Department of Economics and Finance

Gordon S. Lang School of Business and Economics

University of Guelph

Discussion Paper 2024-01

Multi-Objective Frequentistic Model Averaging with an Application to Economic Growth

Stelios Arvanitis
Athens University
stelios@aueb.gr

Mehmet Pinar
Universidad de Sevilla
mpinar@us.es

Thanasis Stengos
University of Guelph
tstengos@uoguelph.ca

Nikolas Topaloglou
Athens University
nikolas@aueb.gr

Department of Economics and Finance
Gordon S. Lang School of Business and Economics | University of Guelph
50 Stone Road East | Guelph, Ontario, Canada | N1G 2W1
www.uoguelph.ca/economics

Multi-Objective Frequentistic Model Averaging with an Application to Economic Growth

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

March 2024

Abstract

In the Frequentistic Model Averaging framework and 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. We consider an estimator arising from 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. Characterizations of the averaging estimators introduced in the context of conservative optimization are also provided. Monte Carlo experiments suggest that the averaging estimators proposed here occasionally provide with bias and/or MSE/MAE reductions. 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 factors like geography or institutions. *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, core regressors, auxiliary regressors.

[*]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 deal with model uncertainty and as such there have been two main approaches in the literature, the Bayesian model averaging (BMA) and the Frequentist model averaging (FMA). Model uncertainty naturally stems from the presence of many competing theories that tend to examine different sources for possible data generating processes (DGP) that are intrinsically 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 (2019) for a very comprehensive survey of the BMA methods. However, in this paper we will examine an approach that rests within the FMA methodological camp and we will provide a comparison with other existing methods within this area. For example an important area that highlights the importance of model uncertainty is in the are of growth theory, where there are different valid approaches based on pure standard economic arguments based on a production function approach (Solow type), an institutional approach as in Acemoglu et al. (2001) [1] and a geography approach as proposed by Sachs (2003) [36]. In fact model uncertainty has been used to shed light on the relative importance of the three main growth theories (geography, integration and institutions), see Kourtellos et al. (2010) [23]. An earlier approach that aimed to shed light on the importance of the multitude of different variables that are being used in growth regressions is Extreme Bounds Analysis (EBA) proposed by Leamer (1983) [24] and applied in the empirics of growth by Levine and Renelt (1992) [25]. In that study variables were classified as "robust" and "fragile" and the ones that survived the different model specifications based on different variable combinations were essentially the "Solow" type variables.

Within the FMA approach there has been a variety of criteria proposed to achieve the averaging such as Mallows model averaging estimator (MMA), the jackknife model averaging estimator (JMA) and its generalized version, focused information criterion model selection (FIC) and plug-in averaging estimator (Plug-In) as well as. The literature on FMA, starting with Hansen (2007) [19], Liu (2015) [27] and Zhang and Liu (2019) [42] focuses on the limiting distributions of least squares averaging estimators for linear regression models in a local asymptotic framework.

In this paper we will extend the analysis of the generalized Jacknife Model Averaging (JMA) and the Mallows Model Averaging (MMA) criteria within a stochastic dominance framework. Specifically, and in the spirit of Zhang and Liu (2019) [42], we consider a version of the MMA estimator where the regularization parameter converges to zero with the sample. We then consider a vector valued criterion consisting of both the MMA and JMA criteria with the purpose of constructing averaging estimator via utilizing information from both methodologies. Using a multiobjective 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 that emerge from the minimization of several scalarizations of the vector criterion; we consider scalarizations emerging from the ℓ^p norms of the multi-objective function, as well as a scalarization emerging from the consideration of Nash bargaining solutions in social choice contexts. The proposed estimators are given as ℓ^1, ℓ^2, ℓ^∞ and Nash and their properties will be developed below as well as that of the approximate bound (AB-see Arvanitis et al. (2021) [5]).

We derive the limit theory of the estimators under both a correct specification and a misspecification framework. In the latter case the set of statistical models is globally misspecified due to the erroneous exclusion of DGP regressors from the analysis. The limit theory depends on the rates at which the MMA and JMA regularization parameters are asymptotically nullified, and it is identical to all considered averaging estimators in the correct specification case; the weights converge in probability to the deterministic vector that picks up the minimal correctly specified model. A richer theory emerges in the misspecification case, where even though the weights in all cases are asymptotically deterministic, their limits are in some cases distinct reflecting differing use of information of the dependence between the regressors used in the analysis.

Monte Carlo experiments are also provided suggesting occasional cases where the averaging estimators proposed here provide with provide with bias and/or MSE/MAE reductions in both the correctly specified and the misspecification scenarios, especially for the estimators emerging from the stochastic dominance considerations and the Nash bargaining solution criterion. Characterizations of the averaging estimators above in the context of conservative optimization are also provided; the robustness analysis is conformable to the findings of the Monte Carlo experiments regarding the behavior of the stochastic dominance based and the Nash averaging estimators in the context of misspecification.

An empirical application from economic growth is also provided; our findings are contrasted with the ones from existing methods-like the much earlier study of Levine and Renelt (1992) [25]. Overall, our proposed methods allocate heavier weights towards models formed by the fundamental Solow growth regressors (Solow, (1956) [38]; Mankiw et al., (1992) [29]), but do not dismiss models that contain auxiliary regressors that underpin the importance of geography (Diamond, (1997) [12]; Gallup et al., (1999) [36]; Sachs, 2003) and institutional quality (Acemoglu et al., (2001) [1];

Rodrik et al., (2004) [35]).

The remaining structure of the paper is the following: Section 2 describes the regression models background, Section 3 discusses the basis averaging criteria, namely the modified MMA and the Zhang and Liu (2019) [42] modification of the JMA criterion. Section 4 derives the limit theory of the modified MMA estimator and contrasts it with the analogous derivations of Zhang and Liu (2019) [42] for the JMA case. Section 5 introduces the proposed averaging estimators based on the multi-objective function that contains both the aforementioned criteria. Their conservative optimization characterizations are derived along with their limit theory under correct specification and the consistency of a sub-sampling estimator for their asymptotic variance. Section 6 contains the limit theory in the misspecification case, and Section 7 discusses potential extensions. Section 8 contains the Monte Carlo experiments. Section 9 contains the empirical application on growth and Section 10 concludes. The appendices contain the proofs of our theoretical results and the tables with the results of our Monte Carlo experiments and our empirical application.

2 Background

The linear model background of Zhang and Liu (2019) [42] is considered. Specifically, the following linear regression in matrix form is examined:

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

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 above 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. The ordering of regressors is considered irrelevant. As in Zhang and Liu (2019) [42], 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 $\mathbf{P}_m := \mathbf{X}_m(\mathbf{X}'_m\mathbf{X}_m)^{-1}\mathbf{X}'_m$, $m = 1, \cdots, M$, the averaging (across the models) projection is $\mathbf{P}(\mathbf{w}) := \sum_{m=1}^{M} \mathbf{w}_m\mathbf{P}_m$. Furthermore, $\sigma_n^2 := (n-K)^{-1}\|\mathbf{y} - \mathbf{X}\beta_M\|^2$, where $\|\cdot\|$ denotes the Euclidean norm, ϕ_n, $\phi_n^\star$-potentially stochastic or data dependent-regularization parameters that depend on n. $\mathbf{P}(i,i)_m$ is the i^{th} diagonal element of $\mathbf{P}_m$, and $\mathbf{D}_m$ is the diagonal matrix with i^{th} diagonal element equal to $\frac{1}{1-\mathbf{P}(i,i)_m}$. Finally, $\mathbf{K} := (k_m)_{m=1,\cdots,M}$.

The Mallows Model Averaging (MMA) criterion is defined as

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

The original criterion (see Hansen (2007) [19]) is recovered when $\phi_n^\star = 2$. The Mallows weights are then defined by the optimization problem specified as $\mathbf{w}_{\text{MMA}} \in \arg\min_{\mathbf{w}\in\Delta^{M-1}} \mathcal{M}_n(\mathbf{w})$, where Δ^{M-1} is the standard $M-1$ dimensional unit simplex. The Mallows averaging estimator of β is then $\beta_{\text{MMA}} := \beta(\mathbf{w}_{m,\text{MMA}})$. Existence and uniqueness is ensured by standard arguments-see Hansen (2007) [19].

The generalized Jacknife Model Averaging (JMA) criterion-see Zhang and Liu (2019) [42]-is defined as

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

The JMA weights are then defined by the optimization problem $\mathbf{w}_{\text{JMA}} \in \arg\min_{\mathbf{w}\in\Delta^{M-1}} \mathcal{J}_n(\mathbf{w})$. The JMA estimator of β is then $\beta_{\text{JMA}} := \beta(\mathbf{w}_{m,\text{JMA}})$. Essentially the JMA methodology rests on the leave one out cross validation technique-see Racine (1997) [33], while due to the form of the $\mathbf{D}_m$ matrix, it takes into account the possibility of forms of conditionally heteroskedasticity for the errors.

4 Limit Theory for the Basis Averaging Estimators

The limit theory (as $n \to \infty$) of the generalized MMA estimator is obtained. The analogous limit theory for the generalized JMA is derived in Zhang and Liu (2019)-see Th. 5 there. A mild assumption framework is first introduced; it is almost identical to conditions **C.1-C.4** of Zhang and Liu (2019) [42]; in what follows $\rightsquigarrow$ denotes 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 $\sup_{i\le n}\sup_{m\le M}\mathbf{P}_m(i,i)=o_p(n^{-1/2})$.

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) [42] 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 of Zhang and Liu (2019) [42]; 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 $k\in\{M_0+1,\dots,M\}$, $\mathbf{w}^{(k)}$ 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,*

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

Furthermore,

$$n^{1/2}(\beta_{\text{MMA}}-\beta)\rightsquigarrow V^\star\mathbf{z}\sim N(\mathbf{0}_{K\times1}V^\star VV^\star), \tag{5}$$

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

Th. 5 of Zhang and Liu (2019) [42] 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 $\frac{\phi_n}{\sqrt{n}}\to0$. Then,*

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

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

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

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

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 $\sqrt{n}(\mathbf{w}_{\text{JMA}} - \mathbf{w}^{(M_0+1)})$.

Both estimators thus share the limit theory of the OLSE for the (latent) minimal well-specified model. Asymptotic normality is the case due to the incorporation of the diverging penalization parameters in their definition.

5 Multi-objective Model Averaging

The MMA and JMA averaging criteria can be jointly used in order to construct a stochastic dominance relation on the Δ^{M-1} simplex. The rationale of the relation is that convex combination $\mathbf{w}$ dominates another $\mathbf{w}^\star$, iff the first attains lower values when evaluated-and thereby is "preferred"-by both $\mathcal{M}_n$, and $\mathcal{J}_n$. Formally,

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

The (pre-) order reflects the following problem of multi-objective optimization (hereafter MOOP-see for example Hwang and Masud (2012) [21]): $\min_\mathbf{w} \mathcal{W}_n(\mathbf{w})$, where the minimum is considered w.r.t. the pointwise order on $\mathbb{R}^2$, for the $\mathbb{R}^2$-valued objective $\mathbf{w} \to \mathcal{W}_n(\mathbf{w}) := (\mathcal{M}_n(\mathbf{w}), \mathcal{J}_n(\mathbf{w}))$ comprised of both the MMA and JMA criteria.

Any maximal element of the (pre-) order, i.e. any solution of the aforementioned optimization problem-which would seldomly exist-corresponds to an element that simultaneously minimizes both criteria, hence dominates every other possible weight.

Any non-dominated-or (Pareto) efficient-weight, must be necessarily preferred over any other weight, by at least one of the associated criteria; the latter generally depends on the alternative weight choice to which the efficient element is compared. Thus, the maximal-when existent, or the weights that minimize at least one of the criteria, i.e. $\mathbf{w}_{\text{MMA}}$, or $\mathbf{w}_{\text{JMA}}$ are examples of efficient elements of the order. Those however may not be the only cases of efficiency. In general, $\mathbf{w}$ is efficient iff for any weight $\mathbf{w}^\star$, for which $\mathcal{M}_n(\mathbf{w}) > \mathcal{M}_\mathbf{n}(\mathbf{w}^\star)$ (respectively $\mathcal{J}_n(\mathbf{w}) > \mathcal{J}_\mathbf{n}(\mathbf{w}^\star)$) holds, then $\mathcal{J}_n(\mathbf{w}) < \mathcal{J}_\mathbf{n}(\mathbf{w}^\star)$ (respectively $\mathcal{M}_n(\mathbf{w}) < \mathcal{M}_\mathbf{n}(\mathbf{w}^\star)$). In general, the location of the set of efficient elements of a (pre-) order and the derivation of its properties lies within the scope of the theory of MOOP.

A type of efficient element of stochastic dominance (pre-) orders, that is of general interest in portfolio selection applications, is that of the approximate bound (AB-see Arvanitis et al. (2021) [5]). This-in the present context-is defined via the following optimization problem:

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

Negativity at the minimum of $\sup_{G\in\{\mathcal{M}_n,\mathcal{J}_n\},\mathbf{w}^\star\in\Delta^{M-1}}(G(\mathbf{w}) - G(\mathbf{w}^\star)$ is equivalent to the existence of a maximal element. More generally, any weight $\mathbf{w}$ that solves this problem is an efficient element of the order; if this were dominated then any dominant weight would further diminish the criterion $\sup_{G\in\{\mathcal{M}_n,\mathcal{J}_n\},\mathbf{w}^\star\in\Delta^{M-1}}(G(\mathbf{w})-G(\mathbf{w}^\star)$ due to transitivity.

The optimization problem above can be equivalently re-expressed as:

$$\min_{\mathbf{w}\in\Delta^{M-1}} \max\left\{(\mathcal{M}_n(\mathbf{w}) - \mathcal{M}_n(\mathbf{w}_{\mathrm{MMA}})), (\mathcal{J}_n(\mathbf{w}) - \mathcal{J}_n(\mathbf{w}_{\mathrm{JMA}}))\right\}, \tag{8}$$

the optimal value of which can be easily seen to be less than or equal to the optimal value of

$$\min_{\mathbf{w}\in\Delta^{M-1}} \max\left\{(\mathcal{M}_n(\mathbf{w}) - \min_{G\in\{\mathcal{M}_n,\mathcal{J}_n\}} \min_{\mathbf{w}} G(\mathbf{w})), (\mathcal{J}_n(\mathbf{w}) - \min_{G\in\{\mathcal{M}_n,\mathcal{J}_n\}} \min_{\mathbf{w}} G(\mathbf{w}))\right\},$$

and greater than or equal to the optimal value of

$$\min_{\mathbf{w}\in\Delta^{M-1}} \max\left\{(\mathcal{M}_n(\mathbf{w}) - \max_{G\in\{\mathcal{M}_n,\mathcal{J}_n\}} \min_{\mathbf{w}} G(\mathbf{w})), (\mathcal{J}_n(\mathbf{w}) - \max_{G\in\{\mathcal{M}_n,\mathcal{J}_n\}} \min_{\mathbf{w}} G(\mathbf{w}))\right\}.$$

Both the optimization problems that act as bounds above share the same set of optimizers, notably the minimizers of the ℓ^∞ norm of $\mathcal{W}_n(\mathbf{w})$. Minimization of this norm w.r.t. $\mathbf{w}$ would also lead to an efficient element of the order; if the optimizer were dominated, the dominant weight would challenge the optimality of the dominated.

The aforementioned optimization problem are both examples of scalarization (see Hwang and Masud (2012) [21]); the vector criterion is transformed to a real valued criterion. The use of the ℓ^∞ norm above can be generalized to ℓ^p for any $p \geq 1$ on $\mathbb{R}^2$. The following definitions summarize those constructions:

Definition 1. The AB weights are defined by:

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

Subsequently, the AB estimator of β is defined by

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

Definition 2. The ℓ^p weights are defined by:

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

Subsequently, the ℓ^p estimator of β is defined by

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

The solutions of the optimization problems defined above are not invariant to rescaling of $\mathcal{M}_n$ and/or $\mathcal{J}_n$. Scaling invariance holds for the Nash bargaining solution to the associated MOOP (see Ch. 32 in Aumann and Hart (1992) [7]). Here the basis criteria are treated as players in a context of "social" choice; the optimal weights are chosen so as to be efficient, scaling invariant, symmetric and independent of irrelevant alternatives. The following definition is then motivated:

Definition 3. The Nash weights are defined by:

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

$$\text{s.t.}$$

$$\mathcal{M}_n(\mathbf{w}_{\text{Nash}}) \leq \mathcal{M}_n(\mathbf{w}_{\text{JMA}}) \text{ and } \mathcal{J}(\mathbf{w}_{\text{Nash}}) \leq \mathcal{J}(\mathbf{w}_{\text{MMA}}).$$

Subsequently, the Nash estimator of β is defined by

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

Existence of the optimal weights in all the above cases is established by the compactness and separability of the Δ^{M-1} simplex, the continuity of $\mathcal{M}_n$ and $\mathcal{J}_n$, the continuity of the ℓ^p norms and the multiplication operation, and Theorem of Measurable Projections (see Par. 1.7 of van der Vaart and Wellner (1996) [41]).

The transitivity of the dominance relation implies that any weight that conforms to any of the definitions above is an efficient element of the relation. In the following paragraphs statistical properties of the aforementioned methodologies of choosing efficient elements are discussed; a conservative optimization characterization is first documented, based on the penalties present in the basis criteria. The MOOP procedures in some cases imply further conservatism. The relevant limit theories are also derived; to the first order all the considered methodologies are indistinguishable as they asymptotically choose the most parsimonious well specified model. Finally, misspecification in the form of omitted variables is also considered; the MOOP methodologies can in some cases partially alleviate the effect of misspecification on the inconsistency of the OLS estimators for the remaining parameters.

5.1 Conservative Optimization Characterizations

The optimization problems appearing in the optimal weights selection in each of the previous methodologies are characterized as problems of conservative optimization; this is essentially based on the penalization factors that occur in the MMA and JMA objectives, along with results on convex duality for robust optimization (see for example Lemma 1 of Gao, Chen, and Kleywegt (2017) [16]).

Some needed notation is initially established: $\mathbb{F}_n$ denoted the emprical distribution (ecdf) of $(\mathbf{y}, \mathbf{X})$. For $\mathbb{P}$ an arbitrary distribution on $\mathbb{R}^{k_1+M}$, and $q\,(p)$ such that $\frac{1}{p}+\frac{1}{q}=1$, the (first) Wasserstein distance between $\mathbb{P}$ and $\mathbb{F}_n$ is defined by $\mathbb{W}\left(\mathbb{F}_n, \mathbb{P}; p\right) := \min_{\gamma\in\Gamma(\mathbb{F}_T,\mathbb{P})} \int \left\|z-z^{\star}\right\|_q d\gamma\left(z, z^{\star}\right)$, where $\Gamma\left(\mathbb{F}_T, \mathbb{P}\right)$ denotes the set of Borel probability distributions that have respective marginals $\mathbb{F}_T$, $\mathbb{P}$, and also have finite first moment (see Gao, Chen, and Kleywegt (2017) [16]). $\mathbb{W}$ metrizes weak convergence-see Rahimian and Mehrota (2019) [32]. For $\epsilon>0$, $\mathcal{N}_p\left(\mathbb{F}_n, \epsilon\right) := \left\{\mathbb{P} : \mathbb{W}\left(\mathbb{F}_n, \mathbb{P}; p\right) \leq \epsilon\right\}$ is the Wasserstein closed ball centered at $\mathbb{F}_n$ with radius ϵ.

The results are summarized in the following theorem-there unindexed suprema denote elementwise suprema over the support of $(\mathbf{y}_1, \mathbf{X}_1)$:

Theorem 3. *Suppose that (a). the distribution of $(\mathbf{y}_i, \mathbf{X}_i)$ is independent of i, (b). the support of the distribution of $(\mathbf{y}_1, \mathbf{X}_1)$ is compact, and (c). $\min_{j=1,\ldots,M} \mathbf{K}_j \geq 1$. Then, for $\epsilon_1 := {}^{\phi_n^{\star}\sigma_n^2}\!/_{2(\sup|\mathbf{y}_1|(1+\sup_{\mathbf{w}}\sup\|\mathbf{P}(\mathbf{w})_1\|))}$, and $\epsilon_2 := {}^{\phi_n}\!/_{2(\sup|\mathbf{y}_1|(\|D(\mathbf{w})_1\|+\sup_{\mathbf{w}}\sup\|D(\mathbf{w})\mathbf{P}(\mathbf{w})_1\|))}$, and for any $\mathbf{w}\in\Delta^{M-1}$:*

$$\inf_{\mathbf{w}\in\Delta^{M-1}} \mathcal{M}_n(\mathbf{w}) \geq \inf_{\mathbf{w}\in\Delta^{M-1}} \sup_{\mathbb{P}\in\mathcal{N}_1(\mathbb{F}_n,\epsilon_1)} \mathbb{E}_{\mathbb{P}}(\mathbf{y}_1-(\mathbf{P}(\mathbf{w})\mathbf{y})_1)^2, \tag{9}$$

$$\inf_{\mathbf{w}\in\Delta^{M-1}} \mathcal{J}_n(\mathbf{w}) \geq \inf_{\mathbf{w}\in\Delta^{M-1}} \sup_{\mathbb{P}\in\mathcal{N}_1(\mathbb{F}_n,\epsilon_2)} \mathbb{E}_{\mathbb{P}}((\mathbf{D}_m(\mathbf{P}(\mathbf{w})-\mathbf{I}_{n\times n})+\mathbf{I}_{n\times n})\mathbf{y}_1)^2. \tag{10}$$

Furthermore, for $\Phi_n := \min(\phi_n, \phi_n^{\star}\sigma_n^2)$, and $\epsilon_3 := {}^{\Phi_n}\!/_{(\frac{\epsilon_1}{\phi_n^{\star}\sigma_n^2}+\frac{\epsilon_2}{\phi_n})}$:

$$\inf_{\mathbf{w}\in\Delta^{M-1}} \|\mathcal{W}_n(\mathbf{w})\|_1 \geq \inf_{\mathbf{w}\in\Delta^{M-1}} \sup_{\mathbb{P}\in\mathcal{N}_1(\mathbb{F}_n,\epsilon_3)} \left(\begin{array}{c} E_{\mathbb{P}}(\mathbf{y_1}-\mathbf{P}(\mathbf{w})\mathbf{y}_1)^2 + \\ \mathbb{E}_{\mathbb{P}}((\mathbf{D}_m(\mathbf{P}(\mathbf{w})-\mathbf{I}_{n\times n})+\mathbf{I}_{n\times n})\mathbf{y}_1)^2 \end{array} \right). \tag{11}$$

Finally, the infimum of the Nash criterion in greater than or equal to

$$\inf_{\mathbf{w}\in\Delta^{M-1}} \max\Big(\sup_{\mathbb{P}\in\mathcal{N}_1(\mathbb{F}_n,\epsilon_1)} \mathbb{E}_{\mathbb{P}}(\mathbf{y}_1-(\mathbf{P}(\mathbf{w})\mathbf{y})_1)^2, \sup_{\mathbb{P}\in\mathcal{N}_1(\mathbb{F}_n,\epsilon_2)} \mathbb{E}_{\mathbb{P}}((\mathbf{D}_m(\mathbf{P}(\mathbf{w})-\mathbf{I}_{n\times n})+\mathbf{I}_{n\times n})\mathbf{y}_1)^2. \tag{12}$$

Condition (a.) in the previous result is trivially satisfied in strictly stationary time series settings. Condition (b.) is quite strong, it however seems indispensable. Condition (c.) is trivial.

The results indicate that for fixed n, the basis criteria obey more conservative characterizations than similar criteria with penalizations based on ℓ^2 norms of the weights that also have disributionaly robust characterizations over Wasserstein neighborhoods of the empirical distribution. This is due to due to the form of the penalization factors of the basis criteria. The multiobjective ℓ^1 criterion, partially alleviates conservativeness by employing smaller neighborhoods; however it also dominates a strictly larger criterion along with a strictly larger penalization term w.r.t. the criterion it is compared. The Nash criterion is even more conservative than both the basis criteria, as well as than the approximate bound criterion. The latter is obviously by construction more conservative than both the basis criteria. As such, both the Nash criterion, as well as the AB one are expected to have superior robustness properties in cases of misspecification.

5.2 Limit Theory

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

Theorem 4. *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* $\max(\phi_n^\star, \phi_n) = o_p(\sqrt{n})$. *Then,*

$$\mathbf{w}_{\mathrm{AB}} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} \max(\sigma^2 1_{C>0}, C^{-1}1_{C>0} + 1_{C=0})\mathbf{K}'(\mathbf{w} - \mathbf{w}^{(M_0+1)}). \tag{13}$$

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

Under the same conditions,

$$\mathbf{w}_{\ell^p} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} (\sigma^{2p}1_{C>0} + (C^{-1}1_{C>0} + 1_{C=0})^p)^{1/p}\mathbf{K}'\mathbf{w}. \tag{14}$$

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

Suppose now, that ***A.1-A.4*** *hold, and that* $\min(\phi_n, \phi_n^\star) \to +\infty$, *while* $\phi_n = o_p(\sqrt{n})$. *Then,*

$$\mathbf{w}_{\mathrm{Nash}} \rightsquigarrow \arg\min_{\mathbf{w}\in\Delta_0^{M-1}} \sigma^2(\mathbf{K}'(\mathbf{w} - \mathbf{w}^{(M_0+1)}))^2. \tag{15}$$

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

Subsequently, under ***A.1-A.4****, and and that* $\min(\phi_n, \phi_n^\star) \to +\infty$, *while* $\phi_n = o_p(\sqrt{n})$, *then for* $J = \mathrm{Nash}$,

$$n^{1/2}(\beta_J - \beta) \rightsquigarrow V^\star \mathbf{z}. \tag{16}$$

If moreover $\frac{\phi_n^\star}{\phi_n} \to C \in [0, +\infty]$, *while* $\phi_n^\star = o_p(\sqrt{n})$ *then* (16) *also holds for* $J = \mathrm{AB}, \ \ell^p$.

The AB and the ℓ^p cases require the $\phi_n^\star = o_p(\sqrt{n})$ condition, due to the scale variance of the associated criterion. The scale invariance of the Nash criterion avoids such restrictions. The weights estimators, due to their convergence to the minimal well specified model-become asymptotically scale invariant. The parameters' estimators are thus asymptotically normal. The limit theory is not fine enough so as to be able to discriminate between the estimators involved under the particular assumption framework.

5.3 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 the sub-sample sequence $((\mathbf{y}, \mathbf{X})_{j,\dots,j+b-1})_{j=1,n-b+1}$. 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 5. *Under **A.1-A.4**, and,* $\min(\phi_n, \phi_n^\star) \to +\infty$, *and for* $J = \text{AB}$, ℓ^p, $\frac{\phi_n^\star}{\phi_n} \to C \in [0, +\infty]$, $\max(\phi_n^\star, \phi_n) = o_p(\sqrt{n})$, *while for* $J = \text{Nash}$ $\phi_n = o_p(\sqrt{n})$, *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:*

$$\mathrm{V}_{b,n,J} \rightsquigarrow \mathrm{Var}(V^\star \mathbf{z}). \tag{17}$$

$\mathrm{V}_{b,n,J}$ is thus a weakly (actually and L^2) consistent estimator of the asymptotic variance, and can therefore be used for inference. 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) [6] 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) [8]).

6 Misspecification

The effect of globally omitted variables misspecification on the averaging procedures is now considered. For notational simplicity and without loss of generality it is supposed that $K_1 < M_0$ and the $\mathbf{X}_1$ regressors submatrix is omitted from the statistical models at hand. Thereby every model considered is now misspecified. Under **A.1** and **A.2** it is readily seen that $\beta_m \rightsquigarrow \Pi^{\star\prime}_m(\Pi^\star_m + (\Pi^\star_m Q\Pi^{\star\prime}_m)^{-1}\Pi^\star_m Q\Pi'_{K_1}\Pi_{K_1})\beta$, with $\Pi^\star_m := (\mathbf{0}_{(k_m-K_1)\times K_1}, \mathbf{I}_{(k_m-K_1)\times(k_m-K_1)}, \mathbf{0}_{(k_m-K_1)\times(K-k_m)})$. Inconsistency is then the case for every model included. Partial consistency, i.e. consistency for the OLSE for the non-omitted regressors slopes occurs iff $\Pi^\star_m Q\Pi'_{K_1}$ is a zero matrix, i.e. every included regressor is "asymptotically orthogonal" to every omitted regressor. In the present framework $M = K_2$. The definitions of the model averaging estimators for β are accordingly modified in all cases. The interest lies in the limiting behavior of model averaging under this framework of global misspecification. Would they for example chose models for which the term $(\Pi^\star_m Q\Pi^{\star\prime}_m)^{-1}\Pi^\star_m Q\Pi'_{K_1}\Pi_{K_1})$ is "minimal", approximating thereby partial consistency for the remaining correct regressors?

The following result concerns the limiting behavior of the MMA estimator under **A.1**, **A.2** and **A.4**; there we have that $\mathbf{K}^\star := (1, 2, \ldots, K_2)'$, and $\sigma^2_\star := \sigma^2 + \beta'(Q - Q\Pi^{\star\prime}_{K_2}(\Pi^\star_{K_2}Q\Pi^{\star\prime}_{K_2})^{-1}\Pi^\star_{K_2}Q)\beta$:

Theorem 6. *Suppose that **A.1**, **A.2** and **A.4** hold, in the global misspecification framework. Suppose also that* $\frac{\phi^\star_n}{n} \to C \in [0, +\infty]$*. Then,*

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

where $\mathcal{M}_C(\mathbf{w}) := \Lambda(\mathbf{w})1_{C<+\infty} + \sigma^2_\star(C1_{0<C<+\infty} + 1_{C=+\infty})\mathbf{K}^{\star\prime}\mathbf{w}$*, where now,* $\Lambda(\mathbf{w}) := \sum_{m,m^\star=1}^{K_2} \mathbf{w}_m\mathbf{w}_{m^\star} \begin{matrix} \beta' Q\Pi^{\star\prime}_m(\Pi^\star_m Q\Pi^{\star\prime}_m)^{-1}\Pi^\star_m Q\Pi^{\star\prime}_{m^\star} \\ (\Pi^\star_{m^\star}Q\Pi^{\star\prime}_{m^\star})^{-1}\Pi^\star_{m^\star}Q\beta \end{matrix} - 2\sum_{m=1}^{K_2}\mathbf{w}_m\beta' Q\Pi^{\star\prime}_m(\Pi^\star_m Q\Pi^{\star\prime}_m)^{-1}\Pi^\star_m Q\beta$.
Finally,

$$\beta_{\text{MMA}} \rightsquigarrow \sum_{m=1}^{K_2} \mathbf{w}^\infty_{\text{MMA}}(m)\Pi^{\star\prime}_m(\Pi^\star_m + (\Pi^\star_m Q\Pi^{\star\prime}_m)^{-1}\Pi^\star_m Q\Pi'_{K_1}\Pi_{K_1})\beta. \tag{18}$$

When $C = +\infty$ it is obtained that $\beta_{\text{MMA}} \rightsquigarrow \Pi^{\star\prime}_1(\Pi^{\star\prime}_1 + (\Pi^\star_1 Q\Pi^{\star\prime}_1)^{-1}\Pi^\star_1 Q\Pi'_{K_1}\Pi_{K_1})\beta$, i.e. the simple regression model is asymptotically picked. When $C \neq +\infty$ then the limiting covariance between the regressors that appear in the underlying models is taken into account, and larger weights are attributed to models with more pronounced asymptotic signals $\Pi^\star_m Q\Pi^{\star\prime}_m$.

The proofs of Th. 3-5 of Zhang and Liu (2019) [42] 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) [42]). This then implies the following result:

Theorem 7. *Suppose that* ***A.1-A.4*** *hold in the global misspecification framework. Suppose also that* $\phi_n = o_p(\sqrt{n})$. *Then,*

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

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

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

and

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

Furthermore, when $p < +\infty$,

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

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

$$\mathbf{w}_{\text{Nash}} \rightsquigarrow \mathbf{w}^{\infty}_{\text{Nash}} := \arg\min_{\mathbf{w}\in\Delta^{M-1}} (\mathcal{M}_C(\mathbf{w}) - \mathcal{M}_C(\mathbf{w}^{\infty}_{\text{JMA}}))(\Lambda(\mathbf{w}^{\infty}_{\text{MMA}}) - \Lambda(\mathbf{w}))$$
$$\text{s.t. } \mathcal{M}_C(\mathbf{w}) \leq \mathcal{M}_C(\mathbf{w}^{\infty}_{\text{JMA}}),\ \Lambda(\mathbf{w}^{\infty}_{\text{MMA}}) \geq \Lambda(\mathbf{w}).$$

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

$$\beta_J \rightsquigarrow \sum_{m=1}^{K_2} \mathbf{w}^{\infty}_J(m)\Pi^{\star\prime}_m(\Pi^{\star}_m + (\Pi^{\star}_m Q\Pi^{\star\prime}_m)^{-1}\Pi^{\star}_m Q\Pi^{\prime}_{K_1}\Pi_{K_1})\beta. \tag{19}$$

The regularization constraints become asymptotically negligible for the JMA methodology since the restriction $\phi_n = o_p(\sqrt{n})$ is retained in order for a unified, with the case of well specification, statistical methodology to be possible.

The limiting behavior of the parameter $\phi^{\star}_n$ influences the asymptotics of the multi-objective methodologies, as those asymptotically take into account the regularized MMA criterion. When $C = 0$ then the asymptotic weights are identically equal to

$\mathbf{w}^{\infty}_{\mathrm{MMA}}$ in all cases considered in the previous theorem. The regularization parameters are diverging slowly enough so that the scaled criteria considered essentially converge to Λ. When $C = +\infty$, so that $\phi^{\star}_{n}$ diverges quickly, the AB and the ℓ^{p} methodologies are again asymptotically equivalent to the MMA; the weights asymptotically pick up the minimal simple regression model. This is not the case for JMA and Nash; the asymptotic optimal weights retain information on the asymptotic covariances between the regressors appearing in the misspecified models. When $0 < C < +\infty$, the latter is also true for the JMA, ℓ^{p} and the Nash methodologies. Compared to the MMA methodology, the ℓ^{p}-for finite p-and Nash methodologies seem to asymptotically attach more significance to the aforementioned limiting covariances, due to the presence of the $\Lambda(\mathbf{w})$ term in their limiting criteria, without completely asymptotically denouncing the regularization, as in the case of the JMA methodology. Hence, they search for a combination of model parsimony with the greater possible signal for the well specified regressors.

The asymptotic framework above also provides for some motivation for the consideration of the multi-objective optimization methodologies. There are cases, $0 < C < +\infty$, and methodologies, namely the ℓ^{p}, for $p < +\infty$, and Nash, that attribute more significance, compared to the MMA, to well specified regressors signal, yet do not abandon parsimony.

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; 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. 19 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 $n \to \infty$, D_m, under some appropriate topology, converges to a tight random operator, then the scaling of the JMA criterion by $(n1_{C<+\infty} + \phi^{\star}_{n}1_{C=+\infty})^{-1}$ would produce asymmptotic 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) [42]. 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 core regressor. The analyst erroneously considers as core regressors $\mathbf{X}_{1,1}$ and $\mathbf{X}_{2,1}$, and 7 auxiliary regressors emerging from $\mathbf{X}_2$ by deleting its' first collumn. 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 $0.001 \times n$ and ,$0.05 \times \ln(n)$ respectively, whereas ${}^{\phi_n^\star}/_{n} \to 0.001$, ${}^{\phi_n}/_{\sqrt{n}} \to 0$, 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 1000. Similarly to Zhang and Liu (2019) [42], 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 for β_4 in the case of "correct specification", and for β_3 in the case of "misspecification". 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 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 1-5. 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 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. There 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 $1 < m$ 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 do not seem to favour uniformly any of the considered estimators. The MMA estimator seems to in several cases induce less bias compared to the simple averaging one, usually at the cost of (sometimes significantly higher) variance/MSE/MAE. The JMA seems to usually improve the variance/MSE/MAE of the simple averaging estimator at the cost of higher bias (except for the first and the third cases in the misspecification setup for $n = 100$ where it seems to also improve on the bias). The multi-objective estimators seem to lie between those two cases; they usually improve the bias of the JMA by incurring higher variance/MSE/MAE, that is usually significantly less than the ones for the MMA;consider though the bias improvements compared to the MMA that appear for the multi-objective estimators in the correct specification heteroskedastic set-up for $n = 400$; in several instances the multi-objective estimators provide with significant bias improvements compared to the MMA achieving simultaneously in several cases significantly less variance or MSE-see for example the performances of the ℓ^p and Nash estimators in Case 2, or the one of the AB estimator in Case 1. The ℓ^∞ estimator seems uniformly close to the JMA in terms of the aforementioned properties. The JMA seems to have a uniformly better performance-in terms of bias and variance/MSE/MAE-in Case 1 of the global misspecification scenario for both $n = 100, 400$, followed closely by the ℓ^∞.

In terms of concentration the results show that the JMA typically is comparatively the one with maximal concentration, again followed shortly by the ℓ^∞. The MMA seems also to have high concentration for $n = 100$ that drops significantly-even compared to the multi-objective estimators- for $n = 400$. In the $n = 100$ case the Nash appears as the one of minimal concentration, something that is in several cases also true for $n = 400$. The remaining multi-objective estimators concentrations lie usually between those two case, mainly with significantly lower concentration compared to the JMA-except for the ℓ^∞. The MMA-especially for $n = 100$-seems to favor the large models. This is partly alleviated when $n = 400$ in which cases its model choices seem close to the ones of the AB. The JMA and the ℓ^∞ seem to quite frequently favor the second model, while the remaining estimators seem to also lie between the aforementioned case usually favoring the fourth model.

The results depend on the choice of the penalization parameters $\phi_n^\star$ and ϕ_n. Auxiliary results that are not reported here-yet are available upon request- suggest that a choice of parameters equal to $\sqrt{\ln(n)}$ would imply, especially for $n = 100$, that the multi-objective estimators would be greatly affected by the sparse selection behavior induced by the large penalization terms, and become maximally concentrated on the minimal models. They would induce some bias compared to the MMA and JMA basis case, while in several cases achieving impressive MSE/MAE improvements. The optimal-in terms of finite sample properties-choice of the penalization terms is a non-trivial task, that may be benefited by out of sample analyses, and it is left as an interesting issue for further research.

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). 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 uncertainty, the existing literature has been using model averaging methods (see e.g., Fernandez et al., (2001) [14]; Sala-i Martin et al., (2004) [37]; Durlauf et al., (2008) [13]; 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) [9], 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) [14] and Liu (2015) [27], and the following cross-section growth regression model is used:

$$\mathbf{growth} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon, \tag{20}$$

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, C, D, and E are reported in Tables 6, 7, 8, 9, and 10 respectively. Finally, Table 11 provides the weights assigned to the coefficients of each model under the respective scenario and Table 12 lists 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) [38]; Mankiw et al., (1992) [29]) and auxiliary regressors highlighting the importance of the geography (Diamond, (1997) [12]; Gallup et al., (1999) [36]; Sachs, 2003) and institutional quality (Acemoglu et al., (2001) [1]; Rodrik et al., (2004) [35]). 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 averaging estimators proposed here occasionally provide with bias and/or 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) [10], 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., Post, T., & Topaloglou, N., (2021). Stochastic Bounds for reference sets in portfolio analysis. Management Science, 67(12), 7737-7754.

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

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

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

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

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

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

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

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

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

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

[16] Gao, R., Chen, X. and Kleywegt, A.J., (2017). Distributional robustness and regularization in statistical learning. arXiv preprint arXiv:1712.06050.

[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] Politis, D. N., Romano, J. P., & Wolf, M. (1999). Subsampling. Springer Science & Business Media.

[32] Rahimian, H. and Mehrotra, S., (2019). Distributionally robust optimization: A review. arXiv preprint arXiv:1908.05659.

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

[34] Rockafellar, R. T., & Wets, R. J. B., (2009). Variational analysis (Vol. 317). Springer Science & Business Media.

[35] 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.

[36] 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.

[37] 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.

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

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

[40] van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press.

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

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

Appendix

Proofs

This appendix contains the proofs of the results.

Proof of Theorem 1. Skorokhod representations verified by Th. 3.7.25 of Gine and Nickl (2021) [17] and the proofs of Th. 3-4 of Zhang and Liu (2019) [42] imply the locally uniform convergence in distribution of $\frac{1}{\phi_n^\star}\mathcal{M}_n$ to $\sigma^2\mathbf{K}'\mathbf{w}$. Then (4) follows from Cor. 5.8 of Van Der Vaart (2000) [40], the compactness of Δ_0^{M-1}, and the fact that the limiting criterion is uniquely minimized over Δ_0^{M-1}, at $\mathbf{w}_{M_0+1}$. The latter then directly imply (5). The rate follows from the following argument: suppose that $\phi_n^\star(\mathbf{w}_{\text{MMA}} - \mathbf{w}^{(M_0+1)}) \neq o_p(1)$. Let $h := \phi_n^\star(\mathbf{w} - \mathbf{w}^{M_0+1})$, $H_n := \left\{h; \mathbf{w} \in \Delta_0^{M-1}\right\}$, and notice that as $n \to \infty$, H_n converges in the Painleve-Kuratwoski topology (see for example Par. 4.B in Rockafellar and Wetts (2009) [34]) to the convex cone $H_\infty = \mathbb{R}_+^{M_0} \times \mathbb{R}_- \times \mathbb{R}_+^{M-M_0-1}$. Furthermore, for the modified criterion we have, $\mathcal{M}(\frac{h}{\phi_n^\star} - \mathbf{w}^{(M_0+1)}) - \mathcal{M}(\mathbf{w}^{(M_0+1)}) = \sum_{m,m^\star=1}^{M}(\frac{h_{(m)}}{\phi_n^\star}\frac{h_{(m^\star)}}{\phi_n^\star} - \frac{h_{(m)}}{\phi_n^\star}\mathbf{w}_m^{(M_0+1)} - \frac{h_{(m)^\star}}{\phi_n^\star}\mathbf{w}_m^{(M_0+1)})\varepsilon' P_m P_{m^\star}\varepsilon - 2\sum_{m=1}^{M}\frac{h_{(m)}}{\phi_n^\star}\varepsilon' P_m\varepsilon + \sigma_n^2\mathbf{K}'h$, weakly converges locally uniformly over H_∞ to $\sigma^2\mathbf{K}'h$ due to Assumptions **A**.1-2. Here $h_{(m)}$ denotes the respective component of h. Furthermore, h_n is by construction the minimizer of $\mathcal{M}(\frac{h}{\phi_n^\star} - \mathbf{w}^{(M_0+1)}) - \mathcal{M}(\mathbf{w}^{(M_0+1)})$. Since h_n is not further restricted to lie inside the o_1 sequences inside H_∞, Cor. 5.8 of Van Der Vaart (2000) [40] implies that it weakly converges to the minimizer of $\sigma^2\mathbf{K}'h$ which is the extended element $(0, 0, \ldots, 0, -\infty, 0, \ldots, 0)$, while the limiting criterion evaluated at its minimizer equals $-\infty$. This is a contradiction since due to the proof of (4), and for $h_n := \phi_n^\star(\mathbf{w}_{\text{MMA}} - \mathbf{w}^{M_0+1})$, $\mathcal{M}(\frac{h_n}{\phi_n^\star} - \mathbf{w}^{(M_0+1)}) - \mathcal{M}(\mathbf{w}^{(M_0+1)}) = o_p(1)$. □

Proof of Theorem 2. (6) and (7) follow directly from Th. 5 of Zhang and Liu (2019) [42]. The final part follows from the proofs of Th.4 and Th.5 of Zhang and Liu (2019) [42], and a similar contradiction argument to the one in the proof of Theorem 1. □

Proof of Theorem 3. The results follow directly form Th. 1 of Gao, Chen, and Kleywegt (2017) [16], the fact that $\mathbf{K}'\mathbf{w} - \|\mathbf{w}\| \geq 0$ due to the Cauchy-Schwarz inequality, the relations between ℓ^p norms, and (c.), and due to the elementary inequality $\inf(A+B) \geq \inf A + \inf B$. □

Proof of Theorem 4. For the case of AB, Skorokhod representations, Theorems 1-2, and the Lipschitz continuity property of the max imply the locally uniform convergence in distribution of the AB criterion scaled by $1/g_n$ to $\max(\sigma^2 1_{C>0}, C^{-1}1_{C>0} + 1_{C=0})\mathbf{K}'(\mathbf{w} - \mathbf{w}^{(M_0+1)})$, for $g_n := \begin{cases} \phi_n^\star, & C > 0, \\ \phi_n, & C = 0 \end{cases}$. The rate follows from the Lipschitz continuity property of max and the in contradiction arguments in the proofs of the analogous results in Theorems 1-2. The case of ℓ^p follows analogously via the Lipschitz property of the ℓ^p norm. For the case of the Nash estimator the result is analogously obtained by considering the original criterion scaled by $(\phi_n\phi_n^\star)^{-1}$, using arguments like the above, by also taking into account the locally uniform convergence of the MMA and JMA criteria and the consistency of the respective optimal weights. The rate follows from the boundedness of the MMA and JMA criteria, and the subsequent Lipschitz continuity property of multiplication along with the in contradiction arguments in the proofs of the analogous results in Theorems 1-2. □

Proof of Theorem 5. The result follows from Lemma 3.8.1 of Politis, Romano and Wolf (1999) [31], via the CMT and the Cramer Wold device, by noting that condition (b.) and stationarity-ergodicity imply the uniform integrability condition of the aforementioned lemma. □

Proof of Theorem 6. The proof of Th. 1 of Zhang and Liu (2019) [42] and the annihilating properties of the $\mathbf{I}_{n\times n} - \Pi_m$ matrices imply that in the present context of misspecification, $\mathbf{w}_{\mathrm{MMA}}$ equivalently minimizes

$$\Lambda_n(\mathbf{w}) := \begin{array}{c} \sum_{m,m^\star=1}^{K_2} \mathbf{w}_m \mathbf{w}_{m^\star} \begin{array}{c} (X\beta+\varepsilon)'X\Pi_m^{\star\prime}(\Pi_m^\star X'X\Pi_m^{\star\prime})^{-1}\Pi_m^\star X' \\ \times X\Pi_{m^\star}^{\star\prime}(\Pi_{m^\star}^\star X'X\Pi_{m^\star}^{\star\prime})^{-1}\Pi_{m^\star}^\star X'(X\beta+\varepsilon) \end{array} \\ -2\sum_{m=1}^{K_2} \mathbf{w}_m (X\beta+\varepsilon)'X\Pi_m^{\star\prime}(\Pi_m^\star X'X\Pi_m^{\star\prime})^{-1}\Pi_m^\star X'(X\beta+\varepsilon) \\ +\sigma_n^2\phi_n^\star \mathbf{K}^{\star\prime}\mathbf{w}. \end{array}$$

Due to **A.1**, **A.2** and **A.4**, $\sigma_n^2 = \frac{y_n' y_n}{n-K_2} - \frac{y_n' X \Pi_{K_2}^{\star'}}{n-K_2} (\frac{\Pi_{K_2}^{\star} X' X \Pi_{K_2}^{\star'}}{n})^{-1} \frac{\Pi_{K_2}^{\star} X' y_n}{n} \rightsquigarrow \sigma_\star^2$. Due to **A.1**, **A.2**, we have that

$$\begin{array}{c} \frac{1}{n}(X\beta+\varepsilon)' X \Pi_m^{\star'} (\Pi_m^\star X' X \Pi_m^{\star'})^{-1} \Pi_m^\star X' \\ \times X \Pi_{m^\star}^{\star'} (\Pi_{m^\star}^\star X' X \Pi_{m^\star}^{\star'})^{-1} \Pi_{m^\star}^\star X' (X\beta+\varepsilon) \end{array} \rightsquigarrow \begin{array}{c} \beta' Q \Pi_m^{\star'} (\Pi_m^\star Q \Pi_m^{\star'})^{-1} \Pi_m^\star Q \Pi_{m^\star}^{\star'} \\ (\Pi_{m^\star}^\star Q \Pi_{m^\star}^{\star'})^{-1} \Pi_{m^\star}^\star Q \beta \end{array},$$

and

$$(X\beta+\varepsilon)' X \Pi_m^{\star'} (\Pi_m^\star X' X \Pi_m^{\star'})^{-1} \Pi_m^\star X' (X\beta+\varepsilon) \rightsquigarrow \beta' Q \Pi_m^{\star'} (\Pi_m^\star Q \Pi_m^{\star'})^{-1} \Pi_m^\star Q \beta.$$

Thus, for $g_n := n 1_{C<+\infty} + \phi_n^\star 1_{C=+\infty}$, locally uniformly on Δ^{M-1},

$$\frac{1}{g_n} \Lambda_n(\mathbf{w}) \rightsquigarrow \Lambda(\mathbf{w}) 1_{C<+\infty} + \sigma_\star^2 (C 1_{0<C<+\infty} + 1_{C=+\infty}) \mathbf{K}^{\star'} \mathbf{w}.$$

Then, the first result follows from the above, Cor. 5.8 of Van Der Vaart (2000) [40], the compactness of Δ^{M-1}, and the fact that the limiting criterion is uniquely minimized over Δ^{M-1}. The second result follows then readily. □

Proof of Theorem 7. The decomposition $\mathcal{J}_n = \mathcal{M}_n + \mathcal{K}_n - (\sigma_n^2 \phi_n^\star \mathbf{K}^{\star'} - \phi_n \mathbf{K}^{\star'}) \mathbf{w}$ that is essentially established in the proof of Th. 3 of Zhang and Liu (2019) [42] is also valid in the misspecification setting. K_n is a quadratic form that becomes asymptotically negligible, locally uniformly over Δ^{M-1}, when scaled by $(n 1_{C<+\infty} + \phi_n^\star 1_{C=+\infty})^{-1}$ as readily seen from the proofs of Th. 4-5 of Zhang and Liu (2019) [42]. As indicated by the proof of Th. 4 this scaling is necessary in order for the MMA part of the criterion to become asymptotically tight in the present misspecification setting. The result then follows from the fact that $\phi_n = o_p(\sqrt{n})$ and the limiting behavior of $\Lambda_n - \sigma_n^2 \phi_n^\star \mathbf{K}^{\star'} \mathbf{w}$ as established in the proof of Th. 6. The limiting behavior of $\mathbf{w}_{\text{JMA}}$ then follows. The rest then follow via similar arguments to the ones establishing weights consistency to the minimal model in the proof of Th. 4, where now the AB and ℓ^p criteria are scaled by $(n 1_{C<+\infty} + \phi_n^\star 1_{C=+\infty})^{-1}$, while the Nash criterion is scaled by $(n 1_{C<+\infty} + \phi_n^\star 1_{C=+\infty})^{-2}$. □

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	1.9312	1.9300	1.0454	-0.0257	1.7441	1.7437	1.0411	-0.0367
	MMA	3.1511	3.1481	1.3207	-0.0127	2.8700	2.8671	1.3186	0.0004
	JMA	0.7451	0.7551	0.5577	-0.1039	0.7595	0.7668	0.5658	-0.0894
	ℓ^1	1.5423	1.5439	0.8609	-0.0559	1.5108	1.5105	0.8730	-0.0349
	ℓ^2	1.4627	1.4647	0.8352	-0.0591	1.4396	1.4396	0.8472	-0.0377
	ℓ^∞	0.7482	0.7582	0.5590	-0.1036	0.7627	0.7698	0.5672	-0.0889
	Nash	1.8210	1.8211	0.9646	-0.0438	1.7654	1.7643	0.9789	-0.0255
	AB	1.7703	1.7705	0.9391	-0.0444	1.7226	1.7215	0.9554	-0.0252
Case 2	SimAver	1.7550	1.7534	1.0395	0.0137	1.6688	1.6672	1.0177	0.0063
	MMA	2.9217	2.9189	1.3020	0.0122	2.7175	2.7148	1.2741	0.0026
	JMA	0.7566	0.7577	0.5214	-0.0430	0.6493	0.6517	0.4749	-0.0552
	ℓ^1	1.5193	1.5182	0.8527	-0.0190	1.3799	1.3794	0.8075	-0.0306
	ℓ^2	1.4465	1.4455	0.8260	-0.0211	1.3104	1.3102	0.7809	-0.0323
	ℓ^∞	0.7598	0.7609	0.5229	-0.0429	0.6523	0.6547	0.4765	-0.0551
	Nash	1.7754	1.7737	0.9610	-0.0116	1.6026	1.6013	0.9197	-0.0191
	AB	1.7343	1.7328	0.9357	-0.0136	1.5762	1.5750	0.8940	-0.0200
Case 3	SimAver	1.5791	1.5779	0.9939	-0.0213	1.8200	1.8198	1.0616	-0.0410
	MMA	2.6855	2.6834	1.2902	0.0243	3.0758	3.0732	1.3597	-0.0205
	JMA	0.6641	0.6799	0.5534	-0.1283	0.7639	0.7785	0.5767	-0.1241
	ℓ^1	1.3675	1.3698	0.8492	-0.0606	1.5914	1.5949	0.8986	-0.0713
	ℓ^2	1.2991	1.3020	0.8240	-0.0654	1.5089	1.5130	0.8707	-0.0745
	ℓ^∞	0.6671	0.6828	0.5548	-0.1279	0.7672	0.7817	0.5781	-0.1238
	Nash	1.5937	1.5933	0.9424	-0.0338	1.8195	1.8206	0.9913	-0.0547
	AB	1.5602	1.5602	0.9210	-0.0395	1.7975	1.7991	0.9732	-0.0585

Entries report the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value of β_4, 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	0.4487	0.4491	0.5266	-0.0302	0.4472	0.4545	0.5320	-0.0879
	MMA	0.5228	0.5230	0.5432	-0.0263	0.4877	0.4946	0.5366	-0.0857
	JMA	0.2215	0.2337	0.3644	-0.1116	0.1826	0.2010	0.3514	-0.1363
	ℓ^1	0.3833	0.3863	0.4618	-0.0578	0.3377	0.3479	0.4507	-0.1026
	ℓ^2	0.3800	0.3831	0.4598	-0.0586	0.3344	0.3447	0.4487	-0.1030
	ℓ^∞	0.2221	0.2343	0.3648	-0.1113	0.1832	0.2016	0.3518	-0.1361
	Nash	0.4128	0.4149	0.4718	-0.0500	0.3623	0.3713	0.4625	-0.0965
	AB	0.5473	0.5473	0.5531	-0.0229	0.4880	0.4939	0.5400	-0.0799
Case 2	SimAver	0.4391	0.4387	0.5229	0.0060	0.4288	0.4297	0.5175	0.0371
	MMA	0.5177	0.5172	0.5308	-0.0069	0.5026	0.5025	0.5187	0.0201
	JMA	0.2362	0.2369	0.3146	-0.0306	0.2239	0.2240	0.3116	-0.0179
	ℓ^1	0.3923	0.3922	0.4401	-0.0168	0.3803	0.3800	0.4305	0.0046
	ℓ^2	0.3893	0.3891	0.4378	-0.0170	0.3773	0.3769	0.4284	0.0043
	ℓ^∞	0.2369	0.2376	0.3151	-0.0305	0.2246	0.2247	0.3122	-0.0178
	Nash	0.4265	0.4263	0.4526	-0.0128	0.4104	0.4101	0.4447	0.0076
	AB	0.5451	0.5447	0.5384	-0.0095	0.5284	0.5281	0.5256	0.0168
Case 3	SimAver	0.4513	0.4527	0.5392	-0.0426	0.4734	0.4734	0.5236	-0.0223
	MMA	0.5134	0.5141	0.5490	-0.0348	0.5388	0.5386	0.5350	-0.0194
	JMA	0.2257	0.2368	0.3701	-0.1066	0.2121	0.2242	0.3488	-0.1107
	ℓ^1	0.3849	0.3880	0.4706	-0.0597	0.3703	0.3725	0.4475	-0.0508
	ℓ^2	0.3817	0.3850	0.4686	-0.0603	0.3669	0.3692	0.4454	-0.0516
	ℓ^∞	0.2264	0.2375	0.3706	-0.1064	0.2127	0.2247	0.3492	-0.1104
	Nash	0.4185	0.4209	0.4837	-0.0534	0.4032	0.4050	0.4589	-0.0472
	AB	0.5425	0.5429	0.5589	-0.0307	0.5285	0.5281	0.5390	-0.0122

Entries report the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value of β_4, 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	2.2674	2.3133	1.1795	0.2193	2.1156	2.1508	1.1568	0.1931
	MMA	3.2431	3.2711	1.3844	0.1767	2.9783	2.9960	1.3551	0.1441
	JMA	1.1270	1.1308	0.7181	0.0699	1.1863	1.1899	0.7283	0.0690
	ℓ^1	1.9158	1.9307	1.0093	0.1295	1.8453	1.8582	0.9933	0.1214
	ℓ^2	1.8479	1.8621	0.9883	0.1267	1.7892	1.8017	0.9740	0.1197
	ℓ^∞	1.1300	1.1338	0.7195	0.0702	1.1892	1.1928	0.7296	0.0694
	Nash	2.1342	2.1526	1.0928	0.1436	2.0286	2.0426	1.0706	0.1266
	AB	2.1122	2.1294	1.0771	0.1387	2.0038	2.0183	1.0553	0.1282
Case 2	SimAver	2.1075	2.1720	1.1407	0.2581	2.0545	2.0913	1.1252	0.1970
	MMA	2.9867	3.0038	1.3424	0.1419	2.9344	2.9407	1.3139	0.0959
	JMA	1.1718	1.2002	0.7076	0.1720	1.0711	1.0898	0.6675	0.1408
	ℓ^1	1.8595	1.8849	0.9873	0.1653	1.7852	1.8009	0.9587	0.1322
	ℓ^2	1.7992	1.8253	0.9669	0.1670	1.7226	1.7390	0.9384	0.1344
	ℓ^∞	1.1745	1.2029	0.7089	0.1720	1.0738	1.0926	0.6689	0.1409
	Nash	2.0505	2.0765	1.0690	0.1674	1.9553	1.9702	1.0380	0.1297
	AB	2.0266	2.0526	1.0527	0.1674	1.9457	1.9603	1.0234	0.1286
Case 3	SimAver	2.1343	2.1425	1.1564	0.1017	2.0810	2.1461	1.1649	0.2591
	MMA	2.9237	2.9214	1.3252	0.0251	2.8666	2.9103	1.3352	0.2158
	JMA	1.1286	1.1277	0.7167	0.0161	1.1914	1.2051	0.7302	0.1221
	ℓ^1	1.8264	1.8256	0.9866	0.0318	1.8363	1.8671	0.9922	0.1806
	ℓ^2	1.7683	1.7676	0.9671	0.0322	1.7858	1.8159	0.9740	0.1785
	ℓ^∞	1.1315	1.1306	0.7180	0.0162	1.1941	1.2079	0.7315	0.1225
	Nash	2.0174	2.0170	1.0621	0.0397	2.0457	2.0812	1.0756	0.1939
	AB	2.0022	2.0017	1.0501	0.0378	2.0150	2.0504	1.0577	0.1934

Entries report the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value of β_3, if the 2nd core regressor is 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	0.5387	0.5765	0.5970	0.1958	0.5152	0.5385	0.5859	0.1543
	MMA	0.6168	0.6404	0.6151	0.1557	0.5892	0.5999	0.6034	0.1061
	JMA	0.3935	0.4024	0.4683	0.0963	0.3657	0.3686	0.4515	0.0569
	ℓ^1	0.5117	0.5291	0.5487	0.1340	0.4825	0.4896	0.5337	0.0872
	ℓ^2	0.5096	0.5270	0.5474	0.1336	0.4804	0.4874	0.5323	0.0869
	ℓ^∞	0.3940	0.4029	0.4686	0.0964	0.3662	0.3691	0.4519	0.0570
	Nash	0.5402	0.5587	0.5620	0.1381	0.5089	0.5166	0.5461	0.0908
	AB	0.6240	0.6456	0.6143	0.1491	0.5940	0.6029	0.6010	0.0974
Case 2	SimAver	0.5470	0.6042	0.6115	0.2402	0.5459	0.6145	0.6238	0.2629
	MMA	0.6727	0.7009	0.6348	0.1700	0.6709	0.7064	0.6487	0.1903
	JMA	0.4519	0.4795	0.4833	0.1675	0.4503	0.4874	0.4937	0.1937
	ℓ^1	0.5728	0.6000	0.5706	0.1666	0.5659	0.6018	0.5831	0.1911
	ℓ^2	0.5707	0.5979	0.5692	0.1668	0.5637	0.5997	0.5817	0.1914
	ℓ^∞	0.4524	0.4800	0.4836	0.1675	0.4508	0.4878	0.4940	0.1937
	Nash	0.6012	0.6279	0.5833	0.1651	0.5929	0.6271	0.5941	0.1868
	AB	0.6802	0.7036	0.6321	0.1553	0.6707	0.7013	0.6448	0.1769
Case 3	SimAver	0.4833	0.5220	0.5784	0.1977	0.5042	0.5507	0.5793	0.2168
	MMA	0.5421	0.5648	0.5856	0.1525	0.5870	0.6133	0.5922	0.1639
	JMA	0.3528	0.3611	0.4441	0.0928	0.3906	0.4002	0.4660	0.1000
	ℓ^1	0.4533	0.4701	0.5221	0.1314	0.4940	0.5125	0.5337	0.1379
	ℓ^2	0.4515	0.4682	0.5208	0.1310	0.4921	0.5105	0.5326	0.1376
	ℓ^∞	0.3532	0.3615	0.4444	0.0930	0.3909	0.4006	0.4663	0.1002
	Nash	0.4729	0.4907	0.5322	0.1354	0.5157	0.5352	0.5429	0.1415
	AB	0.5476	0.5678	0.5825	0.1441	0.5907	0.6131	0.5890	0.1517

Entries report the Monte Carlo variance, bias and the MSE-MAE divergences from the DGP value of β_3, if the 2nd core regressor is 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.

Table 5: Monte Carlo Concentration of the Averaging Estimators

Correct Specification			MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
$n = 100$									
Case 1	Homosk.	Sum Sq.	0.6563	0.6782	0.5268	0.5308	0.6772	0.4419	0.4687
		Max. Conc.	8,4	2,4	4,4	4,4	2,4	5,4	5,4
	Heterosk.	Sum Sq.	0.6504	0.6772	0.5214	0.5252	0.6760	0.4388	0.4632
		Max. Conc.	8,4	2,4	4,4	4,4	2,4	5,4	5,4
Case 2	Homosk.	Sum Sq.	0.6651	0.6783	0.5155	0.5198	0.6771	0.4415	0.4659
		Max. Conc.	8,4	2,4	4,4	4,4	2,4	5,5	5,4
	Heterosk.	Sum Sq.	0.6574	0.6865	0.5258	0.5299	0.6854	0.4426	0.4699
		Max. Conc.	8,4	2,4	4,4	4,4	2,4	5,4	5,4
Case 3	Homosk.	Sum Sq.	0.6675	0.6632	0.5149	0.5185	0.6621	0.4415	0.4643
		Max. Conc.	8,4	2,4	4,4	4,4	2,4	5,4	5,4
	Heterosk.	Sum Sq.	0.6707	0.6718	0.5150	0.5185	0.6706	0.4387	0.4655
		Max. Conc.	8,3	2,4	5,4	4,4	2,4	5,4	5,4
$n = 400$									
Case 1	Homosk.	Sum Sq.	0.4925	0.6359	0.5142	0.5149	0.6352	0.4943	0.5273
		Max. Conc.	6,3	2,4	4,4	4,4	2,4	4,4	6,4
	Heterosk.	Sum Sq.	0.5003	0.6458	0.5229	0.5237	0.6450	0.5024	0.5297
		Max. Conc.	6,3	2,4	4,4	4,4	2,4	4,4	5,4
Case 2	Homosk.	Sum Sq.	0.4955	0.6457	0.5100	0.5107	0.6448	0.4962	0.5301
		Max. Conc.	6,3	2,4	4,4	4,4	2,5	4,4	6,4
	Heterosk.	Sum Sq.	0.4948	0.6428	0.5241	0.5249	0.6420	0.5000	0.5313
		Max. Conc.	6,3	2,4	4,4	4,4	2,5	4,4	6,4
Case 3	Homosk.	Sum Sq.	0.4938	0.6391	0.5261	0.5269	0.6384	0.4984	0.5285
		Max. Conc.	6,3	2,4	4,4	4,4	2,4	4,4	5,4
	Heterosk.	Sum Sq.	0.4959	0.6380	0.5131	0.5136	0.6372	0.4940	0.5263
		Max. Conc.	6,3	2,4	4,4	4,4	2,4	4,4	6,4
Misspecification			**MMA**	**JMA**	ℓ^1	ℓ^2	ℓ^∞	**Nash**	**AB**
$n = 100$									
Case 1	Homosk.	Sum Sq.	0.6667	0.6821	0.5445	0.5477	0.6810	0.4524	0.4807
		Max. Conc.	7,3	2,4	4,3	4,3	2,4	4,4	4,4
	Heterosk.	Sum Sq.	0.6711	0.6869	0.5350	0.5386	0.6859	0.4521	0.4786
		Max. Conc.	7,3	2,4	4,4	4,4	2,4	4,4	4,4
Case 2	Homosk.	Sum Sq.	0.6786	0.6559	0.5286	0.5314	0.6550	0.4504	0.4760
		Max. Conc.	7,3	2,4	4,3	4,3	2,4	5,4	5,4
	Heterosk.	Sum Sq.	0.6773	0.6639	0.5196	0.5230	0.6629	0.4418	0.4671
		Max. Conc.	7,3	2,4	4,3	4,3	2,4	5,4	4,4
Case 3	Homosk.	Sum Sq.	0.6770	0.6796	0.5267	0.5303	0.6785	0.4498	0.4743
		Max. Conc.	7,3	2,4	4,3	4,3	2,4	5,4	4,4
	Heterosk.	Sum Sq.	0.6686	0.6823	0.5351	0.5384	0.6813	0.4508	0.4784
		Max. Conc.	7,3	2,4	4,3	4,3	2,4	4,4	4,4
$n = 400$									
Case 1	Homosk.	Sum Sq.	0.5135	0.6287	0.5279	0.5282	0.6280	0.5088	0.5458
		Max. Conc.	5,3	2,3	4,3	4,3	2,4	4,4	5,4
	Heterosk.	Sum Sq.	0.5102	0.6184	0.5222	0.5225	0.6177	0.5070	0.5490
		Max. Conc.	5,3	2,3	4,3	4,3	2,4	4,4	5,4
Case 2	Homosk.	Sum Sq.	0.5112	0.5783	0.5113	0.5113	0.5779	0.4973	0.5478
		Max. Conc.	6,3	3,3	4,3	4,3	3,4	4,4	5,4
	Heterosk.	Sum Sq.	0.5055	0.5821	0.5054	0.5057	0.5815	0.4938	0.5460
		Max. Conc.	6,3	3,3	4,3	4,3	3,4	4,4	6,4
Case 3	Homosk.	Sum Sq.	0.5090	0.6400	0.5269	0.5275	0.6393	0.5103	0.5422
		Max. Conc.	5,3	2,4	4,4	4,3	2,4	4,4	5,4
	Heterosk.	Sum Sq.	0.5151	0.6352	0.5295	0.5299	0.6345	0.5144	0.5593
		Max. Conc.	5,3	2,4	4,3	4,3	2,4	4,4	5,4

Scalar entries report the Monte Carlo mean of the sum of squares of weights of each averaging estimator that appears in the main text. The vector entries' 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[S] which the second maximum weight is attributed. There 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 $1 < m$ 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. In the case of misspecification the second core regressor is erroneously dropped from the analysis.

Empirical Application Results

Table 6: Coefficient estimates with Model A scenario

	SimAver	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
CONSTANT	0.0489	0.0609	0.0018	0.0593	0.0583	0.0567	0.0568	0.0567
	(0.0154)	(0.0193)	(0.0115)	(0.0182)	(0.0176)	(0.0173)	(0.0173)	(0.0173)
GDP60	-0.0123	-0.0155	0.0014	-0.0152	-0.0148	-0.0144	-0.0144	-0.0144
	(0.0022)	(0.003)	(0.0014)	(0.0027)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
INV	0.1942	0.1368	0.1686	0.155	0.1613	0.1708	0.1718	0.1712
	(0.0312)	(0.0399)	(0.015)	(0.0384)	(0.0369)	(0.0369)	(0.037)	(0.0369)
SCHOOL60	0.0175	0.017		0.018	0.0178	0.0182	0.0182	0.0182
	(0.0059)	(0.0085)		(0.0078)	(0.0076)	(0.0074)	(0.0074)	(0.0074)
LIFE60	0.0006	0.0008		0.0008	0.0008	0.0007	0.0007	0.0007
	(0.0002)	(0.0003)		(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
POP	0.1486	0.346		0.3063	0.2762	0.2629	0.2632	0.2629
	(0.0973)	(0.1908)		(0.1685)	(0.1558)	(0.1479)	(0.1482)	(0.148)
LAW	0.009	0.0173		0.0162	0.0147	0.0144	0.0145	0.0144
	(0.0024)	(0.0058)		(0.0051)	(0.0045)	(0.0045)	(0.0045)	(0.0045)
TROPICS	-0.0028	-0.0075		-0.0064	-0.0057	-0.0053	-0.0053	-0.0053
	(0.0012)	(0.0036)		(0.003)	(0.0027)	(0.0025)	(0.0025)	(0.0025)
ETHNO	-0.0021	-0.0077		-0.0056	-0.0055	-0.0047	-0.0047	-0.0047
	(0.0015)	(0.0066)		(0.0048)	(0.0045)	(0.004)	(0.004)	(0.004)
CONFUC	0.0062	0.0561		0.0404	0.0336	0.0339	0.034	0.0339
	(0.0014)	(0.0128)		(0.0092)	(0.0077)	(0.0078)	(0.0078)	(0.0078)

Note: Standard errors are reported in parentheses.

Table 7: Coefficient estimates with Model B scenario

	SimAver	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
CONSTANT	0.0573	0.0609	0.0242	0.0597	0.0602	0.0575	0.0574	0.0575
	(0.0161)	(0.0193)	(0.0117)	(0.0182)	(0.0177)	(0.0173)	(0.0173)	(0.0173)
GDP60	-0.0145	-0.0155	-0.0026	-0.0153	-0.0153	-0.0146	-0.0145	-0.0146
	(0.0024)	(0.003)	(0.0015)	(0.0028)	(0.0027)	(0.0026)	(0.0026)	(0.0026)
INV	0.2184	0.1369	0.36	0.1584	0.1691	0.1771	0.177	0.1771
	(0.0351)	(0.04)	(0.032)	(0.0386)	(0.0378)	(0.0374)	(0.0374)	(0.0374)
SCHOOL60	0.0197	0.017		0.018	0.0183	0.0182	0.0183	0.0182
	(0.0066)	(0.0085)		(0.0079)	(0.0077)	(0.0074)	(0.0074)	(0.0074)
LIFE60	0.0007	0.0008		0.0008	0.0008	0.0007	0.0007	0.0007
	(0.0002)	(0.0003)		(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
POP	0.1672	0.346		0.3063	0.279	0.263	0.2633	0.2631
	(0.1094)	(0.1908)		(0.1684)	(0.1576)	(0.148)	(0.1481)	(0.148)
LAW	0.0102	0.0174		0.0162	0.0149	0.0144	0.0145	0.0144
	(0.0027)	(0.0058)		(0.0051)	(0.0046)	(0.0045)	(0.0045)	(0.0045)
TROPICS	-0.0032	-0.0075		-0.0064	-0.0058	-0.0053	-0.0053	-0.0053
	(0.0013)	(0.0036)		(0.003)	(0.0027)	(0.0025)	(0.0025)	(0.0025)
ETHNO	-0.0023	-0.0077		-0.0056	-0.0055	-0.0047	-0.0047	-0.0047
	(0.0017)	(0.0066)		(0.0048)	(0.0045)	(0.004)	(0.004)	(0.004)
CONFUC	0.007	0.0561		0.0404	0.0337	0.0339	0.034	0.0339
	(0.0016)	(0.0128)		(0.0092)	(0.0077)	(0.0078)	(0.0078)	(0.0078)

Note: Standard errors are reported in parentheses.

Table 8: Coefficient estimates with Model C scenario

	SimAver	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
CONSTANT	0.062	0.0609	0.0479	0.0608	0.062	0.0595	0.0594	0.0595
	(0.0169)	(0.0193)	(0.0128)	(0.0183)	(0.018)	(0.0174)	(0.0174)	(0.0174)
GDP60	-0.0161	-0.0155	-0.0095	-0.0156	-0.0159	-0.0152	-0.0151	-0.0152
	(0.0026)	(0.003)	(0.002)	(0.0028)	(0.0027)	(0.0027)	(0.0027)	(0.0027)
INV	0.1982	0.1368	0.312	0.1559	0.1622	0.1724	0.1724	0.1723
	(0.0361)	(0.04)	(0.0319)	(0.0387)	(0.0382)	(0.0375)	(0.0375)	(0.0375)
SCHOOL60	0.0225	0.017	0.0379	0.0196	0.0195	0.0212	0.0213	0.0212
	(0.0076)	(0.0085)	(0.0064)	(0.0081)	(0.0081)	(0.0078)	(0.0078)	(0.0078)
LIFE60	0.0008	0.0008		0.0008	0.0008	0.0007	0.0007	0.0007
	(0.0002)	(0.0003)		(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
POP	0.1911	0.3461		0.3067	0.2857	0.2641	0.2645	0.2628
	(0.1251)	(0.1909)		(0.1687)	(0.1623)	(0.1489)	(0.1491)	(0.149)
LAW	0.0116	0.0174		0.0162	0.0153	0.0146	0.0146	0.0146
	(0.0031)	(0.0058)		(0.0051)	(0.0047)	(0.0045)	(0.0045)	(0.0045)
TROPICS	-0.0036	-0.0075		-0.0064	-0.0059	-0.0053	-0.0053	-0.0053
	(0.0015)	(0.0036)		(0.003)	(0.0027)	(0.0025)	(0.0025)	(0.0025)
ETHNO	-0.0027	-0.0077		-0.0056	-0.0056	-0.0047	-0.0046	-0.0047
	(0.0019)	(0.0066)		(0.0048)	(0.0046)	(0.004)	(0.004)	(0.004)
CONFUC	0.008	0.0561		0.0403	0.0339	0.0337	0.0338	0.0339
	(0.0018)	(0.0128)		(0.0092)	(0.0078)	(0.0077)	(0.0077)	(0.0077)

Note: Standard errors are reported in parentheses.

Table 9: Coefficient estimates with Model D scenario

	SimAver	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
CONSTANT	0.0644	0.0609	0.0563	0.0615	0.0629	0.0608	0.0607	0.0612
	(0.0179)	(0.0193)	(0.0131)	(0.0184)	(0.0185)	(0.0175)	(0.0175)	(0.0177)
GDP60	-0.0173	-0.0155	-0.0159	-0.0161	-0.0164)	-0.0161	-0.0161	-0.0162
	(0.0027)	(0.003)	(0.0026)	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)
INV	0.1793	0.1367	0.2406	0.1501	0.1546	0.1614	0.1609	0.1595
	(0.0375)	(0.04)	(0.0353)	(0.0391)	(0.0388)	(0.0383)	(0.0383)	(0.0384)
SCHOOL60	0.0199	0.017	0.0181	0.0181	0.0184	0.0182	0.0182	0.0183
	(0.008)	(0.0085)	(0.0077)	(0.0083)	(0.0082)	(0.0081)	(0.0081)	(0.0082)
LIFE60	0.0009	0.0008	0.0011	0.0009	0.0009	0.0009	0.0009	0.0009
	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
POP	0.2229	0.3462		0.3077	0.2976	0.265	0.2668	0.2726
	(0.1459)	(0.191)		(0.1695)	(0.1706)	(0.1493)	(0.1498)	(0.1545)
LAW	0.0135	0.0174		0.0163	0.016	0.0146	0.0147	0.015
	(0.0036)	(0.0058)		(0.0051)	(0.0049)	(0.0045)	(0.0045)	(0.0046)
TROPICS	-0.0042	-0.0075		-0.0064	-0.0061	-0.0054	-0.0054	-0.0055
	(0.0018)	(0.0036)		(0.003)	(0.0028)	(0.0025)	(0.0025)	(0.0026)
ETHNO	-0.0031	-0.0077		-0.0056	-0.0057	-0.0047	-0.0047	-0.0049
	(0.0022)	(0.0066)		(0.0048)	(0.0047)	(0.004)	(0.004)	(0.0042)
CONFUC	0.0094	0.0561		0.0404	0.0345	0.0339	0.0342	0.0343
	(0.0021)	(0.0128)		(0.0092)	(0.0079)	(0.0078)	(0.0078)	(0.0078)

Note: Standard errors are reported in parentheses.

Table 10: Coefficient estimates with Model E scenario

	SimAver	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
CONSTANT	0.066	0.0609	0.0587	0.0618	0.0635	0.0615	0.0626	0.0627
	(0.0193)	(0.0193)	(0.0202)	(0.0192)	(0.0191)	(0.0192)	(0.0191)	(0.0191)
GDP60	-0.0175	-0.0155	-0.016	-0.0162	-0.0165	-0.0163	-0.0164	-0.0164
	(0.0027)	(0.003)	(0.0028)	(0.0029)	(0.0028)	(0.0028)	(0.0028)	(0.0028)
INV	0.167	0.1367	0.2405	0.1463	0.1509	0.1535	0.1511	0.1517
	(0.0384)	(0.04)	(0.0353)	(0.0395)	(0.0392)	(0.0391)	(0.0392)	(0.0392)
SCHOOL60	0.0203	0.017	0.0184	0.0183	0.0185	0.0185	0.0185	0.0186
	(0.0081)	(0.0085)	(0.0079)	(0.0083)	(0.0083)	(0.0082)	(0.0083)	(0.0083)
LIFE60	0.0009	0.0008	0.001	0.0008	0.0009	0.0009	0.0009	0.0009
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
POP	0.2675	0.3464	-0.0341	0.3156	0.3112	0.282	0.3037	0.3011
	(0.1751)	(0.1911)	(0.1635)	(0.1843)	(0.1828)	(0.1803)	(0.1822)	(0.1819)
LAW	0.0163	0.0174		0.0173	0.0169	0.0167	0.0168	0.0168
	(0.0044)	(0.0058)		(0.0054)	(0.0051)	(0.0051)	(0.0051)	(0.0051)
TROPICS	-0.0051	-0.0075		-0.0065	-0.0064	-0.0054	-0.0062	-0.0061
	(0.0021)	(0.0036)		(0.003)	(0.0029)	(0.0025)	(0.0028)	(0.0028)
ETHNO	-0.0037	-0.0077		-0.0056	-0.0059	-0.0047	-0.0055	-0.0054
	(0.0027)	(0.0066)		(0.0048)	(0.0048)	(0.004)	(0.0046)	(0.0045)
CONFUC	0.0112	0.0561		0.0404	0.0348	0.0341	0.0355	0.0346
	(0.0026)	(0.0128)		(0.0092)	(0.008)	(0.0078)	(0.0081)	(0.0079)

Note: Standard errors are reported in parentheses.

Table 11: 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	0.04	0.04	0.06	0.07	0.07
3	0.00	0.00	0.04	0.04	0.08	0.08	0.07
4	0.00	0.00	0.02	0.05	0.05	0.04	0.05
5	0.00	0.00	0.00	0.04	0.00	0.00	0.00
6	0.00	0.00	0.07	0.06	0.10	0.11	0.11
7	0.00	0.00	0.10	0.06	0.08	0.08	0.08
8	0.00	0.00	0.00	0.08	0.01	0.00	0.00
9	1.00	0.00	0.72	0.60	0.60	0.61	0.60

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	0.08
2	0.00	0.00	0.04	0.05	0.08	0.08	0.08
3	0.00	0.00	0.02	0.05	0.05	0.05	0.05
4	0.00	0.00	0.00	0.04	0.00	0.00	0.00
5	0.00	0.00	0.07	0.06	0.11	0.11	0.11
6	0.00	0.00	0.10	0.07	0.08	0.08	0.08
7	0.00	0.00	0.01	0.08	0.00	0.00	0.00
8	1.00	0.00	0.72	0.60	0.60	0.60	0.60

Panel C. Scenario C

Model	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
1	0.00	1.00	0.08	0.06	0.15	0.15	0.15
2	0.00	0.00	0.02	0.06	0.05	0.04	0.05
3	0.00	0.00	0.00	0.05	0.00	0.00	0.00
4	0.00	0.00	0.07	0.07	0.11	0.12	0.12
5	0.00	0.00	0.10	0.07	0.09	0.09	0.08
6	0.00	0.00	0.01	0.09	0.00	0.00	0.00
7	1.00	0.00	0.72	0.60	0.60	0.60	0.60

Panel D. Scenario D

Model	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
1	0.00	1.00	0.10	0.08	0.20	0.19	0.17
2	0.00	0.00	0.00	0.06	0.00	0.00	0.01
3	0.00	0.00	0.07	0.08	0.11	0.11	0.11
4	0.00	0.00	0.10	0.08	0.09	0.09	0.08
5	0.00	0.00	0.01	0.09	0.00	0.00	0.02
6	1.00	0.00	0.72	0.61	0.60	0.61	0.61

Panel E. Scenario E

Model	MMA	JMA	ℓ^1	ℓ^2	ℓ^∞	Nash	AB
1	0.00	1.00	0.05	0.10	0.10	0.09	0.09
2	0.00	0.00	0.12	0.08	0.20	0.12	0.13
3	0.00	0.00	0.10	0.09	0.09	0.10	0.10
4	0.00	0.00	0.01	0.11	0.00	0.06	0.06
5	1.00	0.00	0.72	0.62	0.61	0.63	0.62

Table 12: 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

Panel C. Regressors for model C scenario

Model	Regressors
1	CONSTANT,GDP60,INV, SCHOOL60
2	CONSTANT,GDP60,INV,SCHOOL60, LIFE60
3	CONSTANT,GDP60,INV,SCHOOL60,LIFE60, POP
4	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP, LAW
5	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW, TROPICS
6	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS, ETHNO
7	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS,ETHNO,CONFUC

Panel D. Regressors for model D scenario

Model	Regressors
1	CONSTANT,GDP60,INV,SCHOOL60, LIFE60
2	CONSTANT,GDP60,INV,SCHOOL60,LIFE60, POP
3	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP, LAW
4	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW, TROPICS
5	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS, ETHNO
6	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS,ETHNO,CONFUC

Panel E. Regressors for model E scenario

Model	Regressors
1	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP
2	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP, LAW
3	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW, TROPICS
4	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS, ETHNO
5	CONSTANT,GDP60,INV,SCHOOL60,LIFE60,POP,LAW,TROPICS,ETHNO,CONFUC