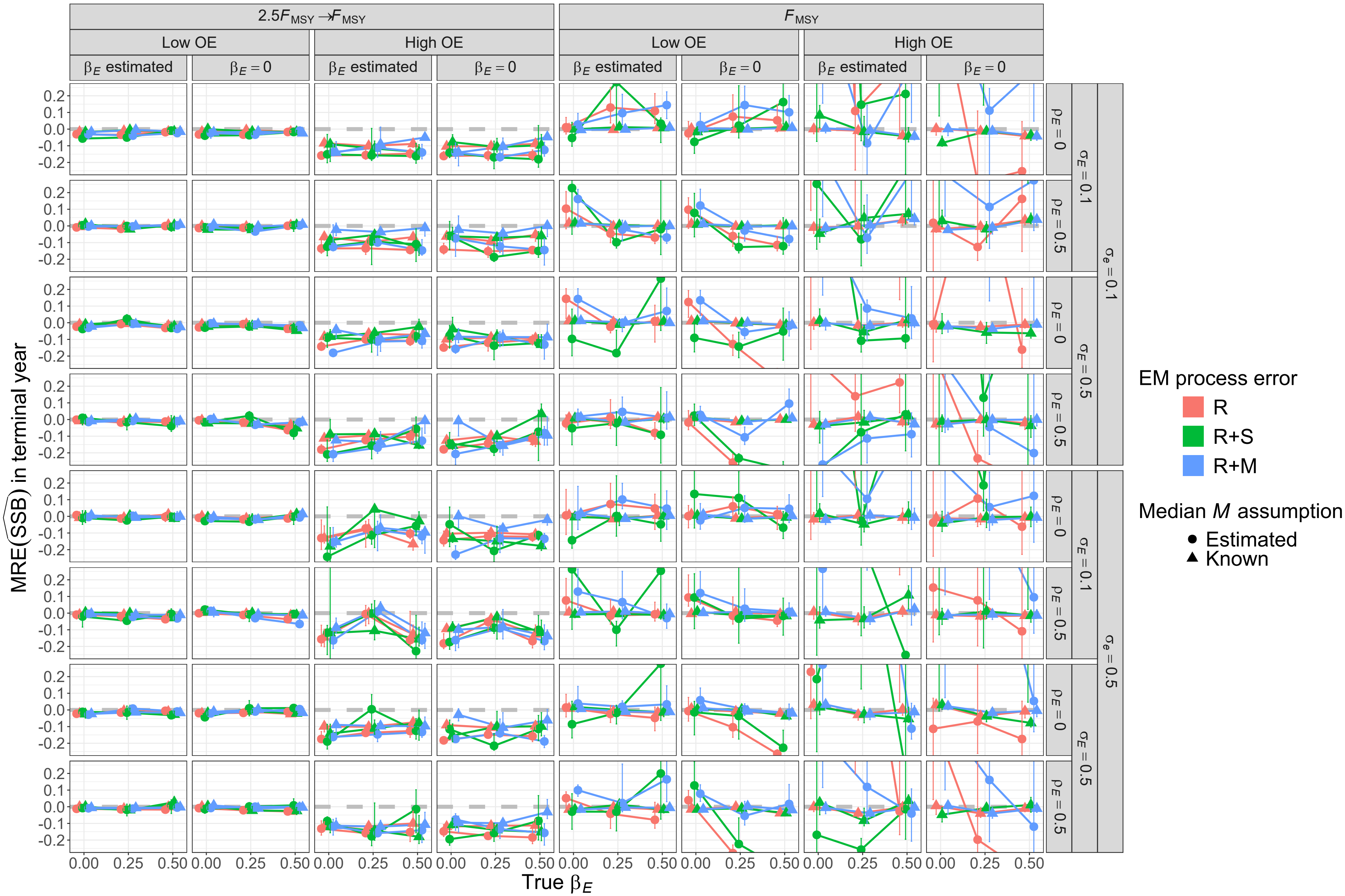



Fig. S55. For R OMs, median relative error (MRE) of estimates of spawning stock biomass (SSB) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

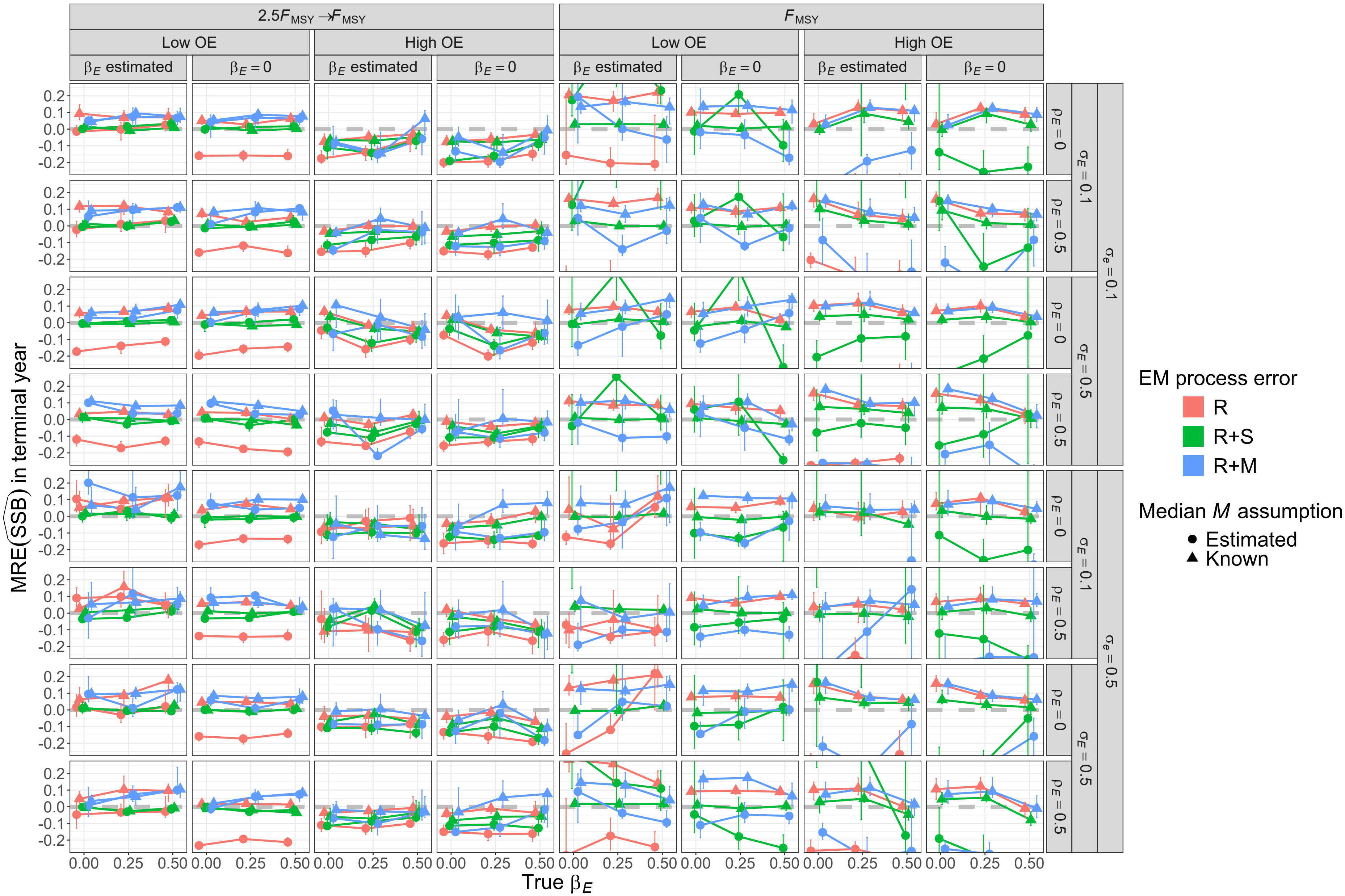



Fig. S56. For R+S OMs, median relative error (MRE) of estimates of spawning stock biomass (SSB) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

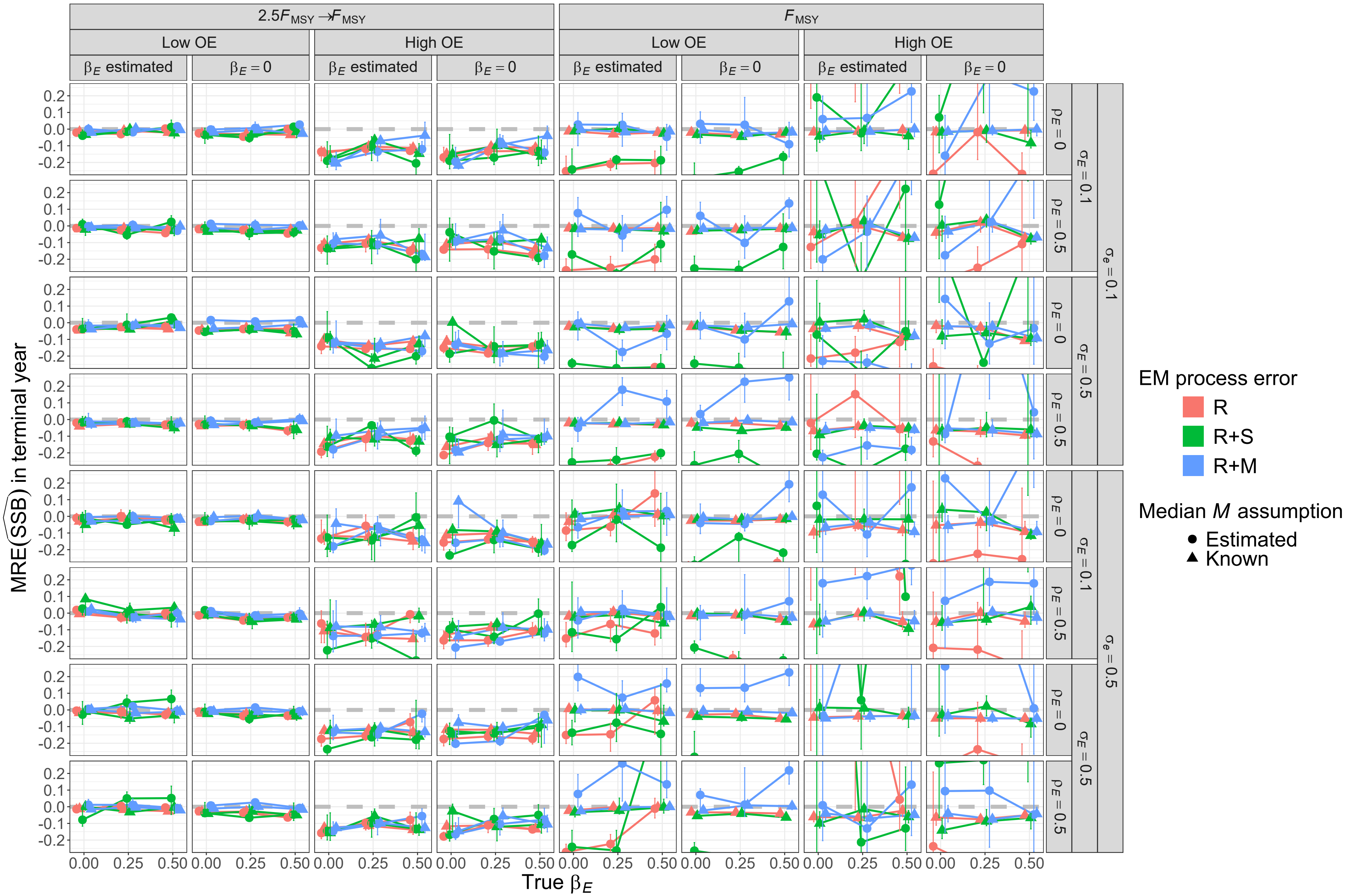



Fig. S57. For R+M OMs, median relative error (MRE) of estimates of spawning stock biomass (SSB) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

## Terminal year spawning stock biomass RMSE

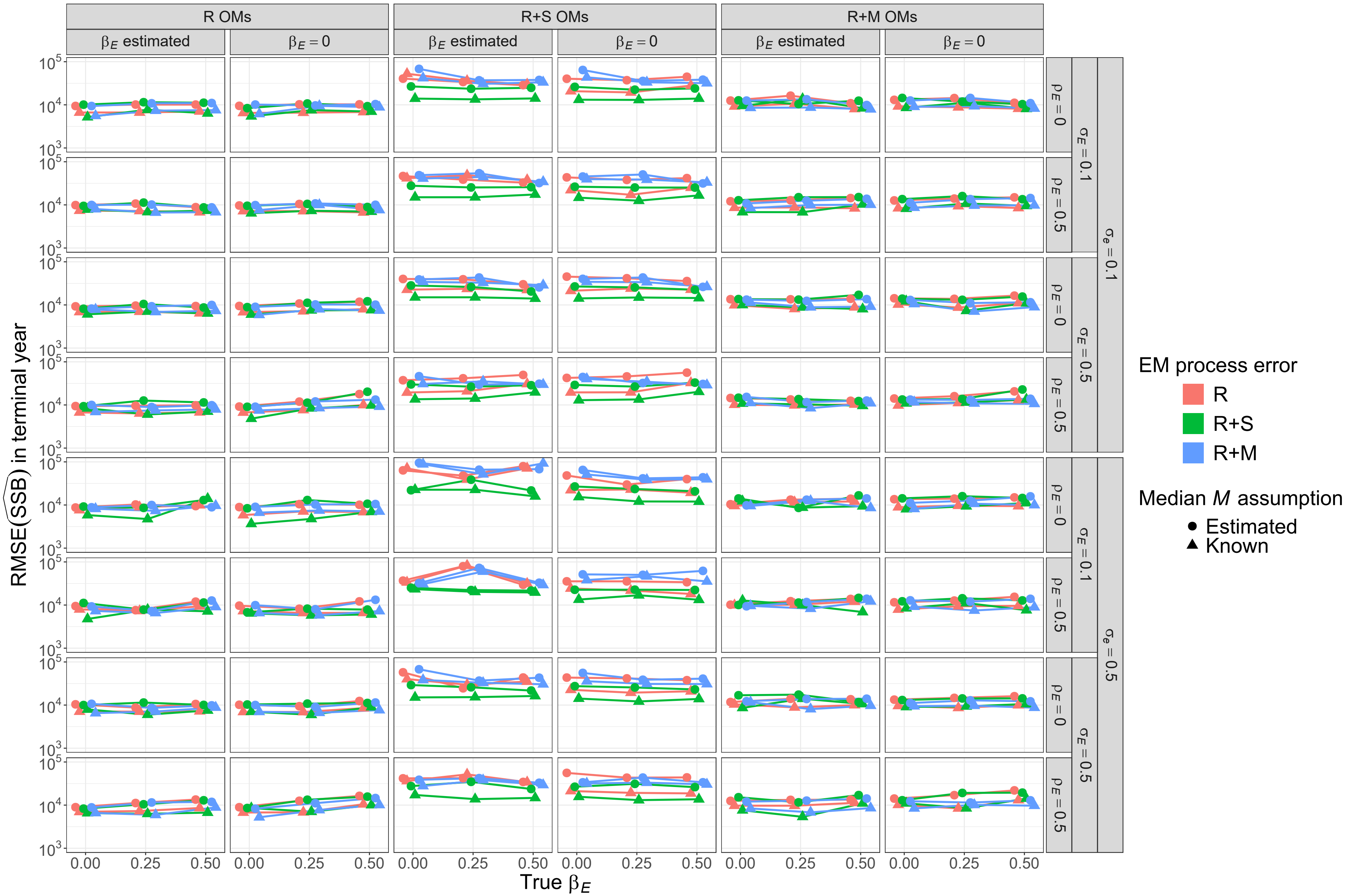



Fig. S58. Root mean square error (RMSE) of estimates of spawning stock biomass in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low population observation error and contrast in fishing mortality.

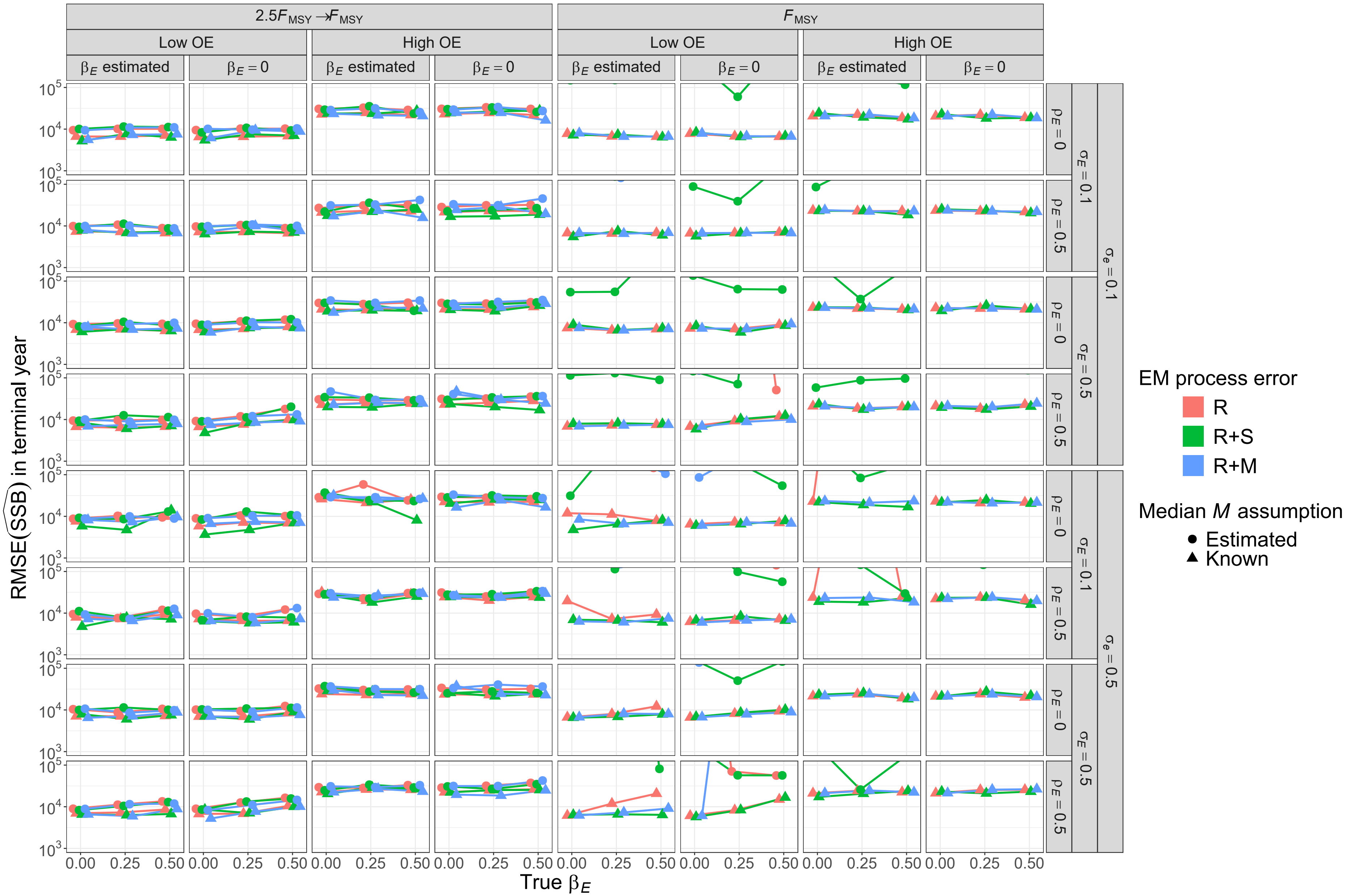



Fig. S59. For R OMs, root mean square error (RMSE) of estimates of spawning stock biomass in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

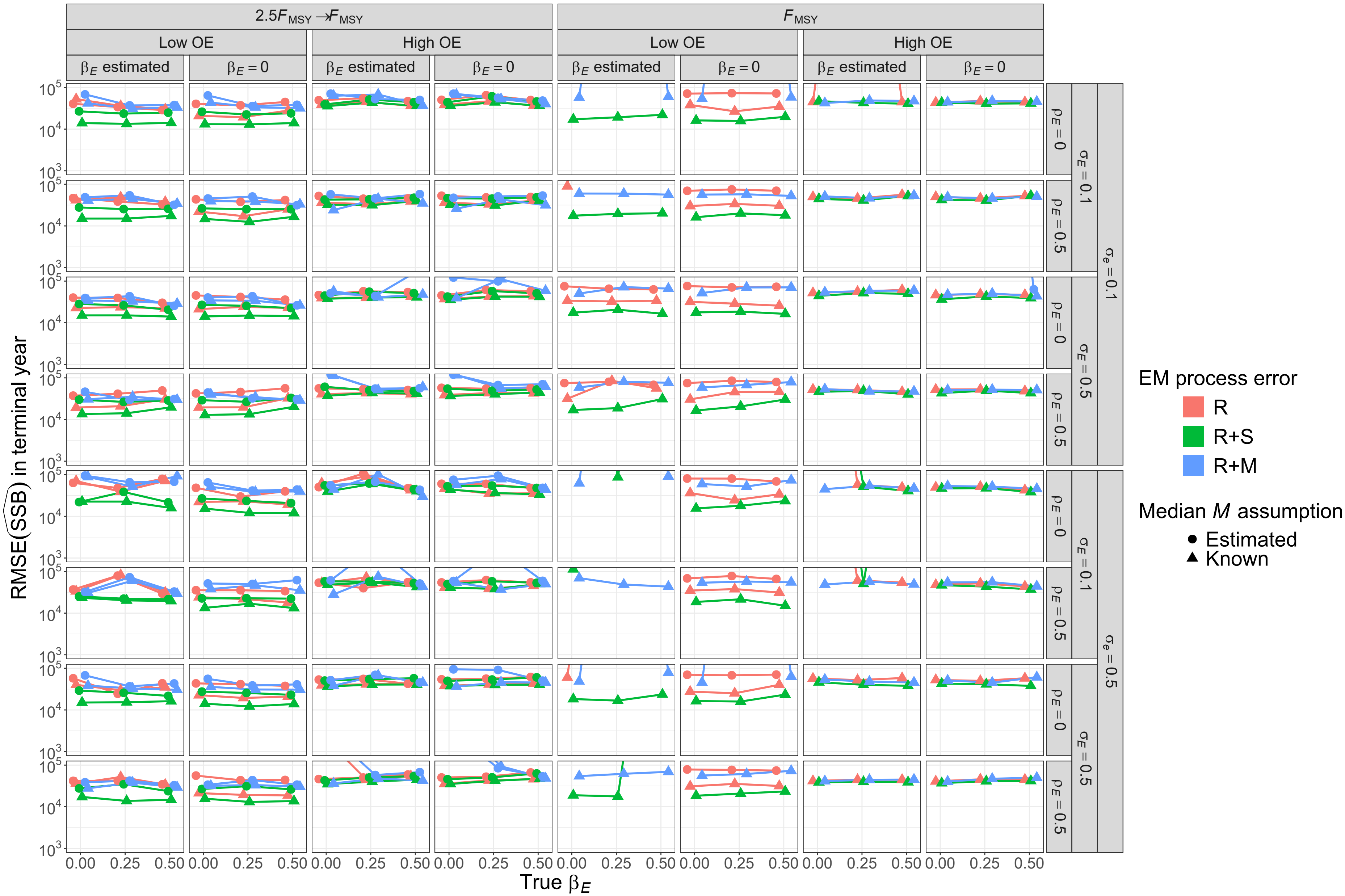



Fig. S60. For R+S OMs, root mean square error (RMSE) of estimates of spawning stock biomass in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

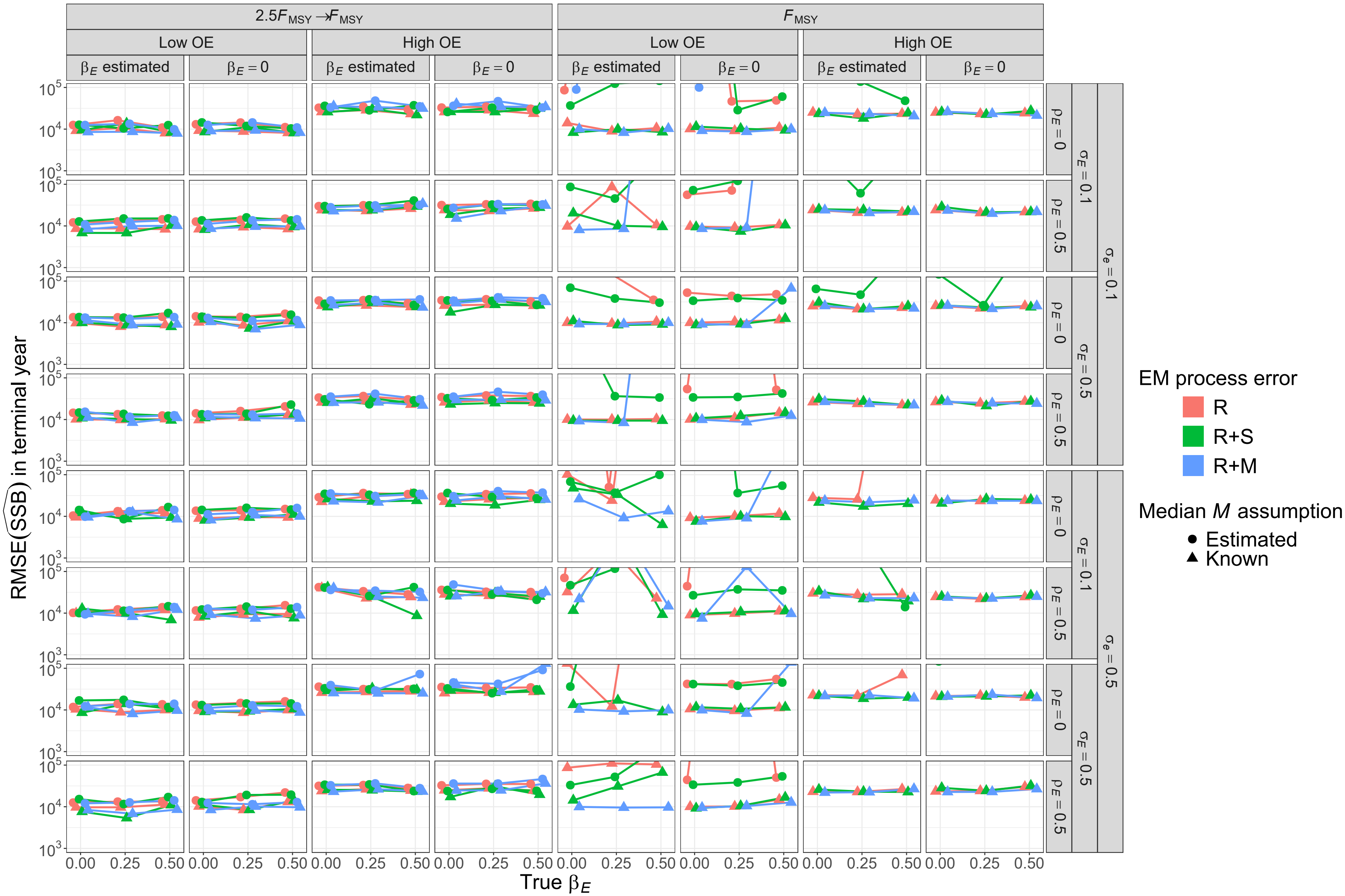



Fig. S61. For R+M OMs, root mean square error (RMSE) of estimates of spawning stock biomass in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

## Terminal year fishing mortality bias

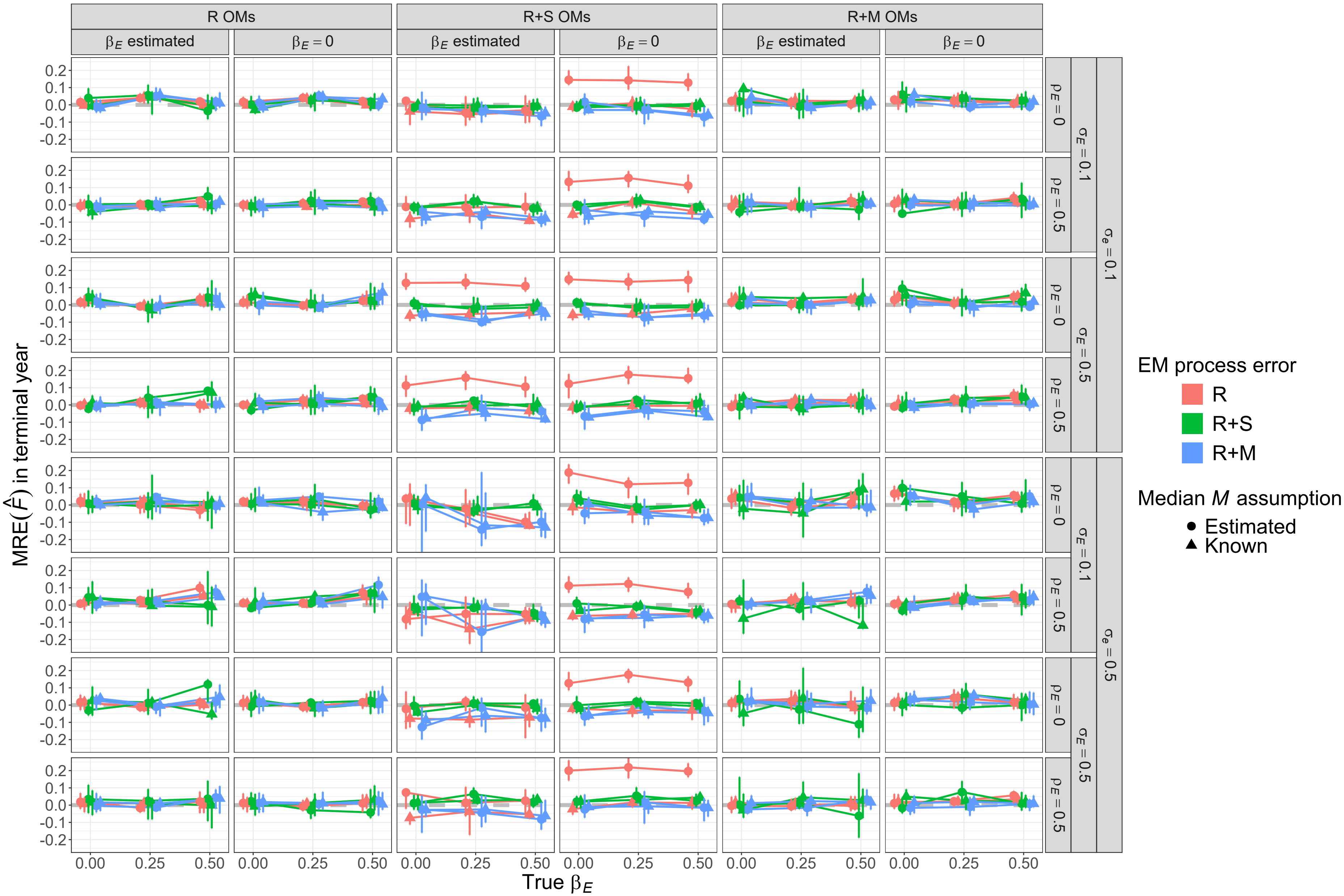



Fig. S62. Median relative error (MRE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low population observation error and contrast in fishing mortality.

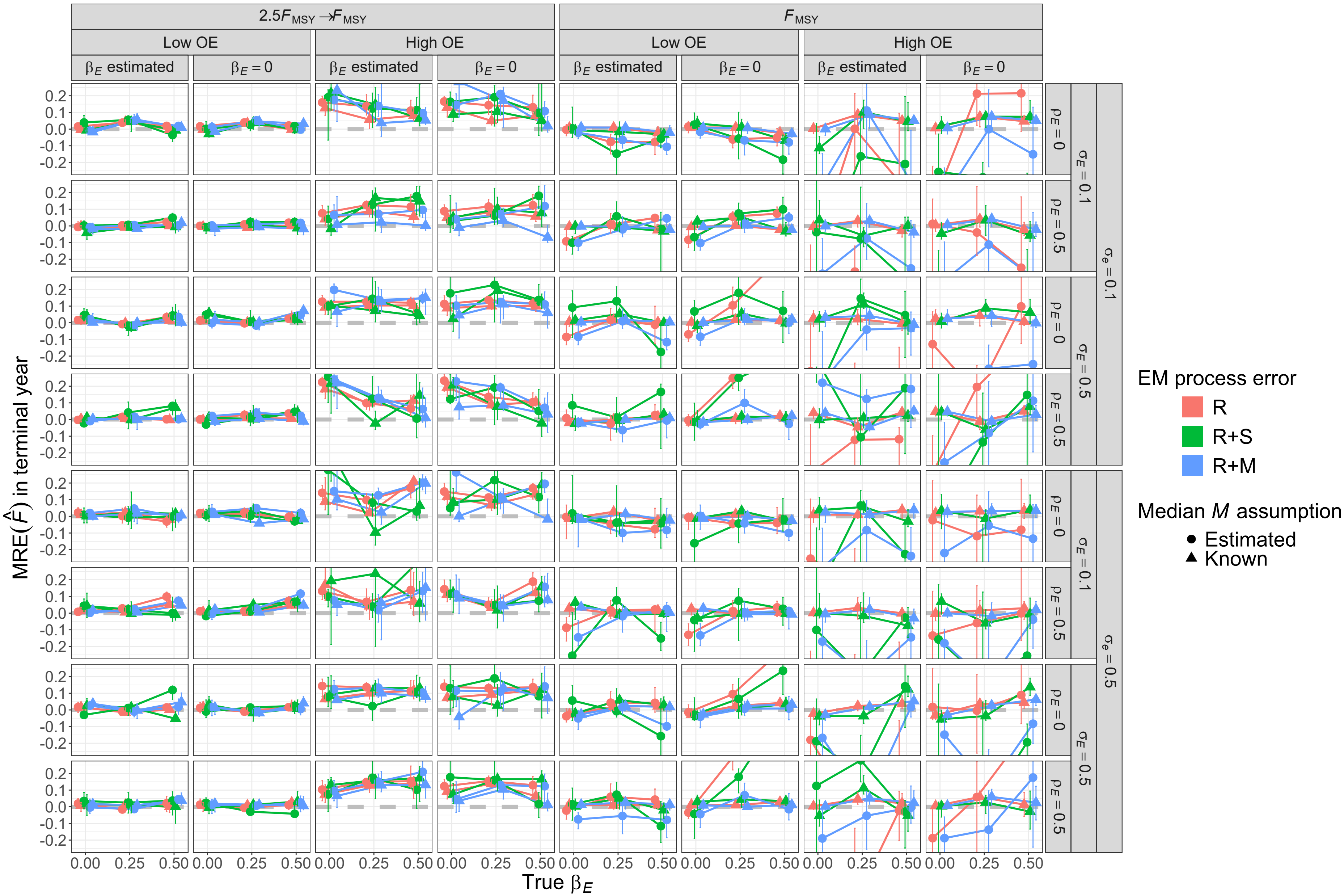



Fig. S63. For R OMs, median relative error (MRE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low observation error and contrast in fishing mortality.

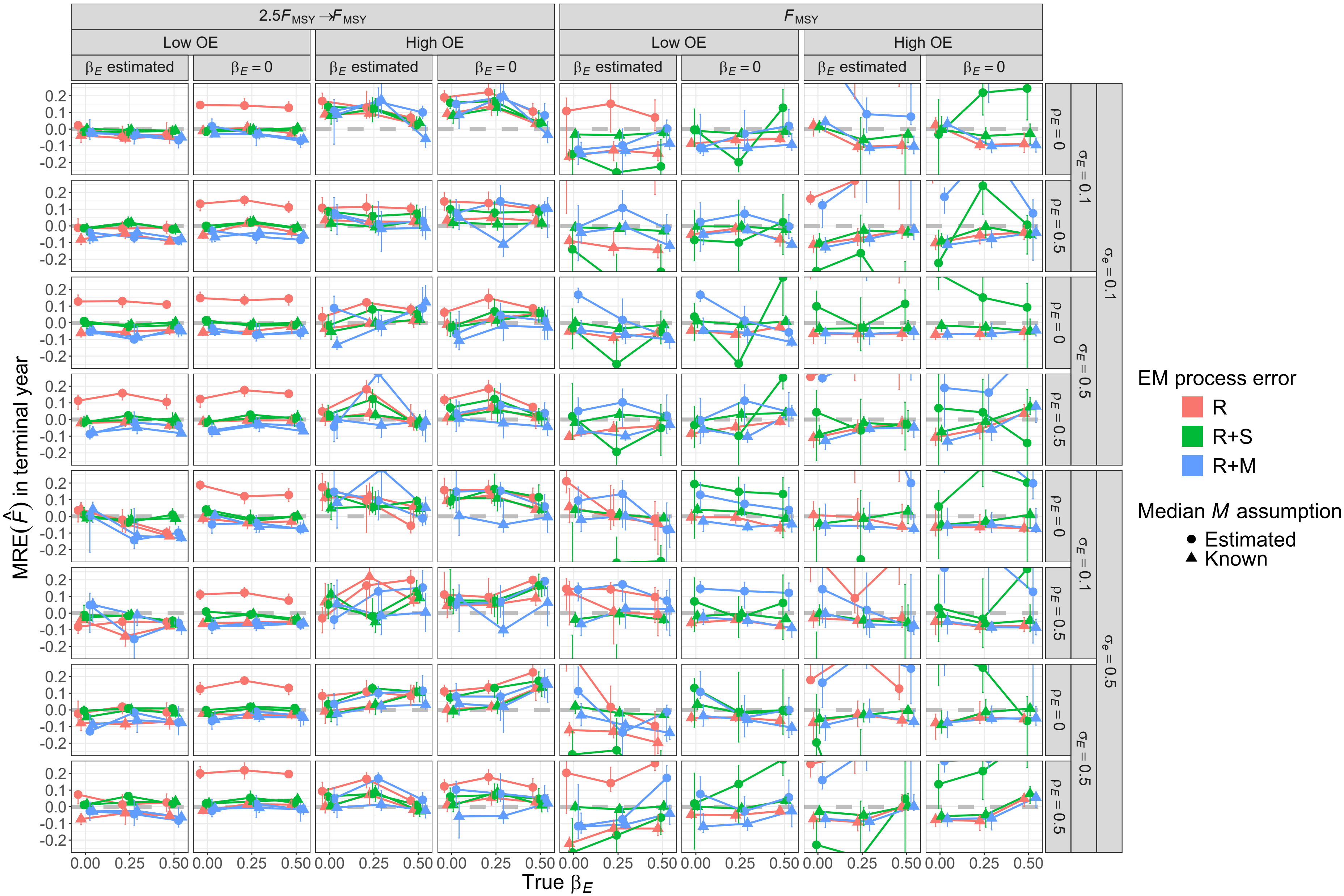



Fig. S64. For R+S OMs, median relative error (MRE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low observation error and contrast in fishing mortality.

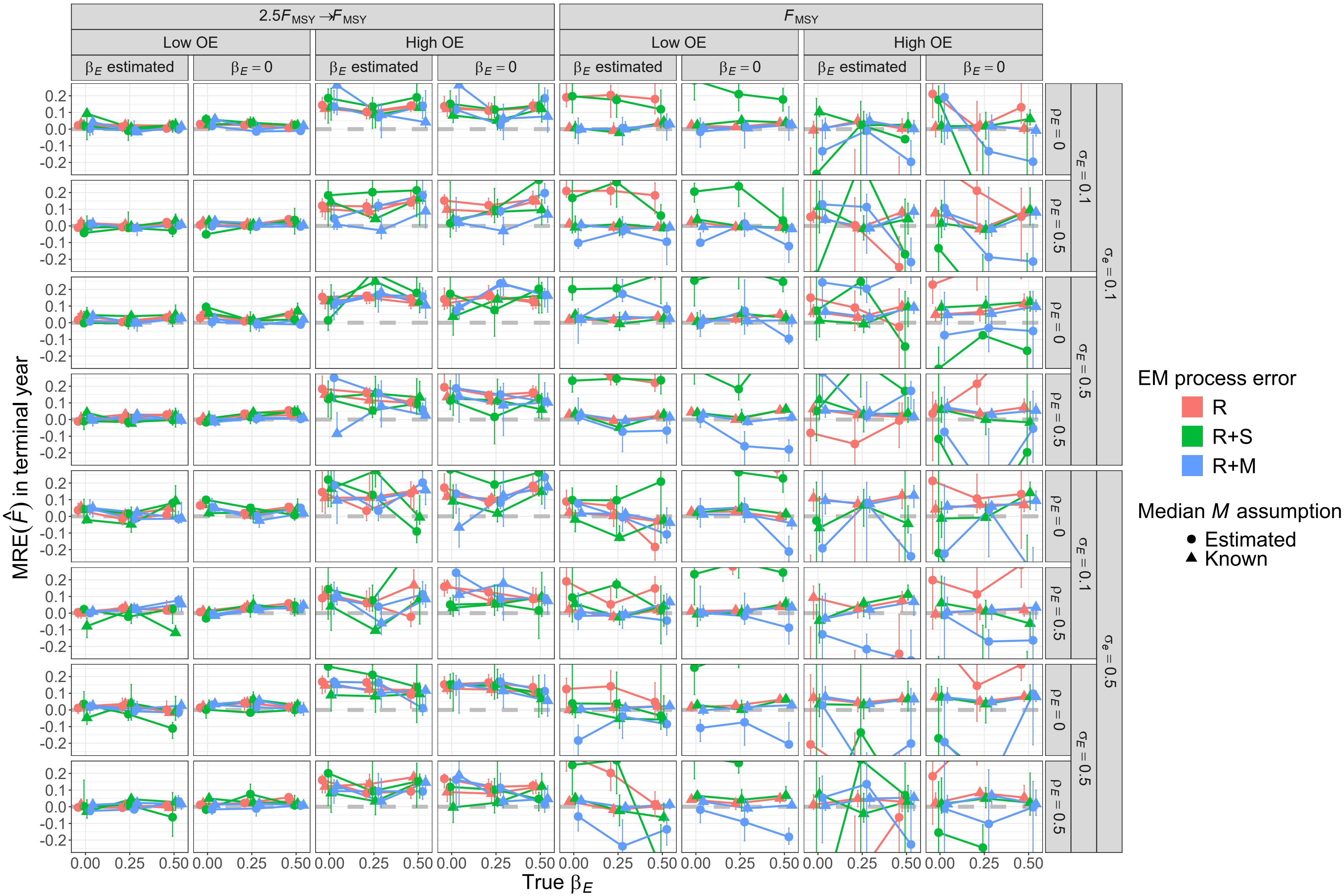



Fig. S65. For R+M OMs, median relative error (MRE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low observation error and contrast in fishing mortality.

## Terminal year fishing mortality RMSE

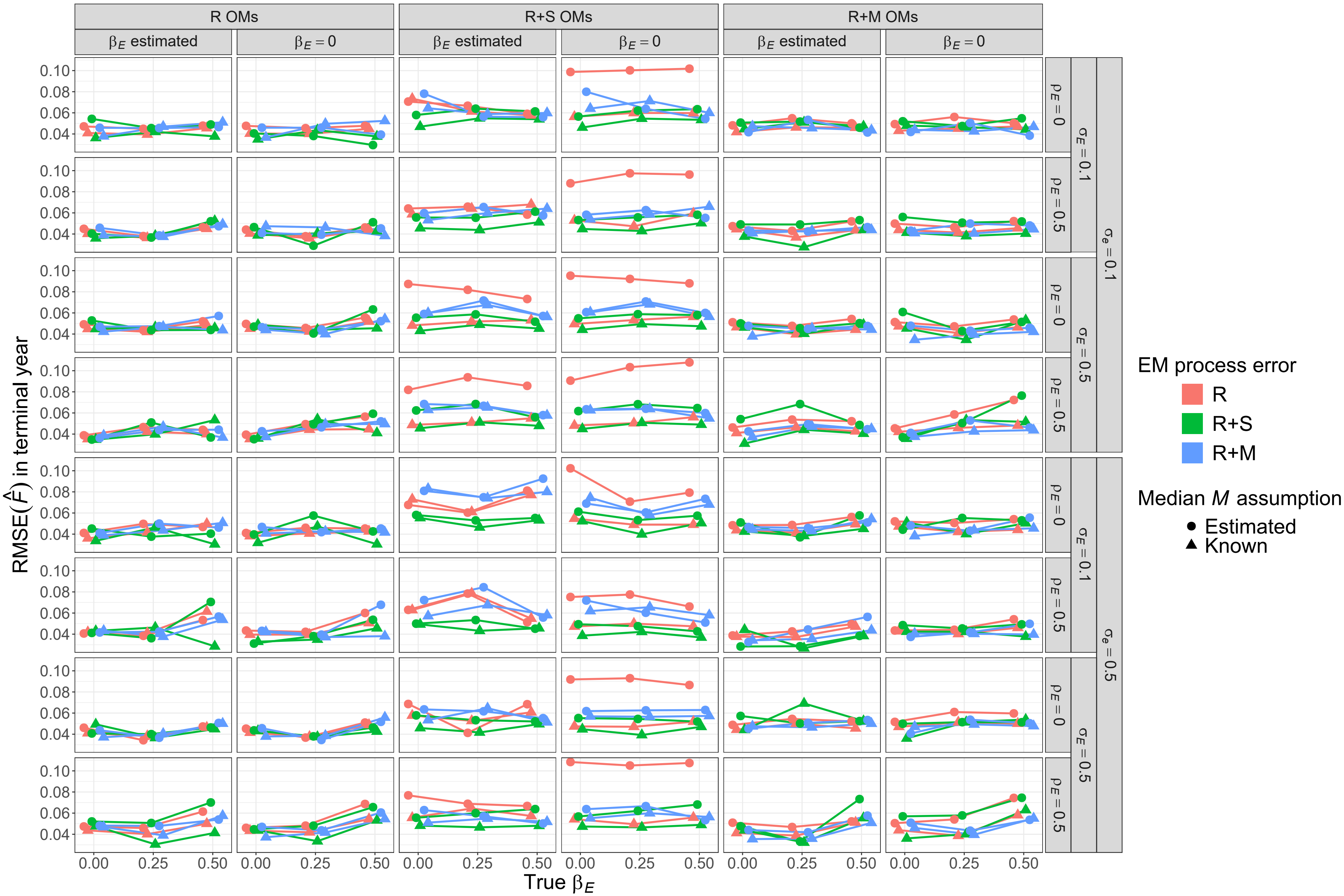



Fig. S66. Root mean square error (RMSE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known). All OMs had low population observation error and contrast in fishing mortality.

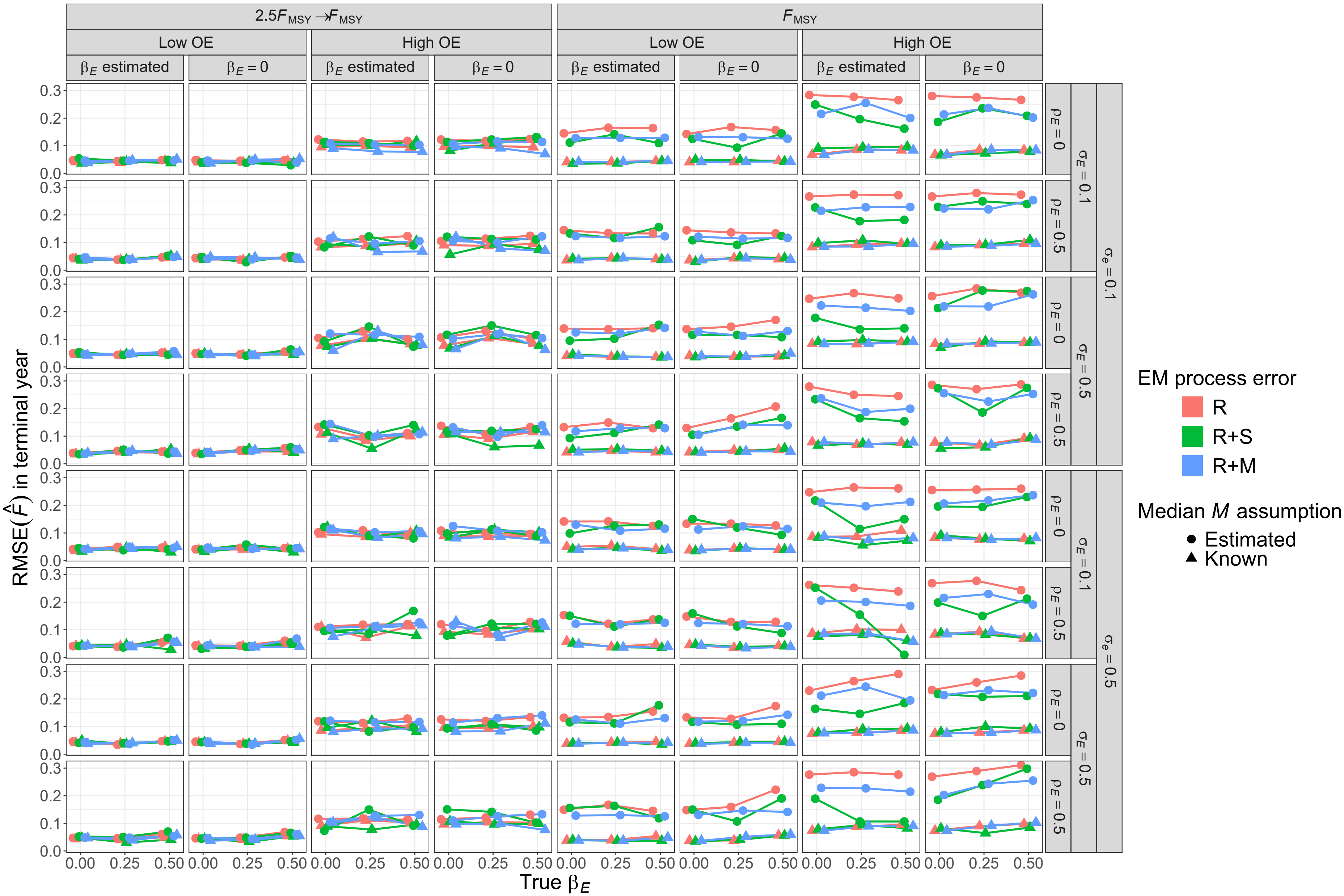



Fig. S67. For R OMs, root mean square error (RMSE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

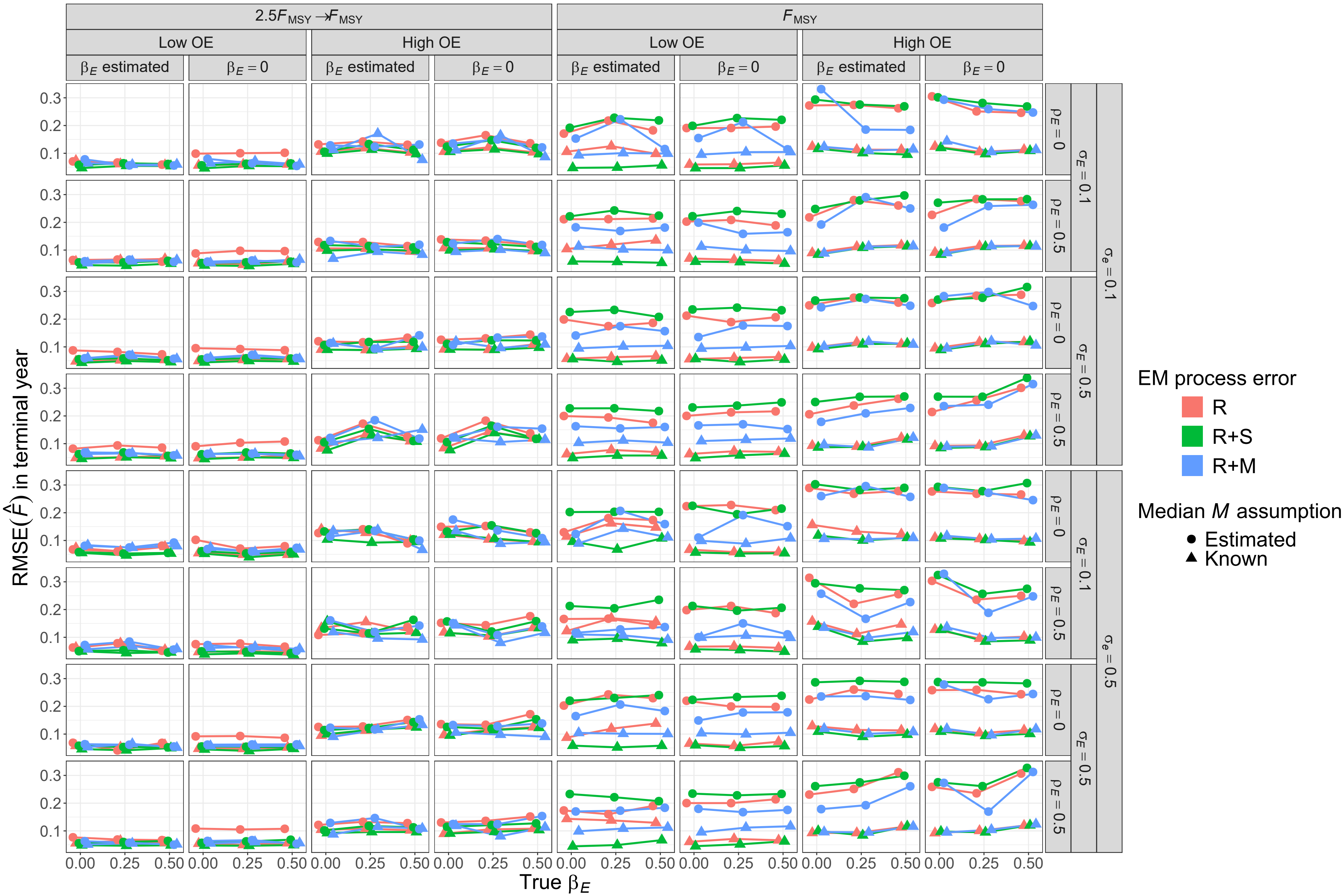



Fig. S68. For R+S OMs, root mean square error (RMSE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).

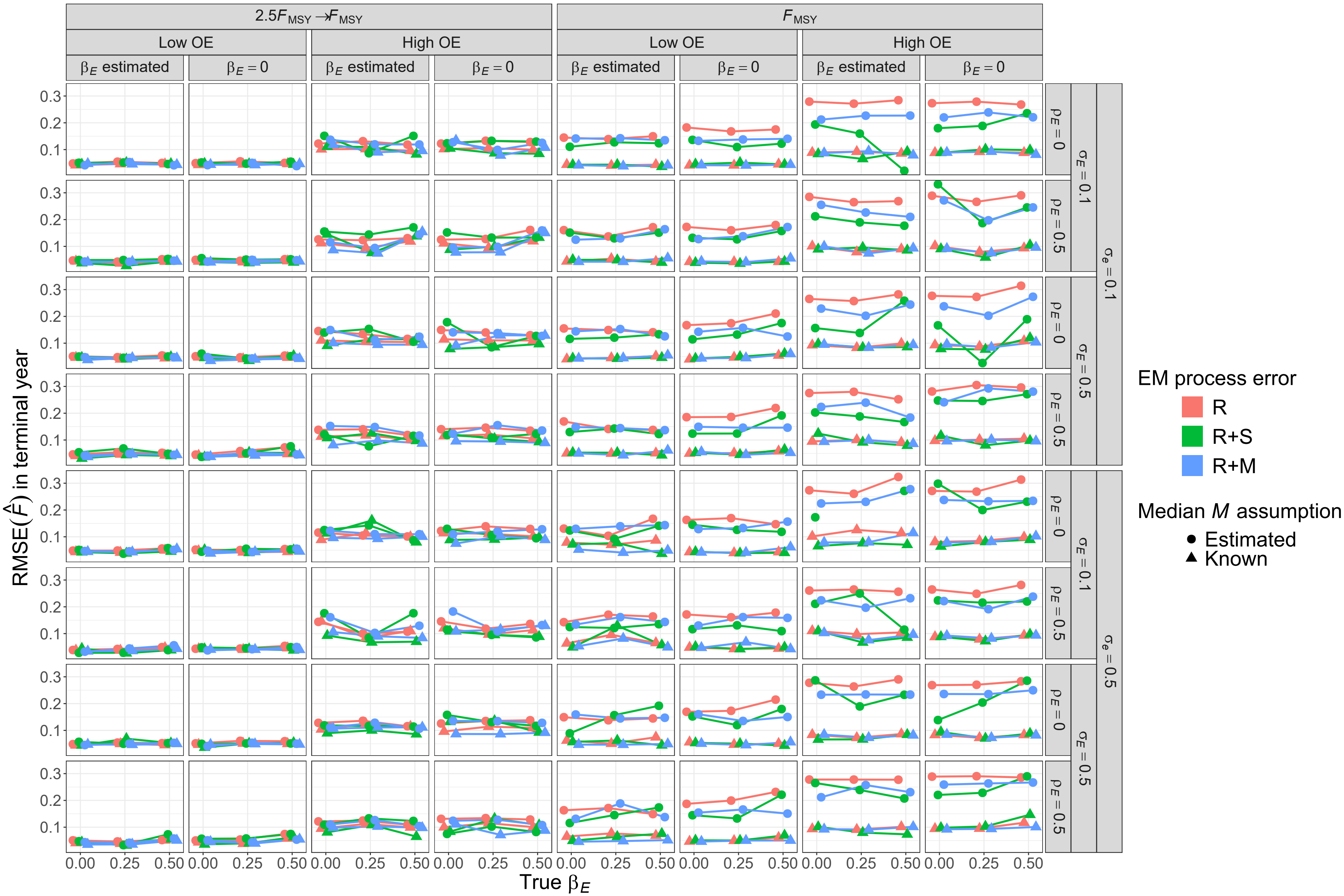



Fig. S69. For R+M OMs, root mean square error (RMSE) of estimates of fully-selected fishing mortality ($F$) in the terminal year for EMs with alternative process error assumptions, treatment of covariate effect ($\beta_E = 0$ or estimated), and treatment of median natural mortality parameter ($\beta_M$ estimated or known).