Unit Averaging for Heterogeneous Panels

Christian Brownlees

Vladislav Morozov

Problem: Estimation of Individual Parameter

Object of interest: parameter μ in a potentially nonlinear and dynamic model

μ could be anything:

- Structural parameters: individual marginal effects, elasticities, ...
- Forecasts: realized volatility, GDP, ...

We have a panel of time series, but is is heterogeneous — every unit has its own μ_i

Example: cross-country/sectoral/etc heterogeneity

Goal

Estimate μ of a given unit with minimal MSE

Interest in a unit-specific parameter, not a population-wide one!

Example: Prediction in Linear Model

Linear model with unit-specific heterogeneity:

$$y_{it} = \boldsymbol{\theta}_i' \boldsymbol{x}_{it} + u_{it}, \quad \mathbb{E}[u_{it}|\boldsymbol{x}_{it}] = 0, \quad i = 1, \dots, t = 1, \dots, T.$$

Goal is MSE-optimal prediction of y for unit 1
$\Rightarrow$ the parameter of interest is

$$\mu(\boldsymbol{\theta}_1) = \mathbb{E}[y_{1T+1}|\boldsymbol{x}_{iT+1}] = \boldsymbol{\theta}_1 \boldsymbol{x}_{1T+1}.$$

Parameter specific to unit 1

Formal Setting: Individual Parameters

Unit differ in some individual parameters $\boldsymbol{\theta}_i$ that satisfy

$$\boldsymbol{\theta}_i = \underset{\boldsymbol{\theta}\in\Theta}{\arg\max}\, \mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T} m(\boldsymbol{\theta}, \boldsymbol{z}_{it})\right)$$

M-estimation problem with

- m Some known smooth function, may be nonlinear
- z_{it} Observed data, may include lags of variables

Object of interest: $\mu(\boldsymbol{\theta}_1)$, where μ is some known smooth function

This presentation: $\mu(\theta_1) \equiv \theta_1$ and θ_1 scalar for simplicity; general case – in the paper

How to estimate $\boldsymbol{\theta}_1$ with minimal MSE?

Is there is a nontrivial bias-variance trade-off between individual-specific and panel-wide information:

Yes: moderate-T case — the setting of this paper
Heuristic criterion: t-statistic on unit-specific estimates between 1 and 5
$\Rightarrow$ unit averaging approach we propose

No: Either bias or variance is large relative to the other one (large-T and small-T settings)
$\Rightarrow$ (large-T): just use the individual estimator of the target unit
$\Rightarrow$ (small-T): pool data in estimation and/or Bayesian approaches (some results for linear models given by Maddala et al. (1997); Wang et al. (2019); Liu et al. (2021))

Our Solution: Unit Averaging

Our estimator for parameter of interest θ for the fixed unit of interest: a compromise unit averaging estimator:

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

where $\hat{\theta}_i$ is the individual estimator of unit i:

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

Fairly broad class of estimators: nests individual-specific, mean group, Stein-like, information criteria-weighted, etc.

Motivation for Unit Averaging

Averaging estimator:

$$\hat{\theta}(\boldsymbol{w}) = \sum_{i=1}^{N} w_i \hat{\boldsymbol{\theta}}_i.$$

Why should unit averaging lower the MSE?

- Every individual θ_i can be written as

$$\theta_i = \mathbb{E}[\theta_i] + \eta_i, \quad \mathbb{E}[\eta_i] = 0,$$

- If $\mathbb{E}[\theta_i] = \theta_0$ for some common θ_0, every unit i has info about the component component θ_0
  $\Rightarrow$ Using info from other units lowers uncertainty about θ_0

Approach for Choosing Weights

Averaging estimator:

$$\hat{\theta}(\boldsymbol{w}) = \sum_{i=1}^{N} w_i \hat{\theta}_i.$$

How to pick weights $\boldsymbol{w}$ to minimize MSE? Target the unit of interest.

- We derive leading terms of the exact moderate-T MSE of $\hat{\theta}(\boldsymbol{w})$ and provide a plug-in estimator
- Feasible weights are obtained by minimizing estimated MSE.
- We propose two schemes: one uses prior information about unit similarity; the other one agnostic

Overview of Results: Theory

Contribution

We give a tractable approach for unit-specific parameters in potentially nonlinear and dynamic models and show that it has good properties

We discuss theoretical properties in two cases:

- Moderate-T (limited information regime)
- Large-T (growing information regime)

Results in moderate-T:

- Formal derivation of leading terms of the MSE (and some other risk functions)
- Asymptotic distribution of averaging estimator and feasible weights. Inference on the target parameter

Results in large-T case: estimator is safe to use:

- The estimator does not use units with $\theta_i \neq$ target value

Overview of Results: Applications

Application: does unit averaging work in simulations and in practice? Yes!
We do two applications:

- Forecasting regional unemployment rates for a panel of German labor market districts
- Nowcasting GDP for a panel of European countries (Online Appendix)

In both cases our methodology performs favorably:

- Our MSE-optimal weights improve on individual estimator (38% percent average improvement for unemployment; 9% average improvement for nowcasting)
- Gains in performance stronger for shorter panels
- Other weighting schemes including equal weights (mean group) – generally worse

Relation to the Literature

Builds on two strands of literature:

1. Estimation of unit-specific parameters (Maddala et al., 1997, 2001; Baltagi et al., 2008; Issler and Lima, 2009; Baltagi, 2013; Zhang et al., 2014; Wang et al., 2019; Liu et al., 2020)
   $\Rightarrow$ We focus on a moderate-T setting (not small-T or large-T)
   $\Rightarrow$ We consider not just linear models, but possible nonlinear ones
2. Unit averaging has similarities to model averaging Hjort and Claeskens (2003); Claeskens and Hjort (2008); Zhang et al. (2014); Liu (2015); Yin et al. (2021)
   $\Rightarrow$ Unit averaging may be viewed as model averaging where every unit is a model, and every model is estimated on a different sample

Theoretical Results

Probabilistic Framework

Unit parameters differ according to mean-zero idiosyncratic random variables η_i:

- η_i can be cross-sectionally heterogeneous
- Must have $\sup_i \mathbb{E}[\eta_i^{12}] < \infty$

Interest in realized value for unit 1 $\Rightarrow$ work conditionally on $\{\eta_1, \eta_2, \dots\}$

Important: we show that all our results hold for almost all realization of $(\eta_1, \eta_2, \dots)$ (with η-probability 1)

Approximating Moderate-T: Limited Information Asymptotics

Moderate-T setting $\Rightarrow$ amount of information in each time series is limited

We reproduce this feature using a local heterogeneity device:

$$\theta_i = \mathbb{E}[\theta_i] + \frac{\eta_i}{\sqrt{T}}.$$

Allows us to do limited information local asymptotics:

- Approximate exact bias and variance of $\hat{\theta}_i$ using asymptotic techniques
- But amount of information in each time series is bounded and finite even as $T \to \infty$

Local Asymptotic Properties of Individual Estimators

Basic building block of averaging – things to be averaged.

Lemma

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

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

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

Limit mean and variance — local approximation to exact moderate-T bias and variance of $\hat{\theta}_i$ for estimating θ_1

Leading Terms of MSE of Unit Averaging Estimator

Theorem

Let units be independent and let

1. $\boldsymbol{w}_N$ be a given N-vector of weights (non-negative and sums to one)
2. $\sup_i |w_{iN} - w_i| = o(N^{-1/2})$ for some $w \in \mathbb{R}^\infty$ with $w_i \geq 0$ and $\sum w_i \leq 1$

Then as $N, T \to \infty$

$$T \times MSE\left(\hat{\theta}(\boldsymbol{w}_N)\right) \to \left(\sum_{i=1}^{\infty} w_i \eta_i - \eta_1\right)^2 + \sum_{i=1}^{\infty} w_i^2 V_i$$

Right hand side – local approximation, leading terms of the bias and the variance.

Towards Feasible Weights

Local approximation to the MSE:

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

To obtain feasible weights:

- Select class of weights to minimize over
- Replace η_i and V_i by estimators

Two Averaging Approaches

We discuss two ways to specify weights, depending on availability of prior information on which units have θ_i similar to θ_1

1. *Fixed-N* — agnostic approach that imposes no structure on weights
2. *Large-N* (details in the paper) — useful with prior information
   Splits units into two sets — unrestricted and restricted. Unrestricted units: any weight. Restricted units: only the total mass of the restricted set

Name of approaches due to underlying statistical frameworks

Large-N only differs from fixed-N when restricted set at least somewhat large

Feasible Optimal Weights in the Fixed-N Case

Fixed-N case — given fixed collection of $\bar{N}$ units using $\bar{N}$ vector of weights $\boldsymbol{w}^{\bar{N}}$. Then:

1. Can write LA-MSE as

$$LA\text{-}MSE_{\bar{N}}(\boldsymbol{w}^{\bar{N}}) = \boldsymbol{w}^{\bar{N}\prime} \boldsymbol{\Psi}_{\bar{N}} \boldsymbol{w}^{\bar{N}} \ ,$$

where $\boldsymbol{\Psi}_{\bar{N}}$ is an $\bar{N} \times \bar{N}$ matrix with elements

$$[\boldsymbol{\Psi}_{\bar{N}}]_{ij} = (\eta_i - \eta_1)(\eta_j - \eta_1)' + \mathbb{I}\{i = j\} V_i$$

2. Replace unknowns with "best possible" estimators in the moderate-T case:

$$[\hat{\boldsymbol{\Psi}}_{\bar{N}}]_{ij} = T(\hat{\theta}_i - \hat{\theta}_1)(\hat{\theta}_j - \hat{\theta}_1)' + \mathbb{I}\{i = j\} \hat{V}_i.$$

3. Feasible optimal fixed-N weights solve quadratic problem

$$\hat{\boldsymbol{w}}^{\bar{N}} = \underset{\sum w = 1, w \geq 0}{\arg\min} \ \boldsymbol{w}^{\bar{N}\prime} \hat{\boldsymbol{\Psi}}_{\bar{N}} \boldsymbol{w}^{\bar{N}}.$$

Moderate-T (Limited Information) Properties of Fixed-N Weights

Feasible weights solve the correct ideal MSE problem plus bias and zero-mean noise:

$$[\hat{\mathbf{\Psi}}_{\bar{N}}]_{i\,j} = [\mathbf{\Psi}_{\bar{N}}]_{i\,j} + V_1 + \mathbb{I}_{i=j} V_j + e_{i\,j} + o_p(1),$$

Here

- $\mathbb{E}[e_{ij}] = 0$
- Extra variance terms V_1 and V_j – price for having a positive definite finite sample problem

Properties of averaging estimator:

1. Approximately distributed as a randomly weighted sum of Gaussian random variables
2. Show how to construct confidence intervals for the parameter of interest based on the unit averaging estimator

Large-T (Growing Information) Properties of Unit Averaging

It is safe to use the feasible optimal weights even if amount of information in each time series is large. Theoretically: fixed parameter growing information asymptotics with

$$\theta_i = \mathbb{E}[\theta_i] + \eta_i.$$

Recall

$$[\hat{\mathbf{\Psi}}_{\bar{N}}]_{ij} = T(\hat{\theta}_i - \hat{\theta}_1)(\hat{\theta}_j - \hat{\theta}_1)' + \mathbb{I}\{i = j\}\hat{V}_i.$$

In the large-T case:

1. If $\theta_i \neq \theta_1$, the bias estimator $\sqrt{T}(\hat{\theta}_i - \hat{\theta}_1)$ will diverge
2. Variance terms remain bounded.

$\Rightarrow$ Procedure will place asymptotically zero weight on all units with $\theta_i \neq \theta_1$. If θ_i are continuously distributed, unit averaging estimator will converge to $\hat{\theta}_1$

Empirical Application: Unemployment Forecasting

Application: Forecasting Unemployment Rates for German Regions

We apply unit averaging to forecast regional unemployment rates for $N = 150$ German labor market districts:

- Regions strongly heterogeneous (de Graaff et al., 2018)
- But combining data might improve forecasting (Schanne et al., 2010)

Data: monthly data 2007-2024 from German Federal Labor Market Agency

Unemployment rate in region i at period t y_{it} modeled as a function of its past, past unemployment rate y_{it}^{RD} in broader region (RD) and Germany (y_t^{DE}):

$$y_{it} = \theta_{i0} + \theta_{i1} y_{it-1} + \theta_{i2} y_{it-1}^{RD} + \theta_{i3} y_{t-1}^{DE} + u_{it}, \quad t = 1, \dots, T$$

Parameter of interest: conditional mean of y_{it} given observables

We use rolling windows (one-step-ahead out-of-sample) forecasting to estimate the MSE for our approaches + some competing alternatives (individual estimator, mean group, AIC/BIC weights). Estimate MSE for $T = 40, 60, 80$

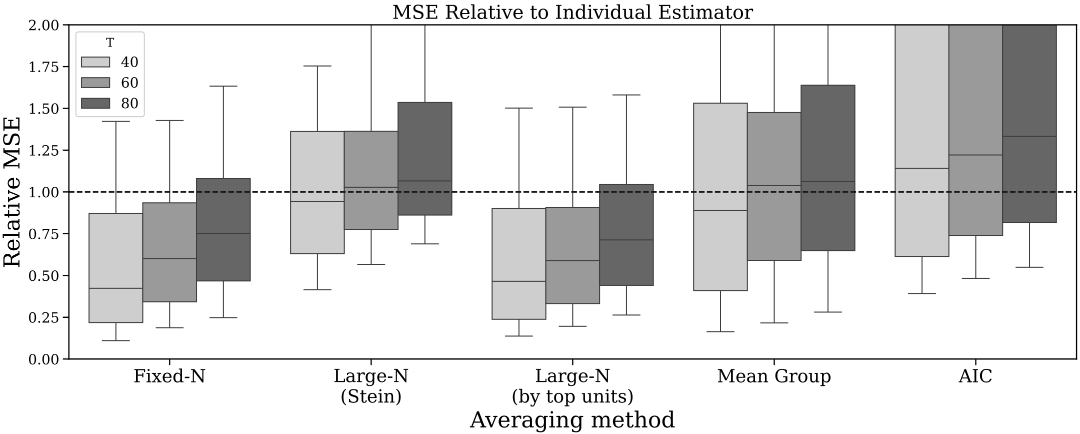

Figure: distribution of relative MSEs across AABs. Split by different averaging strategies and estimation window size. Whiskers – 10th and 90th percentiles; box boundaries – 25th and 75th percentiles; box crossbar – median

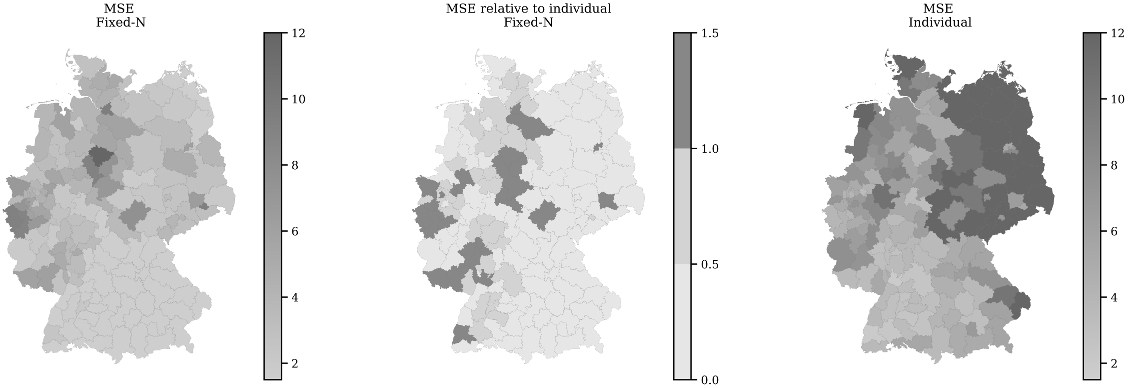

Figure: Geographic distribution of MSE to $T = 40$. Thin lines denote borders of AABs. Left and right panels: MSE of minimum MSE fixed-N and individual estimators respectively. Middle panel: ratio of MSE of fixed-N estimator to individual estimator.

Conclusions

- Even when estimating a unit-specific parameter, there still is value in panel data
- We propose a unit averaging approach
  Fits potentially nonlinear and dynamic models
- Moderate-T: approximate the MSE and provide a feasible optimal weighting scheme
  + characterize the properties of the procedure
- Large-T: unit averaging is safe to use
- Empirical application to unemployment forecasting: gains from using our feasible optimal weights

References I

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

B. H. Baltagi, G. Bresson, and A. Pirotte. To Pool or Not to Pool. In *The Econometrics of Panel Data*, chapter 16, pages 517–546. Springer Berlin Heidelberg, 2008.

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

T. de Graaff, D. Arribas-Bel, and C. Ozgen. Demographic Aging and Employment Dynamics in German Regions: Modeling Regional Heterogeneity. In *Modelling Aging and Migration Effects on Spatial Labor Markets*, chapter 11, pages 211–231. Springer Cham, 2018.

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

References II

J. V. Issler and L. R. Lima. A panel data approach to economic forecasting: The bias-corrected average forecast. *Journal of Econometrics*, 152(2):153–164, 2009.

C.-A. Liu. Distribution Theory of the Least Squares Averaging Estimator. *Journal of Econometrics*, 186(1):142–159, 2015.

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

L. Liu, A. Poirier, and J.-L. Shiu. Identification and Estimation of Average Partial Effects in Semiparametric Binary Response Panel Models. 2021.

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.

G. S. Maddala, H. Li, and V. K. Srivastava. A Comparative Study of Different Shrinkage Estimators for Panel Data Models. *Annals of Economics and Finance*, 2(1):1–30, 2001.

References III

N. Schanne, R. Wapler, and A. Weyh. Regional Unemployment Forecasts with Spatial Interdependencies. *International Journal of Forecasting*, 26(4):908–926, 2010.

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.

S.-Y. Yin, C.-A. Liu, and C.-C. Lin. Focused Information Criterion and Model Averaging for Large Panels with a Multifactor Error Structure. *Journal of Business and Economic Statistics*, 39(1):54–68, 2021.

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