UPPSALA
UNIVERSITET

Working Paper 2016:1

Department of Statistics

On the equivalence of confidence interval estimation based on frequentist model averaging and least-squares of the full model in linear regression

Sebastian Ankargren and Shaobo Jin

Working Paper 2016:1
April 2016
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Title: On the equivalence of confidence interval estimation based on frequentist model averaging and least-squares for the full model in linear regression

Author: Sebastian Ankargren & Shaobo Jin
E-mail: Sebastian.ankargren@statistics.uu.se, Shaobo.jin@statistik.uu.se

On the equivalence of confidence interval estimation based on frequentist model averaging and least squares of the full model in linear regression

Sebastian Ankargren* and Shaobo Jin†

Department of Statistics, Uppsala University

April 5, 2016

Abstract

In many applications of linear regression models, model selection is vital. However, randomness due to model selection is commonly ignored in post-model selection inference. In order to account for the model selection uncertainty in these linear models, least squares frequentist model averaging has been proposed recently. In this paper, we show that the confidence interval from model averaging is asymptotically equivalent to the confidence interval from the full model. Furthermore, we demonstrate that this equivalence also holds in finite samples if the parameter of interest is a linear function of the regression coefficients.

Keywords: Asymptotic equivalence · Linear model · Local asymptotics · Model selection uncertainty · Post-selection inference

Mathematics Subject Classification (2000): 62E20 · 62J05 · 62J99 · 62F12

1 Introduction

In many situations in applied statistics, there is no model structure known to the researcher, but a set of plausible alternatives. Commonly, to pick one model from this set of candidates, information criteria such as AIC (Akaike, 1974) and BIC (Schwarz, 1978) are used. This can have detrimental effects on subsequent inference since the model selection step itself is stochastic and thus subject to uncertainty, an uncertainty which inference post-model selection usually fails to account for. This issue is studied

*sebastian.ankargren@statistics.uu.se

†shaobo.jin@statistics.uu.se

in detail by e.g. Pötscher (1991); Kabaila (1995); Kabaila and Leeb (2006); Leeb and Pötscher (2005); also see Berk et al (2013) for a promising way of doing inference post-model selection.

It is often by this inability of model selection to incorporate the randomness attached to it in post-selection inference that frequentist model averaging is motivated, advocating that it may serve as a better, more truthful alternative. For example, it is claimed that model averaging "compromises across a set of competing models, and in doing so, incorporates model uncertainty into the conclusions about the unknown parameters" (Wan et al, 2010, p. 277). Hjort and Claeskens (2003) developed a frequentist model averaging machinery under the likelihood framework. They proposed a confidence interval that captures the randomness in model selection. However, Wang and Zhou (2013) showed that the proposed interval is asymptotically equivalent to the confidence interval obtained from the full model and that the finite sample confidence intervals are also equivalent if the parameter of interest is a linear combination of regression coefficients of a varying-coefficient partially linear model, which is regarded as the semi-parametric framework by the authors.

The focus of this paper is the linear model using least squares. Least squares-based model averaging has previously been studied by for example Hansen (2007), Wan et al (2010), Hansen and Racine (2012), Liu and Okui (2013), Ando and Li (2014) and Cheng et al (2015), with a focus on prediction. Liu (2015) developed asymptotic distribution theory for frequentist model averaging estimators under the least squares framework, including results for Mallows and Jackknife model averaging (Hansen, 2007; Hansen and Racine, 2012), and proposed a confidence interval in the spirit of Hjort and Claeskens (2003). In this article, we show that the inference we can make about the unknown parameters, using the distributional theory of Liu (2015), is in fact equivalent to what we can make just using the model including all possible covariates, either asymptotically or both in finite samples and asymptotically depending on the nature of the parameter of interest. From this perspective, there is no additional gain in turning to model averaging. Thus, this paper extends the work by Kabaila and Leeb (2006) and Wang and Zhou (2013) in that their criticism geared towards interval estimation based on Hjort and Claeskens (2003) is also valid for interval estimation in linear models based on Liu (2015).

The remainder of the paper is organized as follows. Section 2 briefly reviews the key results in Liu (2015), Section 3 discusses the equivalence of the confidence interval and Section 4 concludes. Proofs are placed in Appendix A.

2 Frequentist model averaging

In this section we briefly review the necessary aspects of Liu (2015). Assume the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{e} \tag{1}$$

in which $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is the $n \times 1$ response vector, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ the $n \times p$ matrix of core regressors, $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$ the $n \times q$ matrix of auxiliary regressors, $\boldsymbol{e} = (e_1, \ldots, e_n)'$ the $n \times 1$ vector of errors and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the $p \times 1$ and $q \times 1$ parameter vectors. By the formulation of the model, the core regressors in $\boldsymbol{X}$ are necessarily included in all models, but those in $\boldsymbol{Z}$ are potential and subject to scrutiny. This does not restrict applications, however, as $\boldsymbol{X}$ may be an empty matrix or simply a vector of ones such that the intercept is always included. The error term is assumed to have a zero conditional mean, $E(e_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = 0$, but allowed to be both heteroskedastic and autocorrelated.

Collect the regressors and the parameters in $\boldsymbol{H} = (\boldsymbol{X}, \boldsymbol{Z})$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$, respectively. The regression model (1) may then be written as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{e}$$

Suppose now that there is a set of M candidate models. If this set consists of all nested models, then $M = q+1$, and if it consists of all submodels, then $M = 2^q$. Whatever the choice of model set, each model m is assumed to contain a unique combination of $0 \leq q_m \leq q$ regressors from $\boldsymbol{Z}$ such that the mth model's subset of auxiliary regressors can be written as $\boldsymbol{Z}_m = \boldsymbol{Z}\boldsymbol{\Pi}_m'$, where $\boldsymbol{Z}_m$ is $n \times q_m$ and $\boldsymbol{\Pi}_m'$ is a $q \times q_m$ selection matrix. Since all M models also include the full set of core regressors $\boldsymbol{X}$, model m includes in total $p + q_m$ regressors.

For ease of notation and with no loss of generality, let the full model be ordered last in the set of models such that model M is the full model including all regressors. The least squares estimator of $\boldsymbol{\theta}$ in this model is

$$\hat{\boldsymbol{\theta}}_M = (\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{y},$$

whereas for submodel m, the estimator of $\boldsymbol{\theta}_m = (\boldsymbol{\beta}', \boldsymbol{\gamma}_m')' = (\boldsymbol{\beta}', \boldsymbol{\gamma}'\boldsymbol{\Pi}_m')'$, a $(p+q_m) \times 1$ vector, is

$$\hat{\boldsymbol{\theta}}_m = (\boldsymbol{H}_m'\boldsymbol{H}_m)^{-1}\boldsymbol{H}_m'\boldsymbol{y}, \tag{2}$$

where $\boldsymbol{H}_m = (\boldsymbol{X}, \boldsymbol{Z}_m) = (\boldsymbol{X}, \boldsymbol{Z}\boldsymbol{\Pi}_m')$. For later use we also define $\boldsymbol{\theta}^{(m)} = (\boldsymbol{\beta}', \boldsymbol{\gamma}_m'\boldsymbol{\Pi}_m)'$, a $(p+q) \times 1$ vector in which there are zeros placed in the positions corresponding to excluded variables in model m. Similarly, $\hat{\boldsymbol{\theta}}^{(m)} = (\hat{\boldsymbol{\beta}}_m', \hat{\boldsymbol{\gamma}}^{(m)'})' = (\hat{\boldsymbol{\beta}}_m', \hat{\boldsymbol{\gamma}}_m'\boldsymbol{\Pi}_m)'$ where $\hat{\boldsymbol{\theta}}^{(m)}$ is $(p+q) \times 1$ and $\hat{\boldsymbol{\gamma}}_m$ is the q_m last elements of $\hat{\boldsymbol{\theta}}_m$ in (2).

For the asymptotic analysis local asymptotics are employed. This framework entails the following assumption.

Assumption 1. *$\boldsymbol{\gamma} = \boldsymbol{\gamma}_n = \boldsymbol{\delta}/\sqrt{n}$, where $\boldsymbol{\delta}$ is an unknown local parameter vector.*

Since the omitted variable bias in all models but the full is of order $O(1/\sqrt{n})$ under Assumption 1, $\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m)$ does not diverge. Instead, we have the following:

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}) &\xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{Q}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}^{-1}) \\ \sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) &\xrightarrow{d} N(\boldsymbol{A}_m\boldsymbol{\delta}, \boldsymbol{Q}_m^{-1}\boldsymbol{\Omega}_m\boldsymbol{Q}_m^{-1}) \end{aligned} \tag{3}$$

where $\boldsymbol{A}_m = \boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{Q}\boldsymbol{S}_0(\boldsymbol{I}_q - \boldsymbol{\Pi}_m\boldsymbol{\Pi}_m')$ and

$$\begin{aligned}
\boldsymbol{S}_0 &= \begin{pmatrix} \boldsymbol{0}_{p\times q} \\ \boldsymbol{I}_q \end{pmatrix}, \quad \boldsymbol{S}_m = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0}_{p\times q_m} \\ \boldsymbol{0}_{q\times p} & \boldsymbol{\Pi}_m' \end{pmatrix} \\
\boldsymbol{Q} &= \begin{pmatrix} \boldsymbol{Q}_{xx} & \boldsymbol{Q}_{xz} \\ \boldsymbol{Q}_{zx} & \boldsymbol{Q}_{zz} \end{pmatrix} = \begin{pmatrix} E(\boldsymbol{x}_i\boldsymbol{x}_i') & E(\boldsymbol{x}_i\boldsymbol{z}_i') \\ E(\boldsymbol{z}_i\boldsymbol{x}_i') & E(\boldsymbol{z}_i\boldsymbol{z}_i') \end{pmatrix} \\
\boldsymbol{\Omega} &= \begin{pmatrix} \boldsymbol{\Omega}_{xx} & \boldsymbol{\Omega}_{xz} \\ \boldsymbol{\Omega}_{zx} & \boldsymbol{\Omega}_{zz} \end{pmatrix} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \begin{pmatrix} E(\boldsymbol{x}_i\boldsymbol{x}_j' e_i e_j) & E(\boldsymbol{x}_i\boldsymbol{z}_j' e_i e_j) \\ E(\boldsymbol{z}_i\boldsymbol{x}_j' e_i e_j) & E(\boldsymbol{z}_i\boldsymbol{z}_j' e_i e_j) \end{pmatrix} \\
\boldsymbol{Q}_m &= \boldsymbol{S}_m'\boldsymbol{Q}\boldsymbol{S}_m, \quad \boldsymbol{\Omega}_m = \boldsymbol{S}_m'\boldsymbol{\Omega}\boldsymbol{S}_m.
\end{aligned}$$

The limiting distributions in (3) are results of the following assumption.

Assumption 2. *As $n\to\infty$, $n^{-1}\boldsymbol{H}'\boldsymbol{H} \xrightarrow{p} \boldsymbol{Q}$ and $n^{-1/2}\boldsymbol{H}'\boldsymbol{e} \xrightarrow{d} \boldsymbol{R} \sim N(\boldsymbol{0},\boldsymbol{\Omega})$.*

Suppose the ultimate goal is to make inference regarding the true value of a known, smooth function $\mu : \mathbb{R}^{p+q} \mapsto \mathbb{R}$ evaluated at the (unknown) parameters; that is, we are interested in the true value of $\mu(\boldsymbol{\theta})$. This is called the focus parameter. The model averaging estimator is a weighted average of the estimator for the focus parameter $\mu(\boldsymbol{\theta})$ in the individual submodels. The averaging estimator is

$$\bar{\mu} = \sum_{m=1}^{M} w(m|\hat{\boldsymbol{\delta}})\hat{\mu}_m,$$

where $w(m|\hat{\boldsymbol{\delta}})$ denotes the data-dependent weight for model m and $\hat{\mu}_m$ should be understood as

$$\hat{\mu}_m := \mu(\hat{\boldsymbol{\theta}}^{(m)}) = \mu(\hat{\boldsymbol{\beta}}_m, \hat{\gamma}^{(m)}),$$

i.e. the function μ at the estimated values of $\boldsymbol{\beta}$ and γ_m, with the $q-q_m$ elements in γ not included in γ_m set to 0.

The local parameter, $\boldsymbol{\delta}$, can be unbiasedly, but not consistently, estimated asymptotically in the sense that

$$\hat{\boldsymbol{\delta}} = \sqrt{n}\hat{\gamma}_M \xrightarrow{d} \boldsymbol{R}_{\boldsymbol{\delta}} = \boldsymbol{\delta} + \boldsymbol{S}_0'\boldsymbol{Q}^{-1}\boldsymbol{R} \sim N(\boldsymbol{\delta}, \boldsymbol{S}_0'\boldsymbol{Q}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}^{-1}\boldsymbol{S}_0).$$

Let $\boldsymbol{D}_\theta$ be the derivative of μ with respect to the full parameter vector $\boldsymbol{\theta}$ evaluated at the null points $(\boldsymbol{\beta}', \boldsymbol{0})'$. The key theorem in Liu (2015) underlying the results for a frequentist model averaging confidence interval for μ is stated next.

Theorem 1 (Liu, 2015)**.** *If $w(m|\hat{\boldsymbol{\delta}}) \xrightarrow{d} w(m|\boldsymbol{R}_{\boldsymbol{\delta}})$ as $n\to\infty$ and Assumptions 1-2 hold, then*

$$\sqrt{n}(\bar{\mu}-\mu) \xrightarrow{d} \boldsymbol{D}_{\boldsymbol{\theta}}'\boldsymbol{Q}^{-1}\boldsymbol{R} + \boldsymbol{D}_{\boldsymbol{\theta}}'\left(\sum_{m=1}^{M} w(m|\boldsymbol{R}_{\boldsymbol{\delta}})\boldsymbol{C}_m\right)\boldsymbol{R}_{\boldsymbol{\delta}}$$

where $\boldsymbol{C}_m = (\boldsymbol{P}_m\boldsymbol{Q} - \boldsymbol{I}_{p+q})\boldsymbol{S}_0$ and $\boldsymbol{P}_m = \boldsymbol{S}_m(\boldsymbol{S}_m'\boldsymbol{Q}\boldsymbol{S}_m)^{-1}\boldsymbol{S}_m'$.

Proof. See Liu (2015). □

The proposed confidence interval rests on the idea of replacing limit quantities by finite sample counterparts. Thus, for sufficiently large n, it should be that

$$\frac{\sqrt{n}(\bar{\mu}-\mu)-\hat{\boldsymbol{D}}_{\boldsymbol{\theta}}'\left(\sum_{m=1}^{M} w(m|\hat{\boldsymbol{\delta}})\hat{\boldsymbol{C}}_m\right)\hat{\boldsymbol{\delta}}}{\hat{\kappa}} \overset{\text{app.}}{\sim} N(0,1)$$

where $\hat{\kappa}$ is a consistent estimator of

$$\kappa=\sqrt{\boldsymbol{D}_{\boldsymbol{\theta}}'\boldsymbol{Q}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}^{-1}\boldsymbol{D}_{\boldsymbol{\theta}}}.$$

Therefore, a confidence interval with asymptotic coverage of $100(1-\alpha)\%$ is

$$\left(\bar{\mu}-b\left(\hat{\boldsymbol{\delta}}\right)-z_{1-\alpha/2}\frac{\hat{\kappa}}{\sqrt{n}},\bar{\mu}-b\left(\hat{\boldsymbol{\delta}}\right)+z_{1-\alpha/2}\frac{\hat{\kappa}}{\sqrt{n}}\right), \tag{4}$$

where $b\left(\hat{\boldsymbol{\delta}}\right)=\hat{\boldsymbol{D}}_{\boldsymbol{\theta}}'\sum w\left(m|\hat{\boldsymbol{\delta}}\right)\hat{\boldsymbol{C}}_m\hat{\gamma}_M$ and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

3 Equivalence of confidence intervals

In this section we establish the equivalence of frequentist model averaging and full model confidence intervals.

3.1 Asymptotic equivalence

Consider the case where there is a fixed model and that this model is the full model. By (3) and the delta theorem, the asymptotic distribution of the estimator of the focus parameter is

$$\sqrt{n}\left(\mu\left(\hat{\boldsymbol{\theta}}_M\right)-\mu\left(\boldsymbol{\theta}\right)\right)\xrightarrow{d}N\left(0,\boldsymbol{D}_{\boldsymbol{\theta}}'\boldsymbol{Q}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}^{-1}\boldsymbol{D}_{\boldsymbol{\theta}}\right).$$

The confidence interval for μ in the full model based upon a finite-sample approximation to this asymptotic distribution is

$$\left(\mu\left(\hat{\boldsymbol{\theta}}_M\right)-z_{1-\alpha/2}\frac{\hat{\kappa}}{\sqrt{n}},\mu\left(\hat{\boldsymbol{\theta}}_M\right)+z_{1-\alpha/2}\frac{\hat{\kappa}}{\sqrt{n}}\right), \tag{5}$$

where z is the $1-\alpha/2$ quantile of a standard normal distribution. Note that the interval in (5) has the same length as the interval in (4). It therefore follows that if the centers of the two intervals are asymptotically the same, then they are asymptotically equivalent. That is, equivalence holds if $\bar{\mu}=\hat{\mu}_M+b\left(\hat{\boldsymbol{\delta}}\right)+o_p\left(n^{-1/2}\right)$. The existence of this particular relation is our main result, and it is established in the following theorem. The proof is placed in Appendix A.

Theorem 2. *The frequentist model averaging and full model confidence intervals in* (4) *and* (5)*, respectively, are asymptotically equivalent.*

Theorem 2 suggests that for large sample sizes, there are no gains in terms of interval estimation in using model averaging. A confidence interval based on least squares estimation of the full model instead offers the more compelling alternative, given its simplicity.

3.2 Small sample equivalence

Theorem 2 shows that the confidence interval (5) is asymptotically equivalent to (4) for a general smooth and real-valued function $\mu(\boldsymbol{\theta})$. However, nothing is said about the finite sample properties, in which there may be substantial differences. In this section, we establish also small sample equivalence when $\mu(\boldsymbol{\theta}) = \boldsymbol{c}'\boldsymbol{\theta}$, where $\boldsymbol{c}$ is a vector of known scalars. That is, the focus parameter is some fixed linear function of the regression coefficients.

If the focus parameter is $\mu(\boldsymbol{\theta}) = \boldsymbol{c}'\boldsymbol{\theta}$, the center of the full model confidence interval (5) is $\boldsymbol{c}'\hat{\boldsymbol{\theta}}_M$ and the center of the model averaging confidence interval (4) is

$$\begin{aligned} &\boldsymbol{c}'\left[\sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\hat{\boldsymbol{\theta}}^{(m)} - \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\hat{\boldsymbol{C}}_m\hat{\gamma}_M\right] \\ &= \boldsymbol{c}'\left[\hat{\boldsymbol{\theta}}_M + \sum_{m=1}^{M-1} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\hat{\boldsymbol{\theta}}^{(m)} - \hat{\boldsymbol{\theta}}_M - \hat{\boldsymbol{C}}_m\hat{\gamma}_M\right)\right], \end{aligned} \tag{6}$$

since $\hat{\boldsymbol{D}}'_{\boldsymbol{\theta}} = \boldsymbol{c}'$. The equality holds because $\hat{\boldsymbol{C}}_M = \boldsymbol{0}$ and $\hat{\boldsymbol{\theta}}^{(M)} = \hat{\boldsymbol{\theta}}_M$. Based on the center (6), we can establish the following theorem.

Theorem 3. *If $\mu(\boldsymbol{\theta}) = \boldsymbol{c}'\boldsymbol{\theta}$, the confidence interval* (5) *is equivalent to the confidence interval* (4) *in finite samples, where* $\mathbf{c}$ *is a* $(p+q)\times 1$ *constant vector.*

Theorem 3 says that the full model produces the same confidence interval as model averaging under some restricted cases. However, Theorem 3 says nothing about the small sample equivalence for other cases, e.g., β_1/β_2 or $\exp\beta_1$. We conjecture that the small sample equivalence is likely to fail for focus parameters not of the form $\mu(\boldsymbol{\theta}) = \boldsymbol{c}'\boldsymbol{\theta}$. This is illustrated in the following example.

Example 1. Consider a model with two regressors

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{\epsilon},$$

where $\gamma = \delta/\sqrt{n}$. From the above discussion, the confidence interval of β from the full model is the same as that from model averaging, both asymptotically and in a finite sample. In particular, from Theorem 2 we know that, for any differentiable function h, the confidence interval of $h(\beta)$ from the full model is asymptotically the same as that from model averaging. However, the above discussion says nothing about the finite sample equivalence. In this example, consider $h(\beta) = \beta^2$. The center from the model averaging confidence interval is

$$w\left(1|\hat{\delta}\right)\hat{\beta}_1^2 + \left[1 - w\left(1|\hat{\delta}\right)\right]\hat{\beta}_M^2 - 2w\left(1|\hat{\delta}\right)\hat{\beta}_M\hat{Q}_{xx}^{-1}\hat{Q}_{xz}\hat{\gamma}_M.$$

Observe that $\hat{\beta}_1 = \hat{Q}_{xx}^{-1}\boldsymbol{X}'\boldsymbol{y}/n$,

$$\hat{\beta}_M = \frac{\hat{Q}_{zz}\boldsymbol{X}'\boldsymbol{y}/n - \hat{Q}_{xz}\boldsymbol{Z}'\boldsymbol{y}/n}{\hat{Q}_{xx}\hat{Q}_{zz} - \hat{Q}_{xz}^2} \text{ and } \hat{\gamma}_M = \frac{-\hat{Q}_{xz}\boldsymbol{X}'\boldsymbol{y}/n + \hat{Q}_{xx}\boldsymbol{Z}'\boldsymbol{y}/n}{\hat{Q}_{xx}\hat{Q}_{zz} - \hat{Q}_{xz}^2}.$$

Then, the difference in two centers is $w\left(1|\hat{\boldsymbol{\delta}}\right)\left[\hat{\beta}_1^2 - \hat{\beta}_M^2 - 2\hat{\beta}_M\hat{Q}_{xx}^{-1}\hat{Q}_{xz}\hat{\gamma}_M\right]$, which is zero if and only if $w\left(1|\hat{\boldsymbol{\delta}}\right) = 0$ or $\hat{\beta}_1^2 - \hat{\beta}_M^2 - 2\hat{\beta}_M\hat{Q}_{xx}^{-1}\hat{Q}_{xz}\hat{\gamma}_M = 0$. The former means the full model receives the total weight. Plugging in the expressions of $\hat{\beta}_1$, $\hat{\beta}_M$, and $\hat{\gamma}_M$, the latter satisfies

$$\hat{\beta}_1^2 - \hat{\beta}_M^2 - 2\hat{\beta}_M\hat{Q}_{xx}^{-1}\hat{Q}_{xz}\hat{\gamma}_M = \frac{1}{n^2} \times \frac{\hat{Q}_{xz}^2\left(\hat{Q}_{xz}\boldsymbol{X}'\boldsymbol{y} - \hat{Q}_{xx}\boldsymbol{Z}'\boldsymbol{y}\right)^2}{\hat{Q}_{xx}^2\left(\hat{Q}_{xx}\hat{Q}_{zz} - \hat{Q}_{xz}^2\right)^2},$$

which is zero only if $\hat{Q}_{xz}\boldsymbol{X}'\boldsymbol{y} = \hat{Q}_{xx}\boldsymbol{Z}'\boldsymbol{y}$ or $\hat{Q}_{xz} = 0$. If X and Z are correlated, the two intervals are likely to be different.

4 Conclusion

In this work, we have studied the frequentist model averaging confidence interval for linear models proposed by Liu (2015). Using similar techniques as Wang and Zhou (2013), our key result is that the model averaging confidence interval is asymptotically equivalent to the interval obtained by use of least squares estimation in the full model. Furthermore, we also provide the stronger result that the intervals are exactly equivalent in finite samples if μ, the function defining the focus parameter, is linear.

While the results are disadvantageous to model averaging, they should be interpreted with care. Numerous studies have illustrated impressive performance of model averaging in terms of prediction. This strand of the literature is separate from what we consider and thus remains unaffected by the conclusion herein. In fact, the good predictive performance may instead suggest that with other methods, we may be able to increase the efficiency of our confidence intervals by resorting to model averaging. This, however, we leave for further research.

A Proofs

Proof of Theorem 2. A proof technique similar to that used by Wang and Zhou (2013) is also employed here, but applied to the different problem setting that we have.

As already noted in the text, the intervals have the same length and thus what needs to be proved is that the centers are asymptotically equal. To this end, let $\boldsymbol{R} \sim N(\boldsymbol{0}, \boldsymbol{\Omega})$. Lemma 1 in Liu (2015) indicates that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\right) = \boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{Q}\boldsymbol{S}_0\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + \boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{R} + o_p(1), \quad (7)$$

where $\boldsymbol{\theta}_m = (\boldsymbol{\beta}', \boldsymbol{\delta}_m'/\sqrt{n})'$. For simplicity and without loss of generality, assume that the set of models is ordered in such a way that M denotes the full model. If $m = M$, then $\boldsymbol{\Pi}_M = \boldsymbol{I}$, $\boldsymbol{S}_M = \boldsymbol{I}$, and $\boldsymbol{Q}_M^{-1} = \boldsymbol{Q}^{-1}$, so then what remains is

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}\right) = \boldsymbol{Q}^{-1}\boldsymbol{R} + o_p(1),$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}'/\sqrt{n})'$. Thus, $\boldsymbol{R} = \sqrt{n}\boldsymbol{Q}\left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}\right) + o_p(1)$ and (7) becomes

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\right) = \boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{Q}\boldsymbol{S}_0\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + \boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{Q}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}\right) + o_p(1) \quad (8)$$

By rules for inverses of block matrices (e.g., Theorem 8.5.11 in Harville, 1997),

$$\boldsymbol{Q}_m^{-1} = \begin{pmatrix} \boldsymbol{Q}_{xx}^{-1} + \boldsymbol{Q}_{xx}^{-1}\boldsymbol{Q}_{xz}\boldsymbol{\Pi}_m'\boldsymbol{K}_m\boldsymbol{\Pi}_m\boldsymbol{Q}_{zx}\boldsymbol{Q}_{xx}^{-1} & -\boldsymbol{Q}_{xx}^{-1}\boldsymbol{Q}_{xz}\boldsymbol{\Pi}_m'\boldsymbol{K}_m \\ -\boldsymbol{K}_m\boldsymbol{\Pi}_m\boldsymbol{Q}_{zx}\boldsymbol{Q}_{xx}^{-1} & \boldsymbol{K}_m \end{pmatrix}, \quad (9)$$

where

$$\boldsymbol{K}_m = \left[\boldsymbol{\Pi}_m\left(\boldsymbol{Q}_{zz} - \boldsymbol{Q}_{zx}\boldsymbol{Q}_{xx}^{-1}\boldsymbol{Q}_{xz}\right)\boldsymbol{\Pi}_m'\right]^{-1} = \left(\boldsymbol{\Pi}_m\boldsymbol{K}^{-1}\boldsymbol{\Pi}_m'\right)^{-1}. \quad (10)$$

with $\boldsymbol{K} = \left(\boldsymbol{Q}_{zz} - \boldsymbol{Q}_{zx}\boldsymbol{Q}_{xx}^{-1}\boldsymbol{Q}_{xz}\right)^{-1}$. Plugging (9) into $\boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{Q}$ yields

$$\boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{A}_m \\ \mathbf{0} & \boldsymbol{B}_m \end{pmatrix},$$

where $\boldsymbol{A}_m = \boldsymbol{Q}_{xx}^{-1}\boldsymbol{Q}_{xz}\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{K}_m\boldsymbol{\Pi}_m\boldsymbol{K}^{-1}\right)$ and $\boldsymbol{B}_m = \boldsymbol{K}_m\boldsymbol{\Pi}_m\boldsymbol{K}^{-1}$. Thus, (8) can be expressed as

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\right) = \begin{pmatrix} \boldsymbol{A}_m \\ \boldsymbol{B}_m \end{pmatrix}\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{A}_m \\ \mathbf{0} & \boldsymbol{B}_m \end{pmatrix}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}\right) + o_p(1).$$

Recall that $\hat{\boldsymbol{\theta}}_m = \left(\hat{\boldsymbol{\beta}}_m', \hat{\boldsymbol{\delta}}_m'/\sqrt{n}\right)'$ and $\boldsymbol{\theta}_m = (\boldsymbol{\beta}', \boldsymbol{\delta}_m'/\sqrt{n})'$, where $\boldsymbol{\delta}_m = \boldsymbol{\Pi}_m\boldsymbol{\delta}$. Then

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}\right) = \boldsymbol{A}_m\boldsymbol{\delta} + \sqrt{n}\left(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}\right) + \boldsymbol{A}_m\sqrt{n}\left(\hat{\gamma}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) + o_p(1), \quad (11)$$

$$\sqrt{n}\left(\hat{\gamma}_m - \frac{\boldsymbol{\delta}_m}{\sqrt{n}}\right) = \boldsymbol{B}_m\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + \boldsymbol{B}_m\sqrt{n}\left(\hat{\gamma}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) + o_p(1), \quad (12)$$

where we have used that $\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{B}_m\right)\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right) = \boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{B}_m$.

The Taylor expansion of $\sqrt{n}\left(\mu\left(\hat{\boldsymbol{\beta}}_m, \boldsymbol{\Pi}_m'\hat{\gamma}_m\right) - \mu\left(\boldsymbol{\beta}, \boldsymbol{\delta}/\sqrt{n}\right)\right)$ around $(\boldsymbol{\beta}, \boldsymbol{\delta}/\sqrt{n})$ is

$$\begin{aligned}
&\sqrt{n}\left[\mu\left(\hat{\boldsymbol{\beta}}_m, \boldsymbol{\Pi}_m'\hat{\gamma}_m\right) - \mu\left(\boldsymbol{\beta}, \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right] \\
&= \left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\sqrt{n}\left(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}\right) + \left(\frac{\partial\mu}{\partial\gamma}\right)'\sqrt{n}\left(\boldsymbol{\Pi}_m'\hat{\gamma}_m - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) + o_p(1),
\end{aligned} \quad (13)$$

where the partial derivatives are evaluated at $\left(\boldsymbol{\beta}, \mathbf{0}_{q\times 1}\right)$ because of the continuous mapping theorem. Likewise, for the full model,

$$\begin{aligned}
&\sqrt{n}\left[\mu\left(\hat{\boldsymbol{\beta}}_M, \hat{\gamma}_M\right) - \mu\left(\boldsymbol{\beta}, \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right] \\
&= \left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\sqrt{n}\left(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}\right) + \left(\frac{\partial\mu}{\partial\gamma}\right)'\sqrt{n}\left(\hat{\gamma}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) + o_p(1).
\end{aligned} \quad (14)$$

We then get

$$
\begin{aligned}
\sqrt{n}\bar{\mu} &= \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\sqrt{n}\mu\left(\hat{\boldsymbol{\beta}}_m, \boldsymbol{\Pi}_m'\hat{\gamma}_m\right) \\
&= \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left[\sqrt{n}\mu\left(\boldsymbol{\beta}, \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) + \left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\sqrt{n}\left(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}\right)\right. \\
&\left. + \left(\frac{\partial\mu}{\partial\gamma}\right)'\sqrt{n}\left(\boldsymbol{\Pi}_m'\hat{\gamma}_m - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right] + o_p(1) \quad \text{(by (13))} \\
&= \sqrt{n}\mu\left(\boldsymbol{\beta}, \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) + \left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\sqrt{n}\left(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}\right) \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\boldsymbol{A}_m\left[\boldsymbol{\delta} + \sqrt{n}\left(\hat{\gamma}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right] \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left[\left(\frac{\partial\mu}{\partial\gamma}\right)'\sqrt{n}\left(\boldsymbol{\Pi}_m'\hat{\gamma}_m - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right] + o_p(1) \ \text{(by (11))}.
\end{aligned}
$$

Rewriting $\left(\boldsymbol{\Pi}_m'\hat{\gamma}_m - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) = \boldsymbol{\Pi}_m'\left(\hat{\gamma}_m - \frac{\boldsymbol{\delta}_m}{\sqrt{n}}\right) - \left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\frac{\boldsymbol{\delta}}{\sqrt{n}}$ and using (12)

$$
\begin{aligned}
\sqrt{n}\bar{\mu} = &\sqrt{n}\mu\left(\boldsymbol{\beta}, \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) + \left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\sqrt{n}\left(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}\right) \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\boldsymbol{A}_m\left[\boldsymbol{\delta} + \sqrt{n}\left(\hat{\gamma}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right] \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left[\left(\frac{\partial\mu}{\partial\gamma}\right)'\left(\boldsymbol{\Pi}_m'\boldsymbol{B}_m\left\{\left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + \sqrt{n}\left(\hat{\gamma}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right\} - \left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta}\right)\right] + o_p(1)
\end{aligned}
$$

Applying (14) yields

$$\begin{aligned}
\sqrt{n}\bar{\mu} &= \sqrt{n}\mu\left(\hat{\boldsymbol{\beta}}_M,\hat{\boldsymbol{\gamma}}_M\right) - \left(\frac{\partial\mu}{\partial\boldsymbol{\gamma}}\right)'\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\boldsymbol{A}_m\left[\boldsymbol{\delta}+\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right] \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left[\left(\frac{\partial\mu}{\partial\boldsymbol{\gamma}}\right)'\left(\boldsymbol{\Pi}_m'\boldsymbol{B}_m\left\{\left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta}+\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right\} - \left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta}\right)\right] + o_p(1) \\
&= \sqrt{n}\mu\left(\hat{\boldsymbol{\beta}}_M,\hat{\boldsymbol{\gamma}}_M\right) \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left[\left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\boldsymbol{A}_m + \left(\frac{\partial\mu}{\partial\boldsymbol{\gamma}}\right)'\left(\boldsymbol{\Pi}_m'\boldsymbol{B}_m-\boldsymbol{I}_q\right)\right]\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left[\left(\frac{\partial\mu}{\partial\boldsymbol{\beta}}\right)'\boldsymbol{A}_m + \left(\frac{\partial\mu}{\partial\boldsymbol{\gamma}}\right)'\boldsymbol{\Pi}_m'\boldsymbol{B}_m\left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\right]\boldsymbol{\delta} \\
&- \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\frac{\partial\mu}{\partial\boldsymbol{\gamma}}\right)'\left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + o_p(1) \\
&= \sqrt{n}\mu\left(\hat{\boldsymbol{\beta}}_M,\hat{\boldsymbol{\gamma}}_M\right) + \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\boldsymbol{D}_{\boldsymbol{\theta}}'\begin{pmatrix}\boldsymbol{A}_m\\ \boldsymbol{\Pi}_m'\boldsymbol{B}_m-\boldsymbol{I}_q\end{pmatrix}\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\boldsymbol{D}_{\boldsymbol{\theta}}'\begin{pmatrix}\boldsymbol{A}_m\\ \boldsymbol{\Pi}_m'\boldsymbol{B}_m\left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\end{pmatrix}\boldsymbol{\delta} \\
&- \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\frac{\partial\mu}{\partial\boldsymbol{\gamma}}\right)'\left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + o_p(1). \qquad (15)
\end{aligned}$$

Further, using

$$\boldsymbol{C}_m = \left(\boldsymbol{S}_m\boldsymbol{Q}_m^{-1}\boldsymbol{S}_m'\boldsymbol{Q}-\boldsymbol{I}\right)\boldsymbol{S}_0 = \begin{pmatrix}\boldsymbol{A}_m\\ \boldsymbol{\Pi}_m'\boldsymbol{B}_m-\boldsymbol{I}_q\end{pmatrix} \qquad (16)$$

in (15) yields

$$\begin{aligned}
\sqrt{n}\bar{\mu} &= \sqrt{n}\mu\left(\hat{\boldsymbol{\beta}}_M,\hat{\boldsymbol{\gamma}}_M\right) + \left[\boldsymbol{D}_{\boldsymbol{\theta}}'\sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\boldsymbol{C}_m\right]\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_M - \frac{\boldsymbol{\delta}}{\sqrt{n}}\right) \\
&+ \boldsymbol{D}_{\boldsymbol{\theta}}'\sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left[\begin{pmatrix}\boldsymbol{A}_m\\ \boldsymbol{\Pi}_m'\boldsymbol{B}_m\left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\end{pmatrix}\right]\boldsymbol{\delta} \\
&- \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\frac{\partial\mu}{\partial\boldsymbol{\gamma}}\right)'\left(\boldsymbol{I}_q-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right)\boldsymbol{\delta} + o_p(1). \qquad (17)
\end{aligned}$$

The center of the confidence interval proposed in Liu (2015) is $\bar{\mu} - b\left(\hat{\boldsymbol{\delta}}\right)$, where $b\left(\hat{\boldsymbol{\delta}}\right) = \hat{\boldsymbol{D}}_{\boldsymbol{\theta}}'\sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right)\hat{\boldsymbol{C}}_m\hat{\boldsymbol{\gamma}}_M$ with $\hat{\boldsymbol{\delta}} = \sqrt{n}\hat{\boldsymbol{\gamma}}_M$. Thus, using (17) together with (16), we have

$$
\begin{aligned}
\sqrt{n}\bar{\mu} &= \sqrt{n}\mu\left(\hat{\beta}_M, \hat{\gamma}_M\right) + \left[\boldsymbol{D}_{\boldsymbol{\theta}}' \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \boldsymbol{C}_m\right] \sqrt{n}\hat{\gamma}_M - \left[\boldsymbol{D}_{\boldsymbol{\theta}}' \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \boldsymbol{C}_m\right] \boldsymbol{\delta} \\
&+ \boldsymbol{D}_{\boldsymbol{\theta}}' \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \left[\begin{pmatrix} \boldsymbol{A}_m \\ \boldsymbol{\Pi}_m' \boldsymbol{B}_m \left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right) \end{pmatrix}\right] \boldsymbol{\delta} \\
&- \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \left(\frac{\partial \mu}{\partial \gamma}\right)' \left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right) \boldsymbol{\delta} + o_p(1) \\
&= \sqrt{n}\mu\left(\hat{\beta}_M, \hat{\gamma}_M\right) + \left[\boldsymbol{D}_{\boldsymbol{\theta}}' \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \boldsymbol{C}_m\right] \sqrt{n}\hat{\gamma}_M \\
&+ \boldsymbol{D}_{\boldsymbol{\theta}}' \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{I} - \boldsymbol{\Pi}_m' \boldsymbol{B}_m \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m \end{pmatrix} \boldsymbol{\delta} \\
&- \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \left(\frac{\partial \mu}{\partial \gamma}\right)' \left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right) \boldsymbol{\delta} + o_p(1) \\
&= \sqrt{n}\mu\left(\hat{\beta}_M, \hat{\gamma}_M\right) + \left[\boldsymbol{D}_{\boldsymbol{\theta}}' \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \boldsymbol{C}_m\right] \sqrt{n}\hat{\gamma}_M \\
&+ \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \left(\frac{\partial \mu}{\partial \gamma}\right)' \left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right) \boldsymbol{\delta} \\
&- \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \left(\frac{\partial \mu}{\partial \gamma}\right)' \left(\boldsymbol{I}_q - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\right) \boldsymbol{\delta} + o_p(1) \quad \text{(by (10))} \\
&= \sqrt{n}\mu\left(\hat{\beta}_M, \hat{\gamma}_M\right) + \left[\boldsymbol{D}_{\boldsymbol{\theta}}' \sum_{m=1}^{M} w\left(m|\hat{\boldsymbol{\delta}}\right) \boldsymbol{C}_m\right] \sqrt{n}\hat{\gamma}_M + o_p(1).
\end{aligned}
$$

Hence, $\bar{\mu} = \hat{\mu}_M + b(\hat{\boldsymbol{\delta}}) + o_p(n^{-1/2})$ and the intervals are asymptotically equivalent. $\square$

Proof of Theorem 3. Also for this proof, it suffices to show that the intervals have the same centers asymptotically as the lengths are equal. Applying equation (9) in the appendix leads to

$$
\hat{\boldsymbol{\theta}}_m = \frac{1}{n}\begin{pmatrix} \left(\hat{\boldsymbol{Q}}_{xx}^{-1} + \hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\boldsymbol{\Pi}_m'\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\right)\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\boldsymbol{\Pi}_m'\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\boldsymbol{Z}'\boldsymbol{y} \\ -\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\boldsymbol{X}'\boldsymbol{y} + \hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\boldsymbol{Z}'\boldsymbol{y} \end{pmatrix},
$$

where the first block corresponds to $\hat{\beta}_m$ and the second block corresponds to $\hat{\gamma}_m$. Note that

$$
\hat{\boldsymbol{\theta}}_M + \hat{\boldsymbol{C}}_m\hat{\gamma}_M = \begin{pmatrix} \hat{\beta}_M + \hat{\boldsymbol{A}}_m\hat{\gamma}_M \\ \boldsymbol{\Pi}_m'\hat{\boldsymbol{B}}_m\hat{\gamma}_M \end{pmatrix},
$$

where $\hat{\boldsymbol{A}}_m = \hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\left(\boldsymbol{I} - \boldsymbol{\Pi}_m'\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{K}}^{-1}\right)$ and $\hat{\boldsymbol{B}}_m = \hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{K}}^{-1}$. Therefore, the difference of the centers satisfies

$$\boldsymbol{c}'\left[\sum_{m=1}^{M-1} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\hat{\boldsymbol{\theta}}^{(m)}-\hat{\boldsymbol{\theta}}_M-\hat{\boldsymbol{C}}_m\hat{\gamma}_M\right)\right]=\boldsymbol{c}'\left[\sum_{m=1}^{M-1} w\left(m|\hat{\boldsymbol{\delta}}\right)\begin{pmatrix}\hat{\boldsymbol{\beta}}_m-\hat{\boldsymbol{\beta}}_M-\hat{\boldsymbol{A}}_m\hat{\gamma}_M\\ \hat{\gamma}^{(m)}-\boldsymbol{\Pi}_m'\hat{\boldsymbol{B}}_m\hat{\gamma}_M\end{pmatrix}\right]. \tag{18}$$

First, for the upper block

$$\begin{aligned}\hat{\boldsymbol{\beta}}_m-\hat{\boldsymbol{\beta}}_M-\hat{\boldsymbol{A}}_m\hat{\gamma}_M&=\left(\hat{\boldsymbol{Q}}_{xx}^{-1}+\hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\boldsymbol{\Pi}_m'\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\right)\boldsymbol{X}'\boldsymbol{y}-\hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\boldsymbol{\Pi}_m'\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\boldsymbol{Z}'\boldsymbol{y}\\&\quad-\left(\hat{\boldsymbol{Q}}_{xx}^{-1}+\hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\hat{\boldsymbol{K}}\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\right)\boldsymbol{X}'\boldsymbol{y}+\hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\hat{\boldsymbol{K}}\boldsymbol{Z}'\boldsymbol{y}\\&\quad-\hat{\boldsymbol{Q}}_{xx}^{-1}\hat{\boldsymbol{Q}}_{xz}\left(\boldsymbol{I}-\boldsymbol{\Pi}_m'\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{K}}^{-1}\right)\left(-\hat{\boldsymbol{K}}\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\boldsymbol{X}'\boldsymbol{y}+\hat{\boldsymbol{K}}\boldsymbol{Z}'\boldsymbol{y}\right)\\&=\boldsymbol{0}_{p\times1}.\end{aligned}$$

Second, for the lower block

$$\begin{aligned}\hat{\gamma}^{(m)}-\boldsymbol{\Pi}_m'\hat{\boldsymbol{B}}_m\hat{\gamma}_M&=\boldsymbol{\Pi}_m'\left(\hat{\gamma}_m-\hat{\boldsymbol{B}}_m\hat{\gamma}_M\right)\\&=\boldsymbol{\Pi}_m'\left[\left(-\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\boldsymbol{X}'\boldsymbol{y}+\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\boldsymbol{Z}'\boldsymbol{y}\right)\right.\\&\quad\left.-\hat{\boldsymbol{B}}_m\left(-\hat{\boldsymbol{K}}\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\boldsymbol{X}'\boldsymbol{y}+\hat{\boldsymbol{K}}\boldsymbol{Z}'\boldsymbol{y}\right)\right]\\&=\boldsymbol{\Pi}_m'\left(\hat{\boldsymbol{B}}_m\hat{\boldsymbol{K}}-\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\right)\left(\hat{\boldsymbol{Q}}_{zx}\hat{\boldsymbol{Q}}_{xx}^{-1}\boldsymbol{X}'-\boldsymbol{Z}'\right)\boldsymbol{y}\\&=\boldsymbol{0}_{q\times1}\end{aligned}$$

where the last equality holds because of $\hat{\boldsymbol{B}}_m\hat{\boldsymbol{K}}=\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m\hat{\boldsymbol{K}}^{-1}\hat{\boldsymbol{K}}=\hat{\boldsymbol{K}}_m\boldsymbol{\Pi}_m$. Thus, the difference between the centers in (18) may be further seen to be

$$\boldsymbol{c}'\left[\sum_{m=1}^{M-1} w\left(m|\hat{\boldsymbol{\delta}}\right)\left(\hat{\boldsymbol{\theta}}_m-\hat{\boldsymbol{\theta}}_M-\hat{\boldsymbol{C}}_m\hat{\gamma}_M\right)\right]=\boldsymbol{c}'\left[\sum_{m=1}^{M-1} w\left(m|\hat{\boldsymbol{\delta}}\right)\begin{pmatrix}\boldsymbol{0}_{p\times1}\\ \boldsymbol{0}_{q\times1}\end{pmatrix}\right]=\boldsymbol{0}_{(p+q)\times1}.$$

Consequently, the confidence intervals for a linear function $\mu(\boldsymbol{\theta})=\boldsymbol{c}'\boldsymbol{\theta}$ of the parameters based on either frequentist model averaging or on least squares in the full model are equivalent, even in finite samples. □

References

Akaike H (1974) A New Look at the Statistical Model Identification. IEEE Automat Contr 19(6):716–723

Ando T, Li KC (2014) A Model-Averaging Approach for High-Dimensional Regression. J Am Stat Assoc 109(505):254–265

Berk R, Brown L, Buja A, Zhang K, Zhao L (2013) Valid post-selection inference. Ann Stat 41(2):802–837

Cheng TCF, Ing CK, Yu SH (2015) Toward optimal model averaging in regression models with time series errors. J Econometrics 189(2):321–334

Hansen BE (2007) Least squares model averaging. Econometrica 75(4):1175–1189

Hansen BE, Racine JS (2012) Jackknife model averaging. J Econometrics 167(1):38–46

Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, New York

Hjort NL, Claeskens G (2003) Frequentist Model Average Estimators. J Am Stat Assoc 98(464):879–899

Kabaila P (1995) The Effect of Model Selection on Confidence Regions and Prediction Regions. Economet Theor 11(3):537–549

Kabaila P, Leeb H (2006) On the Large-Sample Minimal Coverage Probability of Confidence Intervals After Model Selection. J Am Stat Assoc 101(474):619–629

Leeb H, Pötscher BM (2005) Model Selection and Inference: Facts and Fiction. Economet Theor 21(1):21–59

Liu CA (2015) Distribution theory of the least squares averaging estimator. J Econometrics 186(1):142–159

Liu Q, Okui R (2013) Heteroscedasticity-robust C_p model averaging. Economet J 16(3):463–472

Pötscher BM (1991) Effects of Model Selection on Inference. Economet Theor 7(2):163–185

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464

Wan ATK, Zhang X, Zou G (2010) Least squares model averaging by Mallows criterion. J Econometrics 156(2):277–283

Wang H, Zhou SZF (2013) Interval Estimation by Frequentist Model Averaging. Commun Stat Theory 42(23):4342–4356