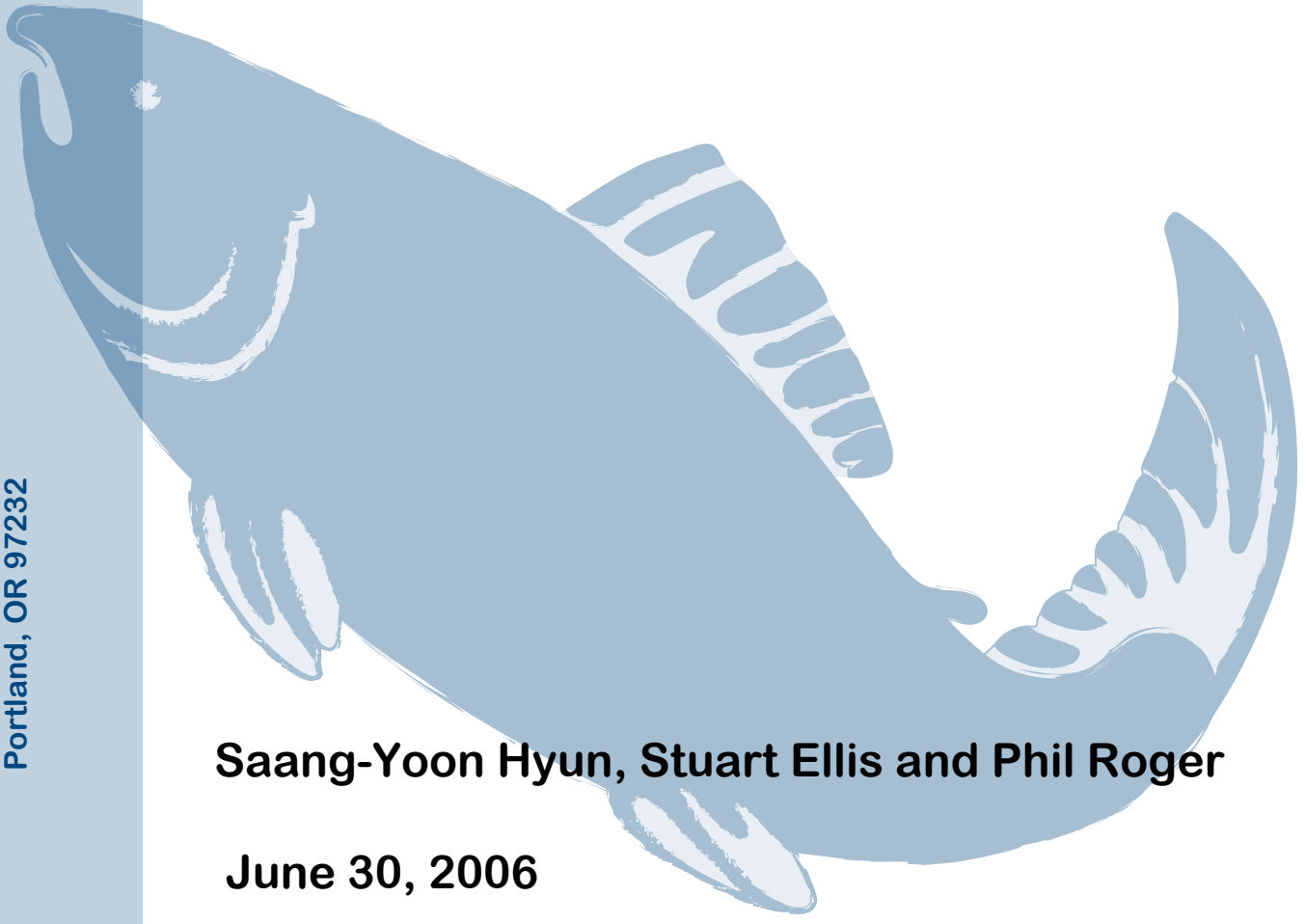
CRITFC

TECHNICAL REPORT 06-11

Preseason Forecasts of Ocean Escapements of Columbia River Chinook Salmon (*Oncorhynchus tshawytscha*) Populations



Saang-Yoon Hyun, Stuart Ellis and Phil Roger

June 30, 2006

Columbia River Inter-Tribal Fish Commission
729 NE Oregon, Suite 200
Portland, OR 97232
503.238.0667
www.critfc.org

Final Report - PSC Southern Fund Project:

Preseason forecasts of ocean escapements of Columbia River Chinook salmon (*Oncorhynchus tshawytscha*) populations

Saang-Yoon Hyun[1], Stuart Ellis, and Phil Roger

Columbia River Inter-Tribal Fish Commission

729 NE Oregon St. Suite 200

Portland, OR 97232 USA

[1]Corresponding author (email: hyus@critfc.org)

30 June 2006

TABLE OF CONTENTS

List of Tables

List of Figures

statistics: (a)-(d) = significance of model M6 for forecasts of URB age-3, -4, -5 and -6 runs; (e)-(h) = that of model M10 for forecasts of BUB age-3, -4, -5 and -6 runs; (i)-(l) = that of model M4 for forecasts of LRW age-3, -4, -5 and -6 runs; (m)-(p) = that of model M14 for forecasts of LRH age-3, -4, -5 and -6 runs; (q)-(s) = that of model M16 for forecasts of BPH age-3, -4, and -5 runs. Vertical dots over each sample size each panel are p-values of F-statistics of the corresponding model used for hind-casting forecasts of population- and age-runs in 10 years (1995-2004). Horizontal broken line in each panel indicates p-value of 0.1. This summary for PUB and UCS populations is not shown because of lack of data on these populations (Figs. 4 and 8).

List of Appendix Table

List of Appendix Figures

indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's three-year hind-casting forecasts (2002-2004) successfully cover actual runs.

1 Abstract

Preseason forecasts for ocean escapements of Columbia River summer and fall Chinook salmon *Oncorhynchus tshawytscha* populations are not only scientifically but also politically important. These forecasts affect salmon fishery quotas in the northeast Pacific Ocean. At present, the preseason forecasts are made by the U.S. v. Oregon Technical Advisory Committee (TAC), and are used as an input value for the Pacific Salmon Commission (PSC) and Pacific Fishery Management Council (PFMC) models that calculate fish abundance in the ocean. Annual harvest quotas for the Chinook salmon ocean fisheries are determined on the basis of these ocean abundance estimates. Although the traditional forecasts by the TAC were not seriously biased, they tended to be under-forecasts, and also uncertainty in the forecasts such as prediction interval is not measured. The main objective of this study was to develop robust forecast models and to express the forecast uncertainty. Fish populations of concern were: Upriver Bright (URB), Bonneville Upriver Bright (BUB), Pool Upriver Bright (PUB), Lower River Wild (LRW), Lower River Hatchery (LRH), Bonneville Pool Hatchery (BPH), and Upper Columbia River summer (UCS). In addition to data on fish sibling runs (age-specific runs from the same brood year) that the TAC uses for the traditional forecast, we explored various data including sea surface temperature, upwelling, and coded-wire-tag (CWT) records. The exploration indicated that ocean recovery rate of CWT fish and winter sea surface temperatures (SSTs) in fish's first ocean year from coast waters of British Columbia, Canada significantly contributed to forecasts. Then, using these data for each population, we built both ordinary regression models and autoregressive error models. To evaluate these models, we made extensive hind-casting forecasts of population runs in 10 years of 1994-2005 with different sample sizes except for PUB and UCS populations. Temporal range of data on PUB and UCS

populations was not as long as those on the other populations, and five year (2000-2004) and three year (2002-2004) run forecasts were made respectively. We tested candidate models based on hind-casting forecast performance in terms of relative error, frequency of prediction interval's coverage, model residuals, and model parsimony. Results indicated that an autoregressive model outperformed the traditional (ordinary regression) model for all populations studied. The ordinary regression model was apparently confounded due to autocorrelation in the model residuals, and routine inclusion of the intercept term in the ordinary regression model did not contribute to forecast in some instances. Inclusion of data on SSTs and CWT ocean recovery rate provided additional improvement in forecasts for some, though not all populations. Future years' forecasts would improve if instead of the current routine ordinary regression model, population-specific autoregressive models were adopted. The form of each population's model could also be reexamined on a regular basis as data accumulates over time.

2 Introduction

Columbia River summer and fall Chinook salmon *Oncorhynchus tshawytscha* are a major species to the Pacific Salmon Commission (PSC), because they form the largest contributing fish group to ocean Chinook fisheries north of Cape Falcon, Oregon (Fig. 1). For fishery management purposes, any Chinook salmon, which passes Bonneville Dam or is harvested in Zones 1-6 (Columbia River mouth through McNary Dam) between 16 June and 31 July, is defined as a summer Chinook salmon, and any Chinook salmon between 1 August and 31 December is a fall Chinook salmon. These dates correspond to the management seasons that were defined by the Parties to US v. Oregon which include tribes, states, and federal agencies (U.S. v. Oregon Parties 2005). The Upper Columbia River summer (UCS) Chinook salmon population is the major component of the summer run, and they are primarily from the Wenatchee, Entiat, and Methow Rivers (Fig. 1). Columbia River fall Chinook salmon are comprised of five major populations: Upriver Bright (URB), Bonneville Upriver Bright (BUB), Pool Upriver Bright (PUB), Lower River Wild (LRW), and Lower River Hatchery (LRH). URB wild and hatchery fish originate mainly from Hanford Reach and the Snake River and partially from the Deschutes River (Fig. 1). Combined BUB and PUB populations are sometimes referred as Mid-Columbia Bright (MCB) group, and they are all hatchery fish. In determining in-river harvest rates, URB fish are a key population, because they include wild fish from the Hanford Reach, which is only free-flow (unrestricted by a hydropower dam and reservoir) area in the Columbia River basin, and also because they include Snake River fall Chinook salmon listed as threatened under the US Endangered Species Act (ESA). BUB and PUB are hatchery fish whose harvest is constrained only by brood stock needs. A minor population of Columbia River fall Chinook salmon is early maturing fish called “Tule”, which

are produced at Bonneville Pool Hatchery (BPH) (also called Spring Creek Hatchery (SCH)). Except for LRH and LRW, adults of the other populations pass Bonneville Dam when they return.

The ocean distribution of Columbia River summer and fall Chinook salmon covers Oregon, Washington, British Columbia, and Alaska coasts. Numbers of the fish distributed in the ocean fishery areas are important parameters to management agencies. At present, the US v. Oregon Technical Advisory Committee (TAC) makes preseason forecasts of ocean escapements of Columbia River Chinook salmon populations five or six months before fish populations return in a year (i.e., about February or March each year). Ocean escapement means the number of fish that survive in the ocean to Columbia River mouth. The PSC Chinook Technical Team and the Pacific Fishery Management Council (PFMC) Salmon Technical Team use those preseason forecasts as input values for their own ocean models whose output includes the ocean run estimates in the year. Based on the output of ocean run estimates, the PFMC recommends ocean fishery quotas and seasons in the year to the U.S. National Marine Fisheries Service (NMFS), which regulates the ocean fishery in the south of the US-Canada border. That is, preseason forecasts affect ocean fishery management of Chinook salmon through the fishery management models. Because of this practical importance, preseason forecasts of greater accuracy (less biased and more precise) would be desirable.

The TAC's traditional forecast of ocean escapement for a population is based on relationships between age-specific returns from the same brood year (so-called siblings' returns). They apply an ordinary regression model to data on siblings' ocean escapements. On the basis of the regression model, they predict unknown age-a escapement in a year using age-"a-1" escapement observed in the previous year, and then sum the respective predicted age-a

escapements to total population escapement. Overall the TAC's forecast model's historical performance has not been poor in terms of bias, but its forecasts have been often lower than actual escapements. Such under-forecasts may lead to under-estimates of ocean abundance by way of the PSC and PFMC ocean models, and then result in forcing ocean fishery quotas to be reduced. Also, the TAC's forecast methods do not measure forecast uncertainty such as a forecast interval. Measurement of a forecast interval must be addressed because a forecast of only a point value does not carry information about how likely the forecast is (e.g., how widely or narrowly the forecast distribution is). In this project, we build various candidate forecast models and also evaluate candidate forecast models in terms of not only bias but also a forecast interval.

3 Methods

Hereafter, we express ocean escapement as run or return size. Notations are summarized in Table 1.

3.1 Target populations and data

Forecasts for seven Chinook salmon population runs were studied (Table 2). Data on annual population return sizes by age were available from the TAC. Temporal range of the data varied by population (Table 2). We extracted coded-wire-tag (CWT) data on fall and summer Chinook salmon to calculate annual ocean recovery rate of CWT fish by age, which is proportional to annual ocean fishing mortality by age. Ocean recoveries of CWT fish are usually small, and thus we increased CWT records by including the records on all fall Chinook salmon populations from the entire Columbia River basin. The CWT data on fall Chinook salmon were

available from brood years since 1972. We separately extracted CWT data for summer Chinook salmon because their behavior and distribution in the ocean may be significantly different from fall Chinook salmon. The CWT data on summer Chinook salmon were available from brood years since 1975. As annual ocean condition data, we explored sea surface temperatures (SSTs) and ocean upwelling measured along the west coast of North America. Previous studies reported significant correlations between these variables and cohort strength of Hanford Upriver Bright (URB) fall Chinook salmon and Snake River spring/summer Chinook salmon (Hyun 1996; Scheuerell and Willams 2005). Those data were available from years since 1963.

3.2 Deterministic relationship between age-specific returns

Deterministic relationship between sibling runs of age-"a" run against age-"a-1" run is perfectly multiplicative. To illustrate the multiplicative relationship, we built conceptual models as a discrete time manner. The decrease in fish abundance at age a-1 in the ocean to next age a involves survival from age a-1 to age a: i.e., $O_{a-1}S_{a-1} = O_a$ (Table 1). And return of age a-fish in the ocean to natal river is related to maturation at the age: i.e., $R_a = O_a A_a$ (Table 1). Thus, the ratio of sibling runs is:

$$\begin{aligned} \frac{R_a}{R_{a-1}} &= \frac{O_1 S_1 S_2 \cdots S_{a-1} A_a}{O_1 S_1 S_2 \cdots S_{a-2} A_{a-1}} \\ &= \frac{S_{a-1} A_a}{A_{a-1}} \end{aligned} \tag{1}$$

Lumping unknown quantities of age-specific survival and maturation rates (S_a, and A_a) in eq. 1, R_a is the product of the lumped parameter and R_{a-1}:

$$R_a = K \cdot R_{a-1} \tag{2}$$

where $K \equiv (S_{a-1}A_a/A_{a-1})$. Because S_a and A_a are fractions and positive, K is dimensionless and positive.

We can reduce the uncertainty in ocean survival rate because of available information about CWT ocean recovery rate, which is proportional to ocean mortality. Replacing ocean survival rate in eq. 1 with ocean mortality and also expressing it with ocean CWT recovery rate (i.e., $S_a = (1 - M_a) = (1 - c_a H_a)$), where c is a positive proportional constant, eq. 1 for R_a becomes

$$\begin{aligned} R_a &= (1 - c_{a-1}H_{a-1})\frac{A_a}{A_{a-1}}R_{a-1} \\ &= \frac{A_a}{A_{a-1}}R_{a-1} - c_{a-1}H_{a-1}\frac{A_a}{A_{a-1}}R_{a-1} \end{aligned} \tag{3}$$

where A_a, A_{a-1}, and c_{a-1} are unknown. Lumping unknown parameters ($K_1 \equiv (A_a/A_{a-1})$, and $K_2 \equiv (\alpha_{a-1}A_a/A_{a-1})$), we have the following equation.

$$R_a = K_1 R_{a-1} - K_2 H_{a-1} R_{a-1} \tag{4}$$

K_1 and K_2 are dimensionless and positive because A_a and c are dimensionless and positive. H_a (CWT ocean recovery rate of age-a fish in a year) is the following equation (Hyun 1996): $H_a = V_a/(LP)$ where V_a is ocean recoveries of age-a fish in a year, P is sampling fraction for tagged fish in the ocean catch in the year, and L is release number of tagged fish from release year (= brood year + 1) (Table 1). These three quantities of V_a, P, and L were from CWT database. We do not add a notation of a year for simplicity.

3.3 Stochastic relationship between age-specific returns

Those parameters involved in the above deterministic relationship are unknown and not constant over years. Also even known values of R_a and H_a have observation errors. Thus, we proceed to a stochastic model, allowing an error term and incorporating other explanatory variables that could reduce the model error. We have the following general stochastic model based on the above deterministic relation.

$$R_a = \beta_0 + \beta_1 R_{a-1} + \beta_2 H_{a-1} R_{a-1} + \beta_3 X_3 + \cdots + \beta_n X_n + \varepsilon \tag{5}$$

where X_k is an additional candidate explanatory variable, and β_k is the coefficient of X_k. Although the sign of coefficient ($=\beta_2$) of $H_{a-1}R_{a-1}$ is positive (+) in the above equation for the general expression, one expects the coefficient value will be negative based on eq.4. Exploring candidate elements in $\mathbf{X}$, we found a significant contribution (p-value < 0.05) of November and December SSTs measured from coastal waters of British Columbia, Canada (Amphitrite Point and Kains Island sites) during fish's first ocean year, and considered the mean of the SSTs as another explanatory variable.

Considering combination effects between the four terms of intercept, sibling run, the product of ocean recovery rate and sibling run, and SST in eq. 5, we explored 8 different models (see Table 3: M1, M3, M5, M7, M9, M11, M13, M15). Further, we considered lag-1 autoregressive error models because we found significant lag-1 autocorrelations over annual runs. M2, M4, M6, M8, M10, M12, M14, and M16 in Table 3 are autoregressive versions transformed from M1, M3, M5, M7, M9, M11, M13, and M15, respectively. That is, we examined a total of 16 models. Model 1 is the same as the TAC's traditional model (Table 3). We described a general form of lag-1 autoregressive models in Appendix.

3.4 Forecast of total run

3.4.1 Hind-casting with different sample sizes

We make forecasts of runs, using hind-casting methods, where data used for forecast of run in a year are from years prior to the forecast year. Because of the significant autocorrelation over annual runs, one must not treat runs from each year independently as in cross-validation or retrospective analysis. We made hind-casting forecasts of runs in 10 years of 1995-2004 for all populations except for PUB and UCS populations whose data's temporal range was short (Table 2).

We did not fix a sample size (the number of years of data) for the hind-casting forecasts. To identify the optimal sample size, we used different sample sizes of 10 years and more because a statistically defensible sample size is at least 10.

Allowing data from such different numbers of years led to a large number of hind-casting forecasts. For example, in forecasting age-3 URB runs in the 10 years, we could explore sample sizes of 10 through 19 (i.e., data from 10 - 19 years prior to every forecast year). That is, 100 forecasts (= 10 sample sizes $\times$ 10 forecast years) were made for the age-3 URB run. Including the other age runs (age-4, -5, and -6 runs) and further including the other populations, the number of forecasts was about 2,000. The main goal for performing such a large number of forecasts was to accommodate all forecast performance regardless of outlier effects as well as to identify optimal sample sizes.

3.4.2 Point forecast of total run and its prediction interval

We assumed the error term in each forecast model (Table 3) is a normal (Gaussian) random variable: $\varepsilon \sim N(0, \sigma^2)$, where σ^2 is the variance of the error term, ε. First, we

predicted unknown age-specific runs ($\tilde{R}_a$), where we distinguish unknown run size with a tilde ($\tilde{}$) from observed and fitted runs (Table 1).

After estimating coefficients in a model of interest by fitting the model to data, we applied those estimates to predictors of explanatory variables. Such a model fitting process is routine and it is taken care of by most statistical software. The mean of the predicted value of $\tilde{R}_a$ is calculated as follows.

$$E(\tilde{R}_a) = \mathbf{X_p^t} \cdot \mathbf{b} \tag{6}$$

where vector $\mathbf{X_p}$ indicates predictors of explanatory variables in a model of concern (Table 3), and vector $\mathbf{b}$ are estimates of coefficients of those explanatory variables.

The variance of unknown return size in a forecast year is the sum of error mean square (MSE) of a model of concern fitted to data, and the estimated variance of fitted return size. That is,

$$\begin{aligned} Var(\tilde{R}_a) &= MSE + Var(\hat{R}_a) \\ &= MSE + \mathbf{X_p^t} \cdot \mathbf{\Sigma}(\mathbf{b}) \cdot \mathbf{X_p} \end{aligned} \tag{7}$$

where MSE = error mean square of a model of concern fitted to data, and $\Sigma(\mathbf{b})$ = the estimated variance-covariance matrix of the model coefficients.

Letting α equal Type-I error, $(1-\alpha)\%$ prediction interval for unknown age-a run size is

$$E(\tilde{R}_a) \pm t(1-\alpha/2; n-p) \cdot SD(\tilde{R}_a) \tag{8}$$

where n = the number of years of data; p = the number of coefficients in a model of concern; $(n-p)$ = degrees of freedom; $t(1-\alpha/2; n-p)$ = t-statistic that corresponds to cumulative

probability of $(1-\alpha/2)$ under the degrees of freedom, $(n-p)$; and $SD(\tilde{R}_a)$ = standard deviation of $\tilde{R}_a$.

Our ultimate goal is to predict unknown total return size of a population ($\tilde{R}$). The mean and variance of unknown total return size, $\tilde{R}$ are the sum of eqs. 6 and 7. That is,

$$E(\tilde{R}) = \sum_{a>2} E(\tilde{R}_a) \tag{9}$$

$$Var(\tilde{R}) = \sum_{a>2} Var(\tilde{R}_a)$$

Fish return size at age 2 are precocious jacks and are not predicted. Note that covariance was involved in eq. 7 but it was not in eq. 9. The respective age-a returns in the same year are from different brood year, and thus the runs are independent (Hyun 2002; Hyun et al. 2005).

$(1-\alpha)\%$ prediction interval is

$$E(\tilde{R}) \pm z(1-\alpha/2) \cdot SD(\tilde{R}) \tag{10}$$

where $z(1-\alpha/2)$ is z-statistic that corresponds to cumulative probability of $(1-\alpha/2)$. One may argue that the z-statistic must be replaced by a t-statistic. However, the difference does not concern us because $\tilde{R}$ is the sum of age-specific runs and the sum approaches a normal random variable by the central limit theorem (CLT). The sum of "any" random variables is normally distributed by the CLT. Also as the number of sample size increases, t-distribution approaches z-distribution.

3.5 Model performance

To compare forecast models in forecast power, we used four criteria: (i) relative errors (%) (= 100 × [point forecast - actual run]/[actual run]), (ii) coverage of prediction intervals, (iii) model residuals, and (iv) model parsimony. The first two criteria measure both bias and precision of forecasts. Determination coefficient often denoted as r^2 is a goodness-of-fit index of a model, but it does not represent forecast power. It is easy to get a large value of r^2 (e.g., > 0.90) by putting many explanatory variables in a model, but a large value of r^2 from such over-fitting does not guarantee improvement in forecast power. Some literature exaggerates forecast power of a model, showing such a large value of r^2 (e.g., Scheuerell and Williams 2005). By examining model residuals (third criterion), we check for violation of the assumption that the error term in each model is a normal random variable (see section, "Point forecast of total run and its prediction interval"). If candidate models are comparable based on the first three criteria, then the model parsimony dictates a preference for model with the smaller number of parameters (coefficients).

4 Results & Discussion

4.1 Upriver Bright (URB)

Forecasts of URB total runs in 1995-2004 were made with different sample sizes up to 14 (Fig. 2), whereas those of age-specific runs in this 10-year period were made with sample sizes up to 19 (Figs. A1 and A2). Temporal range of historical data on age-5 and -6 URB runs was not long enough to make hind-casting forecasts of the 10 year runs with a sample size above 14 (Table 2, Fig. A1), and thus forecasts of total runs were made with sample sizes up to only 14 (Fig. 2).

Relative errors in forecasts of URB total runs by non-autoregressive models were often negative (Fig. 2), indicating those models made under-forecasts. The tendency was obvious especially in models M1, M3, M9, M11, M13, and M15 (Fig. 2). When examining the tendency by age (Fig. A1), we found that such under-forecasts were most obvious in forecasts of age-3 runs by non-autoregressive models (first two rows in Fig. A1) and was minor in those of age-4 runs by such models (3rd and 4th rows in Fig. A1). The obvious under-forecasts seemed to be due to serious autocorrelations in model residuals. We found significant autocorrelations over lag-1 in residuals of non-autoregressive models for URB age-3. In Fig. 9, the autocorrelations from the traditional model M1 hover around 0.5 whereas those from autoregressive models do at zero.

Forecast of the URB run in 2001 was worst out of forecasts of runs in the 10 years. Lowest dots over overall sample size (x-axis) in each panel in the first two rows in Fig. 2 are relative error in forecast of 2001 run. The poor forecast of the 2001 run was obtained regardless of model and sample size (first two rows in Fig. 2). Removing the 2001 run forecast, model M6 was most acceptable because relative errors from model M6's forecasts were within a zone that the other competitive models (e.g., M2 and M4 in Fig. 2) allow, and the frequency of its prediction coverage was highest (Fig. 2). Also, autocorrelations were not significant in residuals of model M6 (Fig. 9).

4.2 Bonneville Upriver Bright (BUB)

Forecasts of BUB total runs in 1995-2004 were made with different sample sizes up to 13 (Fig. 3), whereas those of age-specific runs in the 10-year period were made with sample sizes up to 14 (Figs. A3 and A4). The temporal range of historical data for BUB runs was shorter than for URB, LRW, LRH, and BPH populations (Table 2).

The tendency of negative relative errors found in non-autoregressive model forecasts of URB runs was not clear in those of BUB runs (Fig. 3). In terms of forecast relative error, forecast of BUB runs in 1998 was worst. Highest dots over overall sample size (x-axis) in each panel in the first two rows in Fig. 3 are relative error in forecast of 1998 BUB run. The poor forecast was obtained regardless of sample size, but improved in models M9, M10, M11 and M12 by about 100% (first two rows in Fig. 3). Removing the worst forecast, models M10 and M12 were most acceptable because relative errors from their forecasts were smaller than the other models, and the frequency of the prediction intervals' coverage over sample size 13 was secondly highest (as about 80%) following models M6 and M8 (Fig. 3). Because of model parsimony, we preferred M10 over M12. Also, autocorrelations in residuals of model M10 were not significant (Fig. 9).

4.3 Pool Upriver Bright (PUB)

Forecasts of only 2000-2004 PUB total runs were made with sample sizes 10 and 11 because of lack of data (Fig. 4, Table 2). We selected model M14 for forecast of PUB run based on criteria of the magnitude of forecast relative error, the prediction intervals' coverage, and autocorrelations in model residuals. However, this selection needs a caution because of lack of data, and it may change in time as more data accumulate.

4.4 Lower River Wild (LRW)

Forecasts of LRW total runs in 1995-2004 were made with different sample sizes up to 14 (Fig. 5), whereas those of age-specific runs in the 10 years were made with sample sizes up to 19 (Figs. A7 and A8). The temporal range of historical data on age-5 and -6 LRW runs was not long enough to make hind-casting forecasts of the 10 year runs with a sample size above

14 (Table 2, Fig. A7), and thus forecasts of total runs were made with sample sizes up to only 14 (Fig. 5). On the basis of criteria of relative error magnitude, prediction intervals' coverage, and autocorrelations in residuals, model M4 was best (Figs. 5 and 9).

4.5 Lower River Hatchery (LRH)

Forecasts of LRH total runs in 1995-2004 were made with different sample sizes up to 14 (Fig. 6) whereas those of age-specific runs in the 10 years were made with sample sizes up to 19 (Figs. A9 and A10). Model selection for LRH run was difficult because no model satisfied all selection criteria. Although models M10 and M12 appeared strong candidates based on the magnitude of forecast relative errors and the prediction intervals' coverage, residuals from those models, especially for age-5 run forecasts, did not randomly behave. The non-random pattern in residuals violates the assumption that the error term of each model is a normal random variable. Alternatively, models M6 and M14 were considered, because these models do not seriously violate this assumption and also because the frequency of their prediction intervals' coverage was as high as or higher than those of the other strong candidate models (Fig. 6). Finally we selected model M14 instead of M6 because of model parsimony.

4.6 Bonneville Pool Hatchery (BPH)

Forecasts of BPH total runs in 1995-2004 were made with different sample sizes up to 18 (Fig. 7), whereas those of age-specific runs in the 10 years were made with sample sizes up to 19 (Figs. A11 and A12). BPH age-6 runs are not considered (Table 2, Figs. A11 and A12). On the basis of the three criteria of relative error magnitude, prediction intervals' coverage, and autocorrelations in residuals, models M16 were best (Figs. 7 and 9). Forecast performance of model M16 especially with sample sizes 12 and 13 outperformed the others.

4.7 Upriver Columbia River summer (UCS)

Forecasts of only 2002-2004 UCS total runs were made with sample sizes 10 and 11 because of lack of data (Fig. 8, Table 2). UCS age-3 runs are not considered (Table 2, Figs. A13 and A14). On the basis of the three criteria of relative error magnitude, prediction intervals' coverage, and autocorrelations in residuals, models M12 were best (Fig. 8). However, this selection needs a caution because of lack of data, and it may change in time as more data accumulate.

4.8 Overall findings

Selected models with the highest forecast power were autoregressive models for all populations. The frequency at which the prediction intervals from those selected models included actual run sizes was always higher than that from the traditional (ordinary regression) model M1. Also residuals from those selected models did not violate the assumption that error terms of the models are normally distributed, whereas residuals from the traditional model M1 did (Fig. 9). Autocorrelation of residuals from the traditional model M1 was serious; autocorrelation coefficients over lag 1 were often larger than |0.5| (a, f, i, q, and r in Fig. 9). In Fig. 9, autocorrelation examination for PUB and UCS populations is not shown because of lack of data on these populations (Figs. 4 and 8). When autocorrelation of a model was substantial, the prediction interval from the model was too narrow to cover actual runs (Fig. 10). In Fig. 10, we summarize forecast performance for total runs from all 16 models with sample size 13. In forecasts of total runs, sample sizes 12 - 13 were optimal (Figs. 2-8, 11, and 12). Sample sizes more than 14 were not explored because of lack of data on age-6 runs.

Data on ocean recovery rate and SST contributed to model forecast. Explanatory vari-

ables in selected model M6 for URB, M4 for LRW, M14 for LRH and M16 for BPH include the data. The coefficient of the product of age-specific ocean recovery rate and age-run (= $H_a \cdot R_a$) was negative as expected (eq. 4), and SST (winter SST from BC coastal waters) was negatively correlated with fall Chinook salmon run size.

Although we accommodated all outliers in selecting models, most models' goodness of fit was overall satisfactory in terms of determination coefficient (r^2) and p-value of F-statistics (Figs. 11 and 12). If outliers were removed, models would improve further.

4.9 Conclusions

Selected forecast models varied by population, but they were all autoregressive error models. Routine inclusion of intercept term in a regression model of relating sibling runs should be reconsidered. If there were no error in measurements of data, the intercept term should be zero because relationship between sibling runs is multiplicative. Indeed, the intercept term did not contribute to forecasts for BUB, LRH, and BPH populations. We summarize the models by population.

- URB: M6 (autoregressive error version of form, "$R_a = \beta_0 + \beta_1 \cdot R_{a-1} + \beta_2 \cdot SST + \varepsilon$")
- BUB: M10 (autoregressive error version of form, "$R_a = \beta \cdot R_{a-1} + \varepsilon$")
- LRW: M4 (autoregressive error version of form, "$R_a = \beta_0 + \beta_1 \cdot R_{a-1} + \beta_2 \cdot H_{a-1} R_{a-1} + \varepsilon$")
- LRH: M14 (autoregressive error version of form, "$R_a = \beta_1 \cdot R_{a-1} + \beta_2 \cdot SST + \varepsilon$ ")
- BPH: M16 (autoregressive error version of form, "$R_a = \beta_1 \cdot R_{a-1} + \beta_2 \cdot H_{a-1} R_{a-1} + \beta_3 \cdot SST + \varepsilon$ ")

- PUB and UCS: Selected models for these populations were M14 and M12, but data was insufficient to make a specific recommendation.

5 Acknowledgements

This study was funded from the Pacific Salmon Commission. We thank the US v. Oregon Technical Advisory Committee (TAC), the Regional Mark Processing Center, the Canadian Department of Fisheries and Oceans, and the US National Oceanographic and Atmospheric Administration (NOAA) for sharing data. Rishi Sharma, Peter Galbreath, David Graves and Jeff Fryer at Columbia River Inter-Tribal Fish Commission were consulted for the PSC salmon management, discussion, and map figure.

6 References

1. Durbin, J., and G.S. Watson. 1951. Testing for serial correlation in least squares regression. II. Biometrika 38: 159-178.

2. Hyun, S. 1996. Ocean distributions of the Columbia Rive Hanford Reach and the Snake River fall Chinook salmon (*Oncorhynchus tshawytscha*) stocks and the effect of interannual ocean conditions on their survival. M.S. thesis, University of Washington, Seattle, Wash.

3. Hyun, S. 2002. In-season forecasts of sockeye salmon returns to the Bristol Bay districts of Alaska. Ph.D. thesis, University of Washington, Seattle, Wash.

4. Hyun, S., R. Hilborn, J.J. Anderson, and B. Ernst. 2005. A statistical model for in-season forecasts of sockeye salmon (*Oncorhynchus nerka*) returns to the Bristol Bay districts of Alaska. Can. J. Fish. Aquat. Sci. 62: 1665-1680.

5. Neter, J., M.H., Kutner, C.J. Nachtcheim, and W. Wasserman. 1996. Applied linear statistical models. Fourth edition. WCB McGraw-Hill, Massachusetts, USA. 1408 pp.

6. Scheuerell, M.D., and J.G. Williams. 2005. Forecasting climate-induced changes in the survival of Snake River spring/summer Chinook salmon (*Oncorhynchus tshawytscha*). Fish. Oceanogr. 14: 448-457.

7. U.S. v. Oregon Parties. 2005. 2005-2007 Interim management agreement for upriver Chinook, sockeye, steelhead, coho, and white sturgeon.

Table 1. Notations.

Indices	
a	Age
R	Return size (the number of fish)
O_a	Abundance of fish at age a in the ocean in a year
S_a	Fish survival rate (fraction) from age a to age a+1
A_a	Maturation rate (fraction) of fish at age a
M_a	Ocean mortality rate (fraction) of fish at age a (i.e., $M_a = 1 - S_a$)
c_a	A positive proportional coefficient between ocean mortality of fish at age a, and ocean recovery rate of tagged fish at age a
t	Superscript t denotes transpose operation for a vector (e.g., $\mathbf{X}^t$ is the transpose of vector $\mathbf{X}$)
^	Point estimate of unknown quantity (e.g., $\hat{R}$)
~	Unknown predictive variable. Tilde mark (~) notation helps to distinguish between observed and predictive quantities. For example, $\tilde{R}$ is predictive (unknown) variable whereas R without tilde mark is random variable or observed data
Data	
R_a	Return size (number) of fish at age a to river from the ocean in a year
V_a	Ocean recoveries of tagged fish at age a in a year
P	Sampling fraction for tagged fish in the ocean catch in an ocean recovery year

L	Release number of tagged fish
H_a	Ocean recovery rate of tagged fish at age a in a year relative to release size number of those tagged fish.
SST	Mean of November and December sea surface temperatures measured from coastal waters of British Columbia, Canada during fish's first ocean year.
Parameters	
$\boldsymbol{\beta}$	Vector of coefficients of explanatory variables in a regression model of interest
ε	Error term in a regression model of interest
σ^2	Variance of error term in a regression model of interest
Predictive variable	
$\tilde{R}_a$	Unknown return size at age a in a forecast year
$\tilde{R}$	Unknown total return size in a forecast year

Table 2. List of Columbia River Chinook salmon (*Oncorhynchus tshwytscha*) populations of forecast interest. These are all fall run stocks except for Upper Columbia River summer (UCS) population (see Fig. 1 for origin locations).

Population name	Origin	Available data
Upriver Bright (URB)	Hanford Reach, Snake and Deschutes Rivers	Age-3, -4, and -5 returns (brood years since 1962); age-6 returns (brood years since 1975)
Bonneville Upriver Bright (BUB)	Mid-Columbia River (Bonneville, Little White Salmon national fish, Klickitat and Umatilla hatcheries)	Age-2 returns (brood years since 1978); age-3 and -4 returns (brood years since 1977); age-5 and -6 returns (brood years since 1976)
Pool Upriver Bright (PUB)	Mid-Columbia River (Bonneville, Little White Salmon national fish, Klickitat and Umatilla hatcheries)	Age-2 and -3 returns (brood years since 1983); age-4 returns (brood years since 1982); age-5 returns (brood years since 1981); age-6 returns (brood years since 1983)
Lower River Wild (LRW)	Lewis and Cowlitz Rivers	Age-2, -3, -4, and -5 returns (brood years since 1962); age-6 returns (brood years since 1975)

Lower River Hatchery (LRH)	Lewis and Cowlitz Rivers	Age-2, -3, -4, and -5 returns (brood years since 1962); age-6 returns (brood years since 1975)
Bonneville Pool Hatchery (BPH), or "Tule"	Bonneville Pool Hatchery also called Spring Creek Hatchery	Age-2, -3, -4 and -5 returns (brood years since 1962)
Upper Columbia River summer (UCS)	Wenatchee, Entiat and Methow Rivers	Age-3 returns (brood years since 1987); age-4 returns (brood years since 1986); age-5 returns (brood years since 1985); age-6 returns (brood years since 1984)

Table 3. Summary of candidate models for forecasts of age-return sizes. Column vector **X** in eq. 5 contains explanatory variables, where "1" denotes intercept term. Measurement year of sea surface temperature (SST) variable is fish's ocean first year.

Model	**Form**	**Vector $\mathbf{X}^t$**
M1	$R_a = \beta_0 + \beta_1 \cdot R_{a-1} + \varepsilon$	(1, R_{a-1})
M2	Autoregressive error version of M1	
M3	$R_a = \beta_0 + \beta_1 \cdot R_{a-1} + \beta_2 \cdot H_{a-1} \cdot R_{a-1} + \varepsilon$	(1, R_{a-1}, $H_{a-1}R_{a-1}$)
M4	Autoregressive error version of M3	
M5	$R_a = \beta_0 + \beta_1 \cdot R_{a-1} + \beta_2 \cdot SST + \varepsilon$	(1, R_{a-1}, SST)
M6	Autoregressive error version of M5	
M7	$R_a = \beta_0 + \beta_1 \cdot R_{a-1} + \beta_2 \cdot H_{a-1} \cdot R_{a-1} + \beta_3 \cdot SST + \varepsilon$	(1, R_{a-1}, $H_{a-1}R_{a-1}$, SST)
M8	Autoregressive error version of M7	
M9	$R_a = \beta \cdot R_{a-1} + \varepsilon$	(R_{a-1})
M10	Autoregressive error version of M9	
M11	$R_a = \beta_1 \cdot R_{a-1} + \beta_2 \cdot H_{a-1} \cdot R_{a-1} + \varepsilon$	(R_{a-1}, $H_{a-1}R_{a-1}$)
M12	Autoregressive error version of M11	
M13	$R_a = \beta_1 \cdot R_{a-1} + \beta_2 \cdot SST + \varepsilon$	(R_{a-1}, SST)
M14	Autoregressive error version of M13	
M15	$R_a = \beta_1 \cdot R_{a-1} + \beta_2 \cdot H_{a-1} \cdot R_{a-1} + \beta_3 \cdot SST + \varepsilon$	(R_{a-1}, $H_{a-1}R_{a-1}$, SST)
M16	Autoregressive error version of M15	

Figure 1. Columbia River basin and locations of fish population natal rivers and hatcheries. Bold letters denote hatchery locations: B = Bonneville dam hatchery, L = Little White Salmon National fish hatchery; S = Spring Creek hatchery; K = Klickitat hatchery; and U = Umatilla hatchery. Populations and origin locations are listed in Table 2.

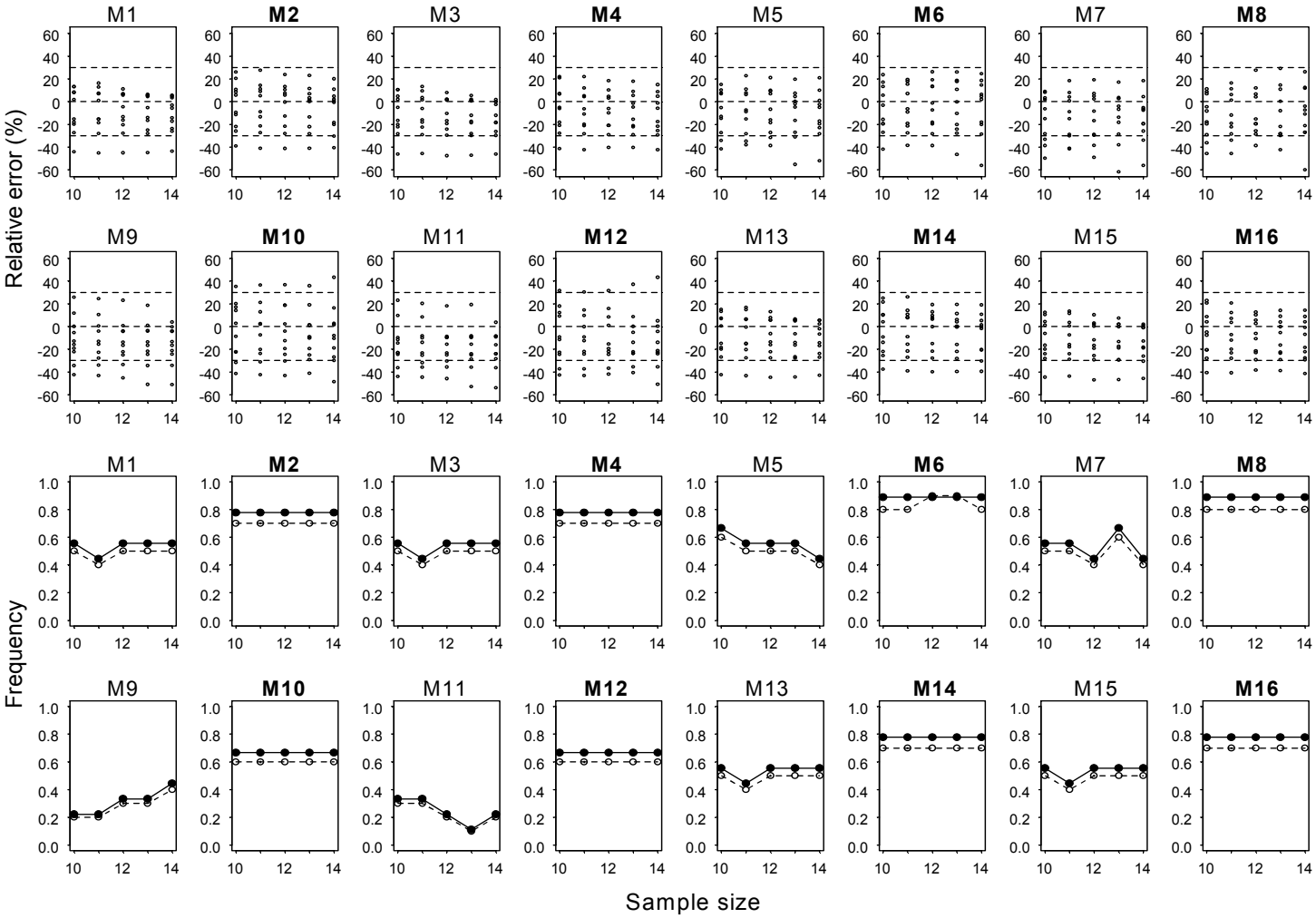


Figure 2. Forecast model performance for total run size of Upriver Bright (URB) population in terms of forecast relative error (%) and 90% prediction interval. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in panels of first two rows are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model. Horizontal broken lines are added at $\pm$ 30% for help to compare models in forecast relative errors. The y-axis in panels of the last two rows is frequency that the prediction intervals from the corresponding model successfully cover actual runs. For example, the frequency of 0.4 means that the prediction intervals cover actual values 40% and miss those values 60%. Black dot on solid line is the frequency with forecast of 2001 run being removed, whereas open dot on broken line is that with all forecasts being included.

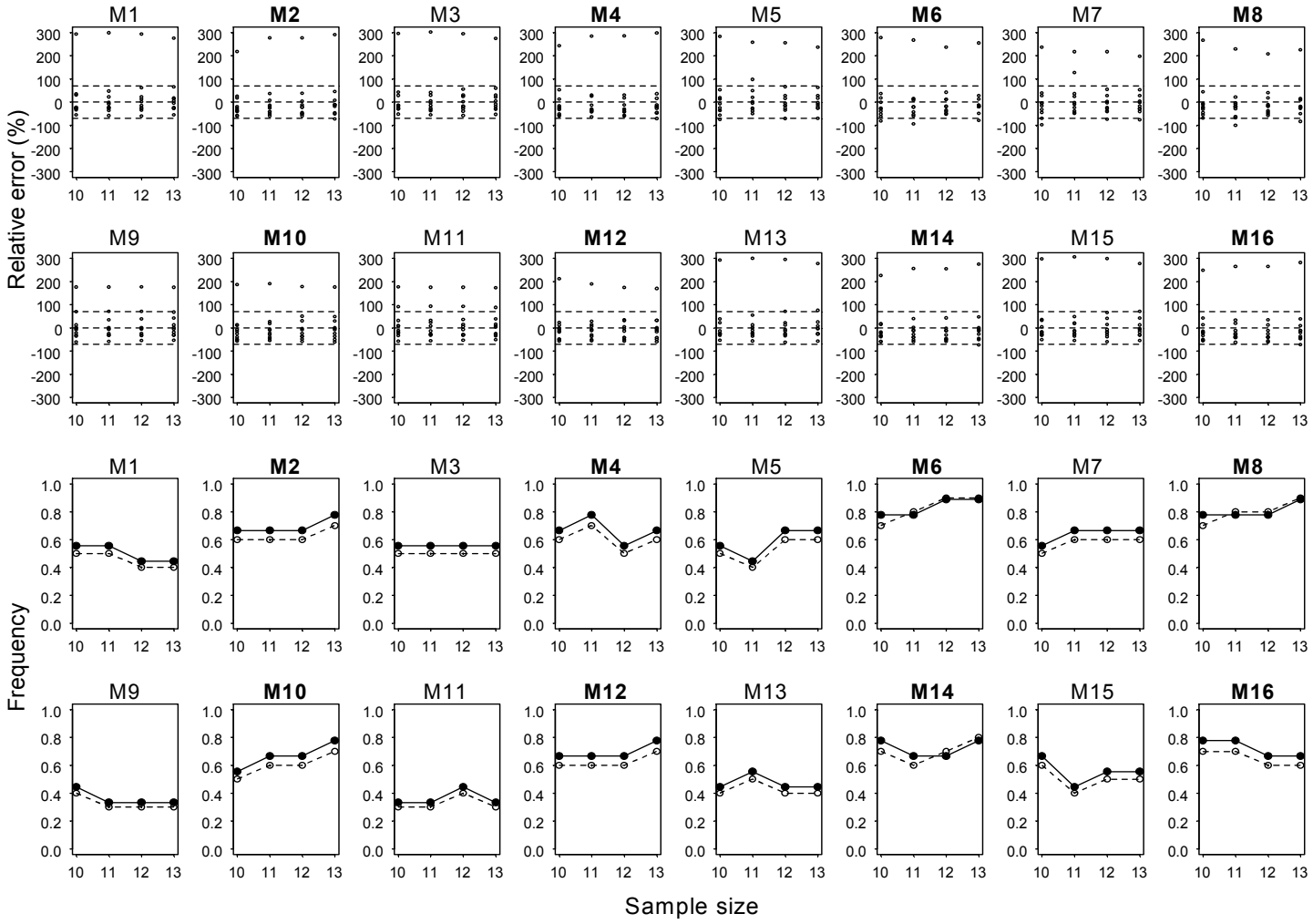


Figure 3. Forecast model performance for total run size of Bonneville Upriver Bright (BUB) population in terms of forecast relative error (%) and 90% prediction interval. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in panels of first two rows are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model. Horizontal broken lines are added at ± 70% for help to compare models in forecast relative errors. The y-axis in panels of the last two rows is frequency that the prediction intervals from the corresponding model successfully cover actual runs. For example, the frequency of 0.4 means that the prediction intervals cover actual values 40% and miss those values 60%. Black dot on solid line is the frequency with forecast of 1998 run being removed, whereas open dot on broken line is that with all forecasts being included.

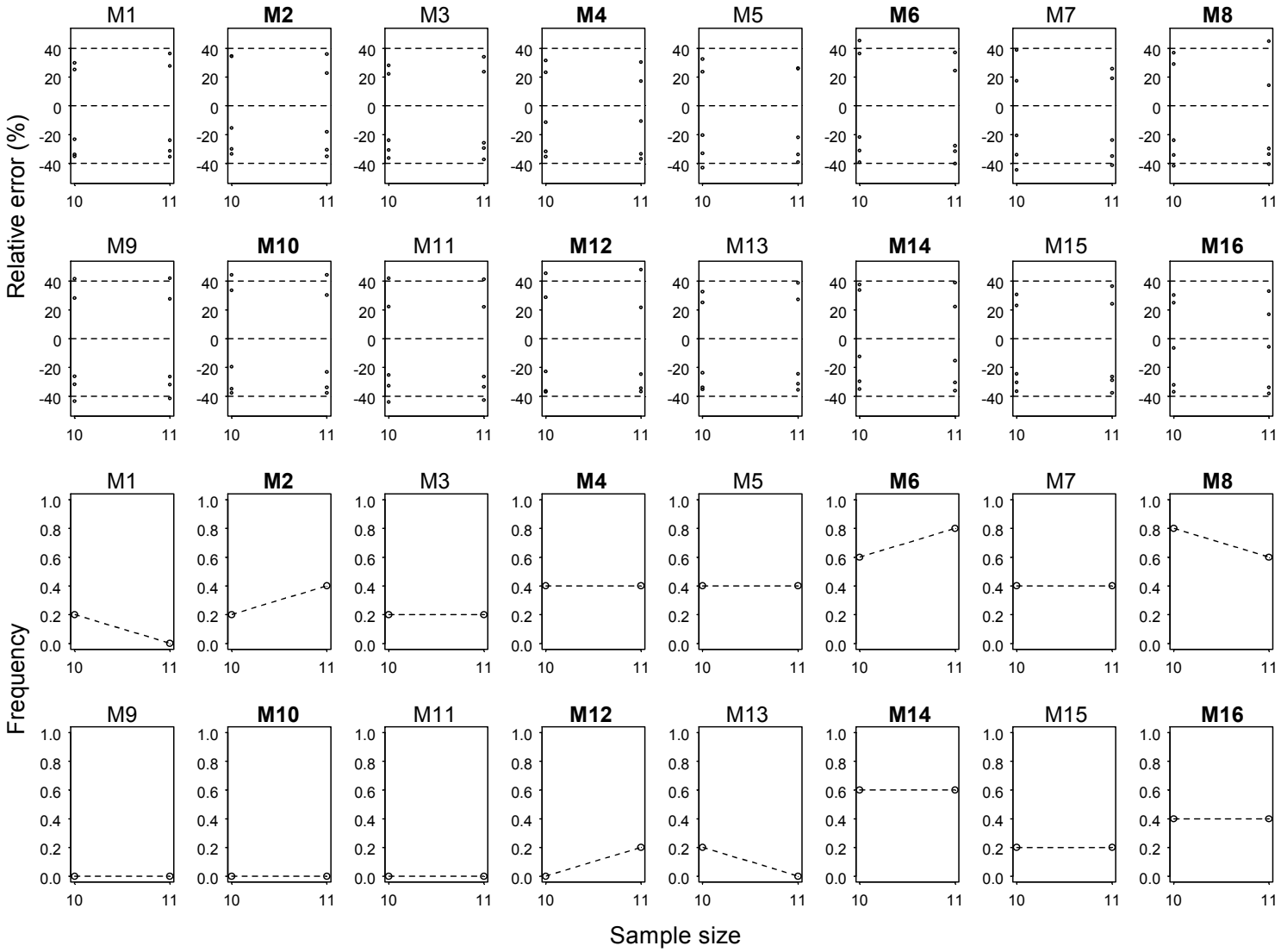


Figure 4. Forecast model performance for total run size of Pool Upriver Bright (PUB) population in terms of forecast relative error (%) and 90% prediction interval. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in panels of first two rows are relative errors (%) in forecasts of five years (2000-2004)' runs made with the corresponding forecast model. Horizontal broken lines are added at $\pm 40\%$ for help to compare models in forecast relative errors. The y-axis in panels of the last two rows is frequency that the prediction intervals from the corresponding model successfully cover actual runs. For example, the frequency of 0.4 means that the prediction intervals cover actual values 40% and miss those values 60%.

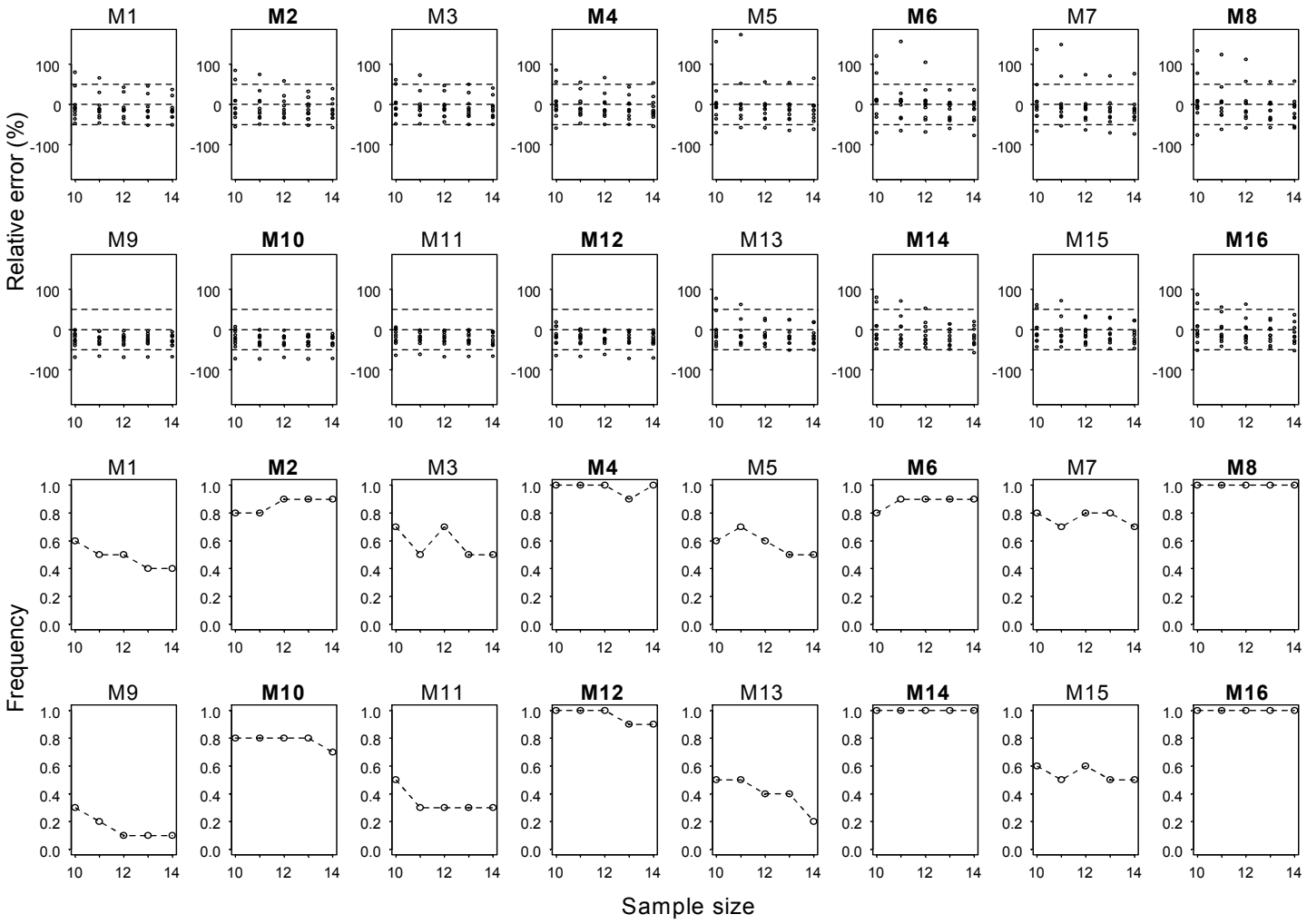


Figure 5. Forecast model performance for total run size of Lower River Wild (LRW) population in terms of forecast relative error (%) and 90% prediction interval. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in panels of first two rows are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model. Horizontal broken lines are added at $\pm$ 50% for help to compare models in forecast relative errors. The y-axis in panels of the last two rows is frequency that the prediction intervals from the corresponding model successfully cover actual runs. For example, the frequency of 0.4 means that the prediction intervals cover actual values 40% and miss those values 60%.

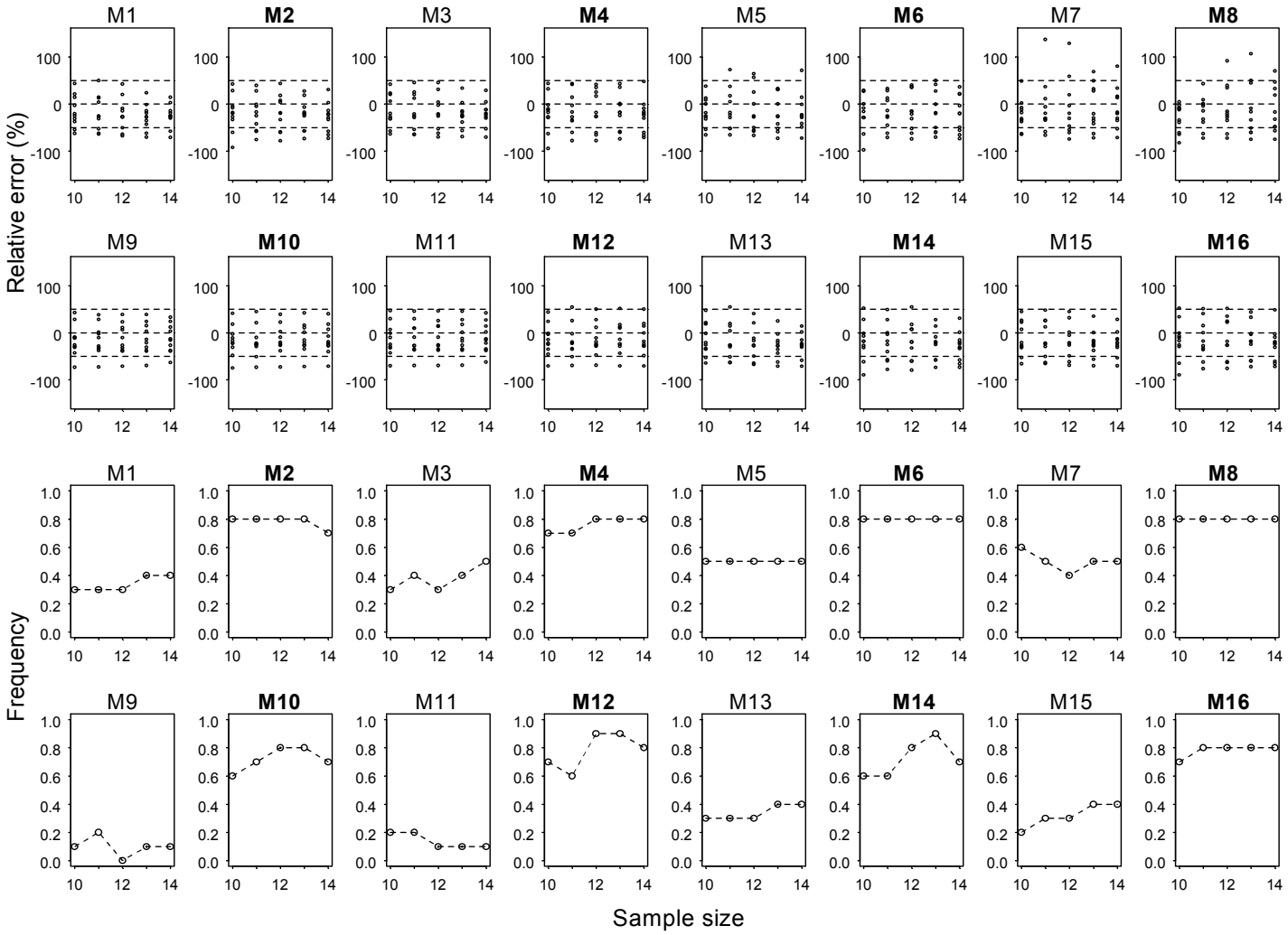


Figure 6. Forecast model performance for total run size of Lower River Hatchery (LRH) population in terms of forecast relative error (%) and 90% prediction interval. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in panels of first two rows are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model. Horizontal broken lines are added at ± 50% for help to compare models in forecast relative errors. The y-axis in panels of the last two rows is frequency that the prediction intervals from the corresponding model successfully cover actual runs. For example, the frequency of 0.4 means that the prediction intervals cover actual values 40% and miss those values 60%.

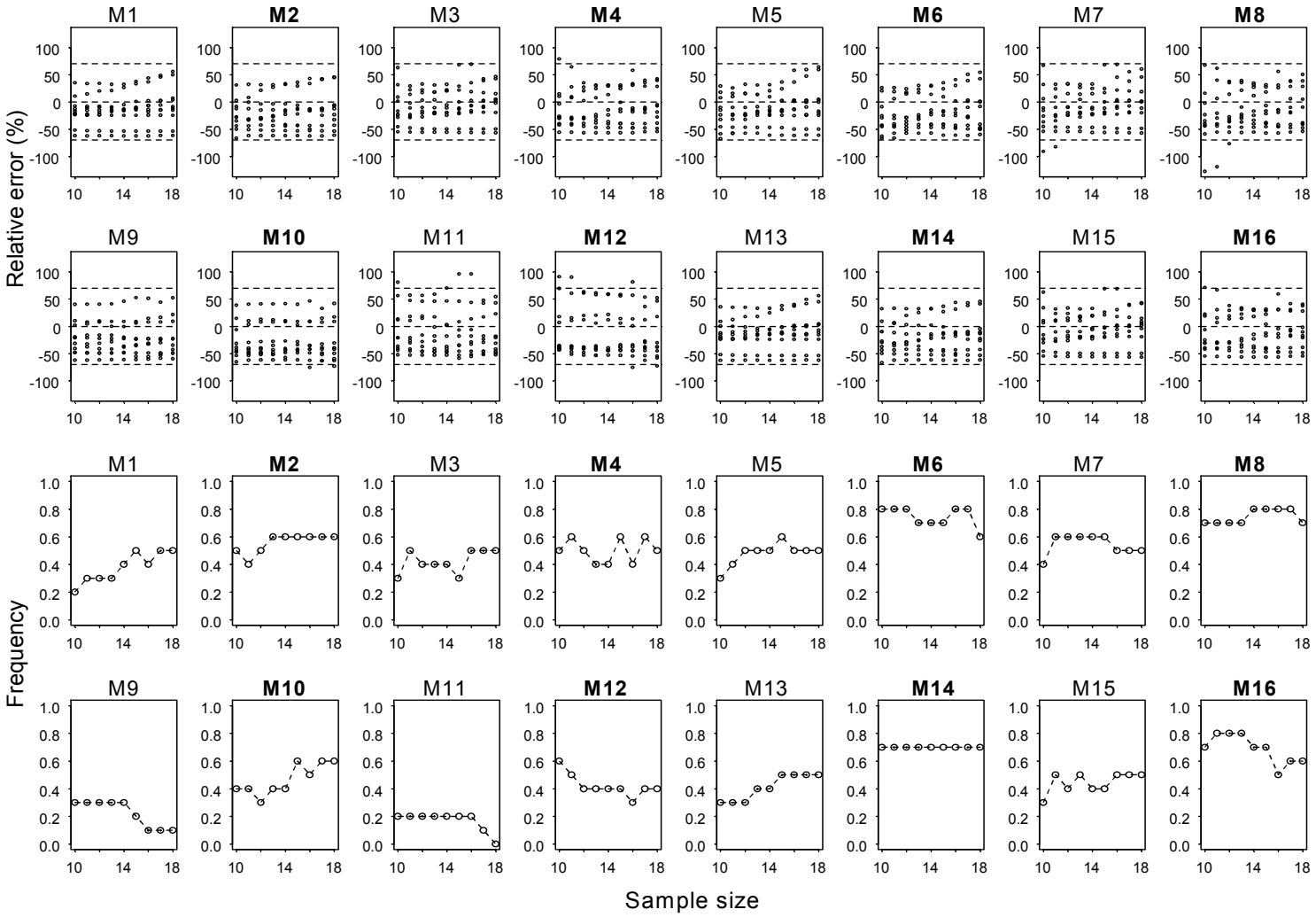


Figure 7. Forecast model performance for total run size of Bonneville Pool Hatchery (BPH) population in terms of forecast relative error (%) and 90% prediction interval. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in panels of first two rows are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model. Horizontal broken lines are added at ± 70% for help to compare models in forecast relative errors. The y-axis in panels of the last two rows is frequency that the prediction intervals from the corresponding model successfully cover actual runs. For example, the frequency of 0.4 means that the prediction intervals cover actual values 40% and miss those values 60%.

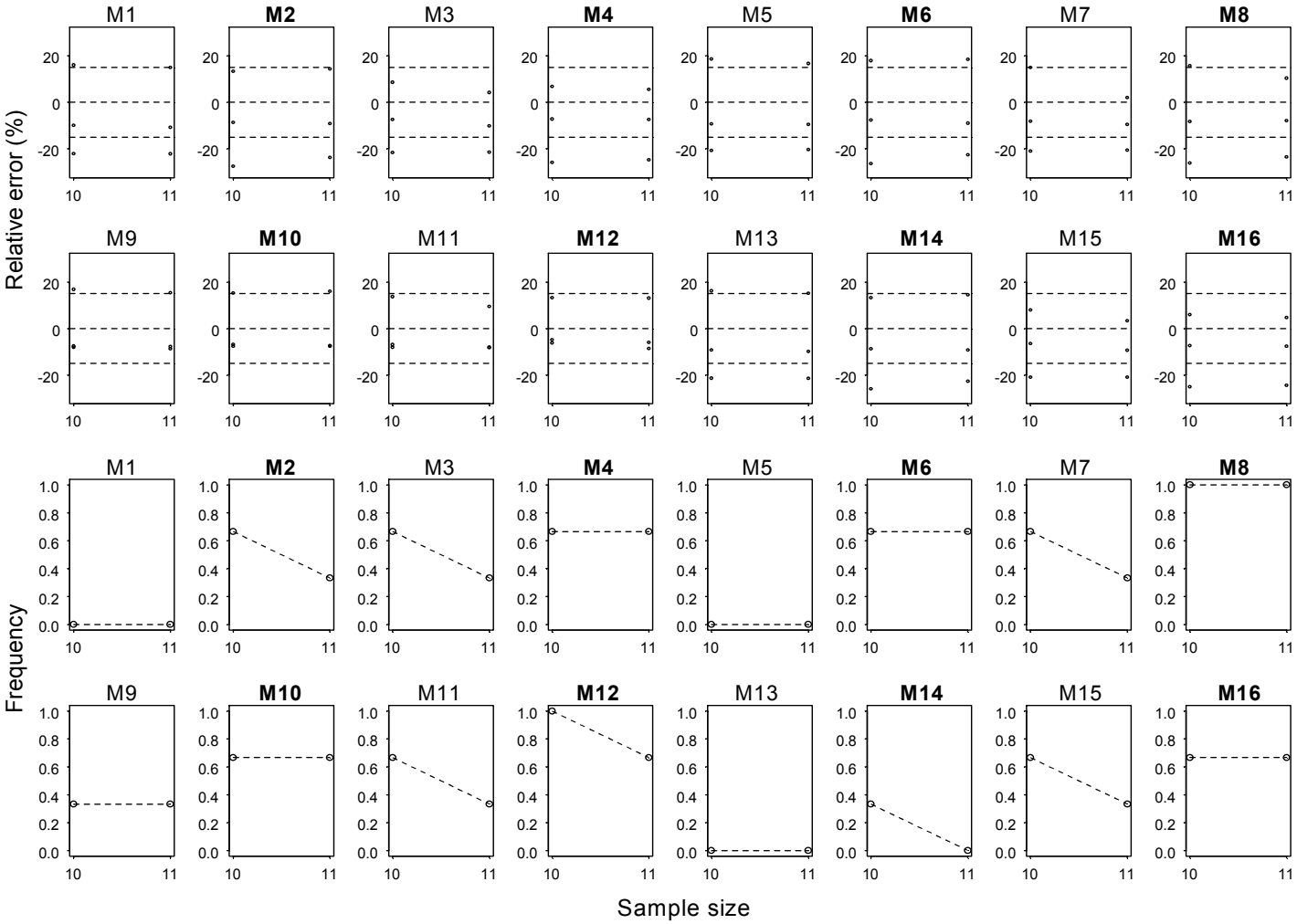


Figure 8. Forecast model performance for total run size of Upper Columbia River summer (UCS) population in terms of forecast relative error (%) and 90% prediction interval. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in panels of first two rows are relative errors (%) in forecasts of three years (2002-2004)' runs made with the corresponding forecast model. Horizontal broken lines are added at ± 15% for help to compare models in forecast relative errors. The y-axis in panels of the last two rows is frequency that the prediction intervals from the corresponding model successfully cover actual runs. For example, the frequency of 0.4 means that the prediction intervals cover actual values 40% and miss those values 60%.

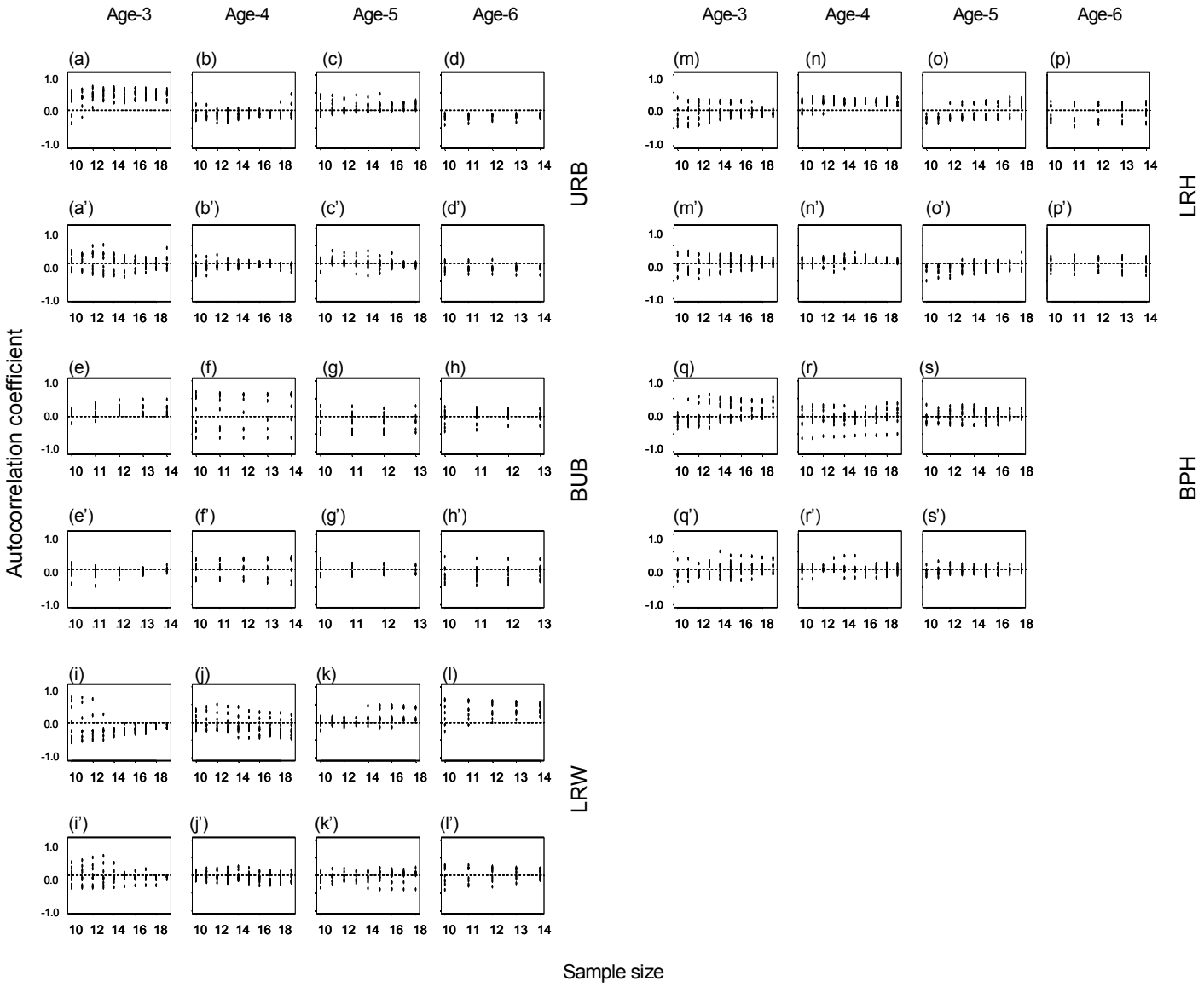


Figure 9. Examination of autocorrelations of model residuals over lag 1, comparing traditional and selected models. Dots over each sample size (data years) in a panel are autocorrelation coefficients of residuals over lag-1 from the corresponding model for forecasting population- and age- runs in 1995-2004. (a)-(d) are such autocorrelation coefficients from forecasts of URB age-3, -4, -5, and -6 runs by the traditional model M1 while (a')-(d') are those by model M6 selected for URB population. In the same arrangement, (e)-(h) = those of BUB age-3, -4, -5, and -6 runs by the traditional model M1; (e')-(h') = those by model M10 selected for BUB population; (i)-(l) = those of LRW age-3, -4, -5 and -6 runs by the traditional model M1; (i')-(l') = those by model M4 selected for LRW population; (m)-(p) = those of LRH age-3, -4, -5 and -6 runs by the traditional model M1; (m')-(p') = those by model M14 selected for LRH population; (q)-(s) = those of BPH age-3, -4, and -5 runs by the traditional model M1; (m')-(p') = those by model M16 selected for BPH population.

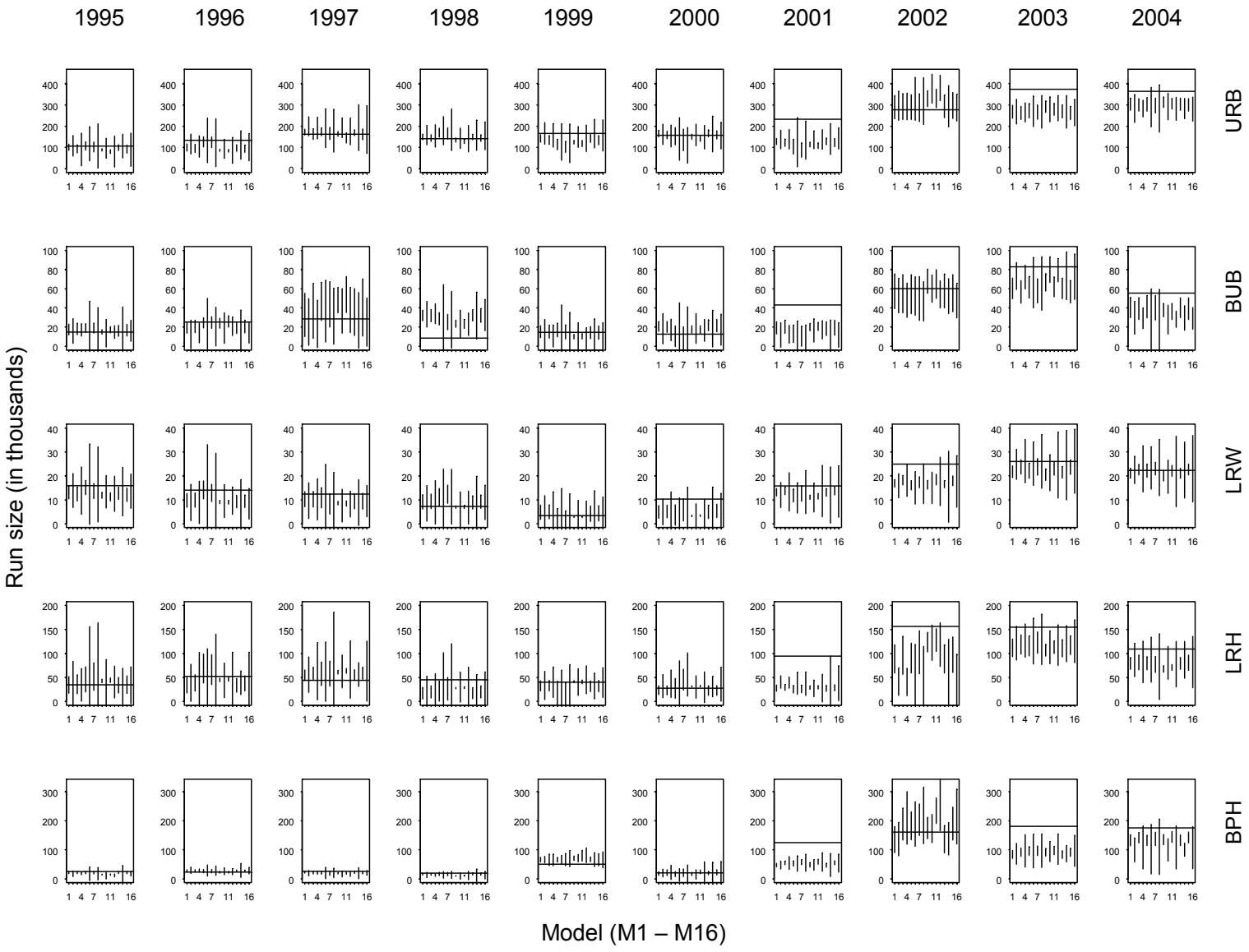


Figure 10. Comparison of 90% prediction interval from 16 models (x-axis) and actual runs, where sample size (years of data) used for each hind-casting forecast was fixed as 13. 10 years' hind-casting forecasts (columns) were made for five populations (rows). Vertical line over each model in each panel is 90% prediction interval (i.e., from lower bound to upper bound), and horizontal line in each panel indicates actual run size. This summary for PUB and UCS populations is not shown because of lack of data on these populations (Figs. 4 and 8).

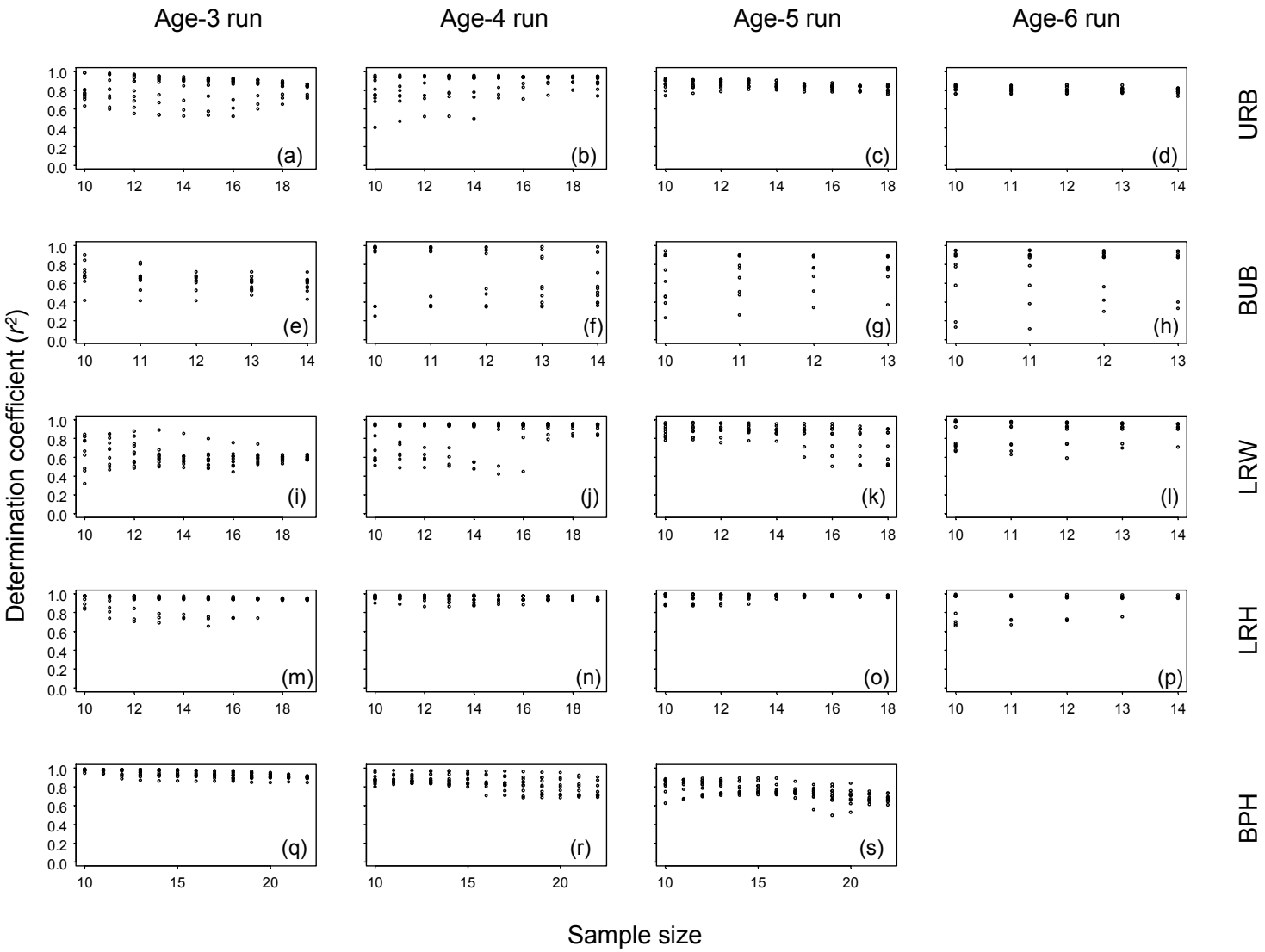


Figure 11. Goodness-of-fit of selected model for each population in terms of determination coefficients (r^2): (a)-(d) = goodness of fit of model M6 for forecasts of URB age-3, -4, -5 and -6 runs; (e)-(h) = that of model M10 for forecasts of BUB age-3, -4, -5 and -6 runs; (i)-(l) = that of model M4 for forecasts of LRW age-3, -4, -5 and -6 runs; (m)-(p) = that of model M14 for forecasts of LRH age-3, -4, -5 and -6 runs; (q)-(s) = that of model M16 for forecasts of BPH age-3, -4, and -5 runs. Vertical dots over each sample size each panel are determination coefficients (r^2) of the corresponding model used for hind-casting forecasts of population- and age-runs in 10 years (1995-2004). This summary for PUB and UCS populations is not shown because of lack of data on these populations (Figs. 4 and 8).

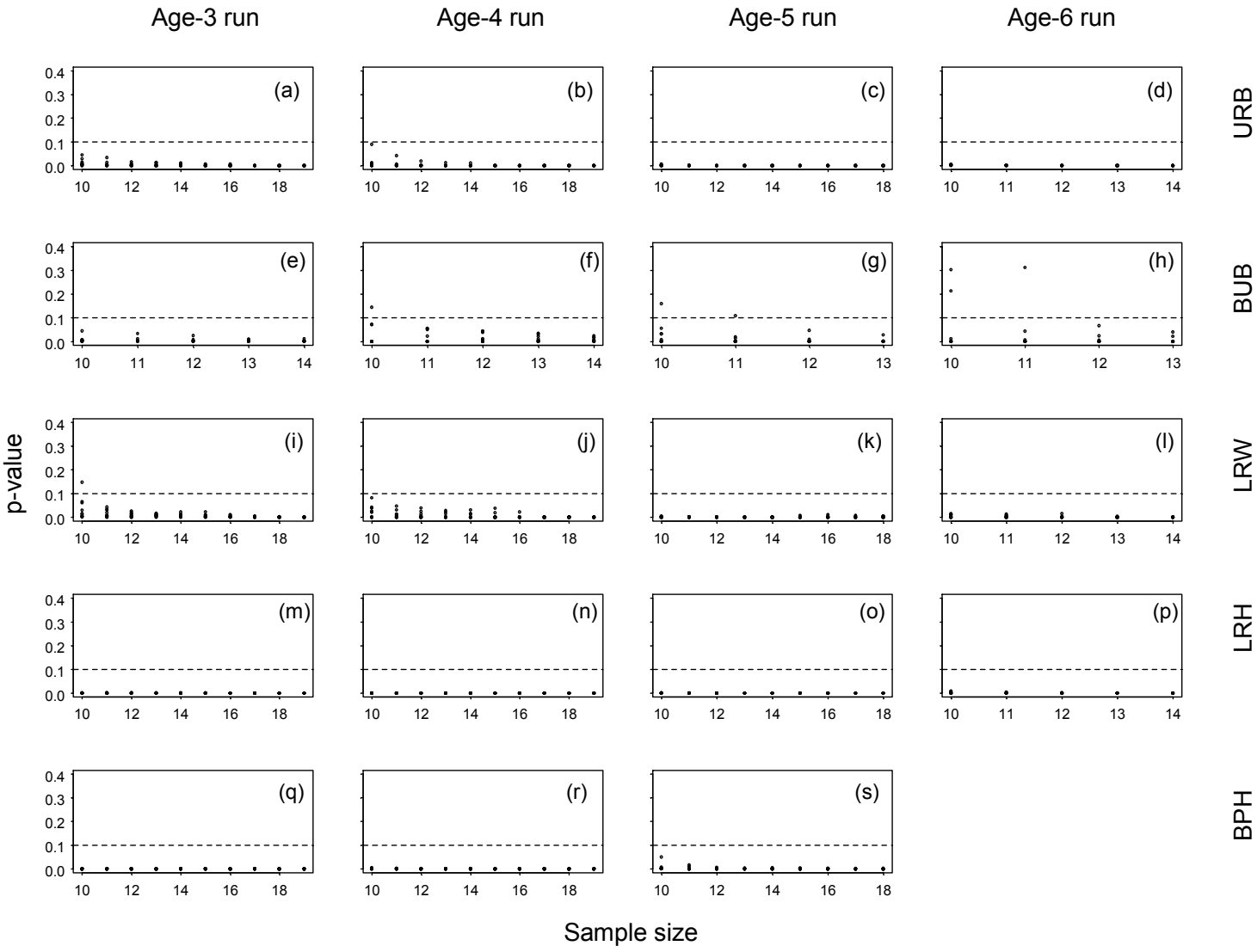


Figure 12. Significance of selected model for each population in terms of p-value of F-statistics: (a)-(d) = significance of model M6 for forecasts of URB age-3, -4, -5 and -6 runs; (e)-(h) = that of model M10 for forecasts of BUB age-3, -4, -5 and -6 runs; (i)-(l) = that of model M4 for forecasts of LRW age-3, -4, -5 and -6 runs; (m)-(p) = that of model M14 for forecasts of LRH age-3, -4, -5 and -6 runs; (q)-(s) = that of model M16 for forecasts of BPH age-3, -4, and -5 runs. Vertical dots over each sample size each panel are p-values of F-statistics of the corresponding model used for hind-casting forecasts of population- and age-runs in 10 years (1995-2004). Horizontal broken line in each panel indicates p-value of 0.1. This summary for PUB and UCS populations is not shown because of lack of data on these populations (Figs. 4 and 8).

APPENDIX

A Autoregressive error model

This description is lag-1 autoregressive error model, where Cochrane-Orcutt procedure is used for estimation of the regression parameters. Neter et al. (1996) shows the description for the regression model with one explanatory variable (i.e., $Y = \beta_0 + \beta_1 \cdot X + \varepsilon$). Here, we describe a generalized form where two explanatory variables are considered.

$$Y_t = (\beta_0 + \beta_1 \cdot X_{t,1} + \beta_2 \cdot X_{t,2} + \varepsilon_t) \tag{A-1}$$

Letting ρ be the autocorrelation parameter between errors over lag 1, we introduce the transformed dependent variable, Y_t' like the following:

$$\begin{aligned} Y_t' &= Y_t - \rho \cdot Y_{t-1} \\ &= (\beta_0 + \beta_1 \cdot X_{t,1} + \beta_2 \cdot X_{t,2} + \varepsilon_t) - \rho(\beta_0 + \beta_1 \cdot X_{t-1,1} + \beta_2 \cdot X_{t-1,2} + \varepsilon_{t-1}) \\ &= \beta_0(1-\rho) + \beta_1(X_{t,1} - \rho \cdot X_{t-1,1}) + \beta_2(X_{t,2} - \rho \cdot X_{t-1,2}) + (\varepsilon_t - \rho \cdot \varepsilon_{t-1}) \end{aligned} \tag{A-2}$$

Letting

$$\begin{aligned} X_{t,2}' &\equiv (X_{t,1} - \rho \cdot X_{t-1,1}) \\ X_{t,2}' &\equiv (X_{t,2} - \rho \cdot X_{t-1,2}) \\ \beta_0' &\equiv \beta_0(1-\rho) \\ \beta_1' &\equiv \beta_1 \\ \beta_2' &\equiv \beta_2 \\ u_t &\equiv (\varepsilon_t - \rho \cdot \varepsilon_{t-1}) \end{aligned}$$

eq. A-2 is

$$Y_t' = \beta_0' + \beta_1' \cdot X_{t,1}' + \beta_2' \cdot X_{t,2}' + u_t \tag{A-3}$$

First, one needs to estimate ρ from

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t \tag{A-4}$$

Equation A-4 is an ordinary regression model without the intercept term, where ρ is the slope of the regression line and u_t is the error term.

Once the estimate of ρ is obtained, now transformed data are obtained using the following equations.

$$\begin{aligned} Y_t' &= Y_t - \hat{\rho} \cdot Y_{t-1} \\ X_{t,1}' &= X_{t,1} - \hat{\rho} \cdot X_{t-1,1} \\ X_{t,2}' &= X_{t,2} - \hat{\rho} \cdot X_{t-1,2} \end{aligned}$$

Transformed model of eq. A-3 is fitted to the above transformed data, yielding an estimated regression function:

$$\hat{Y}' = b_0' + b_1' \cdot X_1' + b_2' \cdot X_2' \tag{A-5}$$

$$
\begin{aligned}
b_0' &= b_0(1-\hat{\rho}) \\
b_1' &= b_1 \\
b_2' &= b_2
\end{aligned}
$$

Thus, estimates of parameters in the original model (before transformation), and variances of those estimates are

$$
\begin{aligned}
b_0 &= b_0'/(1-\hat{\rho}) \\
b_1 &= b_1' \\
b_2 &= b_2' \\
Var(b_0) &= Var(b_0')/(1-\hat{\rho})^2 \\
Var(b_1) &= Var(b_1') \\
Var(b_2) &= Var(b_2')
\end{aligned}
\tag{A-6}
$$

The fitted regression function in the original variables is

$$
\hat{Y} = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 \tag{A-7}
$$

A.1 Forecasting

$$
\tilde{Y}_{n+1} = \beta_0 + \beta_1 \cdot X_{n+1,1} + \beta_2 \cdot X_{n+1,2} + \rho \cdot \varepsilon_n + u_{n+1} \tag{A-8}
$$

Estimates of β_0, β_1, and β_2 are shown in eq. A-6, and the estimate of ρ is from eq. A-4. ε_n is estimated by the residual e_n. Also, $E(u_{n+1}) = 0$. That is,

$$
\begin{aligned}
e_n &= Y_n - \hat{Y}_n \\
&= Y_n - (b_0 + b_1 \cdot X_{n,1} + b_2 \cdot X_{n,2})
\end{aligned}
\tag{A-9}
$$

Thus, the expected valule of forecast, $\tilde{Y}_{n+1}$ is

$$
E(\tilde{Y}_{n+1}) = b_0 + b_1 \cdot X_{n+1,1} + b_2 \cdot X_{n+1,2} + \hat{\rho} \cdot e_n \tag{A-10}
$$

Variance of forecast, $\tilde{Y}_{n+1}$ is the sum of MSE and variance of the fitted value.

$$
Var(\tilde{Y}_{n+1}) = MSE + Var(\hat{Y}_{n+1}) \tag{A-11}
$$

where MSE is based on eq. A-3. For calculation of $Var(\hat{Y}_{n+1})$, the following matrix form is convenient when there are multiple explanatory variables.

$$
Var(\hat{Y}_{n+1}) = \mathbf{X_{n+1}^t} \cdot \mathbf{\Sigma(b)} \cdot \mathbf{X_{n+1}} \tag{A-12}
$$

where $\Sigma(\mathbf{b})$ is the covariance-variance matrix of coefficients in the regression model of eq. A-3, and $\mathbf{X_{n+1}^t}$ is the transpose form of the column vector, $\mathbf{X_{n+1}}$. That is, the transpose form is the row vector (= [1, $X_{n+1,1}$, $X_{n+1,2}$]).

Table A1. Summary timetable of project activities.

Although the contract term started on 1 March 2005, the research fund was not awarded until June 2005.

	June 2005 – June 2006											
Tasks	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	**3**	**4**	**5**	**6**
Data collection and filtering raw data	X	X	X	X									
Literature review	X	X	X	X									
Data analysis, model development, model validation	X	X	X	X	X	X	X	X	X	X	X		
Bi-annual report in an oral presentation and/or written format in related meetings (e.g., the TAC meetings, Statistics for Aquatic Resources)				X	X	X	X	X	X	X			
Incorporating feedback from colleagues, and writing final report in a scientific literature format									X	X	X	X	X

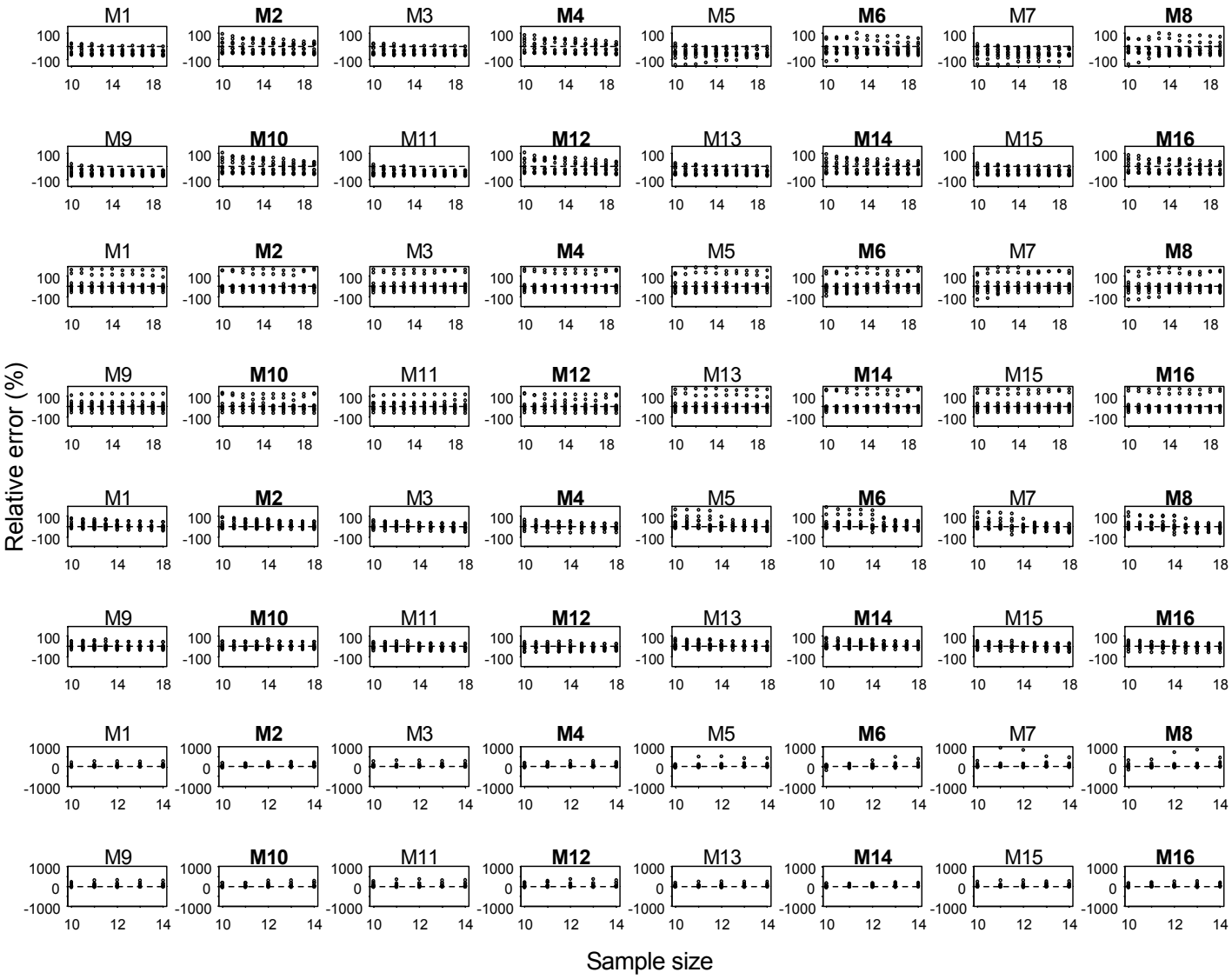


Figure A1. Forecast model performance for age-specific run size of Upriver Bright (URB) population in terms of forecast relative error (%) where the errors from the respective model are arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in each panel are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model.

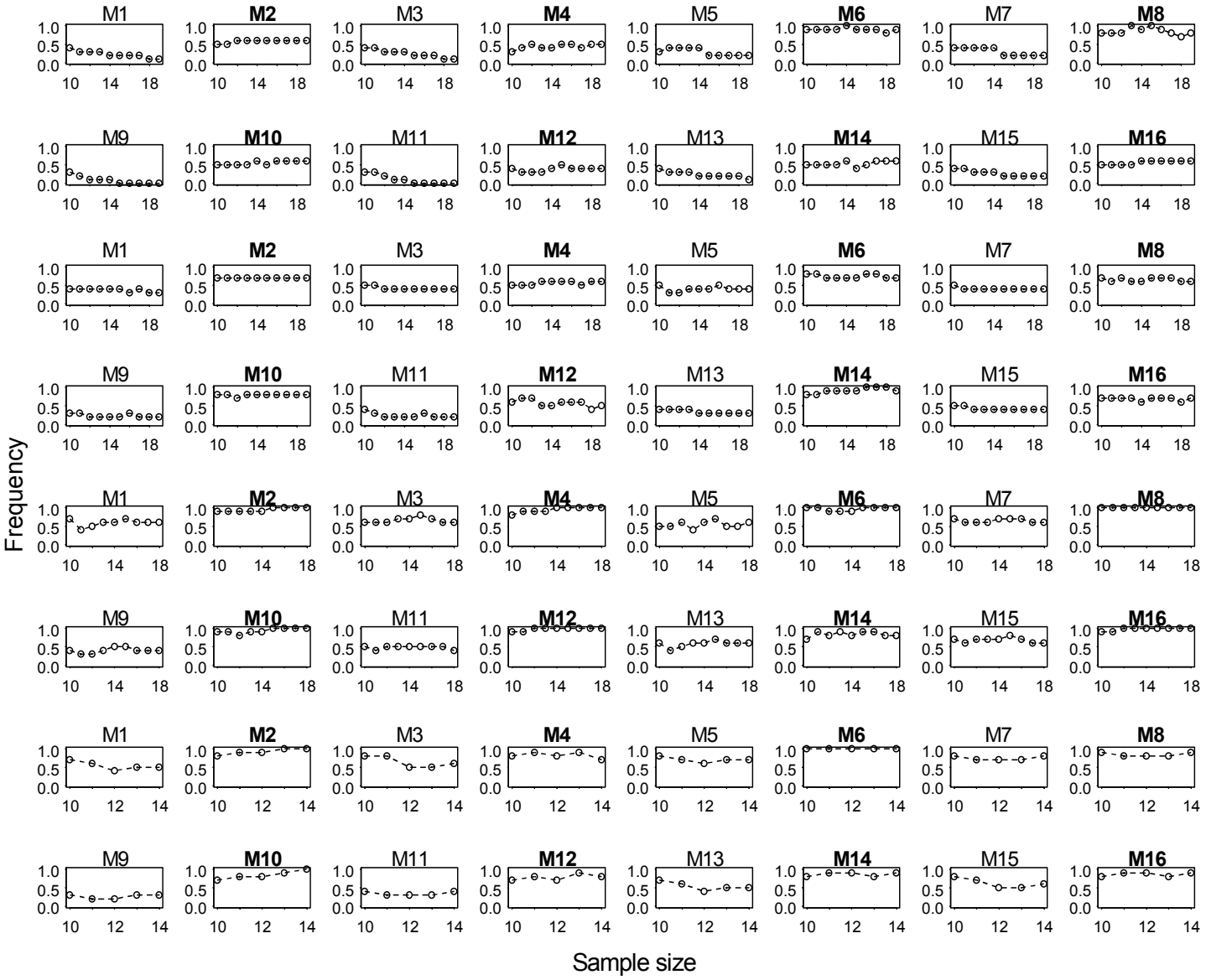


Figure A2. Forecast model performance for age-specific run size of Upriver Bright (URB) population in terms of 90% prediction interval where the prediction coverage from the respective model is arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's 10-year hind-casting forecasts (1995-2004) successfully cover actual runs.

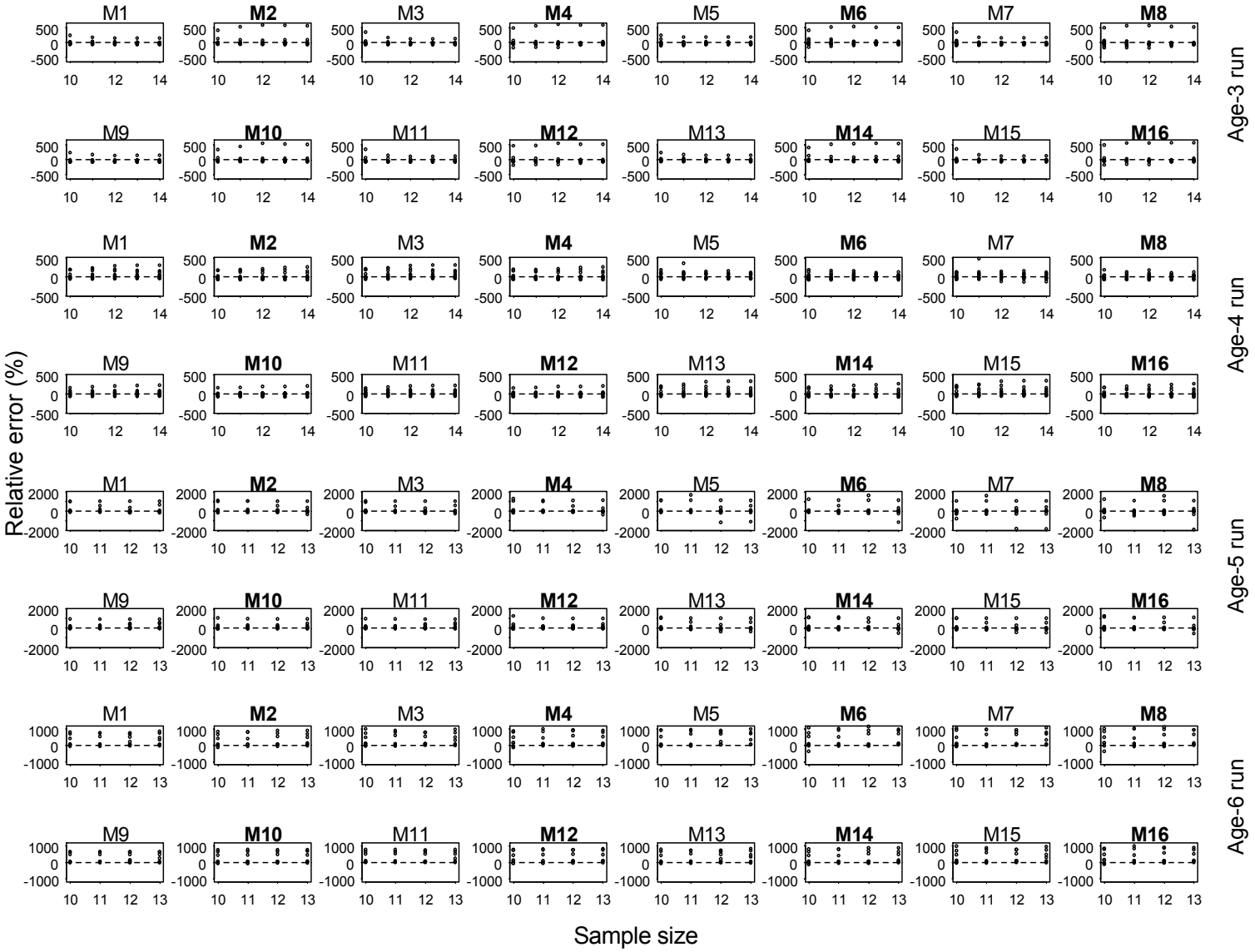


Figure A3. Forecast model performance for age-specific run size of Bonneville Upriver Bright (BUB) population in terms of forecast relative error (%) where the errors from the respective model are arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in each panel are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model.

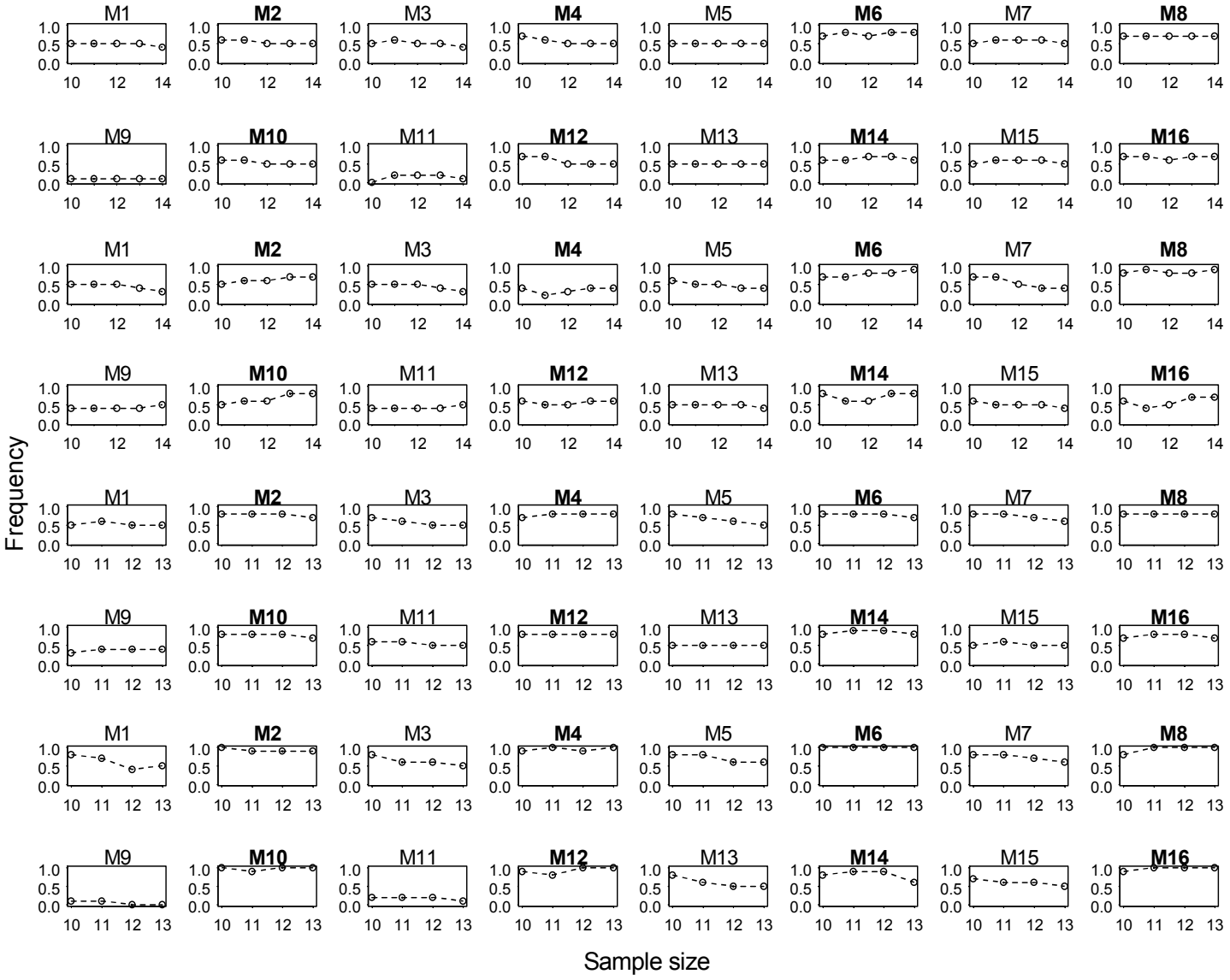


Figure A4. Forecast model performance for age-specific run size of Bonneville Upriver Bright (BUB) population in terms of 90% prediction interval where the prediction coverage from the respective model is arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's 10-year hind-casting forecasts (1995-2004) successfully cover actual runs.

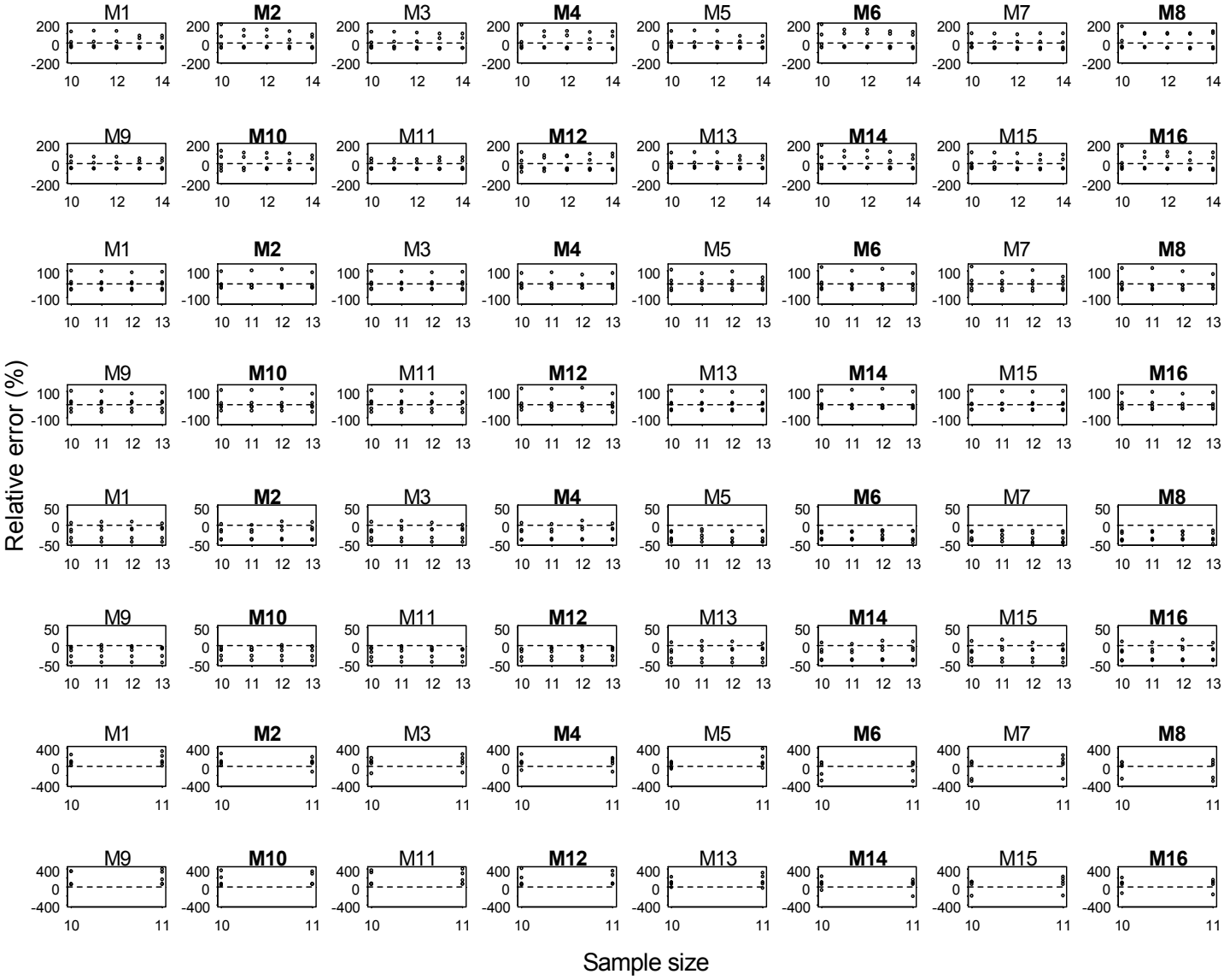


Figure A5. Forecast model performance for age-specific run size of Pool Upriver Bright (PUB) population in terms of forecast relative error (%) where the errors from the respective model are arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in each panel are relative errors (%) in forecasts of five years (2000-2004)' runs made with the corresponding forecast model.

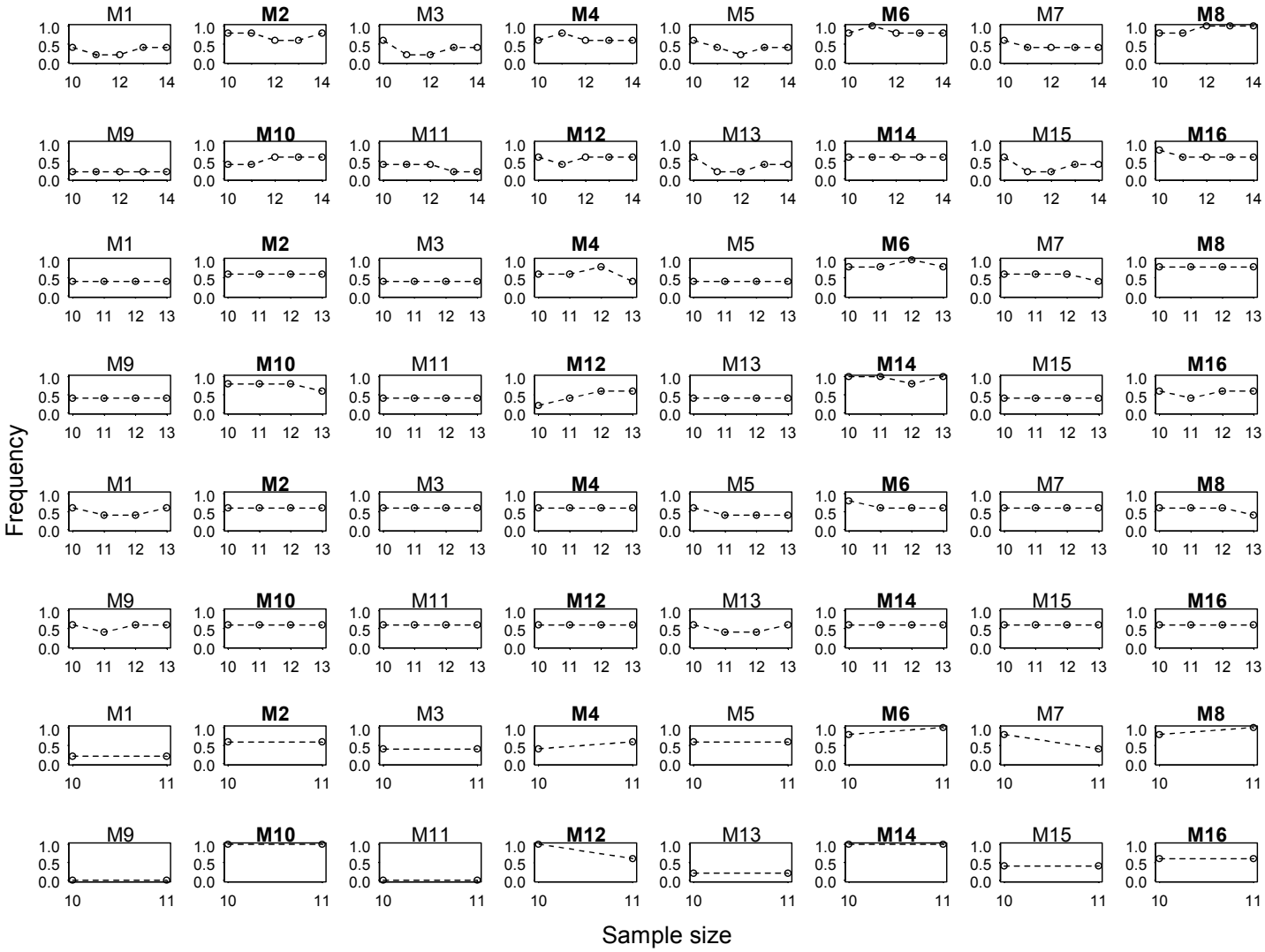


Figure A6. Forecast model performance for age-specific run size of Pool Upriver Bright (PUB) population in terms of 90% prediction interval where the prediction coverage from the respective model is arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's five-year hind-casting forecasts (2000-2004) successfully cover actual runs.

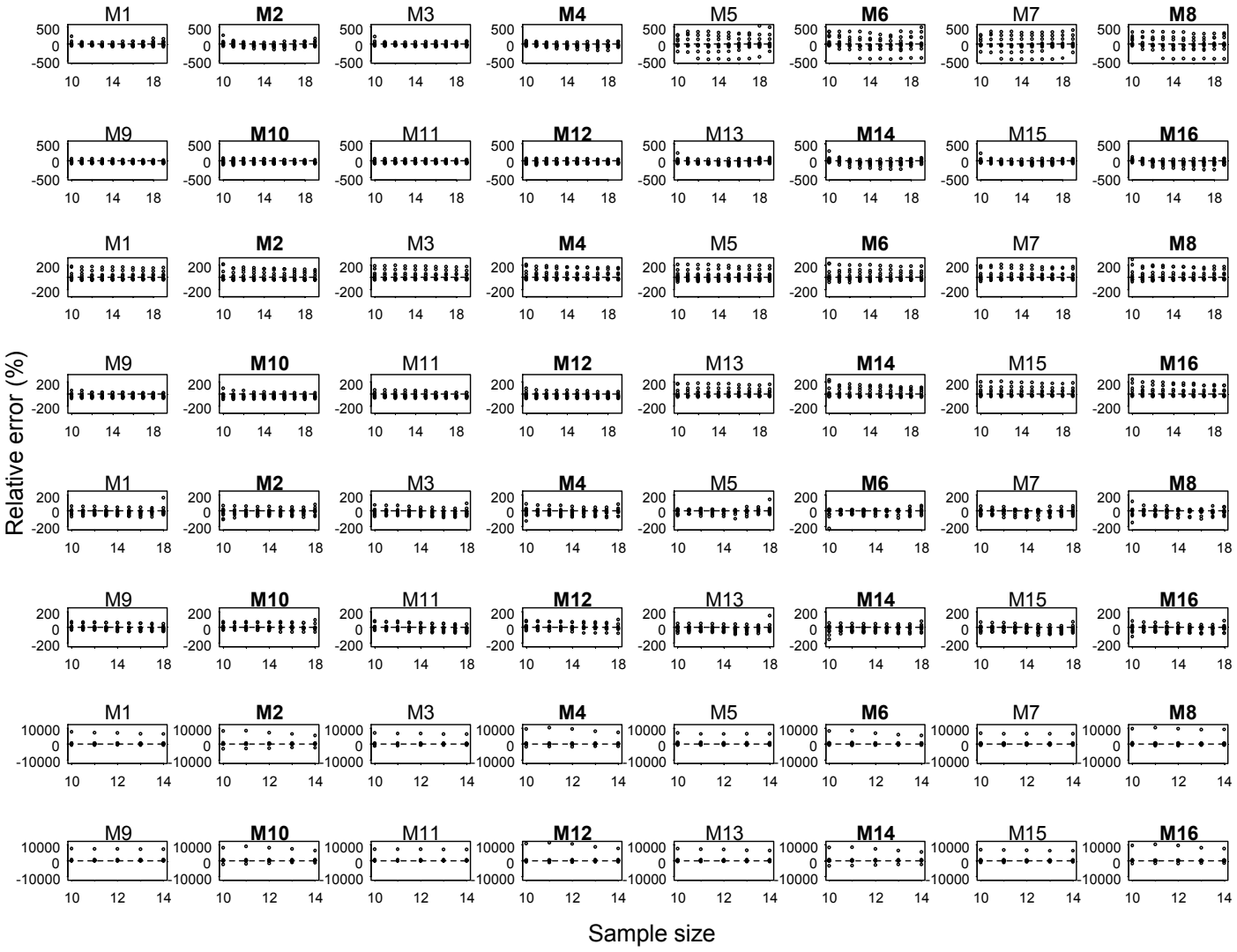


Figure A7. Forecast model performance for age-specific run size of Lower River Wild (LRW) population in terms of forecast relative error (%) where the errors from the respective model are arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in each panel are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model.

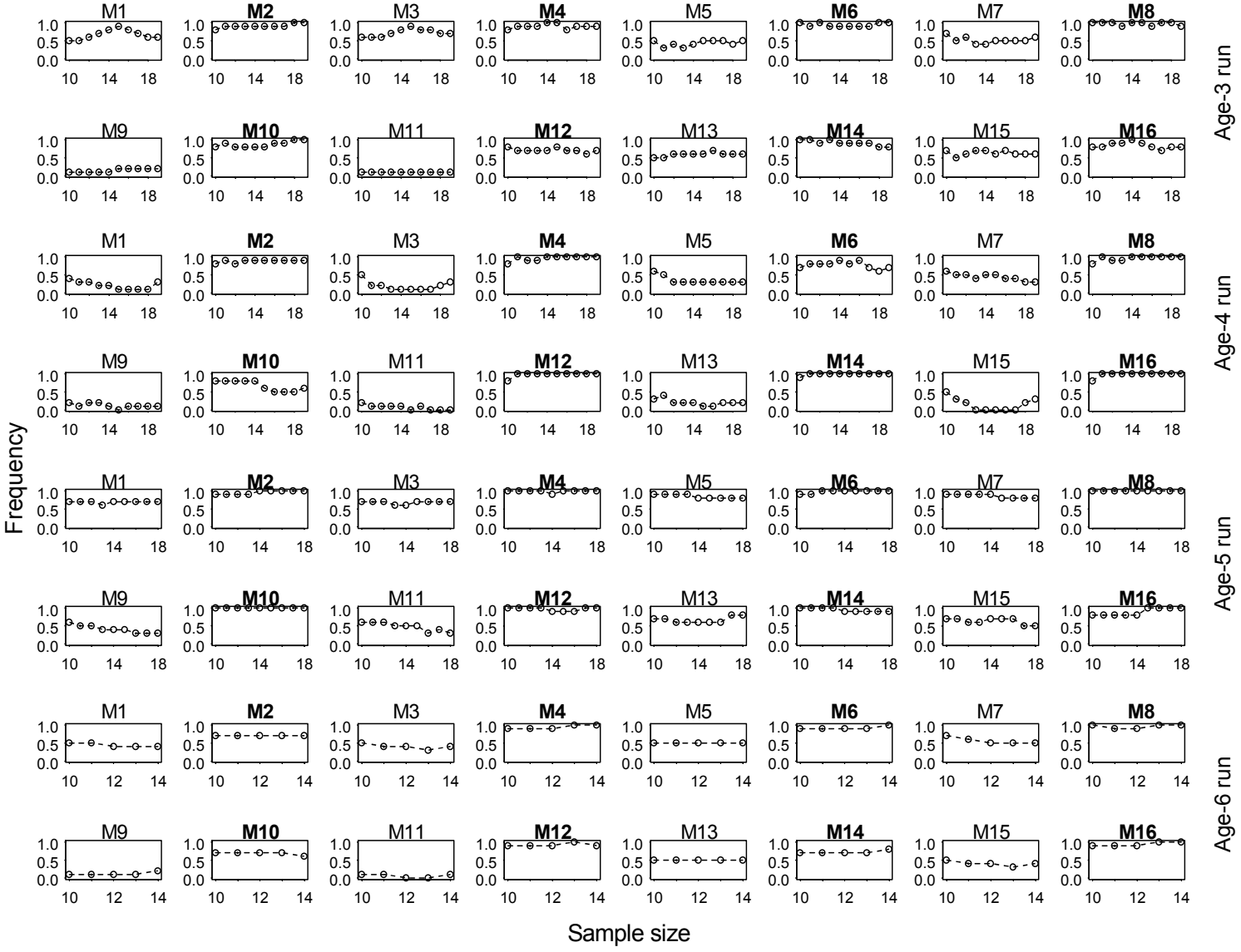


Figure A8. Forecast model performance for age-specific run size of Lower River Wild (LRW) population in terms of 90% prediction interval where the prediction coverage from the respective model is arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's 10-year hind-casting forecasts (1995-2004) successfully cover actual runs.

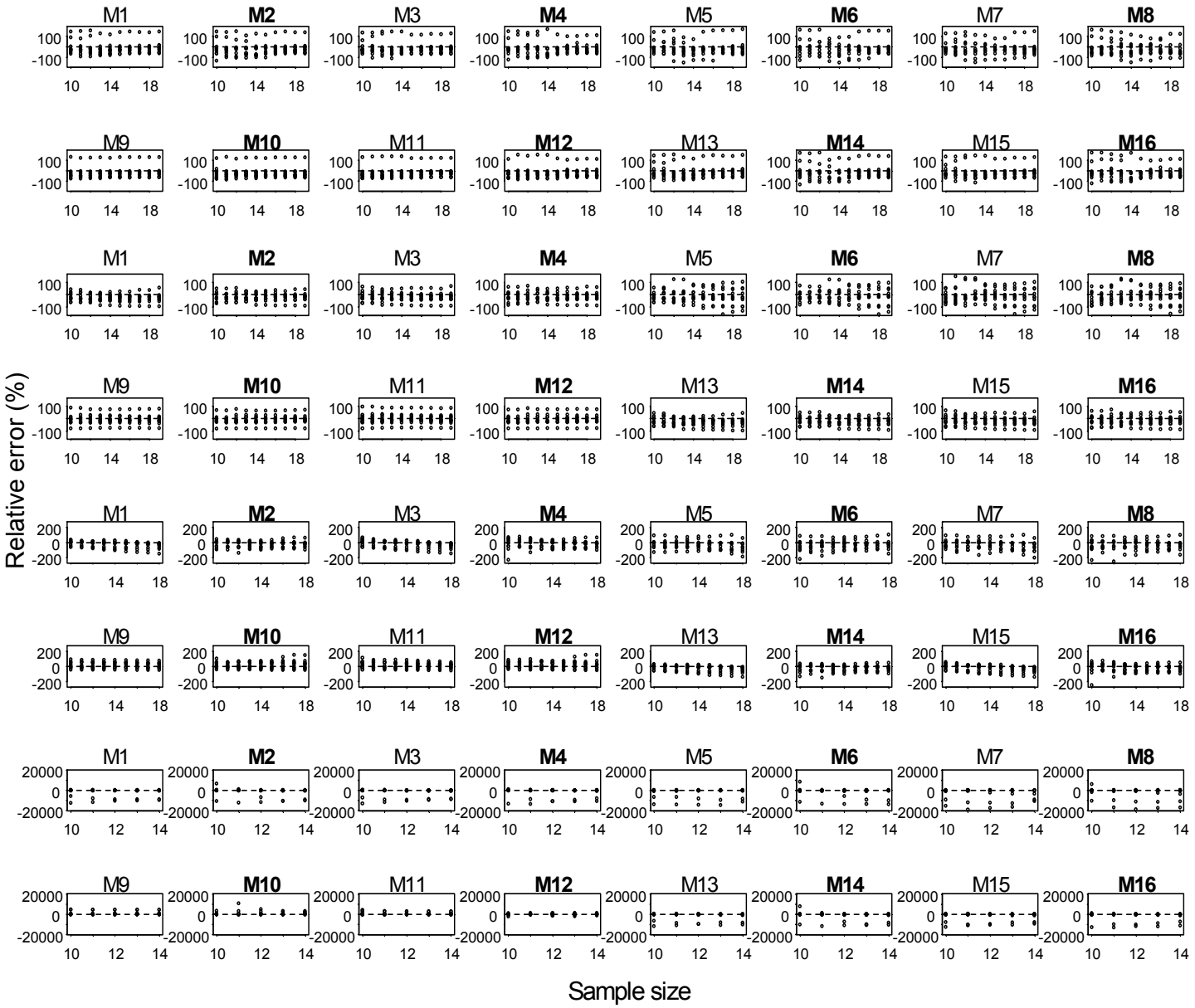


Figure A9. Forecast model performance for age-specific run size of Lower River Hatchery (LRH) population in terms of forecast relative error (%) where the errors from the respective model are arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in each panel are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model.

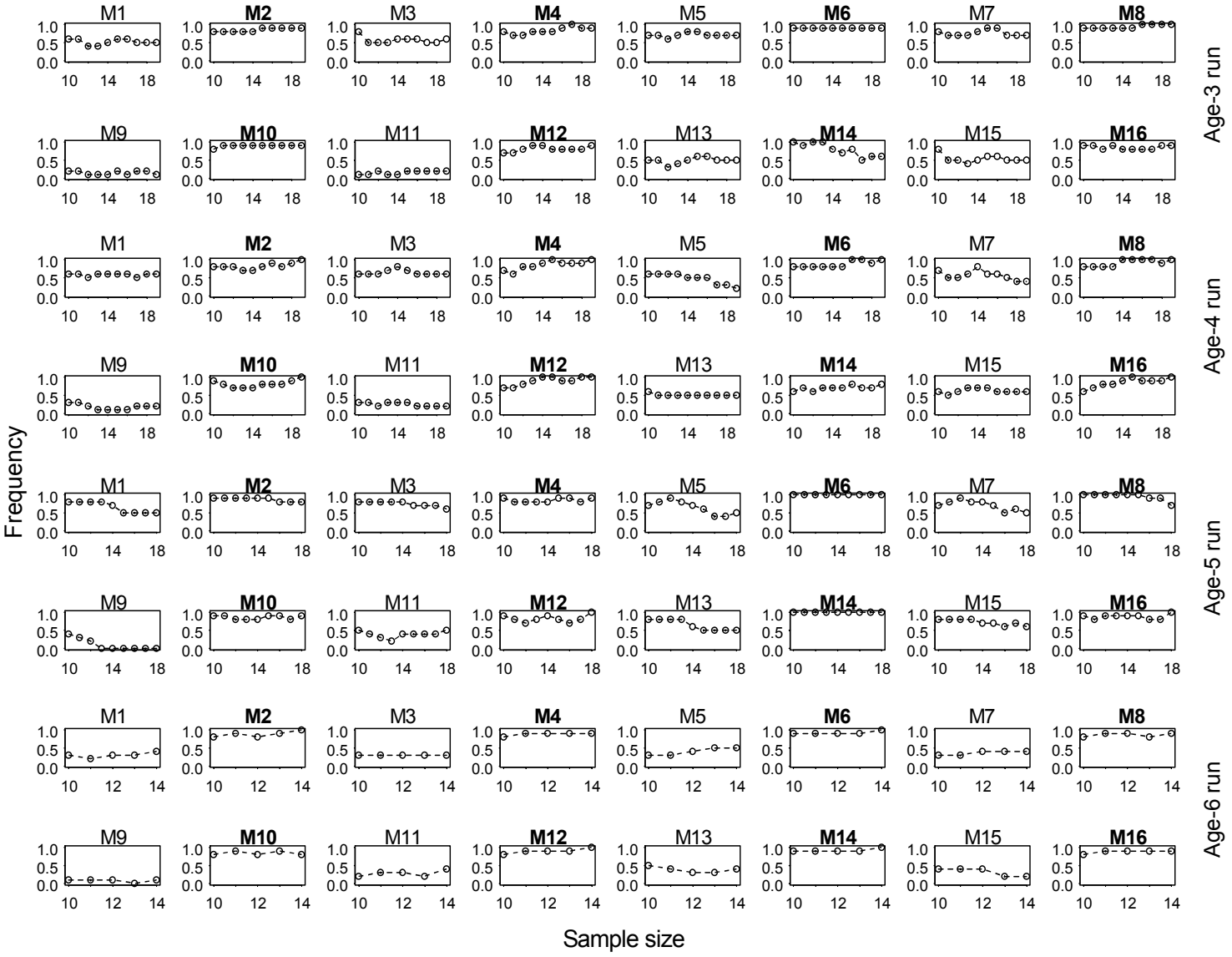


Figure A10. Forecast model performance for age-specific run size of Lower River Hatchery (LRH) population in terms of 90% prediction interval where the prediction coverage from the respective model is arranged by age-3 (first two rows), -4 (3rd and 4th rows), -5 (5th and 6th rows) and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's 10-year hind-casting forecasts (1995-2004) successfully cover actual runs.

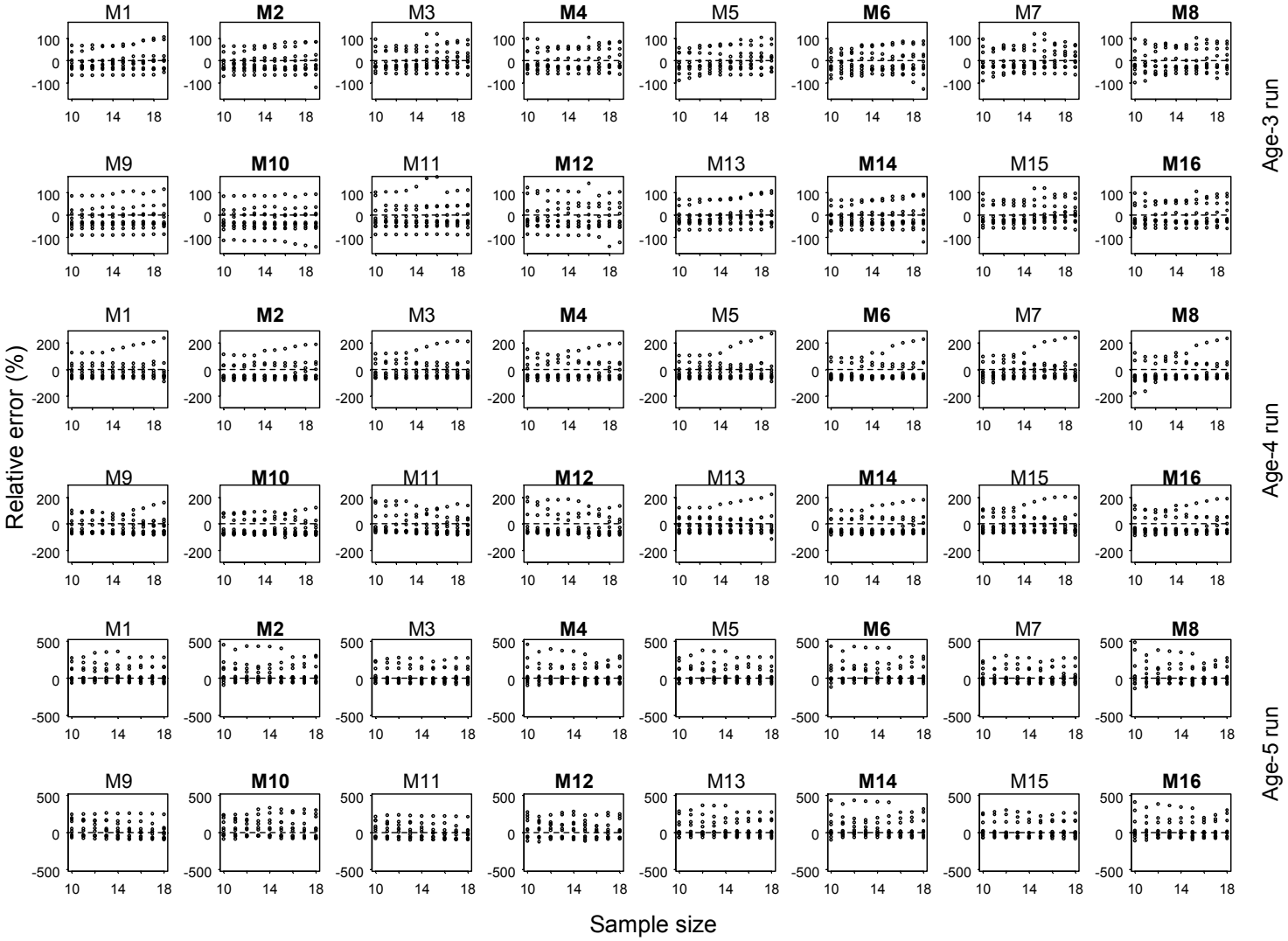


Figure A11. Forecast model performance for age-specific run size of Bonneville Pool Hatchery (BPH) population in terms of forecast relative error (%) where the errors from the respective model are arranged by age-3 (first two rows), -4 (3rd and 4th rows), and -5 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in each panel are relative errors (%) in forecasts of 10 years (1995-2004)' runs made with the corresponding forecast model.

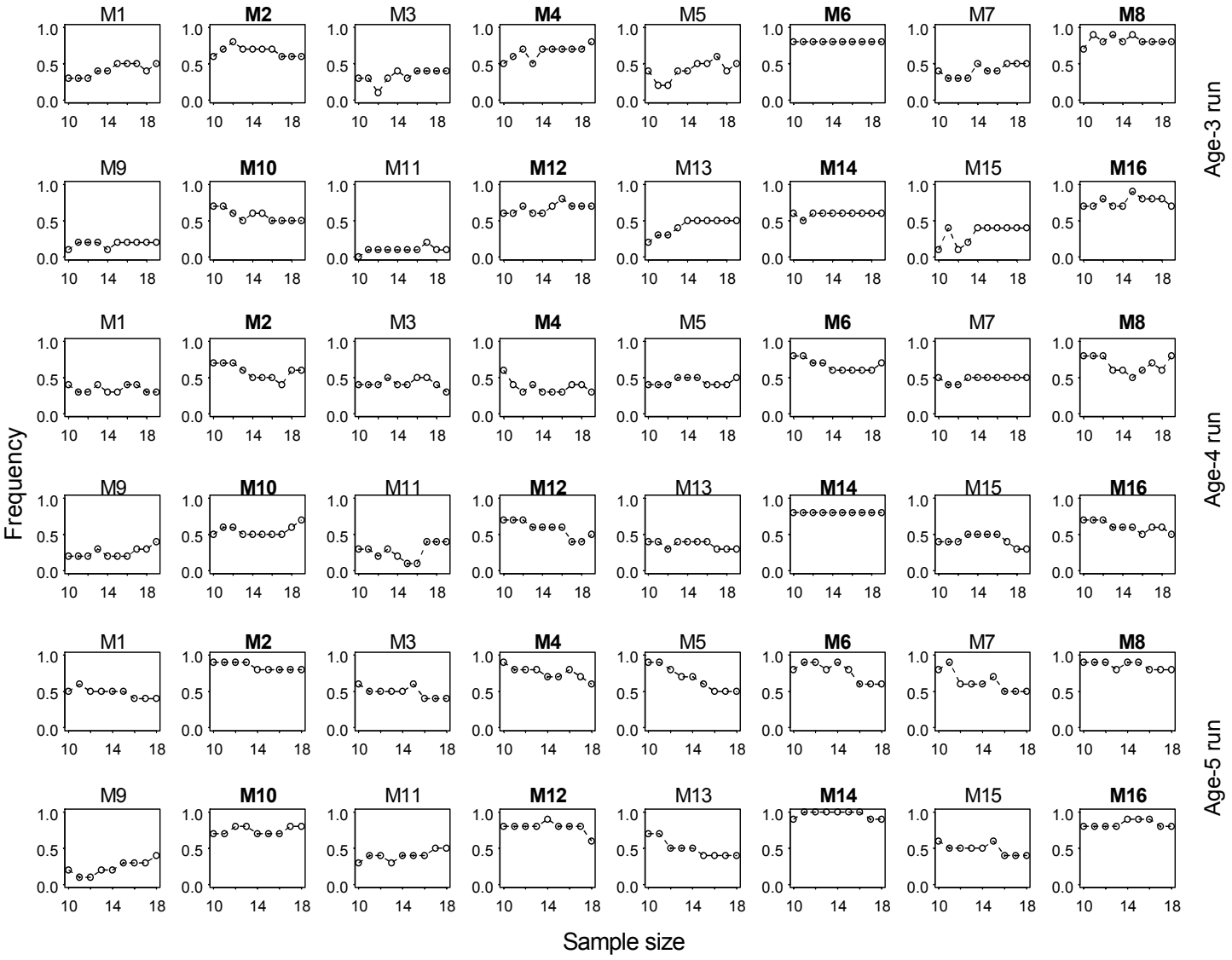


Figure A12. Forecast model performance for age-specific run size of Bonneville Pool Hatchery (BPH) population in terms of 90% prediction interval where the prediction coverage from the respective model is arranged by age-3 (first two rows), -4 (3rd and 4th rows), and -5 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's 10-year hind-casting forecasts (1995-2004) successfully cover actual runs.

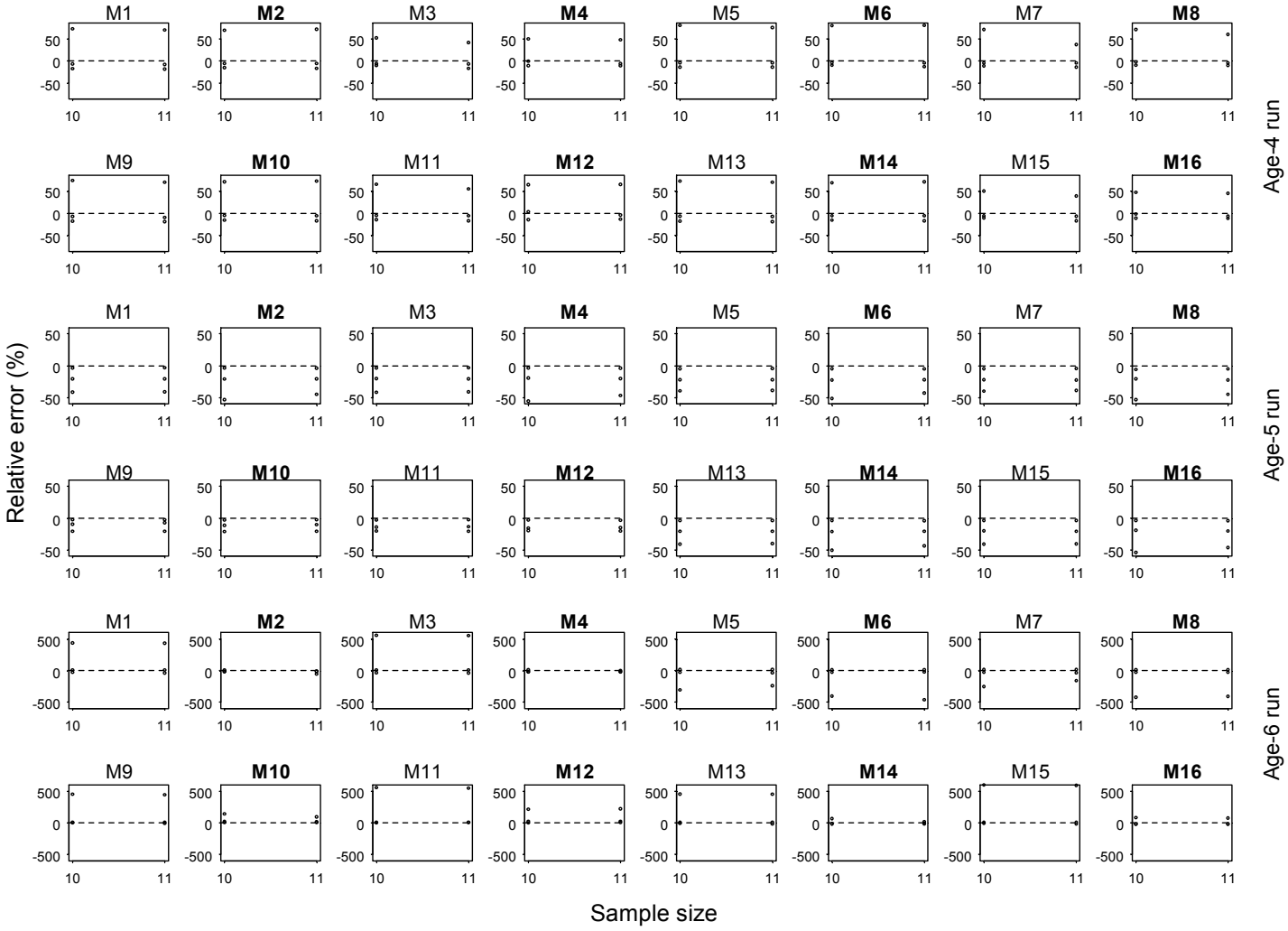


Figure A13. Forecast model performance for age-specific run size of Upper Columbia River summer (UCS) population in terms of forecast relative error (%) where the errors from the respective model are arranged by age-4 (first two rows), -5 (3rd and 4th rows), and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Vertical dots over each sample size in each panel are relative errors (%) in forecasts of three years (2002-2004)' runs made with the corresponding forecast model.

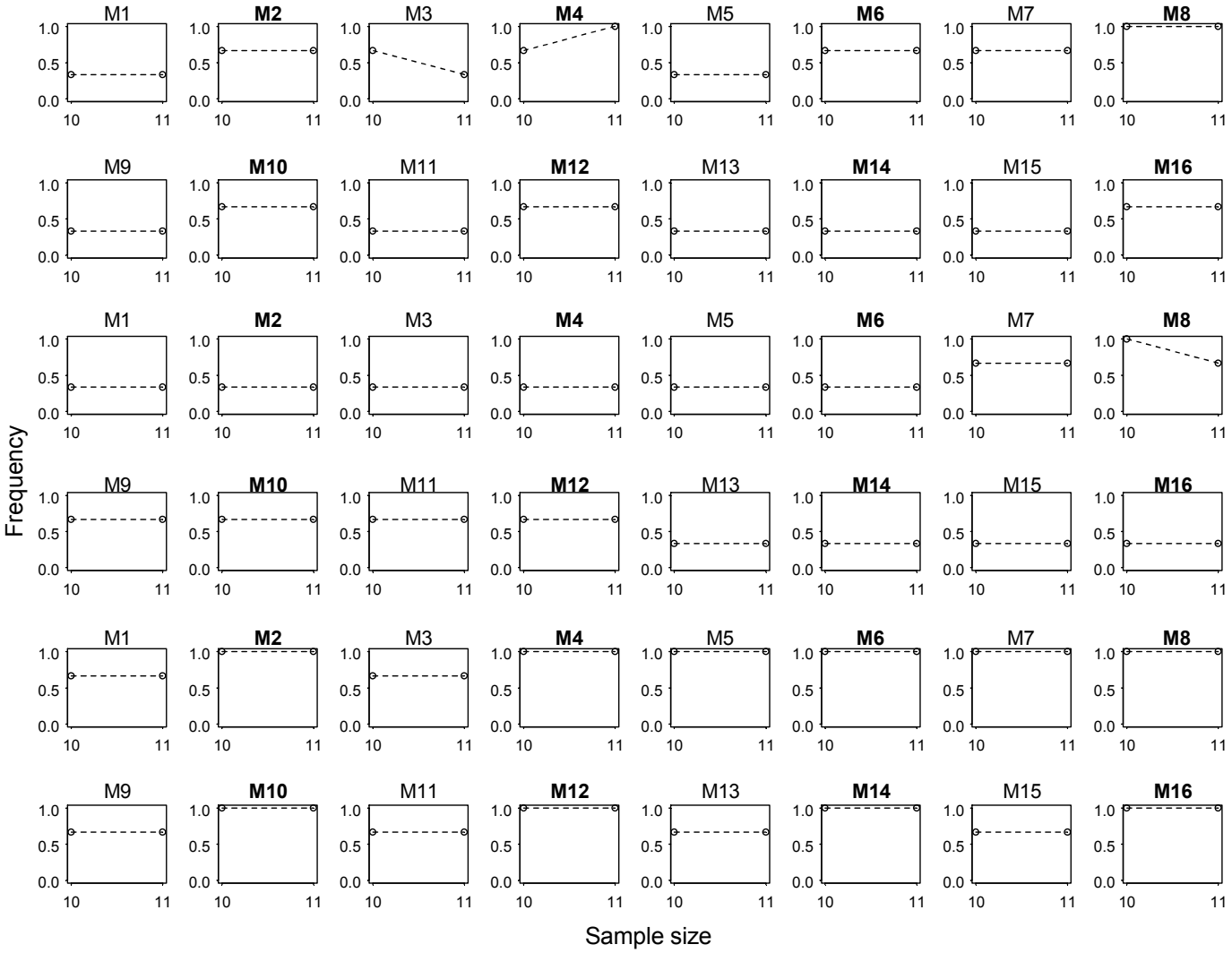


Figure A14. Forecast model performance for age-specific run size of Upper Columbia River summer (UCS) population in terms of 90% prediction interval where the prediction coverage from the respective model is arranged by age-4 (first two rows), -5 (3rd and 4th rows), and -6 (last two rows) runs. Notes of M1-M16 above each panel indicate forecast model used in the panel, where non-bold font models with odd numbers (e.g., M1, M3, etc.) are non-autoregressive models, and bold font models with even numbers (e.g., M2, M4, etc.) are their autoregressive versions (Table 3). The x-axis of each panel is sample size, which is the number of years of data used in hind-casting forecasts. Dot over each sample size in each panel is frequency that the prediction intervals from the corresponding model's three-year hind-casting forecasts (2002-2004) successfully cover actual runs.