Unit Averaging for Heterogeneous Panels

Christian Brownlees
UPF and BSE

Vladislav Morozov
UPF and BSE

Contents

Problem: Estimation of Individual Parameter

- Object of interest: parameter θ in a potentially nonlinear model (can be anything). For example – quarterly GDP nowcast for a fixed country, a multiplier, etc..
- We have a panel of time series, but every unit i has its own θ_i. Example: cross-country heterogeneity (Marcellino et al. 2003)

This is the problem of estimating a unit-specific parameter. Examples include forecasting (e.g. Baltagi (2013); Zhang et al. (2014); Wang et al. (2019); Liu et al. (2020)), slopes (Maddala et al., 1997; Wang et al., 2019), long-run effects (Pesaran and Smith, 1995; Pesaran et al., 1999), etc.

Problem: Estimation of Individual Parameter

- Object of interest: parameter θ in a potentially nonlinear model (can be anything). For example – quarterly GDP nowcast for a fixed country, a multiplier, etc..
- We have a panel of time series, but every unit i has its own θ_i. Example: cross-country heterogeneity (Marcellino et al. 2003)

This is the problem of estimating a unit-specific parameter. Examples include forecasting (e.g. Baltagi (2013); Zhang et al. (2014); Wang et al. (2019); Liu et al. (2020)), slopes (Maddala et al., 1997; Wang et al., 2019), long-run effects (Pesaran and Smith, 1995; Pesaran et al., 1999), etc.

Using Panel Data

How to estimate θ with minimal MSE?

Answer depends on time series length T:

- T large $\Rightarrow$ just use data on unit of interest
- If T is not large, individual estimator is not very precise.
  In this case hope to use panel information to reduce estimation uncertainty without incurring too much bias.

Interesting case: moderate T – when potential bias and variance are of the same magnitude $\leftarrow$ our paper.

Problem and Estimator Considered

In the paper we consider a heterogeneous M-estimation problem. Define the individual estimator for unit i as

$$\hat{\boldsymbol{\theta}}_i = \underset{\theta_i \in \Theta_i \subset \mathbb{R}^p}{\arg\min}\ T^{-1} \sum_{t=1}^{T} m(\boldsymbol{\theta}_i, \boldsymbol{z}_{it})$$

Interest in estimating $\boldsymbol{\theta}_1$ with minimal MSE.

Note: in the paper we discuss $\mu(\boldsymbol{\theta}_1)$ for smooth $\mu(\cdot)$

Unit Averaging Estimation Definition

Individual slope can be written

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_i, \quad \mathbb{E}(\boldsymbol{\eta}_i) = 0$$

Every unit carries information about common mean $\boldsymbol{\theta}_0$. This information is valuable for estimating $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_1$. Bias-variance trade-off: using information on other units reduces uncertainty about θ_0, but creates bias due to η_i.

Idea: consider linear combinations of all units – unit averaging

$$\hat{\theta}_1(\boldsymbol{w}) = \sum_{i=1}^{N} w_i \hat{\theta}_i, \quad w_i \geq 0, \sum_{i=1}^{N} w_i = 1$$

Unit Averaging Estimation Definition

Individual slope can be written

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_i, \quad \mathbb{E}(\boldsymbol{\eta}_i) = 0$$

Every unit carries information about common mean $\boldsymbol{\theta}_0$. This information is valuable for estimating $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_1$. Bias-variance trade-off: using information on other units reduces uncertainty about $\boldsymbol{\theta}_0$, but creates bias due to $\boldsymbol{\eta}_i$.

Idea: consider linear combinations of all units – unit averaging

$$\hat{\theta}_1(\boldsymbol{w}) = \sum_{i=1}^{N} w_i \hat{\theta}_i, \quad w_i \geq 0, \sum_{i=1}^{N} w_i = 1$$

Unit Averaging Estimation Definition

Individual slope can be written

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_i, \quad \mathbb{E}(\boldsymbol{\eta}_i) = 0$$

Every unit carries information about common mean $\boldsymbol{\theta}_0$. This information is valuable for estimating $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_1$. Bias-variance trade-off: using information on other units reduces uncertainty about $\boldsymbol{\theta}_0$, but creates bias due to $\boldsymbol{\eta}_i$.

Idea: consider linear combinations of all units – unit averaging

$$\hat{\boldsymbol{\theta}}_1(\boldsymbol{w}) = \sum_{i=1}^{N} w_i \hat{\boldsymbol{\theta}}_i, \quad w_i \geq 0, \sum_{i=1}^{N} w_i = 1$$

Simple average of slopes – all bias, individual estimator $\hat{\boldsymbol{\theta}}_1$ – all variance. Averaging estimator – compromise

Main Technique: Local Heterogeneity

Question: how to pick weights? To minimize the MSE, we need the MSE

No useful exact finite-sample results at such level of generality. Instead use an approximation by assuming local heterogeneity:

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_0 + \frac{\boldsymbol{\eta}_i}{\sqrt{T}}$$

Allows using asymptotic analysis techniques to approximate a finite-sample setting. Intuitively: overall amount of information is fixed as $T \to \infty$. Bias remains bounded and nontrivial. This creates a bias-variance trade-off asymptotically

Similar to frequentist model averaging approach (used by Hjort and Claeskens (2003); Claeskens and Hjort (2008)) or Hansen (2016, 2017) for shrinkage estimators

Our Results: Theory and Application

Theoretical results: in a moderate-T/local heterogeneity regime:

- Formally justified MSE approximation
- Feasible weights that minimize an MSE estimator and asymptotic distribution of averaging estimator
- Analysis depending on behavior of N: fixed-N and large-N approximations

Application: does unit averaging work in simulations and in practice? Yes! We do nowcasting quarterly GDP for Eurozone members.

- Unit averaging AIC weights on average 5% better than individual estimation.
- Our MSE-optimal weights on average 9% better.
- Equal weights – average 50% worse

Unit averaging with smooth weights leads to improvements.

Our Results: Theory and Application

Theoretical results: in a moderate-T/local heterogeneity regime:

- Formally justified MSE approximation
- Feasible weights that minimize an MSE estimator and asymptotic distribution of averaging estimator
- Analysis depending on behavior of N: fixed-N and large-N approximations

Application: does unit averaging work in simulations and in practice? Yes! We do nowcasting quarterly GDP for Eurozone members.

- Unit averaging AIC weights on average 5% better than individual estimation.
- Our MSE-optimal weights on average 9% better.
- Equal weights – average 50% worse

Unit averaging with smooth weights leads to improvements.

Asymptotic Distribution of Individual Estimators

Basic building block of averaging – things to be averaged.

Lemma

As $T \to \infty$, the individual estimators satisfy

$$\sqrt{T}\left(\hat{\theta}_i - \theta_1\right) \Rightarrow N(\eta_i - \eta_1, \boldsymbol{V}_i) = \boldsymbol{Z}_i$$

$\boldsymbol{Z}_i$ are independent.

Important: $T \to \infty$ is taken in the local approximation sense. Amount of information is each time series is limited and not growing.

Local asymptotic approximation reduces the intractable finite sample problem to the first two moments – bias and variance. These are exactly the components of interest for analyzing the MSE

Asymptotic Distribution of Individual Estimators

Basic building block of averaging – things to be averaged.

Lemma

As $T \to \infty$, the individual estimators satisfy

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1\right) \Rightarrow N(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1, \boldsymbol{V}_i) = \boldsymbol{Z}_i$$

$\boldsymbol{Z}_i$ are independent.

Important: $T \to \infty$ is taken in the local approximation sense. Amount of information is each time series is limited and not growing.
Local asymptotic approximation reduces the intractable finite sample problem to the first two moments – bias and variance. These are exactly the components of interest for analyzing the MSE

MSE of Individual Estimators

From the above lemma, we get that

Lemma

Local asymptotic approximation to the MSE of using $\hat{\boldsymbol{\theta}}_i$ as an estimator for $\boldsymbol{\theta}_1$ is

$$LA\text{-}MSE\left(\hat{\boldsymbol{\theta}}_i\right) = (\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)^2 + \boldsymbol{V}_i$$

Controlled by distance $(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)$ from unit 1 and individual variances $\boldsymbol{V}_i$. Differences between units – ground for trade-off

MSE of Unit Averaging Estimator: Local Asymptotic Approximation

Theorem

Let $\{\boldsymbol{w}_N\}$ satisfy (1) $\sum_{i=1}^N w_{i\,N} = 1$, $w_{j\,N} = 0$ for $j > N$, (2) $\boldsymbol{w}_N$ converges to some weights $\boldsymbol{w}$ such that such that $w_i \geq 0$ and $\sum_{i=1}^{\infty} w_i \leq 1$.

Then as $N, T \to \infty$ jointly

$$T \times MSE\left(\hat{\boldsymbol{\theta}}_N(\boldsymbol{w}_N)\right) \to \left(\sum_{i=1}^{\infty} w_i \boldsymbol{\eta}_i - \boldsymbol{\eta}_1\right)^2 + \sum_{i=1}^{\infty} w_i^2 \boldsymbol{V}_i =: LA\text{-}MSE(\boldsymbol{w}).$$

Note: the limit weights $\boldsymbol{w}$ may sum to less than 1. Example: equal weights $\boldsymbol{w}_N$: $w_{iN} = N^{-1}\mathbb{I}\{i \leq N\}$. $\boldsymbol{w}_N$ converges uniformly to $\boldsymbol{w} = 0$ (this is the mean group estimator)

Towards Optimal Weights

The population MSE-optimal weights are just the minimizer of LA-MSE.

However, these are infeasible for two reasons

1. LA-MSE depends on unknown individual parameters $\{\eta_i\}_{i=1}^{\infty}$ and variances V_i
2. N is not infinite

Estimating Unknown Parameters

The unknown variances $\boldsymbol{V}_i$ are usually straightforward to estimate.

More complex situation for η_i:

1. It cannot be consistently estimated.
   Intuition: locality in T essentially emulates T fixed, and η_i are parameters which can only be estimated from individual time series. However, under locality the amount information in each time series is finite and not growing.
2. Next best thing: use asymptotically unbiased estimators:

$$\sqrt{T}\left(\hat{\theta}_i - \hat{\theta}_1\right) \Rightarrow N\left(\eta_i - \eta_1, \boldsymbol{V}_i + \boldsymbol{V}_1\right) = \boldsymbol{Z}_i - \boldsymbol{Z}_1,$$

$$\sqrt{T}\left(\hat{\theta}_1 - \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i\right) \Rightarrow N(\eta_1, \boldsymbol{V}_1) = \boldsymbol{Z}_1 + \eta_1.$$

Estimating Unknown Parameters

The unknown variances $\boldsymbol{V}_i$ are usually straightforward to estimate.

More complex situation for $\boldsymbol{\eta}_i$:

1. It cannot be consistently estimated.
   Intuition: locality in T essentially emulates T fixed, and $\boldsymbol{\eta}_i$ are parameters which can only be estimated from individual time series. However, under locality the amount information in each time series is finite and not growing.
2. Next best thing: use asymptotically unbiased estimators:

$$\sqrt{T}\left(\hat{\theta}_i - \hat{\theta}_1\right) \Rightarrow N\left(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1, \boldsymbol{V}_i + \boldsymbol{V}_1\right) = \boldsymbol{Z}_i - \boldsymbol{Z}_1,$$

$$\sqrt{T}\left(\hat{\theta}_1 - \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i\right) \Rightarrow N(\boldsymbol{\eta}_1, \boldsymbol{V}_1) = \boldsymbol{Z}_1 + \boldsymbol{\eta}_1.$$

Two Averaging Regimes I

Second problem: N is finite in practice. Two options for dealing with the infinite sums in LA-MSE:

1. Treat N as fixed at some value $\bar{N}$. Then

$$LA\text{-}MSE(\boldsymbol{w}) = \left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{\eta}_i - \boldsymbol{\eta}_1 \right)^2 + \sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{V}_i$$

   Appropriate when the number of cross-sectional units is not large. More generally, when every unit can potentially have a non-negligible weight

Two Averaging Regimes II

2. Treat N is large/growing. Then mechanically some units must have small weights. Let $\bar{N}$ units have potentially non-negligible weights ($\bar{N} \leq N$), put these units first. Suppose the weights of other units satisfy $w_{i\,N} = o(N^{-1/2})$. Then units beyond $\bar{N}$ do not contribute to variance, but contribute to bias:

$$LA\text{-}MSE(\boldsymbol{w}) = \left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{\eta}_i - \boldsymbol{\eta}_1\right)^2 + \sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{V}_i$$
$$+ \left(\left(1 - \sum_{i=1}^{\bar{N}} w_i\right)\boldsymbol{\eta}_1 - 2\sum_{i=1}^{\bar{N}} w_i(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)\right)\left(1 - \sum_{i=1}^{\bar{N}} w_i\right)\boldsymbol{\eta}_1 \, .$$

Potentially all weight mass placed beyond $\bar{N}$, but each weight individually negligible

Feasible Optimal Weights

Regardless of the adopted approach, let $\widehat{LA\text{-}MSE}$ be the corresponding $LA\text{-}MSE$. Define the fixed-N/large-N minimum MSE weights as

$$\hat{\boldsymbol{w}}^{\bar{N}} = \arg\min \widehat{LA\text{-}MSE}(\boldsymbol{w}) \ ,$$

where the minimum is taken over $\bar{N}$-vectors $\boldsymbol{w}$ such that $w_i \geq 0$ and

1. Fixed-N: $\sum_{i=1}^{\bar{N}} w_i = 1$
2. Large-N: $\sum_{i=1}^{\bar{N}} w_i \leq 1$

This is a strictly convex quadratic program.

Asymptotic Properties for Minimum MSE Weights

We show that

1. $\widehat{LA\text{-}MSE}$ converges to a quadratic function of $\boldsymbol{Z}_i(\Leftarrow \sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1))$.
2. The feasible weights converge to the minimizer of that limiting function
3. The minimum MSE unit averaging estimator weakly converges to a randomly weighted sum of normals – a non-standard distribution.
   In the appendix to the paper we discuss how to construct a correctly-sized confidence interval on the basis of the unit averaging estimator with minimum MSE weights.

Minimum MSE weights solve the ideal population problem of minimizing MSE + some zero-mean noise + some bias that preserves the ranking between units in terms of variance (bias is the price of the sample problem always being positive-definite, may be removed)

Asymptotic Properties for Minimum MSE Weights

We show that

1. $\widehat{LA\text{-}MSE}$ converges to a quadratic function of $\boldsymbol{Z}_i(\Leftarrow \sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1))$.
2. The feasible weights converge to the minimizer of that limiting function
3. The minimum MSE unit averaging estimator weakly converges to a randomly weighted sum of normals – a non-standard distribution.
   In the appendix to the paper we discuss how to construct a correctly-sized confidence interval on the basis of the unit averaging estimator with minimum MSE weights.

Minimum MSE weights solve the ideal population problem of minimizing MSE + some zero-mean noise + some bias that preserves the ranking between units in terms of variance (bias is the price of the sample problem always being positive-definite, may be removed)

Minimal MSE Weights In a Large T Setting

It is natural to minimize $\widehat{LA\text{-}MSE}$ is natural even in a non-local setting with growing amount of information:

1. For all i with $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_1$, the bias estimators $\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_1)$ will diverge
2. Variance terms remain bounded.

$\Rightarrow$ Procedure will place asymptotically zero weight on all units with $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_1$.

Parallels a similar result in model averaging where (see Fang et al. (2022))

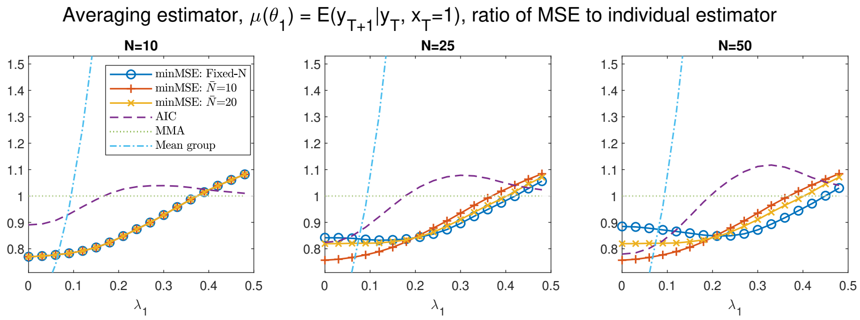
Simulation: Setup

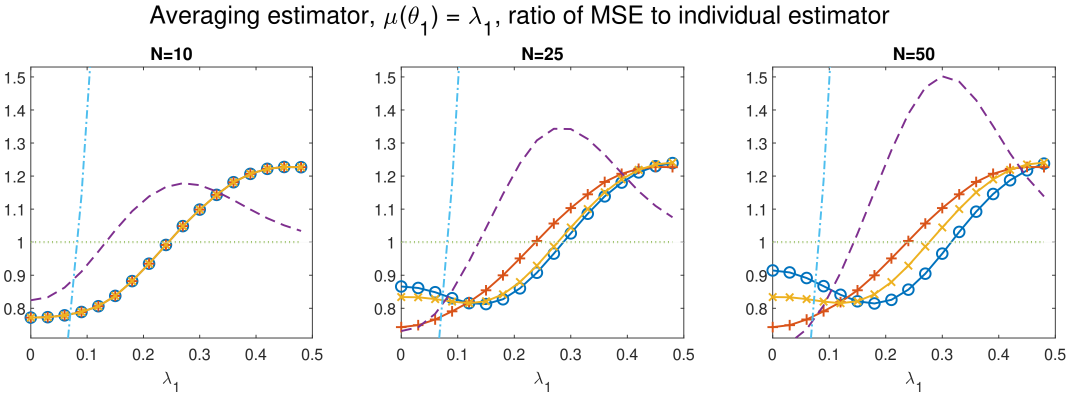
Dynamic panel DGP, similar to the empirical application:

$$y_{it} = \lambda_i y_{it-1} + \beta_i x_{it} + u_{it}$$
$$\lambda_i = \mathbb{E}(\lambda_i) + \eta_i, \quad \mathbb{E}(\lambda_i) = 0$$

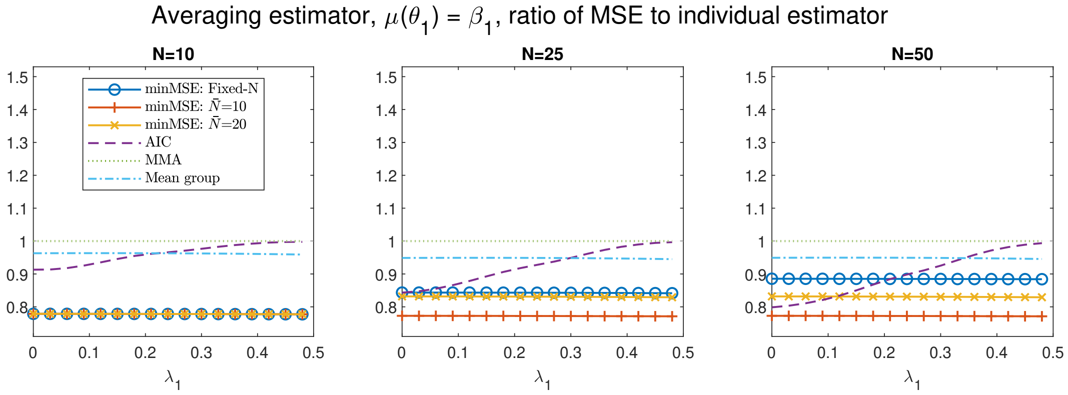
We compare performance of minimum MSE weights compared to AIC/BIC (Buckland et al., 1997), mean group (equal weights), MMA Hansen (2007)

We conduct simulations for the one-step ahead forecast for y, coefficients λ_1 and β_1, and the long-run effect of a change in x: $\beta_1/(1-\lambda_)1$

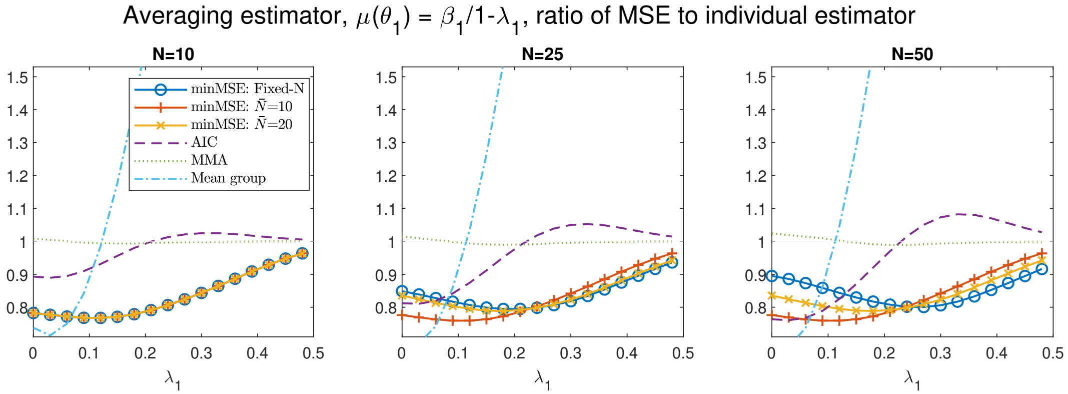

Averaging estimator, $\mu(\theta_1)$ = E(y_{T+1}|y_T, x_T=1), ratio of MSE to individual estimator
N=10
N=25
N=50
minMSE: Fixed-N
minMSE: $\bar{N}$=10
minMSE: $\bar{N}$=20
AIC
MMA
Mean group
λ_1



Averaging estimator, $\mu(\theta_1) = \lambda_1$, ratio of MSE to individual estimator
N=10
N=25
N=50
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0
0.1
0.2
0.3
0.4
0.5
λ_1



Averaging estimator, $\mu(\theta_1) = \beta_1$, ratio of MSE to individual estimator
N=10
N=25
N=50
minMSE: Fixed-N
minMSE: $\bar{N}$=10
minMSE: $\bar{N}$=20
AIC
MMA
Mean group
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0
0.1
0.2
0.3
0.4
0.5
λ_1



Averaging estimator, $\mu(\theta_1)$ = β_1/1-λ_1, ratio of MSE to individual estimator
N=10
N=25
N=50
minMSE: Fixed-N
minMSE: $\bar{N}$=10
minMSE: $\bar{N}$=20
AIC
MMA
Mean group
λ_1
0
0.1
0.2
0.3
0.4
0.5
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5

Empirical Application: GDP Nowcasting

Empirical application – nowcasting GDP for founding members of the Eurozone + UK.

Natural application for unit averaging:

1. Evidence of significant cross-country heterogeneity (Marcellino et al., 2003)
2. Partial pooling of data may improve performance (Garcia-Ferrer et al., 1987; Hoogstrate et al., 2000)

We follow standard practices in nowcasting literature (e.g. Schumacher (2016)):

1. We account for delays in data publication ("ragged edge") and different possible times in the quarter for nowcasting ("vintages").
2. Nowcasting using factor unrestricted MIDAS (Foroni et al., 2015). Factors estimated by EM-PCA (Stock and Watson, 1999).
3. Factors estimated using real, financial, and survey data (up to $\approx$ 160 vars/country)
4. Estimation is done using a rolling window.

Empirical Application: GDP Nowcasting

Empirical application – nowcasting GDP for founding members of the Eurozone + UK. Natural application for unit averaging:

1. Evidence of significant cross-country heterogeneity (Marcellino et al., 2003)
2. Partial pooling of data may improve performance (Garcia-Ferrer et al., 1987; Hoogstrate et al., 2000)

We follow standard practices in nowcasting literature (e.g. Schumacher (2016)):

1. We account for delays in data publication ("ragged edge") and different possible times in the quarter for nowcasting ("vintages").
2. Nowcasting using factor unrestricted MIDAS (Foroni et al., 2015). Factors estimated by EM-PCA (Stock and Watson, 1999).
3. Factors estimated using real, financial, and survey data (up to $\approx$ 160 vars/country)
4. Estimation is done using a rolling window.

Empirical Application: Summary of Results

Unit averaging works!

1. Using smooth data-dependent averaging weights (minimum MSE and AIC) leads to improvement in nowcasting performance.
   AIC: 5% improvement on average.
   mMSE: 9% improvement on average.
2. Using MMA and equal (mean group) weights does not lead to improvements.
   Equal weights: 50% worse on average (no forecast combination puzzle)
3. Averaging is most beneficial for smaller T – magnitude of improvement is shrinking as sample size increases. This is intuitive: averaging estimators converge to the individual estimator
4. Gains from averaging are heterogeneous across countries

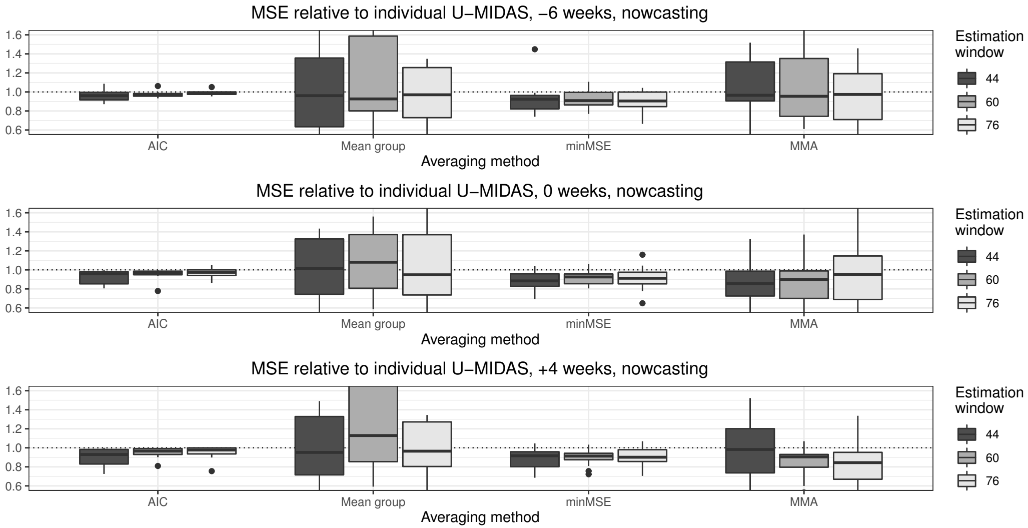


Figure: Distribution of MSEs across countries. MSE relative to individual estimator. Split by different averaging strategies and estimation window size (quarters). Nowcasting done at selected positions relative to end of quarter

Conclusions

- If T is moderate, there is bias-variance tradeoff in using other units.
- We propose and study an averaging estimator that uses all units in the panel
- We characterize a local asymptotic approximation to the MSE of averaging. There are two key regimes of averaging depending on magnitude of N
- Minimum MSE solve the population MSE minimization problem + rank-preserving bias + mean-zero noise.
- Unit averaging with deterministic weights: asymptotic normality. With random weights: randomly weighted sum of normals.
- Estimator performs favorable in simulations.
- Empirical application to GDP nowcasting: there are are gains from using unit averaging, especially minimum MSE weights.

References I

B. H. Baltagi. *Panel data forecasting*, volume 2. Elsevier B.V., 2013.

S. T. Buckland, K. P. Burnham, and N. H. Augustin. Model Selection: An Integral Part of Inference. *Biometrics*, 53(2):603–618, 1997.

G. Claeskens and N. L. Hjort. *Model Selection and Model Averaging*. Cambridge University Press, Cambridge, 2008.

F. Fang, C. Yuan, and W. Tian. An Asymptotic Theory for Least Squares Model Averaging with Nested Models. *Econometric Theory*, pages 1–30, 2022.

C. Foroni, M. Marcellino, and C. Schumacher. Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials. *Journal of the Royal Statistical Society: Series A*, 178(1):57–82, 2015.

References II

A. Garcia-Ferrer, R. A. Highfield, F. C. Palm, and A. Zellner. Macroeconomic Forecasting Using Pooled International Data. *Journal of Business and Economic Statistics*, 5(1):53–67, 1987.

D. Giannone, L. Reichlin, and D. Small. Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, 55(4):665–676, 2008.

B. E. Hansen. Least squares model averaging. *Econometrica*, 75(4):1175–1189, 2007.

B. E. Hansen. Efficient shrinkage in parametric models. *Journal of Econometrics*, 190(1): 115–132, 2016.

B. E. Hansen. Stein-like 2SLS estimator. *Econometric Reviews*, 36(6-9):840–852, 2017.

References III

N. L. Hjort and G. Claeskens. Frequentist Model Average Estimators. *Journal of the American Statistical Association*, 98(464):879–899, 2003.

A. J. Hoogstrate, F. C. Palm, and G. A. Pfann. Pooling in Dynamic Panel-Data Models: An Application to Forecasting GDP Growth Rates. *Journal of Business and Economic Statistics*, 18(3):274–283, 2000.

L. Liu, H. R. Moon, and F. Schorfheide. Forecasting with Dynamic Pane Data Models. *Econometrica*, 88(1):171–201, 2020.

G. S. Maddala, R. P. Trost, H. Li, and F. Joutz. Estimation of Short-Run and Long-Run Elasticities of Energy Demand From Panel Data Using Shrinkage Estimators. *Journal of Business and Economic Statistics*, 15(1):90–100, 1997.

References IV

M. Marcellino, J. H. Stock, and M. W. Watson. Macroeconomic Forecasting in the Euro Area: Country Specific Versus Area-Wide Information. *European Economic Review*, 47 (1):1–18, 2003.

M. H. Pesaran and R. P. Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 6061:473–477, 1995.

M. H. Pesaran, Y. Shin, and R. P. Smith. Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association*, 94(446): 621–634, 1999.

G. Rünstler, K. Barhoum, S. Benk, R. Cristadoro, A. D. Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. van Nieuwenhuyze. Short-Term Forecasting of GDP Using Large Datasets: A Pseudo Real-Time Forecast Evaluation Exercise. *Journal of Forecasting*, 28(7):595–611, 2009.

References V

C. Schumacher. A Comparison of MIDAS and Bridge Equations. *International Journal of Forecasting*, 32(2):257–270, 2016.

J. H. Stock and M. W. Watson. A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In R. F. Engle and H. White, editors, *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*, pages 1–44. Oxford University Press, 1999.

K. F. Wallis. Forecasting with an Econometric Model: The 'Ragged Edge' Problem. *Journal of Forecasting*, 5(1):1–13, 1986.

A. T. K. Wan, X. Zhang, and G. Zou. Least Squares Model Averaging by Mallows Criterion. *Journal of Econometrics*, 156(2):277–283, 2010.

References VI

W. Wang, X. Zhang, and R. Paap. To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions? *Journal of Applied Econometrics*, 34(5):724–745, 2019.

X. Zhang, G. Zou, and H. Liang. Model averaging and weight choice in linear mixed-effects models. *Biometrika*, 101(1):205–218, 2014.