Queen's Economics Department Working Paper No. 1535

Frequentist Model Averaging with Nash Bargaining: A Stochastic Dominance Approach

Stelios Arvanitis
Department of Economics, AUEB

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

3-2025

Frequentist Model Averaging with Nash Bargaining: A Stochastic Dominance Approach

Stelios Arvanitis[*], Mehmet Pinar[†], Thanasis Stengos[‡], Nikolas Topaloglou [§]

March 8, 2025

Abstract

Within the Frequentist Model Averaging framework for linear models, we introduce a multi-objective model averaging methodology that extends both the generalized Jackknife Model Averaging (JMA) and the Mallows Model Averaging (MMA) criteria. Our approach constructs estimators based on stochastic dominance principles and explores averaging methods that minimize multiple scalarizations of the joint criterion integrating MMA and JMA. Additionally, we propose an estimator that can be interpreted as a Nash bargaining solution between the competing scalar criteria. We establish the asymptotic properties of these estimators under both correct specification and global misspecification. Monte Carlo simulations demonstrate that some of the proposed averaging estimators outperform JMA and MMA in terms of MSE/MAE. In an empirical application to economic growth data, our model averaging methods assign greater weight to fundamental Solow-type growth variables while also incorporating regressors that capture the role of geography and institutional quality.

JEL Codes: C51, C52.

Keywords: frequentistic model averaging, Jacknife MA, Mallows MA, multi-objective optimization, stochastic dominance, approximate bound, ℓ^p-scalarization, Nash bargaining solution, growth regressions.

[*]Athens University of Economics and Business, Greece. Email: stelios@aueb.gr
[†]Universidad de Sevilla, Spain. Email:mpinar@us.es
[‡]University of Guelph, Canada. Email: tstengos@uoguelph.ca
[§]Athens University of Economics and Business, Greece. Email: nikolas@aueb.gr

1 Introduction

Model Averaging has been proposed as a general framework to address model uncertainty. Two primary approaches have emerged in the literature: the Bayesian Model Averaging (BMA) and the Frequentist Model Averaging (FMA). Model uncertainty naturally arises due to the presence of multiple competing theories, each emphasizing different potential data-generating processes (DGPs), which are often non-nested.

There is a voluminous amount of work done in the BMA strand, as Bayesian methods are ideally suited for averaging or combining different models using posterior-based weights; see Steel (2020) [36] for a very comprehensive survey of the BMA methods. However, in this paper, we examine an approach that rests within the FMA methodological camp and provide a comparison with other existing methods in this area.

A prominent field where model uncertainty plays a crucial role is economic growth theory, which encompasses multiple competing perspectives. The standard production function-based Solow growth model, the institutional perspective proposed by Acemoglu et al. (2001) [1], and the geographic determinants emphasized by Sachs (2003) [33] all represent valid but distinct explanatory approaches. The importance of model uncertainty in assessing the relative influence of these growth theories has been studied in Kourtellos et al. (2010) [23]. An earlier approach that aimed to evaluate the robustness of various regressors in growth models is the Extreme Bounds Analysis (EBA), introduced by Leamer (1983) [24] and applied to economic growth empirics by Levine and Renelt (1992) [25]. In that study, variables were classified as "robust" or "fragile," with those surviving across different model specifications being primarily associated with the Solow growth framework.

Within the FMA framework, several model averaging estimators have been developed, including the Mallows Model Averaging (MMA) estimator, the Jackknife Model Averaging (JMA) estimator, its generalized variant, the Focused Information Criterion (FIC), and the Plug-in averaging estimator. The theoretical development of these methods was initiated by Hansen (2007) [19], Liu (2015) [27], and Zhang and Liu (2019) [38], who analyzed the asymptotic properties of least-squares-based averaging estimators in linear regression models under local asymptotic frameworks.

Our main contribution in this paper is to introduce a model averaging methodology that expands on both the generalized Jackknife Model Averaging (JMA) and the Mallows Model

Averaging (MMA) criteria in a multi-objective context within a stochastic dominance framework. Specifically, and in the spirit of Zhang and Liu (2019) [38], we consider a variant of the MMA estimator where the regularization parameter asymptotically vanishes as the sample size increases. We then construct a multi-objective optimization problem incorporating both the MMA and JMA criteria, aiming to construct averaging estimators that leverage information from both methodologies.

Using a multi-objective optimization approach, we consider averaging estimators that approximate a potentially infeasible solution that jointly minimizes both criteria, emerging from a stochastic dominance perspective. We also consider averaging estimators derived from different scalarization techniques of the multi-objective function, including those based on ℓ^p-norms and a Nash bargaining solution approach from cooperative game theory. The suggested estimators, denoted as ℓ^1, ℓ^2, ℓ^∞, and Nash, are analyzed alongside their approximate bound (AB), as referenced in Arvanitis et al. (2021) [6]. Methodologically, the framework for multi-objective optimization and the previously mentioned estimators can be easily expanded to incorporate additional basis averaging estimators, beyond the MMA and the JMA.

We derive the limit theory of the estimators under both correctly specified and globally misspecified scenarios. In the correctly specified case, all proposed averaging estimators converge in probability to the deterministic vector that selects the minimal well-specified model. However, in the presence of global misspecification—where all models exclude at least one DGP regressor—the theory exhibits richer dynamics. While weights eventually stabilize in all cases, their asymptotic limits vary, reflecting differences in how auxiliary regressors influence estimation. Moreover, the Nash approach appears to provide a systematic method for increasing the weight given to well-defined regressors with minimal covariance with the omitted ones, while still preserving aspects of parsimony. Overall, the Nash estimator strikes a balance between the regularized MMA criterion and the structural properties captured by JMA. This distinction highlights the Nash estimator's potential advantage in misspecified stiings, where it retains greater information on the asymptotic dependence stucture among regressors while still controlling overfitting.

Monte Carlo experiments in both correctly specified and misspecified frameworks are conducted. The results suggest that for every considered case, at least one of the proposed multi-objective model averaging estimators outperforms the base MMA and JMA estimators in terms of MSE/MAE criteria. In many cases, the dominant estimator is the Nash averaging estimator.[1]

An empirical application in economic growth analysis is also provided. Our findings are compared with those from existing methods, including the classic study of Levine and Renelt (1992) [25]. Overall, our proposed methods assign heavier weights to models constructed from fundamental Solow growth regressors (Solow, 1956 [35]; Mankiw et al., 1992 [29]) but do not entirely discount models incorporating auxiliary regressors linked to geography (Diamond, 1997 [13]; Gallup et al., 1999 [33]; Sachs, 2003 [33]) and institutional quality (Acemoglu et al., 2001 [1]; Rodrik et al., 2004 [32]).

The remainder of the paper is structured as follows: Section 2 provides the background on regression models. Section 3 introduces the key model averaging criteria, including the modified MMA and the Zhang and Liu (2019) [38] modification of the JMA criterion. Section 4 derives the limit theory for the modified MMA estimator and contrasts it with the analogous derivations of Zhang and Liu (2019) [38] for the JMA case. Section 5 presents the multi-objective model averaging estimators constructed from both criteria, analyzing their asymptotic properties under correct specification. Section 6 extends these results to the case of misspecification due to omitted regressors. Section 7 discusses potential extensions. Section 8 presents Monte Carlo experiments. Section 9 provides the empirical application to economic growth. Section 10 concludes. The appendix contains tables related to the Monte Carlo experiments and empirical application. A supplement [5] contains the proofs of theoretical results and additional tables for further analysis.

2 Background

The linear model background of Zhang and Liu (2019) [38] is considered. In particular, we analyze the linear regression model expressed in matrix notation as follows:

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon, \tag{1}$$

where the dependent variable $\mathbf{y}$ is a random n-vector, $\mathbf{X}_1$ is the core regressors $n \times K_1$ random matrix, $\mathbf{X}_2$ is the auxiliary regressors $n \times K_2$ random matrix, and ε is the random n-vector of errors. β_1 and β_2 are the associated unknown parameter vectors. From the information above

[1] Unreported results from experiments under alternative settings that violate our assumption framework suggest that this dominance is not universally guaranteed.

an array of $M := K_2+1$ statistical models is formed. The m^{th} model is formed by the regression consisting of the core regressors, accompanied by the initial $m-1$ auxiliary regressors, excluding the remaining K_2-m+1 ones. As in Zhang and Liu (2019) [38], it is assumed that there exists a maximal $K_1 \leq M_0 < M$ for which the models $m = 1, \ldots, M_0$ are considered misspecified, in the sense that they have non-zero sloped omitted regressors. M_0 thus represents the minimal well-specified regression. The regressors included in the m^{th} model are $\mathbf{X}_m := \mathbf{X}\Pi'_m$, where $\mathbf{X} := (\mathbf{X}_1, \mathbf{X}_2)$, and $\Pi_m := (\mathbf{I}_{k_m \times k_m}, \mathbf{0}_{k_m \times (K-k_m)})$, where $k_m = K_1 + m - 1$, $K = K_1 + K_2$. The unrestricted OLSE for $\beta := (\beta'_1, \beta'_2)$ in the m^{th} model is $\beta_m := \Pi'_m(\mathbf{X}'_m\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{y}$, and for $\mathbf{w}$ an element of the $M-1$ unit simplex, the resulting OLSE averaging estimator of β is $\beta(\mathbf{w}) := \sum_{m=1}^{M} \mathbf{w}_m\beta_m$.

3 Basis Averaging Criteria

Given the projection matrices

$$\mathbf{P}_m := \mathbf{X}_m(\mathbf{X}'_m\mathbf{X}_m)^{-1}\mathbf{X}'_m, \quad m = 1, \ldots, M,$$

the overall averaging (across models) projection is given by:

$$\mathbf{P}(\mathbf{w}) := \sum_{m=1}^{M} \mathbf{w}_m\mathbf{P}_m.$$

Furthermore, we define the variance estimator:

$$\sigma_n^2 := (n-K)^{-1}\|\mathbf{y} - \mathbf{X}\beta_M\|^2,$$

where $\|\cdot\|$ denotes the Euclidean norm. The terms ϕ_n, $\phi_n^\star$ are potentially stochastic or data-dependent regularization parameters that depend on n.

The notation $\mathbf{P}(i,i)_m$ refers to the i^{th} diagonal element of $\mathbf{P}_m$, and $\mathbf{D}_m$ is a diagonal matrix with elements:

$$\mathbf{D}_m(i,i) = \frac{1}{1-\mathbf{P}(i,i)_m}, \quad i = 1, \ldots, n.$$

Finally, we define the model complexity vector as: $\mathbf{K} := (k_m)_{m=1,\ldots,M}$.

3.1 Mallows Model Averaging (MMA)

The Mallows Model Averaging (MMA) criterion is given by:

$$\mathcal{M}_n(\mathbf{w}) := \|(\mathbf{I}_{n\times n} - \mathbf{P}(\mathbf{w}))\mathbf{y}\|^2 + \phi_n^\star \sigma_n^2 \mathbf{K}'\mathbf{w}. \tag{2}$$

The first term of $\mathcal{M}_n(\mathbf{w})$ represents the residual sum of squares (RSS) under the model-averaged projection, while the second term penalizes model complexity, weighted by $\phi_n^\star$. The original MMA criterion (see Hansen (2007) [19]) is recovered when $\phi_n^\star = 2$.

The MMA weights are obtained as the solution to the following optimization problem:

$$\mathbf{w}_{\text{MMA}} \in \arg\min_{\mathbf{w}\in\Delta^{M-1}} \mathcal{M}_n(\mathbf{w}),$$

where Δ^{M-1} is the standard $M-1$ dimensional unit simplex. The corresponding Mallows averaging estimator of β is given by:

$$\beta_{\text{MMA}} := \beta(\mathbf{w}_{m,\text{MMA}}).$$

The existence of a solution follows from standard arguments, utilizing the compactness of the simplex and the continuity of $\mathcal{M}_n(\mathbf{w})$ with respect to $\mathbf{w}$.

3.2 Generalized Jackknife Model Averaging (JMA)

The generalized Jackknife Model Averaging (JMA) criterion, introduced by Zhang and Liu (2019) [38], is given by:

$$\mathcal{J}_n(\mathbf{w}) := \|(\mathbf{D}_m(\mathbf{P}(\mathbf{w}) - \mathbf{I}_{n\times n}) + \mathbf{I}_{n\times n})\mathbf{y}\|^2 + \phi_n \mathbf{K}'\mathbf{w}. \tag{3}$$

The JMA criterion is designed to minimize predictive risk using a leave-one-out cross-validation approach, adapting the bias correction from the Jackknife resampling method (see Racine (1997) [31]).

The JMA weights are obtained by solving:

$$\mathbf{w}_{\text{JMA}} \in \arg\min_{\mathbf{w}\in\Delta^{M-1}} \mathcal{J}_n(\mathbf{w}).$$

The JMA estimator of β is then:

$$\beta_{\mathrm{JMA}} := \beta(\mathbf{w}_{m,\mathrm{JMA}}).$$

A key advantage of JMA over MMA is its ability to adjust for potential conditional heteroskedasticity in the errors, as reflected in the structure of $\mathbf{D}_m$. This makes JMA particularly suitable for cases where residual variance is model-dependent.

4 Limit Theory for the Basis Averaging Estimators

The asymptotic behavior of the generalized MMA estimator is determined as n approaches infinity. The corresponding asymptotic behavior for the generalized JMA is presented in Zhang and Liu (2019), particularly in Theorem 5. Initially, a set of mild assumptions is introduced, closely resembling conditions **C.1-C.4** outlined in Zhang and Liu (2019) [38]. The symbol $\rightsquigarrow$ is used to signify convergence in distribution:

A.1 $n^{-1}\mathbf{X}'\mathbf{X} \rightsquigarrow Q := \mathbb{E}(\mathbf{X}_1\mathbf{X}_1')$, *with* Q *pd.* $\mathbf{X}_1$ *denotes the first row of* $\mathbf{X}$.

A.2 $n^{-1/2}\mathbf{X}'\varepsilon \rightsquigarrow \mathbf{z} \sim N(\mathbf{0}_{K\times 1}, V)$, *with* $V := \mathbb{E}(\varepsilon_1^2\mathbf{X}_1\mathbf{X}_1')$ *positive definite.*

A.3 $\sup_{i\leq n}\sup_{m\leq M}\mathbf{P}_m(i,i) = o_p(n^{-1/2})$.

A.4 $n^{-1}\sum_{i=1}^n \varepsilon_i^2\mathbf{X}_i\mathbf{X}_i' \rightsquigarrow V$ and $n^{-1}\sum_{i=1}^n \varepsilon_i^2 \rightsquigarrow \sigma^2 > 0$.

In time series settings **A.1** and **A.4** are expected to hold under conditions of stationarity and ergodicity, as well as tail decay conditions for the marginal distributions of the random variables involved. Similarly, **A.2** would follow under additional mixing conditions, as well as linear independence for the random elements that appear in the form of the asymptotic variance. Approrpiate notions of exchangeability could validate the aforementioned assumptions in non-time series settings. As Zhang and Liu (2019) [38] point out, **A.3** is weaker than the analogous condition in Andrews (1991) [3].

A limit theory for the generalized MMA can be obtained via the proofs of Th. 3-4 of Zhang and Liu (2019) [38]; here the fact that the matrices D_m and $\mathbf{I}_{n\times n} - D_m = \mathrm{diag}(\frac{P_{(i,i)m}}{1-P_{(i,i)m}})_i$ do not affect the criterion, which is also self-normalized due to Assumptions **A.1-2**, imply that the rate of divergence of the regularization parameter $\phi_n^{\star}$ can be unrestricted. In what follows, and for $k \in \{M_0+1,\ldots,M\}$, $\mathbf{w}^{(k)}$ denotes the k^{th} element of the Δ^{M-1} simplex, i.e., $\mathbf{w}_m^{(k)} := \mathbb{I}_k(m)$, with $\mathbb{I}_k(\cdot)$ denoting the indicator of the k^{th} coordinate:

Theorem 1. *Suppose that* ***A.1-A.4*** *hold, and that* $\phi_n^\star \to \infty$. *Then,*

$$\mathbf{w}_{\mathrm{MMA}} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} \sigma^2\mathbf{K}'\mathbf{w} = \sigma^2\mathbf{K}'\mathbf{w}^{(M_0+1)}. \tag{4}$$

Furthermore,

$$n^{1/2}(\beta_{\mathrm{MMA}} - \beta) \rightsquigarrow V^\star\mathbf{z} \sim N(\mathbf{0}_{K\times 1}V^\star V V^\star), \tag{5}$$

with $V^\star := \Pi_{M_0+1}(\Pi_{M_0+1}Q\Pi'_{M_0+1})^{-1}\Pi'_{M_0+1}$. *Finally,* $\phi_n^\star(\mathbf{w}_{\mathrm{MMA}} - \mathbf{w}^{(M_0+1)}) = o_p(1)$.

The proof is given in the supplement [5]. Theorem 5 of Zhang and Liu (2019) [38] provides the limit theory for the JMA estimator; the regularization parameters' growth to infinity is restricted due to the dependence of the criterion on the $\mathbf{P}_m$ matrices and their asymptotic behavior as prescribed by **A.3**:

Theorem 2. *Suppose that* ***A.1-A.4*** *hold, and that* $\phi_n \to \infty$, *while* $\frac{\phi_n}{\sqrt{n}} \to 0$. *Then,*

$$\mathbf{w}_{\mathrm{JMA}} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} \mathbf{K}'\mathbf{w} = \mathbf{K}'\mathbf{w}^{(M_0+1)}, \tag{6}$$

where $\Delta_0^{M-1} := \left\{\mathbf{w} \in \Delta^{M-1}, \mathbf{w}_m = 0, \forall m = 1, \ldots, M_0\right\}$. *Furthermore,*

$$n^{1/2}(\beta_{\mathrm{JMA}} - \beta) \rightsquigarrow V^\star\mathbf{z}. \tag{7}$$

Finally, $\phi_n(\mathbf{w}_{\mathrm{JMA}} - \mathbf{w}^{(M_0+1)}) = o_p(1)$.

The proof is also given in the supplement [5]. Essentially, $\mathcal{J}_n$ epi-converges in distribution (see Knight (1999) [22]) to the linear function $\mathbf{w} \to \mathbf{K}'\mathbf{w}$, something that results to (6) from Prop. 3.2 in Ch. 5 of Molchanov (2006) [30] via the use of Skorokhod representations justified by Th. 3.7.25 of Gine and Nickl (2021) [17]. By construction then the limiting criterion is uniquely minimized at $\mathbf{w}^{(M_0+1)}$. The rate result along with the restrition $\phi_n/\sqrt{n} \to 0$, implies that the above are non informative on the issue of asymptotic tightness for $\sqrt{n}(\mathbf{w}_{\mathrm{JMA}} - \mathbf{w}^{(M_0+1)})$.

Both estimators thus share the limit theory of the OLSE for the (latent) minimal well-specified model. Asymptotic normality occurs as a result of integrating the diverging penalization parameters into their definition.

5 Multi-objective Model Averaging

The MMA and JMA averaging criteria can be jointly utilized to construct a stochastic dominance (SD) ordering over the simplex Δ^{M-1}. The fundamental idea behind this ordering is that a convex combination $\mathbf{w}$ dominates another $\mathbf{w}^\star$ if it attains lower values under both the Mallows and Jackknife criteria, and is thus universally preferred. Formally,

$$\mathbf{w} \succeq \mathbf{w}^\star \quad \text{iff} \quad \mathcal{M}_n(\mathbf{w}) \leq \mathcal{M}_n(\mathbf{w}^\star), \quad \text{and} \quad \mathcal{J}_n(\mathbf{w}) \leq \mathcal{J}_n(\mathbf{w}^\star).$$

This (pre-) order defines a multi-objective optimization problem (MOOP) (see Hwang and Masud (2012) [21]):

$$\min_{\mathbf{w} \in \Delta^{M-1}} \mathcal{W}_n(\mathbf{w}),$$

where the objective function is given by

$$\mathcal{W}_n(\mathbf{w}) := (\mathcal{M}_n(\mathbf{w}), \mathcal{J}_n(\mathbf{w})).$$

This formulation naturally arises in multi-criteria decision-making settings, where a trade-off must be made between competing objectives—here, the bias-variance trade-off of MMA and the robustness properties of JMA.

5.1 Efficient Weights and Pareto Optimality

Any maximal element of this (pre-) order—i.e., a solution that simultaneously minimizes both criteria—dominates all other possible weight allocations. Such an element would ideally inherit advantages from both methodologies, balancing Mallows' bias-variance trade-off with Jackknife's robustness against outliers. However, such a solution is expected to be rarely attainable in finite samples due to the nontrivial structure of the objective functions.

Instead, we focus on Pareto-efficient solutions, where no weight allocation can be improved in one criterion without worsening the other. A weight $\mathbf{w}$ is Pareto-efficient if, for any alternative weight $\mathbf{w}^\star$, whenever

$$\mathcal{M}_n(\mathbf{w}) > \mathcal{M}_n(\mathbf{w}^\star)$$

(respectively, $\mathcal{J}_n(\mathbf{w}) > \mathcal{J}_n(\mathbf{w}^\star)$), it must hold that

$$\mathcal{J}_n(\mathbf{w}) < \mathcal{J}_n(\mathbf{w}^\star)$$

(respectively, $\mathcal{M}_n(\mathbf{w}) < \mathcal{M}_n(\mathbf{w}^\star)$).

In other words, efficient weights represent different compromise solutions, reflecting diverse preferences over the trade-off between the characteristics of the two base criteria.

5.2 Approximate Bound (AB) Weights

A particularly relevant concept in stochastic dominance optimization is the Approximate Bound (AB), introduced by Arvanitis et al. (2021) [6]. This weight vector is as close as possible to being maximal, meaning that it minimizes the worst-case deviation from dominance. It is defined as the solution to:

$$\min_{\mathbf{w}\in\Delta^{M-1}} \sup_{G\in\{\mathcal{M}_n,\mathcal{J}_n\},\mathbf{w}^\star\in\Delta^{M-1}} (G(\mathbf{w}) - G(\mathbf{w}^\star)).$$

This formulation ensures that the chosen estimator achieves a balanced performance across both criteria, while also being an efficient element of the stochastic dominance order.

5.3 Scalarization via ℓ^p-Norms

The AB solution is closely related to scalarization techniques (see Hwang and Masud (2012) [21]). Instead of treating the problem as a vector optimization, one can convert it into a single-objective problem by applying an ℓ^p-norm to the objective:

$$\min_{\mathbf{w}\in\Delta^{M-1}} \|\mathcal{W}_n(\mathbf{w})\|_p. \tag{8}$$

This generalizes the previous approach, allowing for different prioritization between the two criteria by varying p. The extreme case $p = \infty$ recovers the AB estimator, while lower values of p emphasize global minimization rather than worst-case performance.

Definition 1. The ℓ^p weights are defined by:

$$\mathbf{w}_{\ell^p} \in \arg\min_{\mathbf{w}\in\Delta^{M-1}} \|\mathcal{W}_n(\mathbf{w})\|_p.$$

The corresponding ℓ^p estimator of β is then:

$$\beta_{\ell^p} = \sum_{m=1}^{M} \mathbf{w}_{m,\ell^p} \beta_m.$$

5.4 The Nash Bargaining Solution and Its Role in Model Averaging

While scalarization provides a structured way to prioritize objectives, it lacks scale-invariance; the results depend on the relative magnitudes of $\mathcal{M}_n$ and $\mathcal{J}_n$. To address this, we turn to the Nash bargaining solution (see Ch. 32 in Aumann and Hart (1992) [8]).

In this framework, Mallows and Jackknife act as cooperative players negotiating over model averaging choices. Each player's utility is defined as the relative improvement over their fallback (least satisfactory) option:

$$\mathcal{M}_n(\mathbf{w}_{\mathrm{JMA}}) - \mathcal{M}_n(\mathbf{w})$$

for the Mallows criterion, and similarly

$$\mathcal{J}_n(\mathbf{w}_{\mathrm{MMA}}) - \mathcal{J}_n(\mathbf{w})$$

for the Jackknife criterion.

The Nash solution selects the weight vector that maximizes the product of these utilities, ensuring that neither criterion receives an unacceptably poor outcome. This results in an estimator that is efficient, scale-invariant, symmetric, and independent of irrelevant alternatives—properties highly desirable in model averaging.

Definition 2. The Nash weights are defined by:

$$\mathbf{w}_{\mathrm{Nash}} \in \arg\min_{\mathbf{w}\in\Delta^{M-1}} (\mathcal{M}_n(\mathbf{w}) - \mathcal{M}_n(\mathbf{w}_{\mathrm{JMA}}))(\mathcal{J}_n(\mathbf{w}_{\mathrm{MMA}}) - \mathcal{J}_n(\mathbf{w})),$$

$$\text{s.t.} \quad \mathcal{M}_n(\mathbf{w}_{\mathrm{Nash}}) \leq \mathcal{M}_n(\mathbf{w}_{\mathrm{JMA}}), \quad \mathcal{J}_n(\mathbf{w}_{\mathrm{Nash}}) \leq \mathcal{J}_n(\mathbf{w}_{\mathrm{MMA}}).$$

The corresponding Nash estimator of β is:

$$\beta_{\mathrm{Nash}} = \sum_{m=1}^{M} \mathbf{w}_{m,\mathrm{Nash}} \beta_m.$$

In the following sections, we establish the asymptotic properties of the proposed estimators, demonstrating that they remain consistent under correct specification while also exhibiting robustness to misspecification.

5.5 Limit Theory

Theorems 1-2 almost directly provide the limit theory of the multi-objective averaging estimators. Specifically, the following results are obtained:

Theorem 3. *Suppose that* ***A.1-A.4*** *hold, and that* $\min(\phi_n, \phi_n^\star) \to +\infty$, $\frac{\phi_n^\star}{\phi_n} \to C \in [0, +\infty]$, *while* $\phi_n = o_p(\sqrt{n})$. *Then,*

$$\mathbf{w}_{\mathrm{AB}} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} \max(\sigma^2 1_{C>0}, C^{-1}1_{C>0} + 1_{C=0})\mathbf{K}'(\mathbf{w} - \mathbf{w}^{(M_0+1)}), \tag{9}$$

and, $\max(\phi_n^\star, \phi_n)(\mathbf{w}_{\mathrm{AB}} - \mathbf{w}^{(M_0+1)}) = o_p(1)$.

Also,

$$\mathbf{w}_{\ell^p} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} (\sigma^{2p} 1_{C>0} + (C^{-1}1_{C>0} + 1_{C=0})^p)^{1/p}\mathbf{K}'\mathbf{w}. \tag{10}$$

and, $\max(\phi_n^\star, \phi_n)(\mathbf{w}_{\ell^p} - \mathbf{w}^{(M_0+1)}) = o_p(1)$.

Finally,

$$\mathbf{w}_{\mathrm{Nash}} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} \sigma^2(\mathbf{K}'(\mathbf{w} - \mathbf{w}^{(M_0+1)}))^2. \tag{11}$$

and, $\max(\phi_n^\star, \phi_n)(\mathbf{w}_{\mathrm{Nash}} - \mathbf{w}^{(M_0+1)}) = o_p(1)$.

Subsequently, under the conditions above and for $J = \mathrm{AB}$, ℓ^p, Nash,

$$n^{1/2}(\beta_J - \beta) \rightsquigarrow V^\star \mathbf{z}. \tag{12}$$

The proof is provided in the supplement [5]. Under the given assumption framework, the estimators of the weights converge to the minimal well-specified model, ensuring their asymptotic scale invariance across all cases. Consequently, the estimators of the parameters exhibit asymptotic normality, confirming their standard large-sample behavior.

However, the derived limit theory does not possess sufficient granularity to distinguish between the considered estimators under correct specification. As a result, asymptotically, the proposed methodologies appear equivalent within the assumed framework.

To further investigate potential differences in finite samples, we employ Monte Carlo experiments. These experiments aim to detect potential discrepancies that may emerge when sample sizes are fixed, thus providing empirical motivation for the introduction of the multi-objective procedures.

5.6 Subsampling Estimation of Asymptotic Variance

The Multi-objective Model Averaging estimators converge to the limiting distribution of the OLSE for the minimally well specified model. Due to the latency of the latter, the limit theory cannot be directly used for inference; the asymptotic variance cannot be directly and consistently estimated via analogy. One way to circumvent this is via resampling. In what follows a subsampling approach for the estimation of the asymptotic covariance matrix is presented in a stationary and ergodic time series setting which conforms to our empirical application later on.

Specifically, let $b < n$ and given the sample $(\mathbf{y}, \mathbf{X})$ consider the b dependent sequence of sub-samples $((\mathbf{y}, \mathbf{X})_{j,\ldots,j+b-1})_{j=1,n-b+1}$. Then, $\beta_{J,j}$ denotes the MA estimator over the j^{th} subsample, for $J = \text{AB}$, ℓ^p, Nash. The subsampling variance is then defined as $\mathrm{V}_{b,n,J} := \frac{b}{n(n-b+1)} \sum_{j=1}^{n-b+1} (\beta_{J,j} - \frac{1}{(n-b+1)} \sum_{j_\star=1}^{n-b+1} \beta_{J,j_\star})(\beta_{J,j} - \frac{1}{(n-b+1)} \sum_{j_\star=1}^{n-b+1} \beta_{J,j_\star})'$. In a time series context of strict stationarity and strong mixing we obtain the following weak consistency result:

Theorem 4. *Under **A.1-A.4**,* $\min(\phi_n, \phi_n^\star) \to +\infty$*, and,* $\phi_n = o_p(\sqrt{n})$*, and if (a.)* $(\mathbf{y}, \mathbf{X})$ *is stationary and strongly mixing, and (b.) for some* $\epsilon > 0$, $\mathbb{E}(\|\mathbf{X}_1 \varepsilon_1\|^{4+\epsilon}) < +\infty$*, and (c.)* $b \to \infty$, ${}^b/_n \to 0$*, then and for* $J = \text{AB}$, ℓ^p, Nash*:*

$$\mathrm{V}_{b,n,J} \rightsquigarrow \mathrm{Var}(V^\star \mathbf{z}). \tag{13}$$

The proof is given given in the supplement [5]. $\mathrm{V}_{b,n,J}$ is a weakly (actually an L^2-) consistent estimator of the asymptotic variance, and can therefore be used for inference. The current linear models framework implies a manageable computational burden for the multi-objective estimators, and the subsequent subsampling procedure, at least for $p = 1, 2, \infty$. A slight modification of the estimator, in which the weights are held constant on the original sample can be also proven consistent in the present framework. This is associated with minimal computational burden (see for example Section 4 and Proposition 2 in Arvanitis, Scaillet and Topaloglou (2023) [7] for a similar approach). The result can also be extended in non-time series contexts involving exchangeability or more generally invariance of the underlying joint distributions under groups of transformations (see for example Austern and Orbanz (2022) [9]).

6 Omitted Variables Misspecification

We now analyze the effects of global misspecification due to omitted variables on the model averaging procedures. For notational simplicity, and without loss of generality, suppose that $K_1 < M_0$ and that the submatrix of regressors $\mathbf{X}_1$ is omitted from all statistical models under consideration. As a result, every model analyzed is inherently misspecified due to omitted variables.

Under **A.1** and **A.2**, it follows that

$$\beta_m \rightsquigarrow \Pi_m^{\star\prime}(\Pi_m^{\star} + (\Pi_m^{\star} Q \Pi_m^{\star\prime})^{-1}\Pi_m^{\star} Q \Pi_{K_1}^{\prime}\Pi_{K_1})\beta,$$

where $\Pi_m^{\star} := (\mathbf{0}_{(k_m-K_1)\times K_1}, \mathbf{I}_{(k_m-K_1)\times(k_m-K_1)}, \mathbf{0}_{(k_m-K_1)\times(K-k_m)})$.

Thus, inconsistency arises in all models, since the omitted variables induce bias in the parameter estimates. However, partial consistency, where the ordinary least squares estimators remain consistent for the non-omitted regressors, occurs if and only if $\Pi_m^{\star} Q \Pi_{K_1}^{\prime}$ is a zero matrix. This condition implies that every included regressor must be asymptotically uncorrelated with every omitted regressor.

The implications of this misspecification framework for model averaging are of key interest. In particular, do model averaging estimators asymptotically select models that minimize the limiting term $(\Pi_m^{\star} Q \Pi_m^{\star\prime})^{-1}\Pi_m^{\star} Q \Pi_{K_1}^{\prime}\Pi_{K_1}$, thereby mitigating the inconsistency for the retained regressors?

6.1 Asymptotics of the MMA Estimator under Misspecification

Consider the limit behavior of the MMA estimator under assumptions **A.1**, **A.2**, and **A.4**. Define $\mathbf{K}^{\star} := (1, 2, \ldots, K_2)'$ and

$$\sigma_{\star}^2 := \sigma^2 + \beta'(Q - Q\Pi_{K_2}^{\star\prime}(\Pi_{K_2}^{\star} Q \Pi_{K_2}^{\star\prime})^{-1}\Pi_{K_2}^{\star} Q)\beta.$$

Also, for each $m = 1, \cdots, k_2$, let $\Pi_m^{c} := (\mathbf{0}_{K\times(K-k_m)}, \mathbf{I}_{(K-k_m)\times(K-k_m)})$. Then the following result is obtained; its proof can be also found in the supplement [5].

Theorem 5. *Suppose that assumptions* **A.1***,* **A.2***, and* **A.4** *hold under the global misspecification* *framework. Suppose also that* $\frac{\phi_n^\star}{n} \to C \in [0, +\infty]$*. Then,*

$$\mathbf{w}_{\mathrm{MMA}} \rightsquigarrow \mathbf{w}_{\mathrm{MMA}}^{\infty} := \arg \min_{\mathbf{w} \in \Delta^{M-1}} \mathcal{M}_C(\mathbf{w}),$$

where

$$\mathcal{M}_C(\mathbf{w}) := \Lambda(\mathbf{w}) 1_{C<+\infty} + \sigma_\star^2 (C 1_{0<C<+\infty} + 1_{C=+\infty}) \mathbf{K}^{\star\prime} \mathbf{w}.$$

Here,

$$\Lambda(\mathbf{w}) := \sum_{m,m^\star=1}^{K_2} \mathbf{w}_m \mathbf{w}_{m^\star} \beta' (\Pi_{k_1}^{'} A_{(m,m^\star)} \Pi_{k_1} + \Pi_{k_1}^{'} B_{(m,m^\star)} \Pi_{m^\star}^c + \Pi_m^{c'} B_{(m^\star,m)}^{'} \Pi_{k_1} + \Pi_m^{c'} C_{(m,m^\star)} \Pi_{m^\star}^c) \beta,$$

where

$$\begin{aligned} A_{(m,m^\star)} := {} & \Pi_{k_1} Q \Pi_{k_1}^{'} - \Pi_{k_1} Q \Pi_m^{'} (\Pi_m Q \Pi_m^{'})^{-1} \Pi_m Q \Pi_{k_1}^{'} - \Pi_{k_1} Q \Pi_{m^\star}^{'} (\Pi_{m^\star} Q \Pi_{m^\star}^{'})^{-1} \Pi_{m^\star} Q \Pi_{k_1}^{'} \\ & + \Pi_{k_1} Q \Pi_m^{'} (\Pi_m Q \Pi_m^{'})^{-1} \Pi_m Q \Pi_{m^\star}^{'} (\Pi_{m^\star} Q \Pi_{m^\star}^{'})^{-1} \Pi_{m^\star} Q \Pi_{k_1}^{'}, \end{aligned}$$

$$\begin{aligned} B_{(m,m^\star)} := {} & \Pi_{k_1} Q \Pi_{m^\star}^{c'} - \Pi_{k_1} Q \Pi_m^{'} (\Pi_m Q \Pi_m^{'})^{-1} \Pi_m Q \Pi_{m^\star}^{c'} - \Pi_{k_1} Q \Pi_{m^\star}^{'} (\Pi_{m^\star} Q \Pi_{m^\star}^{'})^{-1} \Pi_{m^\star} Q \Pi_{m^\star}^{c'} \\ & + \Pi_{k_1} Q \Pi_m^{'} (\Pi_m Q \Pi_m^{'})^{-1} \Pi_m Q \Pi_{m^\star}^{'} (\Pi_{m^\star} Q \Pi_{m^\star}^{'})^{-1} \Pi_{m^\star} Q \Pi_{m^\star}^{c'}, \end{aligned}$$

$$\begin{aligned} C_{(m,m^\star)} := {} & \Pi_m^c Q \Pi_{m^\star}^{c'} - \Pi_m^c Q \Pi_m^{'} (\Pi_m Q \Pi_m^{'})^{-1} \Pi_m Q \Pi_{m^\star}^{c'} - \Pi_m^c Q \Pi_{m^\star}^{'} (\Pi_{m^\star} Q \Pi_{m^\star}^{'})^{-1} \Pi_{m^\star} Q \Pi_{m^\star}^{c'} \\ & + \Pi_m^c Q \Pi_m^{'} (\Pi_m Q \Pi_m^{'})^{-1} \Pi_m Q \Pi_{m^\star}^{'} (\Pi_{m^\star} Q \Pi_{m^\star}^{'})^{-1} \Pi_{m^\star} Q \Pi_{m^\star}^{c'}. \end{aligned}$$

This formulation indicates that the limiting behavior of the MMA weights depends on how regressors interact within the misspecified framework, highlighting the role of dependence among omitted and included variables. Consequently, when $C = +\infty$, the regularization term dominates, and the estimator asymptotically selects the simplest regression model. This suggests that, in extreme regularization cases, model averaging leads to highly parsimonious specifications, potentially sacrificing signal strength for robustness.

6.2 Limit Behavior of the Other Averaging Estimators

The proofs of Th. 3-5 of Zhang and Liu (2019) [38] directly imply that under **A.1**-**A.4**, and if $\phi_n = o_p(\sqrt{n})$, $\frac{1}{n}\mathcal{J}(\mathbf{w})$ converges weakly, and locally uniformly over Δ^{M-1}, modulo constants that do not affect optimization, to $\Lambda(\mathbf{w})$, while $\frac{1}{\phi_n}\mathcal{J}(\mathbf{w})$ is asymptotically non tight, due to the behavior of the MMA part of the JMA criterion (see Th. 3 of Zhang and Liu (2019) [38]). This then implies the following result:

Theorem 6. *Suppose that **A.1**-**A.4** hold in the global misspecification framework. Suppose also that $\phi_n = o_p(\sqrt{n})$. Then,*

$$\mathbf{w}_{\mathrm{JMA}} \rightsquigarrow \mathbf{w}_{\mathrm{JMA}}^{\infty} := \arg\min_{\mathbf{w}\in\Delta^{M-1}} \Lambda(\mathbf{w}).$$

If, furthermore, $\frac{\phi_n^\star}{n} \to C \in [0, +\infty]$, then,

$$\mathbf{w}_{\mathrm{AB}} \rightsquigarrow \mathbf{w}_{\mathrm{MMA}}^{\infty},$$

and

$$\mathbf{w}_{\ell^\infty} \rightsquigarrow \mathbf{w}_{\mathrm{MMA}}^{\infty}.$$

Furthermore, when $p < +\infty$,

$$\mathbf{w}_{\ell^p} \rightsquigarrow \mathbf{w}_{\ell^p}^{\infty} := \arg\min_{\mathbf{w}\in\Delta^{M-1}} \mathcal{R}(\mathbf{w})^{\frac{1}{p}},$$

where, $\mathcal{R}(\mathbf{w}) := (\Lambda(\mathbf{w})1_{C<+\infty} + \sigma_\star^2(C1_{0<C<+\infty} + 1_{C=+\infty})\mathbf{K}^{\star\prime}\mathbf{w})^p + (\Lambda(\mathbf{w})1_{C<+\infty})^p$. Finally,

$$\begin{aligned}\mathbf{w}_{\mathrm{Nash}} \rightsquigarrow \mathbf{w}_{\mathrm{Nash}}^{\infty} := \arg\min_{\mathbf{w}\in\Delta^{M-1}} &(\mathcal{M}_C(\mathbf{w}) - \mathcal{M}_C(\mathbf{w}_{\mathrm{JMA}}^{\infty}))(\Lambda(\mathbf{w}_{\mathrm{MMA}}^{\infty}) - \Lambda(\mathbf{w})) \\ &\text{s.t. } \mathcal{M}_C(\mathbf{w}) \le \mathcal{M}_C(\mathbf{w}_{\mathrm{JMA}}^{\infty}),\ \Lambda(\mathbf{w}_{\mathrm{MMA}}^{\infty}) \ge \Lambda(\mathbf{w}).\end{aligned}$$

Consequently, for $J = \mathrm{JMA},\ \mathrm{AB},\ \ell^p,\ \mathrm{Nash}$,

$$\beta_J \rightsquigarrow \sum_{m=1}^{K_2} \mathbf{w}_J^{\infty}(m)\Pi_m^{\star\prime}(\Pi_m^\star + (\Pi_m^\star Q \Pi_m^{\star\prime})^{-1}\Pi_m^\star Q \Pi_{K_1}'\Pi_{K_1})\beta. \tag{14}$$

The proof is given in the supplement [5]. The regularization constraints become asymptotically negligible for the JMA methodology since the restriction $\phi_n = o_p(\sqrt{n})$ is retained in order to maintain a unified statistical framework that is consistent with the case of correct specification.

The limiting behavior of the parameter $\phi_n^\star$ plays a crucial role in shaping the asymptotics of the multi-objective methodologies, as they asymptotically incorporate the regularized MMA criterion. When $C = 0$, the regularization term vanishes asymptotically, and the weights converge to $\mathbf{w}_{\mathrm{MMA}}^{\infty}$ across all multi-objective estimators considered in Theorem 6. This implies that, in such cases, model averaging prioritizes parsimony while effectively utilizing available information.

When $C = +\infty$, so that $\phi_n^\star$ diverges rapidly, the AB and ℓ^p methodologies remain asymptotically equivalent to the MMA, meaning that the model averaging procedure overwhelmingly selects the minimal simple regression model.

However, this extreme regularization behavior does not apply to JMA and Nash, which continue to incorporate information on the limiting covariance structure between regressors in the misspecified models. In particular, the JMA estimator remains asymptotically non-regularized. However, the Nash estimator achieves an asymptotically persistent regularization. This is due to Theorem 6 implies that, given that the estimator is asymptotically different from the limiting MMA and JMA weights, it asymptotically maximizes $\ln(\Lambda(\mathbf{w}_{\mathrm{MMA}}^{\infty})-\Lambda(\mathbf{w}))+\ln(\mathbf{K}^{\star'}(\mathbf{w}_{\mathrm{JMA}}^{\infty}-\mathbf{w}))$ under the inequality restrictions that appear in the aforementioned result. Extracting the constant term as a common factor from the first logarithm in the previous expression, considering a Taylor expansion for the logarithm, and noticing that Λ is strictly convex, we obtain that:

Corollary 1. *In the premises of Theorem 6, and if $C = +\infty$, then*

$$\mathbf{w}_{\mathrm{Nash}}^{\infty} := \arg\min_{\mathbf{w}\in\Delta^{M-1}} \Lambda(\mathbf{w}) - \Lambda(\mathbf{w}_{\mathrm{MMA}}^{\infty})\ln(\mathbf{K}^{\star'}(\mathbf{w}_{\mathrm{JMA}}^{\infty}-\mathbf{w})) + O(\frac{\Lambda^2(\mathbf{w})}{\Lambda(\mathbf{w}_{\mathrm{MMA}}^{\infty})})$$
$$\text{s.t. } \mathbf{K}^{\star'}(\mathbf{w}_{\mathrm{JMA}}^{\infty}-\mathbf{w}) \geq 0,\ \Lambda(\mathbf{w}_{\mathrm{MMA}}^{\infty}) > \Lambda(\mathbf{w}).$$

Hence, if the MMA regularization parameter diverges fast enough, the Nash methodology limiting criterion is approximately equal to Λ plus a regularization term that "punishes" solutions extremely close to the unique minimizers of Λ, i.e. the limiting JMA weights and advocates for solutions that promote parsimony.

For intermediate cases where $0 < C < +\infty$, the Nash estimator has a similar behavior, albeit via a more complicated limiting criterion. It thus continues to asymptotically achieve weight allocation that balances the trade off between parsimony and well specified regressors. Something analogous holds for the ℓ^p methodology for finite p. The resulting estimators also asymptotically place greater emphasis on the covariance structure between regressors, as captured by the presence of $\Lambda(\mathbf{w})$ in their limiting criteria. They do not entirely discard regularization, as JMA does, but rather seek a balance between model parsimony and maximizing the signal for the well-specified regressors. Notice though that for $p = 1$, the resulting ℓ_1 limiting criterion for the given C, would be equivalent to $\mathcal{M}_{C/2}$, suggesting that the resulting estimator would be asymptotically equivalent to an MMA estimator.

The asymptotic framework above provides strong motivation for the use of multi-objective optimization methodologies in model averaging. In particular, for values of C in the range $0 < C < +\infty$, the ℓ^p (for $1 < p < +\infty$) and Nash methodologies seem to offer a principled way to enhance the weight allocation towards well-specified regressors while still retaining elements of parsimony. This property remains valid for the Nash estimator even when $C = +\infty$.

These findings motivate the need for further exploration of fixed-sample discrepancies among estimators, which is pursued in the Monte Carlo analysis below.

7 Discussion

The limit theory for the case of correct specification allows for a direct extension to data dependent penalization parameters. The results hold unaltered whenever C is a well defined almost everywhere non-negative random variable that could attain extended values with positive probability. This seizes to be generally the case in the global misspecification framework; there it is not difficult to see that when C is a random element, the ℓ^p, for $p < +\infty$, and the Nash weights would have stochastic limits. Further investigation of data dependent regularization is also left for future research.

The misspecification results of Th.6 depend crucially on **A.3** that essentially regulates the asymptotic behavior of P_m. The $o_p(n^{-\frac{1}{2}})$ rate for its diagonal elements, implies locally uniform asymptotic negligibility for the non-MMA part of the JMA criterion apart from the regularization term. Hence, information on potential forms of conditional heteroskedasticity is lost by every averaging estimator considered here. This kind of information may be recoverable under other forms of **A.3**. For example, when as $n \to \infty$, if D_m, under some appropriate topology, converges to a tight random operator, then the scaling of the JMA criterion by $(n1_{C<+\infty} + \phi_n^{\star}1_{C=+\infty})^{-1}$ would produce asymptotic terms that would be associated with this limit, thus analogously affecting the limit theories of the multi-objective averaging procedures. Again, the investigation of such extensions is also delegated to further research.

8 Monte Carlo Experiments

In this section finite sample properties of the averaging estimators considered above are approximated through a Monte Carlo study.

The design follows closely some of the Monte Carlo experiments of Zhang and Liu (2019) [38]. Specifically the DGP has the linear form that appears in (1), where $K_1 = 2$, $K_2 = 8$, $M = 9$, the first column of $\mathbf{X}_{1,1}$, is constant, i.e. $\mathbf{X}_{1,1} = (1, 1, \ldots, 1)'$, for the i^{th} row vector consisting of the i^{th} elements of the remaining regressors, i.e. $(\mathbf{X}_{1,2}, \mathbf{X}_2)_i$, we have that it follows $N(\mathbf{0}_{9\times 1}, \Sigma_X)$, where Σ_X is a 9×9 matrix with diagonal elements equal to 0.7, and off-diagonal elements equal to 0.7^2. Those regressors' row vectors are independent across $i = 1, \ldots, 9$. ε_i has the martingale difference form of $u_i\sigma_i$, with the u_is being iid and independent of the regressors. (A) in the homoskedastic case, $u_i \sim N(0, 1)$ and $\sigma_i = 2.5$ identically over $i = 1, \ldots, n$, while (b) in the heteroskedastic case, $u_i \sim t_4$ and $\sigma_i = {}^{(1+2|\mathbf{X}_{1,2}(i)|+4|\mathbf{X}_{2,8}(i)|)}/_{3}$, $\forall i = 1, \ldots, n$. Assumptions **A.1-A.4** are trivially satisfied in this setup. The following cases for the population regression coefficients are considered:

C.1 $\beta = (1, 1, 0.5, 0.5^2, 0.5^3, 0.5^4, 0, 0, 0, 0)'$,

C.2 $\beta = (1, 1, 0.5^4, 0.5^3, 0.5^2, 0.5, 0, 0, 0, 0)'$, and,

C.3 $\beta = (1, 1, 0.5, 0.5^2, 0, 0, 0.5^3, 0.5^4, 0, 0)'$.

As far as the analysts' choice of the regressors' matrix is concerned, two cases of specification are considered. In the first case of "correct specification" the regressors' matrix used is the full matrix of regressors in the DGP, i.e. $\mathbf{X}$. Thus, M_0-that in this case represents the number of regressors in the minimal correctly specified model-equals 4 in **C.1**-**C.2**, and equals 6 in **C.3**. In the second case of "misspecification", the analyst uses as a regressors' matrix $\mathbf{X}$ without the second and third core regressors. The analyst erroneously considers as core regressors $\mathbf{X}_{1,1}$ and $\mathbf{X}_{4,1}$, and 5 auxiliary regressors emerging from $\mathbf{X}_2$ by deleting its' first and second collumns. Thus in this case, the number of statistical models considered $M = 8$, and every one of them is misspecified, i.e. $M_0 = 8$ for **C.1**, **C.2** and **C.3**.

The sample size, n, is set equal to 100 and 400. The multiplier coefficients that appear in the definitions of MMA and JMA (i.e. $\phi_n^\star$ and ϕ_n respectively) are set equal to $0.01 \times n$ and, $0.05 \times n^{1/3} \times \ln(n)$ respectively, whereas ${}^{\phi_n^\star}/_{n} \to 0.01$, ${}^{\phi_n}/_{\sqrt{n}} \to 0$, and ${}^{\phi_n^\star}/_{\phi_n} \to \infty$. Those choices correspond to convergence to non-stochastic weights and asymptotic selection of the narrowest well-specified model under correct specification for all estimators. Under global misspecification those choices are relevant to the limiting choice of the narrowest regression model for all the considered estimators except for the JMA and Nash.

The numerical evaluation of the MMA and JMA estimators is performed on simple modifications of the Liu (2015) [27] freely available Matlab code that among others involves optimization solvers for quadratic programming. The evaluation of the multi-objective estimators is also performed in Matlab using the fmincon solver for non-linear (interior point or convex) programming. For the sample sizes involved in the experiments and the empirical applications, the cumulative time spent on optimization for all the estimators involved using computers with five-core chip-sets does not exceed 3-5 seconds.

The number of Monte Carlo replications is set equal to 500. The Frobenius norms of the Monte Carlo variance, MSE, MAE and bias are reported for the simple averaging (SimAve), MMA, JMA, ℓ^1, ℓ^2, ℓ^∞, Nash, and AB averaging estimators. In both cases the Monte Carlo mean of the squared Euclidean norm of the weights as well as the Monte Carlo rounded mean of the first two models at which the weights are maximally concentrated for all the methodologies above are also reported. This in order to obtain a sense of the finite sample analogy of the asymptotic concentration of the averaging estimators as reported by the limit theories of the previous sections at least in the case of correct specification, as well as the models at which they maximally concentrate on average.

This information is reported in Tables 1-4 as well as in Table 1 in the supplement [5]. Specifically, Tables 1-2 provide information on the Monte Carlo variance-bias trade-off and the MSE-MAE divergences from the DGP value in the case of correct specification. Tables 3-4 deal with the analogous information regarding misspecification. Finally, Table 1 in the supplement [5] provides the aforementioned information regarding the weights. Specifically, for each averaging estimator it presents the Monte Carlo mean of the sum of squares of its' weights, as well as a vector of two integers. The vector's first component represents the rounded Monte Carlo mean of the statistical model at which the maximum weight is attributed, and the second component the rounded Monte Carlo mean at which the second maximum weight is attributed. Occasional coincidence between the components is due to rounding. Moreover, the number 1 corresponds to the narrow model, i.e. the one that contains only what the analyst considers as core regressors, and the number $m > 1$ corresponds to the model that besides the core contains also the first $m - 1$ regressors from the regressors' matrix. $m \leq s$, where $s = 9$ in the case of "correct specification" and $s = 8$ in the case of misspecification.

The simulation results seem to point that in every case there exists at least one of the proposed multi-objective model averaging estimators that dominate the basis MMA and JMA in terms of the MSE/MAE Frobenius norm. In most of the cases this is the Nash estimator, except for the $n = 400$ case of misspecification, where the ℓ^∞ estimator, and in some instances there the AB estimator appear to dominate the others. The ℓ^1 and ℓ^2 estimators seem to usually lie between the MMA and JMA estimators in terms of the MSE/MAE Frobenius norm performance, except for some instances in the correct specification case of $n = 400$ where every multi-objective estimator seem to be uniformly better in these terms. Notably in several cases the ℓ^∞ estimator seem to fall very close to the JMA estimator.

In terms of concentration the results show that the JMA typically is comparatively the one with maximal concentration, for all the "large sample" cases of $n = 400$, while the Nash estimator replaces it for all the "small sample" cases of $n = 100$. For $n = 100$ all estimators seem to have similar behavior in terms of the average choice of models. For $n = 400$ that is no longer the case with the Nash estimator usually on average favouring narrow models, in contrast to the ℓ^1 and the AB estimators that seem to systematically favor larger models in both the correct and incorrect specification cases. [2]

Overall, the multi-objective model averaging criteria lead to significant small to medium size sample improvements and may prove useful in applied work for the sample sizes that are typically found in empirical applications, such as the empirical growth one below.

9 Empirical Application

This section provides an empirical application of the proposed methodologies alongside the standard model averaging methods. One of the main areas in which the model averaging methods is the cross-section growth regressions (see, e.g., Steel, (2020) [36]). As extensive numbers of explanatory variables are needed to explain the growth differences across countries, model averaging provides a valuable and reliable attempt to provide a selection and combination of models with different numbers of explanatory variables. Therefore, to overcome (or limit) the model

[2] The results depend among others on the choice of the degrees of freedom parameter in the student's t parameterization. Auxiliary results that are not reported here-yet are available upon request- suggest that a choice of degrees of freedom equal to 2, that would be contrary to a large part of the assumption framework, show in most cases relatively poor performance for the MMA, compared to the JMA in terms of the MSE/MAE criteria and concentration. The performance of the multi-objective estimators seems to fall between those two cases.

uncertainty, the existing literature has been using model averaging methods (see e.g., Fernandez et al., (2001) [15]; Sala-i Martin et al., (2004) [34]; Durlauf et al., (2008) [14]; Magnus et al., (2010) [28]; Amini and Parmeter, (2012) [2]; Liu, (2015) [27]; Gunby et al., (2017) [18]; Arin et al., (2019) [4]; Cazachevici et al., (2020) [10], among many others). In this section, we use the following standard model averaging methods: the Mallows model averaging (MMA; Hansen, (2007) [19]), the jackknife model averaging (JMA; Hansen and Racine, (2012) [20]); and estimators proposed in this paper: ℓ^1, ℓ^2, ℓ^∞, Nash and AB. To provide a comparison of different growth regression models, we use the same data set of Magnus et al. (2010) [15] and Liu (2015) [27], and the following cross-section growth regression model is used:

$$\mathbf{growth} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon, \tag{15}$$

where growth is average growth rate of gross domestic product (GDP) per capita between 1960 and 1996. $\mathbf{X}_1$ represents the core regressors used in the classical growth theory. In the application, we use different numbers of core regressors to provide empirical evidence based on different core regressors. Five core regressors include i) the logarithm of GDP per capita in 1960 (GDP60); ii) the share of the equipment investment as a share of the GDP between 1960 and 1985 (INV); iii) the primary school enrolment rate in 1960 (SCHOOL60); iv) the life expectancy at birth in 1960 (LIFE60); and v) the population growth rate between 1960 and 1990 (POP). Finally, a set of auxiliary regressors, $\mathbf{X}_2$, is included: i) the rule of law index (RULE) as a proxy for institutional quality; ii) the proportion of a country's land area within geographical tropics (TROPICS); iii) Average of five different indices of ethnolinguistic fragmentation (ETHNO); and iv) fraction of Confucian population in 1970 and 1980 (CONFUC). A detailed description of the data set could be obtained from Magnus et al. (2010), and the number of countries used in the analysis is 74. For comparison of different model averaging methods, we use different number of core regressors, leading to a set of model setups: i) Model A with one core regressor (GDP60); ii) Model B with two core regressors (GDP60 and INV); iii) Model C with three core regressors (GDP, INV and SCHOOL60); iv) Model D with four core regressors (GDP, INV, SCHOOL60, and LIFE60); and v) Model E with five core regressors (GDP, INV, SCHOOL60, LIFE60 and POP). All of the models also include a constant term.

Our parameter of interest is the log GDP per capita coefficient in 1960 to examine the beta convergence. Our analyses for models A to E consider the core regressors and include each auxiliary regressor one at a time to the model specifications. The penalization parameters are both chosen equal to $10^{-3} \times \sqrt{\ln(n)}$ in order to avoid maximal concentration due to large penalties. The estimation results for Models A, B, are reported Tables 5-6. The ones for C, D, and E are reported in Tables 2-4 in the supplement [5]. Table 7 provides the weights assigned to the coefficients of models A, B (respectively Table 5 in the supplement [5] provides the weights assigned to the coefficients of models C, D and E). Table 8 as well as Table 6 in the supplement [5] list the regressors included in each sub-model in consideration foe each respective scenario.

Overall, with scenarios A-E, in line with the Monte Carlo simulations, JMA assigns full weight to the narrow model (i.e., model that contains core regressors) and MMA allocates full weight to the model that contains all the core and auxiliary regressors. On the other hand, our model averaging methods assign relatively more weights towards the model with full set of regressors (i.e., full model) but also assigns some positive weights to the majority of the rest of the models. Our methods allocate more weights towards to models with all regressors including extended Solow growth regressors (Solow, (1956) [35]; Mankiw et al., (1992) [29]) and auxiliary regressors highlighting the importance of the geography (Diamond, (1997) [13]; Gallup et al., (1999) [33]; Sachs, (2003) [33]) and institutional quality (Acemoglu et al., (2001) [1]; Rodrik et al., (2004) [32]). With respect to the initial GDP per capita coefficients, they tend to be closer to each other.

10 Conclusion

Within a linear model background, we consider averaging methodologies that extend the analysis of both the generalized Jacknife Model Averaging (JMA) and the Mallows Model Averaging (MMA) criteria in a multi-objective setting within the context of a stochastic dominance perspective. We also consider averaging estimators that emerge from the minimization of several scalarizations of the vector criterion consisting of both the MMA and the JMA criteria as well as an estimator that can be represented as a Nash bargaining solution between the competing scalar criteria. We derive the limit theory of the estimators under both a correct specification and a global misspecification framework and our Monte Carlo experiments suggest that the some of thee averaging estimators proposed here seem to systematically provide with MSE/MAE reductions in both the correctly specified and the misspecification scenarios. An empirical application using data from growth theory suggests that our model averaging methods assign relatively higher weights towards the traditional Solow type growth variables, yet they do not seem to exclude regressors that underpin the importance of geography or institutions.

For future research, we would like to extend our framework to some nonlinear settings, such as threshold regression and kink regression models, where the analysis would allow for possible discontinuities and/or kinks in the regression function, something that our current analysis has not considered.

Furthermore, methodologically, the multi-objective optimization framework can be readily extended to include further basis averaging estimators, like the focused information criterion (FIC) with a view towards local misspecification; see Claeskens and Hjort, (2003) [11], and/or modifications of the MMA/JMA procedures so as to incorporate sparsity restrictions in diverging number of regressors frameworks; see Liao et al. (2021) [26]. A general theory of what properties of the basis estimators are retained and/or combined via scalarization methodologies with a view towards the optimal selection of penalization for the basis estimators and scalarization seems like a fascinating issue for further research.

References

[1] Acemoglu, D., Johnson, S., Robinson, J.A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review, 91(5), 1369-1401.

[2] Amini, S.M., & Parmeter, C.F. (2012). Comparison of model averaging techniques: assessing growth determinants. Journal of Applied Econometrics, 27(5), 870-876.

[3] Andrews, D.W.K. (1991). Asymptotic optimality of generalized CL, cross-validation, and generalized cross-validation in regression with heteroskedastic errors. Journal of Econometrics 47, 359–377.

[4] Arin, K.P., Braunfels, E., & Doppelhofer, G. (2019). Revisiting the growth effects of fiscal policy: A Bayesian model averaging approach. Journal of Macroeconomics, 62, 103158.

[5] Arvanitis, S., Pinar, M., Stengos T., & Topaloglou, N., (2024). Supplement for: Multi-Objective Frequentist Model Averaging with an Application to Economic Growth.

[6] Arvanitis, S., Post, T., & Topaloglou, N., (2021). Stochastic Bounds for reference sets in portfolio analysis. Management Science, 67(12), 7737-7754.

[7] Arvanitis, S., Scaillet, O. and Topaloglou, N., (2023). Sparse spanning portfolios and under-diversification with second-order stochastic dominance. mimeo.

[8] Aumann, R. J., & Hart, S. (Eds.). (1992). Handbook of game theory with economic applications (Vol. 2). Elsevier.

[9] Austern, M., & Orbanz, P. (2022). Limit theorems for distributions invariant under groups of transformations. The Annals of Statistics, 50(4), 1960-1991.

[10] Cazachevici, A., Havranek, T., & Horvath, R. (2020). Remittances and economic growth: A meta-analysis. World Development, 134, 105021.

[11] Claeskens, G. and Hjort, N. L., (2003 a). Frequentist model average estimators. Journal of the American Statistical Association, 98(464), 879-899.

[12] Claeskens, G., Hjort, N.L. (2003 b). The focused information criterion. Journal of the American Statistical Association, 98(464), 900-916.

[13] Diamond, J.M. (1997). Guns, germs and steel: The fate of human societies. New York: W.W. Norton & Co.

[14] Durlauf, S., Kourtellos, A., &Tan, C. (2008). Are any growth theories robust? Economic Journal, 118, 329-346.

[15] Fernandez, C., Ley, E., & Steel, M. (2001). Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics, 16, 563-576.

[16] Gallup, J.K., Sachs, J.D., Mellinger, A.D. (1999). Geography and Economic Development. International Regional Science Review, 22(2), 179-232.

[17] Giné, E., & Nickl, R. (2021). Mathematical foundations of infinite-dimensional statistical models. Cambridge university press.

[18] Gunby, P., Jin, Y., & Reed, W.R., 2017. Did FDI Really Cause Chinese Economic Growth? A Meta-Analysis. World Development, 90, 242-255.

[19] Hansen, B.E. (2007). Least squares model averaging. Econometrica 75, 1175–1189.

[20] Hansen, B.E., & Racine, J. (2012). Jackknife model averaging. Journal of Econometrics, 167, 38-46.

[21] Hwang, C. L., & Masud, A. S. M. (2012). Multiple objective decision making—methods and applications: a state-of-the-art survey (Vol. 164). Springer Science & Business Media.

[22] Knight, K., (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. mimeo.

[23] Kourtellos, A., Stengos, T. and Tan, C.M., (2010). Do institutions rule? The role of heterogeneity in the institutions vs. geography debate. Economics Bulletin, 30(3), pp.1710-1719.

[24] Leamer, E.E., (1983). Model choice and specification analysis. Handbook of econometrics, 1, pp.285-330.

[25] Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regression. American Economic Review,82, 942-963.

[26] Liao, J., Zou, G., Gao, Y., & Zhang, X. (2021). Model averaging prediction for time series models with a diverging number of parameters. Journal of Econometrics, 223(1), 190-221.

[27] Liu, C.A. (2015). Distribution theory of the least squares averaging estimator. Journal of Econometrics, 186, 142-159.

[28] Magnus, J., Powell, O., & Prufer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. Journal of Econometrics, 154, 139-153.

[29] Mankiw, N.G., Romer, D., Weil, D.N. (1992). A Contribution to the Empirics of Economic Growth. The Quarterly Journal of Economics, 107(2), 407-437.

[30] Molchanov, I., (2006).Theory of random sets, Springer Science & Business Media.

[31] Racine, J. (1997). Feasible cross-validatory model selection for general stationary processes. Journal of Applied Econometrics, 12(2), 169-179.

[32] Rodrik, D., Subramanian, A., Trebbi, F. (2004). Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. Journal of Economic Growth, 9, 131–165.

[33] Sachs, J.D. (2003). Institutions don't rule: direct effects of geography on per capita income. National Bureau of Economic Research Working Paper No. 9490.

[34] Sala-i Martin, X., Doppelhofer, G., & Miller, R. (2004). Determinants of long-term growth: a Bayesian Averaging of Classical Estimates (BACE) approach. American Economic Review, 94, 813-835.

[35] Solow, R.M. (1956). A Contribution to the Theory of Economic Growth. The Quarterly Journal of Economics, 70(1), 65-94.

[36] Steel, M.F.J. (2020). Model Averaging and Its Use in Economics. Journal of Economic Literature, 58(3), 644-719.

[37] van der Vaart, A. W., & Wellner, J. A. (1996). Weak convergence and empirical processes with applications to statistics. Springer-Verlag New York.

[38] Zhang, X., & Liu, C. A. (2019). Inference after model averaging in linear regression models. Econometric Theory, 35(4), 816-841.

Appendix-Monte Carlo Results

Table 1: Simulation results in three cases for n = 100: Correct Specification.

		Homoskedastic Setup				Heteroskedastic Setup			
	Method	Var	MSE	MAE	Bias	Var	MSE	MAE	Bias
Case 1	SimAver	9.6245	10.5711	3.8826	0.2296	8.4327	9.2018	3.6625	0.1634
	MMA	6.3838	7.1683	2.6812	0.4079	5.3733	6.1149	2.4777	0.3971
	JMA	5.8764	6.6773	2.5861	0.4212	4.9358	5.7230	2.3965	0.4139
	ℓ^1	6.0049	6.8008	2.6121	0.4208	5.0849	5.8513	2.4236	0.4094
	ℓ^2	5.9949	6.7914	2.6101	0.4211	5.0778	5.8452	2.4224	0.4097
	ℓ^∞	5.8764	6.6773	2.5861	0.4212	4.9357	5.7230	2.3964	0.4139
	Nash	5.6997	6.4425	2.5339	0.3884	4.8765	5.5929	2.3566	0.3768
	AB	6.0091	6.8030	2.6140	0.4195	5.0857	5.8513	2.4244	0.4083
Case 2	SimAver	8.0651	8.8856	3.5948	0.2760	9.0979	9.9655	3.7859	0.2463
	MMA	5.3311	6.1383	2.4998	0.4400	5.8847	6.6502	2.6205	0.4291
	JMA	4.9105	5.6983	2.3919	0.4236	5.4273	6.1773	2.5254	0.4283
	ℓ^1	5.0699	5.8692	2.4287	0.4323	5.5338	6.2902	2.5488	0.4334
	ℓ^2	5.0622	5.8614	2.4268	0.4321	5.5250	6.2812	2.5470	0.4335
	ℓ^∞	4.9111	5.6988	2.3919	0.4234	5.4273	6.1773	2.5254	0.4283
	Nash	4.8454	5.6033	2.3700	0.4206	5.3117	6.0391	2.4793	0.4098
	AB	5.0592	5.8574	2.4297	0.4328	5.5473	6.3047	2.5532	0.4332
Case 3	SimAver	9.4367	10.3282	3.7791	0.1591	9.3498	10.2862	3.7232	0.0867
	MMA	6.3465	7.2173	2.5627	0.3500	5.7249	6.4618	2.4956	0.3342
	JMA	5.5417	6.4053	2.4492	0.3295	5.0762	5.7648	2.4163	0.3186
	ℓ^1	5.7734	6.6344	2.4894	0.3390	5.2626	5.9643	2.4334	0.3275
	ℓ^2	5.7550	6.6162	2.4868	0.3387	5.2532	5.9550	2.4320	0.3275
	ℓ^∞	5.5416	6.4053	2.4492	0.3295	5.0762	5.7648	2.4163	0.3186
	Nash	5.5854	6.4242	2.4220	0.3244	5.2025	5.9062	2.3784	0.3125
	AB	5.7782	6.6397	2.4904	0.3389	5.2713	5.9754	2.4360	0.3278

Entries report the Frobenius norm of the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value, in the case of correct specification, for all averaging estimators considered in the text along with the simple averaging (equal weights) averaging estimator. $n = 100$ and all three cases for true parameter values of the auxiliary regressors are considered, in both the homoskedastic and the heteroskedastic scenarios for the regression errors.

Table 2: Simulation results in three cases for n = 400: Correct Specification.

		Homoskedastic Setup				Heteroskedastic Setup			
	Method	Var	MSE	MAE	Bias	Var	MSE	MAE	Bias
Case 1	SimAver	1.9722	2.1508	1.8049	0.0767	1.9075	2.1132	1.7621	0.1140
	MMA	0.8919	1.3982	1.1842	0.5693	0.8406	1.3508	1.1603	0.5703
	JMA	1.3058	1.5216	1.3227	0.2720	1.2052	1.4799	1.2864	0.3145
	ℓ^1	0.9326	1.3419	1.1750	0.4921	0.8729	1.3237	1.1530	0.5153
	ℓ^2	0.9325	1.3414	1.1748	0.4918	0.8731	1.3234	1.1529	0.5149
	ℓ^∞	0.9917	1.2728	1.1568	0.3619	0.9776	1.3263	1.1583	0.3982
	Nash	0.8558	1.3861	1.1736	0.5850	0.8088	1.3210	1.1468	0.5736
	AB	0.9436	1.3079	1.1620	0.4528	0.8882	1.3001	1.1387	0.4774
Case 2	SimAver	2.0053	2.2162	1.8260	0.2543	2.1856	2.4139	1.8951	0.2075
	MMA	0.9404	1.5316	1.2229	0.6430	0.9418	1.6075	1.2315	0.7047
	JMA	1.3579	1.6439	1.3776	0.3822	1.4566	1.7712	1.4326	0.4196
	ℓ^1	0.9952	1.5138	1.2261	0.5898	0.9993	1.5682	1.2412	0.6399
	ℓ^2	0.9951	1.5134	1.2259	0.5897	0.9990	1.5677	1.2411	0.6398
	ℓ^∞	1.0378	1.4426	1.2218	0.5065	1.0699	1.5257	1.2562	0.5531
	Nash	0.8950	1.4876	1.1989	0.6408	0.8891	1.5579	1.2106	0.7031
	AB	1.0052	1.4793	1.2176	0.5569	1.0143	1.5418	1.2413	0.6103
Case 3	SimAver	2.1177	2.3057	1.8580	0.1038	2.0495	2.2482	1.8414	0.1245
	MMA	0.9337	1.3984	1.1842	0.5101	0.9120	1.4644	1.2156	0.6061
	JMA	1.4260	1.6544	1.3500	0.2302	1.4251	1.7045	1.4140	0.3281
	ℓ^1	1.0076	1.3970	1.1836	0.4327	0.9929	1.4679	1.2248	0.5346
	ℓ^2	1.0076	1.3969	1.1836	0.4325	0.9927	1.4674	1.2246	0.5343
	ℓ^∞	1.0899	1.3881	1.1829	0.3274	1.0586	1.4332	1.2292	0.4333
	Nash	0.8646	1.3456	1.1659	0.5325	0.8478	1.4064	1.1913	0.6147
	AB	1.0281	1.3885	1.1743	0.3989	1.0073	1.4486	1.2180	0.5029

Entries report the Frobenius norm of the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value, in the case of correct specification, for all averaging estimators considered in the text along with the simple averaging (equal weights) averaging estimator. $n = 400$ and all three cases for true parameter values of the auxiliary regressors are considered, in both the homoskedastic and the heteroskedastic scenarios for the regression errors.

Table 3: Simulation results in three cases for n = 100: Misspesification.

		Homoskedastic Setup				Heteroskedastic Setup			
	Method	Var	MSE	MAE	Bias	Var	MSE	MAE	Bias
Case 1	SimAver	7.5611	8.4754	3.3566	0.5993	7.4849	8.3189	3.2963	0.4193
	MMA	5.9170	6.9356	2.5603	0.8646	5.7419	6.5929	2.4421	0.7011
	JMA	5.2458	6.2397	2.4665	0.8589	5.3636	6.2161	2.3828	0.6951
	ℓ^1	5.4284	6.4258	2.4982	0.8669	5.4947	6.3530	2.4027	0.7044
	ℓ^2	5.4098	6.4068	2.4959	0.8665	5.4873	6.3458	2.4016	0.7046
	ℓ^∞	5.2458	6.2397	2.4665	0.8589	5.3637	6.2161	2.3828	0.6951
	Nash	5.1935	6.1595	2.4297	0.8466	5.3439	6.1828	2.3545	0.7034
	AB	5.4322	6.4282	2.5003	0.8653	5.4959	6.3532	2.4037	0.7034
Case 2	SimAver	7.3529	8.1112	3.2880	0.3343	8.1189	9.0166	3.4136	0.2919
	MMA	5.9101	6.6535	2.6144	0.5846	6.3354	7.1587	2.6532	0.5405
	JMA	5.5780	6.3126	2.5552	0.5931	5.9138	6.7213	2.6014	0.5598
	ℓ^1	5.6848	6.4265	2.5754	0.5915	6.0379	6.8474	2.6109	0.5547
	ℓ^2	5.6792	6.4208	2.5742	0.5917	6.0264	6.8350	2.6095	0.5551
	ℓ^∞	5.5780	6.3127	2.5552	0.5931	5.9139	6.7213	2.6014	0.5598
	Nash	5.4043	6.1309	2.5095	0.5946	5.7836	6.5569	2.5356	0.5381
	AB	5.6875	6.4278	2.5760	0.5899	6.0426	6.8520	2.6130	0.5539
Case 3	SimAver	7.3822	8.1646	3.3097	0.5541	7.4608	8.3906	3.2666	0.3996
	MMA	5.5613	6.4352	2.4465	0.7826	5.7883	6.7295	2.4496	0.6600
	JMA	5.3043	6.1736	2.4069	0.7718	5.6404	6.5774	2.4305	0.6510
	ℓ^1	5.3820	6.2561	2.4166	0.7786	5.6755	6.6217	2.4287	0.6604
	ℓ^2	5.3769	6.2511	2.4158	0.7786	5.6723	6.6187	2.4282	0.6606
	ℓ^∞	5.3043	6.1736	2.4069	0.7718	5.6404	6.5774	2.4305	0.6509
	Nash	5.1461	6.0027	2.3576	0.7796	5.3865	6.3039	2.3619	0.6697
	AB	5.3805	6.2541	2.4183	0.7783	5.6744	6.6195	2.4296	0.6591

Entries report the Frobenius norm of the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value, if the 2nd and 3rd core regressors are dropped from analysis, for all averaging estimators considered in the text along with the simple averaging (equal weights) averaging estimator. $n = 100$ and all three cases for true parameter values of the auxiliary regressors are considered, in both the homoskedastic and the heteroskedastic scenarios for the regression errors.

Table 4: Simulation results in three cases for n = 400: Misspesification.

		Homoskedastic Setup				Heteroskedastic Setup			
	Method	Var	MSE	MAE	Bias	Var	MSE	MAE	Bias
Case 1	SimAver	1.7871	2.0373	1.7149	0.3821	1.9078	2.1830	1.7388	0.3966
	MMA	0.9295	2.0301	1.3575	1.0270	0.9340	2.1060	1.3910	1.0715
	JMA	1.4456	1.8243	1.3989	0.6011	1.6015	1.9969	1.4388	0.6174
	ℓ^1	1.0171	1.8922	1.3280	0.9228	1.0777	1.9902	1.3609	0.9512
	ℓ^2	1.0171	1.8913	1.3278	0.9223	1.0777	1.9891	1.3605	0.9507
	ℓ^∞	1.1231	1.7100	1.3065	0.7622	1.1905	1.8216	1.3267	0.7954
	Nash	0.8853	2.0568	1.3637	1.0592	0.8729	2.1523	1.3954	1.1179
	AB	1.0335	1.8131	1.3075	0.8737	1.1035	1.9190	1.3396	0.9012
Case 2	SimAver	1.9309	2.2179	1.7051	0.3983	1.7713	2.0426	1.6799	0.4266
	MMA	1.0147	2.4971	1.5367	1.1407	0.9737	2.4913	1.5427	1.1560
	JMA	1.7385	2.1896	1.5462	0.6147	1.6153	2.1105	1.5464	0.6699
	ℓ^1	1.1742	2.2892	1.4866	0.9957	1.1093	2.2759	1.4997	1.0214
	ℓ^2	1.1740	2.2880	1.4863	0.9954	1.1092	2.2749	1.4994	1.0211
	ℓ^∞	1.2481	2.0034	1.4156	0.8396	1.1560	1.9459	1.4203	0.8644
	Nash	0.9386	2.5669	1.5463	1.1914	0.8887	2.5706	1.5596	1.2134
	AB	1.1993	2.1805	1.4558	0.9386	1.1316	2.1670	1.4667	0.9695
Case 3	SimAver	1.7823	2.1212	1.6987	0.5099	1.7312	2.0563	1.6958	0.5223
	MMA	0.9680	2.2470	1.4102	1.1309	0.8299	2.0617	1.3612	1.1115
	JMA	1.4185	1.9565	1.4020	0.7491	1.3877	1.8622	1.4010	0.7230
	ℓ^1	1.0426	2.1238	1.3733	1.0462	0.8995	1.9076	1.3198	1.0174
	ℓ^2	1.0425	2.1228	1.3731	1.0458	0.8996	1.9071	1.3197	1.0171
	ℓ^∞	1.1580	1.9130	1.3421	0.8833	1.0172	1.7563	1.3068	0.8811
	Nash	0.9280	2.2725	1.4195	1.1586	0.7958	2.0819	1.3733	1.1351
	AB	1.0506	2.0368	1.3530	1.0022	0.9182	1.8321	1.3032	0.9724

Entries report the Frobenius norm of the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value, if the 2nd and 3rd core regressors are dropped from analysis, for all averaging estimators considered in the text along with the simple averaging (equal weights) averaging estimator. $n = 400$ and all three cases for true parameter values of the auxiliary regressors are considered, in both the homoskedastic and the heteroskedastic scenarios for the regression errors.

Appendix-Empirical Application Results

Table 5: Coefficient estimates with Model A scenario

	SimAver	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
CONSTANT	0.0489	0.0609	0.0018	0.0596	0.0585	0.0573	0.0568	0.0595
	(0.0154)	(0.0193)	(0.0115)	(0.0183)	(0.0176)	(0.0175)	(0.0173)	(0.0181)
GDP60	-0.0123	-0.0155	0.0014	-0.0153	-0.0149	-0.0146	-0.0144	-0.0152
	(0.0022)	(0.003)	(0.0014)	(0.0028)	(0.0027)	(0.0026)	(0.0026)	(0.0027)
INV	0.1942	0.1368	0.1686	0.1534	0.1605	0.1681	0.1718	0.1531
	(0.0312)	(0.0399)	(0.015)	(0.0386)	(0.0370)	(0.0373)	(0.0370)	(0.0380)
SCHOOL60	0.0175	0.017		0.018	0.0178	0.0182	0.0182	0.0177
	(0.0059)	(0.0085)		(0.0079)	(0.0076)	(0.0074)	(0.0074)	(0.0079)
LIFE60	0.0006	0.0008		0.0008	0.0008	0.0007	0.0007	0.0008
	(0.0002)	(0.0003)		(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
POP	0.1486	0.3460		0.3109	0.2783	0.2722	0.2632	0.3014
	(0.0973)	(0.1908)		(0.1708)	(0.1569)	(0.1529)	(0.1482)	(0.1678)
LAW	0.009	0.0173		0.0165	0.0148	0.0150	0.0145	0.0157
	(0.0024)	(0.0058)		(0.0051)	(0.0045)	(0.0046)	(0.0045)	(0.0049)
TROPICS	-0.0028	-0.0075		-0.0065	-0.0058	-0.0055	-0.0053	-0.0063
	(0.0012)	(0.0036)		(0.003)	(0.0027)	(0.0026)	(0.0025)	(0.0030)
ETHNO	-0.0021	-0.0077		-0.0057	-0.0055	-0.0048	-0.0047	-0.0061
	(0.0015)	(0.0066)		(0.0049)	(0.0045)	(0.0041)	(0.0040)	(0.0051)
CONFUC	0.0062	0.0561		0.0407	0.0339	0.0344	0.0340	0.0398
	(0.0014)	(0.0128)		(0.0093)	(0.0078)	(0.0078)	(0.0078)	(0.0091)

Note: Standard errors are reported in parentheses.

Table 6: Coefficient estimates with Model B scenario

	SimAver	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
CONSTANT	0.0573	0.0609	0.0242	0.0599	0.0604	0.0579	0.0574	0.0608
	(0.0161)	(0.0193)	(0.0117)	(0.0183)	(0.0178)	(0.0175)	(0.0173)	(0.0182)
GDP60	-0.0145	-0.0155	-0.0026	-0.0153	-0.0154	-0.0147	-0.0145	-0.0155
	(0.0024)	(0.0030)	(0.0015)	(0.0028)	(0.0027)	(0.0026)	(0.0026)	(0.0028)
INV	0.2184	0.1369	0.360	0.1565	0.1679	0.1734	0.177	0.1588
	(0.0351)	(0.04)	(0.032)	(0.0388)	(0.0379)	(0.0377)	(0.0374)	(0.0385)
SCHOOL60	0.0197	0.017		0.018	0.0183	0.0182	0.0183	0.0182
	(0.0066)	(0.0085)		(0.0079)	(0.0078)	(0.0074)	(0.0074)	(0.008)
LIFE60	0.0007	0.0008		0.0008	0.0008	0.0007	0.0007	0.0008
	(0.0002)	(0.0003)		(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
POP	0.1672	0.346		0.3109	0.2816	0.2723	0.2633	0.3021
	(0.1094)	(0.1908)		(0.1708)	(0.1588)	(0.1529)	(0.1482)	(0.1684)
LAW	0.0102	0.0174		0.0164	0.015	0.015	0.0145	0.0158
	(0.0027)	(0.0058)		(0.0051)	(0.0046)	(0.0046)	(0.0045)	(0.0049)
TROPICS	-0.0032	-0.0075		-0.0065	-0.0058	-0.0055	-0.0053	-0.0063
	(0.0013)	(0.0036)		(0.003)	(0.0027)	(0.0026)	(0.0025)	(0.003)
ETHNO	-0.0023	-0.0077		-0.0057	-0.0056	-0.0048	-0.0047	-0.0061
	(0.0017)	(0.0066)		(0.0049)	(0.0046)	(0.0041)	(0.004)	(0.0051)
CONFUC	0.007	0.0561		0.0407	0.034	0.0344	0.034	0.0394
	(0.0016)	(0.0128)		(0.0093)	(0.0078)	(0.0079)	(0.0078)	(0.009)

Note: Standard errors are reported in parentheses.

Table 7: Weights on each model in different scenarios

Panel A. Scenario A							
Model	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
1	0.00	0.53	0.01	0.03	0.02	0.01	0.02
2	0.00	0.47	0.03	0.04	0.06	0.07	0.02
3	0.00	0.00	0.03	0.04	0.07	0.08	0.03
4	0.00	0.00	0.02	0.05	0.04	0.04	0.03
5	0.00	0.00	0.00	0.04	0.00	0.00	0.03
6	0.00	0.00	0.07	0.06	0.11	0.11	0.04
7	0.00	0.00	0.11	0.06	0.09	0.08	0.06
8	0.00	0.00	0.01	0.08	0.00	0.00	0.06
9	1.00	0.00	0.72	0.60	0.61	0.61	0.71
Panel B. Scenario B							
Model	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
1	0.00	1.00	0.04	0.05	0.08	0.08	0.03
2	0.00	0.00	0.03	0.05	0.07	0.08	0.03
3	0.00	0.00	0.02	0.05	0.04	0.05	0.04
4	0.00	0.00	0.00	0.04	0.00	0.00	0.03
5	0.00	0.00	0.07	0.06	0.11	0.11	0.05
6	0.00	0.00	0.11	0.07	0.09	0.08	0.06
7	0.00	0.00	0.01	0.08	0.00	0.00	0.06
8	1.00	0.00	0.72	0.60	0.61	0.60	0.70

Table 8: Regressors for different model scenarios

Panel A. Regressors for model A scenario	
Model	Regressors
1	CONSTANT, GDP60
2	CONSTANT,GDP60, INV
3	CONSTANT,GDP60,INV, SCHOOL60
4	CONSTANT,GDP60,INV,SCHOOL60, LIFE60
5	CONSTANT,GDP60,INV,SCHOOL60,LIFE60, POP
6	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP, LAW
7	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW, TROPICS
8	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS, ETHNO
9	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS,ETHNO,CONFUC
Panel B. Regressors for model B scenario	
Model	Regressors
1	CONSTANT,GDP60, INV
2	CONSTANT,GDP60,INV, SCHOOL60
3	CONSTANT,GDP60,INV,SCHOOL60, LIFE60
4	CONSTANT,GDP60,INV,SCHOOL60,LIFE60, POP
5	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP, LAW
6	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW, TROPICS
7	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS, ETHNO
8	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS,ETHNO,CONFUC