Unit Averaging for Heterogeneous Panels
Supplementary Online Appendix

Christian Brownlees[†]
and
Vladislav Morozov[‡*]

May 10, 2024

This supplementary appendix provides additional theoretical, numerical, and empirical results. In section OA.1, we show that the unit averaging distribution is asymptotically normal if the weights do not depend on data, complementing the distributional results for data-dependent weights (theorems 2-3 in the main text). In section OA.2, we discuss optimal estimation under risks that are not the mean squared error. In section OA.3, we propose a practical confidence interval based on the minimal MSE estimator. Section OA.4 is devoted to further simulation results. We consider additional sample sizes and focus parameters. We also analyze the weights of the minimal MSE estimator, and consider the finite sample properties of ours confidence intervals for the focus parameter. Similarly, in section OA.5 we provide further estimation results and analyze averaging weights for the empirical application of section 5. Finally, in section OA.6 we provide an application to GDP nowcasting for a panel of European countries. As in the application of section 5, the minimum MSE estimator improves nowcasting performance relative both to the individual estimator and to competing averaging schemes. The improvement is stronger for shorter panels.

[*†] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona School of Economics; e-mail: christian.brownlees@upf.edu;
[‡] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona School of Economics; e-mail: vladislav.morozov@barcelonagse.eu. *Corresponding author.*
We thank Jan Ditzen, Kirill Evdokimov, Geert Mesters, Luca Neri, Katerina Petrova, Barbara Rossi, Wendun Wang, and the participants at 26th International Panel Data Conference, 7th RCEA Time Series Workshop, 9th SIDE WEEE, and the 2021 Econometric Research in Finance Workshop for comments and discussion. Christian Brownlees acknowledges support from the Spanish Ministry of Science and Technology (Grant MTM2012-37195) and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075).

Contents

OA.1 Asymptotic Distribution For Data-Independent Weights

OA.1.1 Theorem Statement

Theorems 2-3 in the main text characterize the distribution of the unit averaging estimator which uses the optimal weights (4) and (6). Theorems 2-3 establish that the unit averaging estimator with these data-dependent weights is approximately distributed as a randomly weight sum of Gaussian variables.

In contrast, if the weights do not depend on the data, the unit averaging estimator is approximately normally distributed. The mean and variance are given by the weighted sums of biases and variances, respectively. To formalize the result, let $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots,\}$ and $\boldsymbol{w}$ be as defined before theorem 1 in the main text. The following theorem formally states the normality result.

Theorem OA.1.1. *Assume that assumptions A.1–A.5 are satisfied. Let $\{\boldsymbol{w}_N\}$ be such that (i) for each N, $\boldsymbol{w}_N$ is measurable w.r.t. $\sigma(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_N)$, (ii) for each N, $w_{i\,N} \geq 0$ for all i, $\sum_{i=1}^N w_{i\,N} = 1$, $w_{j\,N} = 0$ for $j > N$, (iii) for some $\bar{N} \geq 0$ it holds that $\sup_{i>\bar{N}} w_{i\,N} = o(N^{-1/2})$, and (iv) $\{w_{i\,N}\}_{i=1}^{\bar{N}} \to \{w_i\}_{i=1}^{\bar{N}}$.*

Then as $N, T \to \infty$ jointly it holds that

$$\sqrt{T}\left(\hat{\mu}(\boldsymbol{w}_N) - \mu(\boldsymbol{\theta}_1)\right) \Rightarrow N\left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{d}_0' \boldsymbol{\eta}_i - \boldsymbol{d}_0' \boldsymbol{\eta}_1, \sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0\right).$$

Note that condition (iii) imposes a stronger uniform convergence requirement on the weights than theorem 1. However, this requirement is still compatible the weights assigned to the restricted sets in the large-N approach.

We apply theorem OA.1.1 in two ways. First, in section OA.2, we establish a local approximation to an alternative notion of risk for the unit averaging estimator — the mean average deviation $\mathbb{E}[|\hat{\mu}(\boldsymbol{w}_N) - \mu(\boldsymbol{\theta}_1)|]$. Second, in section OA.3, we use theorem OA.1.1 as a building block to construct valid confidence intervals for the focus parameter based on the minimum MSE estimator.

OA.1.2 Proof of Theorem OA.1.1

Before presenting the proof of theorem OA.1.1, we introduce a number of intermediate results.

We first give a straightforward modification of theorem 1 in Phillips and Moon (1999), which allows us to replace sequential convergence (first taking limits as $T \to \infty$, then as $N \to \infty$) by joint convergence ($N, T \to \infty$ jointly).

Lemma OA.1.2. *Let $Y_{i\,T}$ be random variables indexed by $i = 1, \dots, N$ and $T = 1,2, \dots,$. Suppose $Y_{i\,T}$ are independent over i and that*

(i) $Y_{i\,T} \Rightarrow \Lambda_i$ *as* $T \to \infty$,

(ii) $\sum_{i=1}^{N} w_{i\,N}\Lambda_i \Rightarrow X$ *as* $N \to \infty$,

(iii) $\limsup_{N,T\to\infty} \sum_{i=1}^{N} w_{i\,N}|\mathbb{E}(Y_{i\,T}) - \mathbb{E}(\Lambda_i)| = 0$,

(iv) $\limsup_{N,T\to\infty} \sum_{i=1}^{N} \mathbb{E}|w_{i\,N}Y_{i\,T}| < \infty$,

(v) $\limsup_{N\to\infty} \sum_{i=1}^{N} \mathbb{E}\left[w_{i\,N}|\Lambda_i|\mathbb{I}_{|w_{i\,N}\Lambda_i|>\varepsilon}\right] = 0$ *for any* $\varepsilon > 0$, *and*

(vi) $\limsup_{N,T\to\infty} \sum_{i=1}^{N} \mathbb{E}\left[w_{i\,N}|Y_{i\,T}|\mathbb{I}_{|w_{i\,N}Y_{i\,T}|>\varepsilon}\right] = 0$ *for any* $\varepsilon > 0$.

Then as $N, T \to \infty$

$$\sum_{i=1}^{N} w_{i\,N}Y_{i\,T} \Rightarrow X.$$

In particular, if as $N \to \infty$ *it holds that* $\sum_{i=1}^{N} w_{i\,N}\Lambda_i \xrightarrow{p} A$ *for* A *non-random, then as* $N, T \to \infty$ *it holds that* $\sum_{i=1}^{N} w_{i\,N}Y_{i\,T} \xrightarrow{p} A$.

Proof. The proof is close to that of theorem 1 in Phillips and Moon (1999). The key modification consists in replacing $n^{-1}\zeta_{k,n,T}$ (in their notation) by

$$W_{k\,N\,T} = \sum_{1\le i<k} w_{i\,N}Y_{i\,T} + \sum_{k<i\le N} w_{i\,N}\Lambda_i$$

and every factor $1/n$ by the appropriate weight $w_{i\,N}$. As in their theorem 1, this establishes condition (3.9) of Phillips and Moon (1999): for all bounded continuous f

$$\limsup_{N,T\to\infty}\left|\mathbb{E}\left(f\left(\sum_{i=1}^{N} w_{i\,N}Y_{i\,T}\right)\right)-\mathbb{E}\left(f\left(\sum_{i=1}^{N} w_i\Lambda_i\right)\right)\right|=0$$

By lemma 6 in Phillips and Moon (1999), this implies the result of the theorem. □

To apply lemma OA.1.2, for the remainder of the section define

$$Y_{i\,T}=\sqrt{T}(\mu(\hat{\boldsymbol{\theta}}_i)-\mu(\boldsymbol{\theta}_1)), \tag{OA.1}$$

and note that $Y_{i\,T}\Rightarrow\Lambda_i$ as $T\to\infty$, where $\Lambda_i\sim N\left(\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1),\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0\right)$ is the random variable that appears on the right hand side in lemma 1. As before, let $\boldsymbol{d}_1=\nabla\mu(\boldsymbol{\theta}_1)$, $\boldsymbol{d}_0=\nabla\mu(\boldsymbol{\theta}_0)$.

Lemma OA.1.3. *Let $Y_{i\,T}$ be defined as in eq. (OA.1). Under assumptions of theorem OA.1.1*

$$\sum_{i=1}^{\bar{N}} w_{i\,N}Y_{i\,T}\Rightarrow\sum_{i=1}^{\bar{N}} w_i\Lambda_i \text{ as } N,T\to\infty.$$

Proof. Note that randomness enters only the T dimension here. As $\{Y_{i\,T}\}_{i=1}^{\bar{N}}\Rightarrow\{\Lambda_i\}_{i=1}^{\bar{N}}$ as $N,T\to\infty$ (N does not matter), and as $N,T\to\infty$ $\{w_{i\,N}\}_{i=1}^{\bar{N}}\to\{w_i\}_{i=1}^{\bar{N}}$ as $N,T\to\infty$. Slutsky's theorem gives the result. □

Recall that under assumption (ii) of theorem OA.1.1 it holds that

$$\sup_{i>\bar{N}} w_{i\,N}=o(N^{-\frac{1}{2}})\ .$$

Lemmas OA.1.4-OA.1.8 verify conditions (ii)-(vi) of lemma OA.1.2 for $\sum_{i=\bar{N}+1}^{N} w_{i\,N}Y_{i\,T}$, $N>\bar{N}$.

Lemma OA.1.4. *Let $Y_{i\,T}$ be defined as in eq. (OA.1). Under assumptions of theorem*

OA.1.1

$$\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i \xrightarrow{p} -\left(1-\sum_{i=1}^{\bar{N}} w_i\right) \boldsymbol{d}_0'\boldsymbol{\eta}_1 \ \ as\ N\to\infty\ .$$

Proof. By the triangle inequality

$$\begin{aligned} &\left|\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i - \left(1-\sum_{i=1}^{\bar{N}} w_i\right)(-\boldsymbol{d}_0'\boldsymbol{\eta}_1)\right| \\ &\leq \left|\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i - \sum_{i=\bar{N}+1}^{N} w_{i\,N}\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1)\right| \\ &\quad + \left|\sum_{i=\bar{N}+1}^{N} w_{i\,N}\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1) - \left(1-\sum_{i=1}^{\bar{N}} w_i\right)(-\boldsymbol{d}_0'\boldsymbol{\eta}_1)\right|. \end{aligned} \tag{OA.2}$$

We show that both terms on the right hand side converge to zero in probability. First we show that $\left|\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i - \sum_{i=\bar{N}+1}^{N} w_{i\,N}\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1)\right| \xrightarrow{p} 0$. Consider the variance of $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i$:

$$\begin{aligned} \operatorname{Var}\left(\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i\right) &= \sum_{i=\bar{N}+1}^{N} w_{i\,N}^2\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0 \\ &\leq \left[\sup_{j>\bar{N}} w_{j\,N}\right]\sum_{i=\bar{N}+1}^{N} w_{i\,N}\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0 \\ &\leq \bar{\lambda}_{\boldsymbol{\Sigma}}\underline{\lambda}_{\boldsymbol{H}}^2\left\|\boldsymbol{d}_0\right\|^2\left[\sup_{j>\bar{N}} w_{j\,N}\right], \end{aligned}$$

where we used independence of Λ_i, the expressions for variance of Λ_i given in lemma 1, and the bound on variance $\boldsymbol{V}_i = \boldsymbol{H}_i^{-1}\boldsymbol{\Sigma}_i\boldsymbol{H}_i^{-1}$ implied by assumption A.3 on the bounds of eigenvalues of component variance matrices. Since $\mathbb{E}\left(\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i\right) = \sum_{i=\bar{N}+1}^{N} w_{i\,N}\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1)$, by Chebyshev's inequality and the above bound for variance, for any $\varepsilon>0$ it holds that

$$\begin{aligned} &P\left(\left|\sum_{i=\bar{N}+1}^{N} w_{i\,N}\Lambda_i - \sum_{i=\bar{N}+1}^{N} w_{i\,N}\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1)\right| > \varepsilon\right) \\ &\leq \frac{\bar{\lambda}_{\boldsymbol{\Sigma}}\underline{\lambda}_{\boldsymbol{H}}^2\left\|\boldsymbol{d}_0\right\|^2\left[\sup_{j>\bar{N}} w_{j\,N}\right]}{\varepsilon} = o(1), \end{aligned} \tag{OA.3}$$

by assumption (iii) of theorem OA.1.1. Next we show that

$$\left| \sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) - \left(1 - \sum_{i=1}^{\bar{N}} w_i\right) (-\boldsymbol{d}_0'\boldsymbol{\eta}_1) \right| \to 0$$

by considering two cases depending on whether $\sum_{i=1}^{\bar{N}} w_i$ is equal to 1 or not.

Case I: suppose that $\sum_{i=1}^{\bar{N}} w_i \neq 1$. In this case there exist N_0, $\varepsilon_w > 0$ such that for all $N > N_0$ it holds that $\sum_{i=1}^{\bar{N}} w_{i\,N} \leq 1 - \varepsilon_w$. Note that N_0 is necessarily larger than $\bar{N}$. Define $\tilde{w}_{i\,N} = w_{i\,N} / \left(1 - \sum_{i=1}^{\bar{N}} w_{i\,N}\right)$. For $N > N_0$, $(\tilde{w}_{\bar{N}+1\,N}, \tilde{w}_{\bar{N}+2\,N}, \dots, \tilde{w}_{N-\bar{N}\,N})$ satisfies $\tilde{w}_{i\,N} \geq 0$ and $\sum_{i=\bar{N}+1}^{N} \tilde{w}_{i\,N} = 1$. For all $N > \tilde{N}$ we have that $\tilde{w}_{i\,N} \leq \varepsilon_w^{-1} w_{i\,N}$, which implies that $\sup_{i>\bar{N}} \tilde{w}_{i\,N} \leq \varepsilon_w^{-1} \sup_{j>\bar{N}} w_{i\,N} = o(N^{-1/2})$. By lemma A.2.4 taken with $\gamma = 1/2$, we obtain that $\sum_{i=\bar{N}+1}^{N} \tilde{w}_{i\,N} \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) = \sum_{i=\bar{N}+1}^{N} \tilde{w}_{i\,N} \boldsymbol{d}_0'\boldsymbol{\eta}_i - \boldsymbol{d}_0'\boldsymbol{\eta}_1 \to -\boldsymbol{d}_0'\boldsymbol{\eta}_1$ (a.s. with respect to the distribution of $\boldsymbol{\eta}$). The weights $\tilde{w}$ satisfy the hypothesis of lemma A.2.4 with the limit weights equal to the zero sequence as $\sup_{i>\bar{N}} \tilde{w}_{i\,N} = o(N^{-1/2})$. Since $\sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) = \left(1 - \sum_{i=1}^{\bar{N}} w_{i\,N}\right) \sum_{i=\bar{N}+1}^{N} \tilde{w}_{i\,N} \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)$, we obtain that $\left| \sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) - \left(1 - \sum_{i=1}^{\bar{N}} w_i\right) (-\boldsymbol{d}_0'\boldsymbol{\eta}_1) \right| \to 0$. Together with eqs. (OA.2) and (OA.3), this implies that in this case

$$\sum_{i=\bar{N}+1}^{N} w_{i\,N} \Lambda_i \xrightarrow{p} -\left(1 - \sum_{i=1}^{\bar{N}} w_i\right) \boldsymbol{d}_0'\boldsymbol{\eta}_1 \; .$$

Case II: suppose that $\sum_{i=1}^{\bar{N}} w_i = 1$. We show that $\sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) \to 0$ $\boldsymbol{\eta}$-a.s.. First, $\sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0'\boldsymbol{\eta}_1 = \boldsymbol{d}_0'\boldsymbol{\eta}_1 \sum_{i=\bar{N}+1}^{N} w_{i\,N} \to 0$ by the assumption that $\sum_{i=1}^{\bar{N}} w_{i\,N} \to 1$. Second, $\sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0'\boldsymbol{\eta}_i \to \mathbb{E}_{\boldsymbol{\eta}}(\boldsymbol{d}_0'\boldsymbol{\eta}_i) = 0$ by lemma A.2.3, since $\boldsymbol{d}_0'\boldsymbol{\eta}_i$ are independent variables with uniformly bounded third moments. As above, this argument and eqs. (OA.2) and (OA.3) imply that $\sum_{i=\bar{N}+1}^{N} w_{i\,N} \Lambda_i \xrightarrow{p} 0$.

Combining the two cases yields the assertion. □

Lemma OA.1.5. *Let Y_{iT} be defined as in eq.* (OA.1). *Under assumptions of theorem*

OA.1.1

$$\limsup_{N,T\to\infty} \sum_{i=\bar{N}+1}^{N} w_{iN}|\mathbb{E}(Y_{iT}) - \mathbb{E}(\Lambda_i)| = 0 \ \textit{as} \ N,T \to \infty.$$

Proof. First, from lemma A.2.2 it follows that $\mathbb{E}|Y_{iT}|$ exists for all i and $T > T_0$. By lemma 1, $\mathbb{E}\,\Lambda_i = \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)$. By eq. (A.1.2) of lemma A.1.1, we have

$$\mu(\hat{\boldsymbol{\theta}}_i) = \mu(\boldsymbol{\theta}_1) + \boldsymbol{d}_1'\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1\right) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1), \tag{OA.4}$$

where $\boldsymbol{d}_1 = \nabla\mu(\boldsymbol{\theta}_1)$ and $\acute{\boldsymbol{\theta}}_i$ lies on the segment joining $\hat{\boldsymbol{\theta}}_i$ and $\boldsymbol{\theta}_1$. Then

$$Y_{iT} - \mathbb{E}(\Lambda_i) = \boldsymbol{d}_1'\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1\right) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1) - \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1). \tag{OA.5}$$

We now establish a bound on $|\mathbb{E}(Y_{iT}) - \mathbb{E}(\Lambda_i)|$. Take expectations in eq. (OA.5):

$$\begin{aligned}
&|\mathbb{E}(Y_{iT}) - \mathbb{E}(\Lambda_i)| \\
&= \left|\mathbb{E}\left[\boldsymbol{d}_1'\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1\right) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1) - \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)\right]\right| \\
(*) &\leq \left|\boldsymbol{d}_1'\,\mathbb{E}\left[\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right]\right| + \left|\mathbb{E}\left[\frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)\right]\right| \\
&\quad + |(\boldsymbol{d}_1 - \boldsymbol{d}_0)'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)| \\
(**) &\leq \|\boldsymbol{d}_1\|\frac{C_{Bias}}{\sqrt{T}} + C_{\nabla^2\mu}\left|\mathbb{E}\left(\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)\right)\right| + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\|\boldsymbol{\eta}_1\|\,\|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_1\| \\
(***) &\leq \|\boldsymbol{d}_1\|\frac{C_{Bias}}{\sqrt{T}} + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\mathbb{E}\left\|\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\right\|^2 + \frac{2C_{\nabla^2\mu}}{\sqrt{T}}\mathbb{E}\left\|\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\right\|\,\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\| \\
&\quad + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^2 + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\|\boldsymbol{\eta}_1\|\,\|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_1\| \\
(****) &\leq C_{\nabla\mu}\frac{C_{Bias}}{\sqrt{T}} + \frac{C_{\nabla^2\mu}C_{\hat{\boldsymbol{\theta}},2}}{\sqrt{T}} + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\left(2C_{\hat{\boldsymbol{\theta}},1} + \|\boldsymbol{\eta}_1\|\right)\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\| \\
&\quad + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^2.
\end{aligned} \tag{OA.6}$$

where the constants C do not depend on i. Here

(*) $\boldsymbol{\theta}_1$ is replaced by $\boldsymbol{\theta}_i$ in the first term using $\boldsymbol{d}_1'\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1) - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) = \boldsymbol{d}_1'\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)$

(**) In the first term we apply Hölder's inequality inside the absolute value as

$$\left| \boldsymbol{d}_1' \, \mathbb{E}\left[\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i \right) \right] \right| \leq \|\boldsymbol{d}_1\|_\infty \left\| \mathbb{E}(\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)) \right\|_1 \leq \|\boldsymbol{d}_1\|_2 \sqrt{T} \left\| \mathbb{E}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \right\|_1 .$$

Assumption A.4 bounds $\sqrt{T}\left\| \mathbb{E}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \right\|_1 \leq C_{Bias}/\sqrt{T}$. In the second term apply A.5 to replace the Hessian $\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)$. In the third term apply assumptions A.1 and A.5: $\nabla\mu$ is a differentiable function with norm of the derivative bounded, which implies that $\|\boldsymbol{d}_1 - \boldsymbol{d}_0\| = \left\| \nabla\mu(\boldsymbol{\theta}_0 + T^{-1/2}\boldsymbol{\eta}_1) - \nabla\mu(\boldsymbol{\theta}_0) \right\| \leq C_{\nabla^2\mu} \|\boldsymbol{\eta}_1\|/\sqrt{T}$.

(***) Add and subtract $\boldsymbol{\theta}_1$ in both parentheses in the quadratic term, apply the triangle inequality.

(****) Recall that $\boldsymbol{\theta}_i - \boldsymbol{\theta}_1 = (\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)/\sqrt{T}$ by A.1. Expectations of $\left\| \sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \right\|$ are bounded using lemma A.2.1; by A.5 $\|\boldsymbol{d}_1\| \leq C_{\nabla\mu}$

Last, we can consider the sum $\sum_{\bar{N}+1}^{N} w_{i\,N} |\mathbb{E}(Y_{i\,T}) - \mathbb{E}(\Lambda_i)|$, bounded by the corresponding weighted sum of the right hand side of eq. (OA.6). The first two terms in the bound do not depend on i, and so

$$\sum_{i=\bar{N}+1}^{N} w_{i\,N} \left[\frac{C_{\nabla\mu} C_{Bias}}{\sqrt{T}} + \frac{C_{\nabla^2\mu} C_{\hat{\boldsymbol{\theta}},2}}{\sqrt{T}} \right] \leq \frac{C_{\nabla\mu} C_{Bias}}{\sqrt{T}} + \frac{C_{\nabla^2\mu} C_{\hat{\boldsymbol{\theta}},2}}{\sqrt{T}} \to 0$$

since $w_{i\,N}$ are part of a weight vector. For the third and the fourth term we make use of the conditions on weight decay and the moments of $\boldsymbol{\eta}_i$. Examine

$$\frac{C_{\nabla^2\mu}}{\sqrt{T}} \left[2C_{\hat{\boldsymbol{\theta}},2} + \|\boldsymbol{\eta}_1\| \right] \sum_{i=\bar{N}+1}^{N} w_{i\,N} \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\| + \frac{C_{\nabla^2\mu}}{\sqrt{T}} \sum_{i=\bar{N}+1}^{N} w_{i\,N} \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^2 .$$

By lemma A.2.4 $\sup_N \sum_{i=\bar{N}+1}^{N} w_{i\,N} \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^k$, $k = 1, 2$ are finite. Then for some $M < \infty$ the above display is bounded by $M/\sqrt{T}$ and thus converges to zero as well. Combining the last two results together, we obtain that $\sup_N \sum_{i=\bar{N}+1}^{N} w_{i\,N} |\mathbb{E}(Y_{i\,T}) - \mathbb{E}(\Lambda_i)| \to 0$ as $T \to \infty$, giving the result of the lemma. □

Lemma OA.1.6. *Let Y_{iT} be defined as in eq.* (OA.1)*. Under assumptions of theorem OA.1.1*

$$\limsup_{N,T\to\infty} \sum_{i=\bar{N}+1}^{N} w_{iN}\, \mathbb{E}|Y_{iT}| < \infty \text{ as } N, T \to \infty.$$

Proof. Existence of $\mathbb{E}|Y_{iT}|$ for $T > T_0$ follows from lemma A.2.2. Add and subtract $\mathbb{E}(\Lambda_i)$ under the absolute value in $\mathbb{E}|Y_{iT}|$ to get

$$\begin{aligned}
\mathbb{E}|Y_{iT}| &\leq |\mathbb{E}(\Lambda_i)| + \mathbb{E}|Y_{iT} - \mathbb{E}(\Lambda_i)| \\
&= |\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)| + \mathbb{E}|Y_{iT} - \mathbb{E}(\Lambda_i)| \\
&\leq \|\boldsymbol{d}_0\|\, \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_0\| + \mathbb{E}|Y_{iT} - \mathbb{E}(\Lambda_i)|\ ,
\end{aligned}$$

where we apply the Cauchy-Schwarz inequality in the last line. Take weighted sums

$$\sum_{i=\bar{N}+1}^{N} w_{iN}\, \mathbb{E}|Y_{iT}| \leq \|\boldsymbol{d}_0\| \sum_{i=\bar{N}+1}^{N} w_{iN}\, \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\| + \sum_{i=\bar{N}+1}^{N} w_{iN}\, \mathbb{E}|Y_{iT} - \Lambda_i|.$$

We show that both sums are bounded as $N, T \to \infty$. First, as in lemma OA.1.5, from lemma A.2.4 it follows that $\sup_N \sum_{i=\bar{N}+1}^{N} w_{iN}\, \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\| < \infty$. Now turn to the second sum. Using eq. (OA.4), we proceed similarly to the proof of lemma OA.1.5:

$$\begin{aligned}
&\mathbb{E}|(Y_{iT}) - \mathbb{E}(\Lambda_i)| \\
&= \mathbb{E}\left|\left[\boldsymbol{d}_1'\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1\right) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1) - \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) \pm \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)\right]\right| \\
&\leq \mathbb{E}\left|\boldsymbol{d}_1'\left[\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right]\right| + \mathbb{E}\left|\left[\frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)\right]\right| + |(\boldsymbol{d}_1 - \boldsymbol{d}_0)'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)| \\
&\leq C_{\nabla\mu}C_{\hat{\boldsymbol{\theta}},1} + \frac{C_{\nabla^2\mu}C_{\hat{\boldsymbol{\theta}},2}}{\sqrt{T}} + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\left[2C_{\hat{\boldsymbol{\theta}},1} + \|\boldsymbol{\eta}_1\|\right]\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\| + \frac{C_{\nabla^2\mu}}{\sqrt{T}}\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^2.
\end{aligned}$$

There is one change relative to lemma OA.1.5: by the Cauchy-Schwarz inequality and assumption A.5, $\mathbb{E}\left|\boldsymbol{d}_1'\left[\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right]\right| \leq C_{\nabla\mu}\mathbb{E}\left\|\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\right\|$, to which we then apply lemma A.2.1. The constant C_{Bias} does not appear in the above bound. Take weighted sums in $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}|Y_{i\,T}-\Lambda_i|$, and use the above bound for each term in the sum. The argument proceeds similarly to lemma OA.1.5. The first two terms in the bound satisfy $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\left(C_{\nabla\mu}C_{\hat{\boldsymbol{\theta}},1}+C_{\nabla^2\mu}C_{\hat{\boldsymbol{\theta}},2}/\sqrt{T}\right) \leq C_{\nabla\mu}C_{\hat{\boldsymbol{\theta}},1}+C_{\nabla^2\mu}C_{\hat{\boldsymbol{\theta}},2}/\sqrt{T}$, which is independent of N and convergent in T. Both sums $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\|\boldsymbol{\eta}_i-\boldsymbol{\eta}_1\|$ and $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\|\boldsymbol{\eta}_i-\boldsymbol{\eta}_1\|^2$ are bounded in N regardless of T by lemma A.2.4. We conclude that $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}|Y_{i\,T}-\Lambda_i|$ is bounded in N and T, giving the claim of the lemma. □

Lemma OA.1.7. *Let assumptions of theorem OA.1.1 hold, and let Λ_i be as in lemma 1. Then for any $\varepsilon>0$*

$$\limsup_{N\to\infty}\sum_{i=\bar{N}+1}^{N}\mathbb{E}\left[w_{i\,N}|\Lambda_i|\mathbb{I}_{|w_{i\,N}\Lambda_i|>\varepsilon}\right]=0.$$

Proof. Since $\sup_{i>\bar{N}} w_{i\,N}=o(N^{-1/2})$, there exists some $C_w>0$ and N_0 such that for all $N>N_0$ it holds that $w_{i\,N}<C_w^{-1}N^{-1/2}$ for all $i>\bar{N}$. Also observe that for $p>1$ $\mathbb{E}(|X|\,\mathbb{I}_{X>M})\leq M^{-(p-1)}\,\mathbb{E}(|X|^p)$. Hence for $p>1$

$$\begin{aligned}\sum_{i=\bar{N}+1}^{N}\mathbb{E}\left[w_{i\,N}|\Lambda_i|\mathbb{I}_{|w_{i\,N}\Lambda_i|>\varepsilon}\right] &\leq \sum_{i=\bar{N}+1}^{N}\mathbb{E}\left[w_{i\,N}|\Lambda_i|\,\mathbb{I}_{|\Lambda_i>C_wN^{1/2}\varepsilon|}\right]\\ &\leq \frac{1}{(C_w\varepsilon N^{1/2})^{p-1}}\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}(|\Lambda_i|^p). \qquad \text{(OA.7)}\end{aligned}$$

Pick $p=2$. Since $1/(C_w\varepsilon N^{1/2})\to 0$, it is sufficient to show that $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}(|\Lambda_i|^2)$ is bounded over N.

Since $|\Lambda_i|$ is folded normal, its first two moments are given by (see Elandt (1961)):

$$\begin{aligned}\mathbb{E}|\Lambda_i|^2 &= (\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1))^2+\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0-(\mathbb{E}|\Lambda_i|)^2,\\ \mathbb{E}|\Lambda_i| &= \sqrt{\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0}\sqrt{\frac{2}{\pi}}e^{-\frac{(\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1))^2}{2\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0}}+\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1)\left(1-2\Phi\left(-\frac{\boldsymbol{d}_0'(\boldsymbol{\eta}_i-\boldsymbol{\eta}_1)}{2\sqrt{\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0}}\right)\right).\end{aligned}$$

It is sufficient to establish the boundedness of the weighted sum of each term separately. We proceed in order of appearance in the preceding display.

1. By the Cauchy-Schwarz inequality

$$\sum_{i=\bar{N}+1}^{N} w_{i\,N} (\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1))^2 \leq \|\boldsymbol{d}_0\|^2 \sum_{i=\bar{N}+1}^{N} w_{i\,N} \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^2 .$$

The sum on the right is bounded over N by lemma A.2.4.

2. By the bound on variance of assumption A.3 it holds that

$$\sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0 \leq \bar{\lambda}_{\boldsymbol{\Sigma}} \underline{\lambda}_{\boldsymbol{H}}^2 \|\boldsymbol{d}_0\|^2 \ .$$

3. Consider the first term in $(\mathbb{E}|\Lambda_i|)^2$:

$$\sum_{i=\bar{N}+1}^{N} w_{i\,N} \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0 \frac{2}{\pi} \left[e^{-\frac{(\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1))^2}{2\boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}} \right]^2 \leq \bar{\lambda}_{\boldsymbol{\Sigma}} \underline{\lambda}_{\boldsymbol{H}}^2 \|\boldsymbol{d}_0\|^2 \frac{2}{\pi}.$$

4. Cross-term in $(\mathbb{E}|\Lambda_i|)^2$:

$$\begin{aligned} &\sum_{i=\bar{N}+1}^{N} w_{i\,N} \left| \sqrt{\boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0} \sqrt{\frac{2}{\pi}} e^{-\frac{(\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1))^2}{2\boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}} \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) \left(1 - 2\Phi\left(-\frac{\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)}{2\sqrt{\boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}} \right) \right) \right| \\ \leq &\sqrt{\bar{\lambda}_{\boldsymbol{\Sigma}} \underline{\lambda}_{\boldsymbol{H}}^2} \|\boldsymbol{d}_0\|^2 \sqrt{\frac{2}{\pi}} \sum_{i=\bar{N}+1}^{N} w_{i\,N} \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\| . \end{aligned}$$

The sum in the last line is bounded over N by lemma A.2.4.

5. Square of the second term:

$$\begin{aligned} &\sum_{i=\bar{N}+1}^{N} w_{i\,N} [\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)]^2 \left(1 - 2\Phi\left(-\frac{\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)}{2\sqrt{\boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}} \right) \right)^2 \\ \leq &\sum_{i=\bar{N}+1}^{N} w_{i\,N} [\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)]^2 \leq \|\boldsymbol{d}_0\|^2 \sum_{i=\bar{N}+1}^{N} w_{i\,N} \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^2 , \end{aligned}$$

where the last sum is bounded by lemma A.2.4.

Combining the above arguments, we conclude that $\sup_N \sum_{i=\bar{N}+1}^{N} w_{i\,N} \mathbb{E}(|\Lambda_i|^2) < \infty$. By eq.

(OA.7)

$$\sum_{i=\bar{N}+1}^{N} \mathbb{E}\left[w_{i\,N}|\Lambda_i|\mathbb{I}_{|w_{i\,N}\Lambda_i|>\varepsilon}\right] \leq \frac{1}{C_w \varepsilon N^{1/2}} \sup_N \sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}(|\Lambda_i|^2$$

The right hand side tends to 0 as $N \to \infty$. □

Lemma OA.1.8. *Let $Y_{i\,T}$ be defined as in eq. (OA.1). Under assumptions of theorem OA.1.1, for any $\varepsilon > 0$*

$$\limsup_{N,T\to\infty} \sum_{i=\bar{N}+1}^{N} \mathbb{E}\left[w_{i\,N}|Y_{i\,T}|\mathbb{I}_{|w_{i\,N}Y_{i\,T}|>\varepsilon}\right] = 0.$$

Proof. Existence of $\mathbb{E}|Y_{i\,T}|$ for $T > T_0$ follows from lemma A.2.2. We use the same strategy as in lemma OA.1.7. Since $\sup_{i>\bar{N}} w_{i\,N} = o(N^{-1/2})$, there exists some $C_w > 0$ and N_0 such that for all $N > N_0$ it holds that $w_{i\,N} < C_w^{-1}N^{-1/2}$ for all $i > \bar{N}$. Then for $p > 1$, if $\mathbb{E}|Y_{i\,T}|^p$ exists, we obtain that

$$\begin{aligned}
\sum_{i=\bar{N}+1}^{N} \mathbb{E}\left[w_{i\,N}|Y_{i\,T}|\mathbb{I}_{|w_{i\,N}Y_{i\,T}|>\varepsilon}\right] &\leq \sum_{i=\bar{N}+1}^{N} \mathbb{E}\left[w_{i\,N}|Y_{i\,T}|\mathbb{I}_{|Y_{i\,T}|>C_w N^{1/2}\varepsilon}\right] \\
&\leq \frac{1}{(C_w \varepsilon N^{1/2})^{p-1}} \sum_{i=\bar{N}+1}^{N} w_{i,N}\,\mathbb{E}\left[|Y_{i\,T}|^p\right] \\
&\leq \frac{2^{p-1}}{(C_w \varepsilon N^{1/2})^{p-1}} \sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}|Y_{i\,T} - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^p \\
&\quad + \frac{2^{p-1}}{(C_w \varepsilon N^{1/2})^{p-1}} \sum_{i=\bar{N}+1}^{N} w_{i\,N}|\boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^p. \qquad \text{(OA.8)}
\end{aligned}$$

It is sufficient to establish convergence of the weighted sums for some $p > 1$, since the leading $N^{(p-1)/2}$ will then drive the expression to zero. Take $p = 1 + \delta'$ where $\delta' = \delta/2$ for δ from assumption A.3.

The second sum in eq. (OA.8) is bounded over N by lemma A.2.4, as

$$\sum_{\bar{N}+1}^{N} w_{i\,N}|\boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^{1+\delta'} \leq C_{\nabla\mu}^{1+\delta'} \sum_{i=\bar{N}+1}^{N} w_{i\,N}\left\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\right\|^{1+\delta'}.$$

Now consider $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}|Y_{i\,T} - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^{1+\delta'}$. We proceed similarly to the proof of lemma OA.1.6. First, by lemma A.2.2 $\mathbb{E}|Y_{i\,T} - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^{1+\delta'} < \infty$. It remains to show that the weighted sum is bounded over N. Recall from lemma OA.1.5 that

$$Y_{i\,T} - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) = \boldsymbol{d}_1'\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)$$

for $\acute{\boldsymbol{\theta}}_i$ is intermediate between $\hat{\boldsymbol{\theta}}_i$ and $\boldsymbol{\theta}_1$. Then

$$\begin{aligned}
&|Y_{i\,T} - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^{1+\delta'} \\
&\leq 2^{\delta'}\left|\boldsymbol{d}_1'\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right|^{1+\delta'} + 2^{\delta'}\left|\frac{1}{2}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)'\nabla^2\mu(\acute{\boldsymbol{\theta}}_i)\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_1)\right|^{1+\delta'} \\
&\leq 2^{\delta'}\left|\boldsymbol{d}_1'\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right|^{1+\delta'} + \frac{2^{2\delta'}C_{\nabla^2\mu}^{1+\delta'}}{T^{(1+\delta')/2}}\left\|\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_i)\right\|^{2(1+\delta')} \\
&\quad + 2^{1+3\delta'}C_{\nabla^2\mu}^{1+\delta'}\left|\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)'\frac{(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)}{\sqrt{T}}\right|^{1+\delta'} + \frac{2^{2\delta'}C_{\nabla^2\mu}^{1+\delta'}}{T^{(1+\delta')/2}}|(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^{1+\delta'}.
\end{aligned}$$

Taking expectations, we obtain

$$\begin{aligned}
&\mathbb{E}|Y_{i\,T} - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^{1+\delta'} \\
&\leq 2^{1+3\delta'}\Bigg[C_{\mu}^{1+\delta'}C_{\hat{\boldsymbol{\theta}},1+\delta/2} + \frac{C_{\nabla^2\mu}^{1+\delta'}C_{\hat{\boldsymbol{\theta}},2+\delta}}{T^{(1+\delta')/2}} \\
&\quad + \frac{2C_{\nabla^2\mu}^{1+\delta'}C_{\hat{\boldsymbol{\theta}},1+\delta/2}}{T^{(1+\delta')/2}}\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^{1+\delta'} + \frac{C_{\nabla^2\mu}^{1+\delta'}}{T^{(1+\delta')/2}}\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^{2(1+\delta')}\Bigg],
\end{aligned} \tag{OA.9}$$

where the bounds on $\mathbb{E}\left\|\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\right\|^k, k = 1 + \delta/2, 2 + \delta$ follow from lemma A.2.1.

Take weighted sums $\sum_{i=\bar{N}+1}^{N} w_{i\,N}\,\mathbb{E}|Y_{i\,T} - \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)|^{1+\delta'}$. Then for the first two terms it holds that

$$\sum_{i=\bar{N}+1}^{N} w_{i\,N}C_{\nabla\mu}^{1+\delta}C_{\hat{\boldsymbol{\theta}},1+\delta/2} + \frac{C_{\nabla^2\mu}^{1+\delta'}C_{\hat{\boldsymbol{\theta}},2+\delta}}{T^{(1+\delta')/2}} \leq C_{\nabla\mu}^{1+\delta}C_{\hat{\boldsymbol{\theta}},1+\delta/2} + \frac{C_{\nabla^2\mu}^{1+\delta'}C_{\hat{\boldsymbol{\theta}},2+\delta}}{T^{(1+\delta')/2}},$$

since constants are independent of i. For the third and the fourth term of eq. (OA.9), it is sufficient to observe that by lemma A.2.4 $\sup_N \sum_{i=\bar{N}+1}^{N} w_{i\,N}\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\|^{2(1+\delta')} < \infty$ and

$\sup_N \sum_{i=\bar{N}+1}^{N} w_{i\,N} \left\|\boldsymbol{\eta}_i - \boldsymbol{\eta}_1\right\|^{1+\delta'} < \infty$.

Hence, both sums in eq. (OA.8) are bounded uniformly over N. Taking $N \to \infty$ shows the original sum of interest converges to 0. □

Finally, we present the proof of theorem OA.1.1.

Proof of theorem OA.1.1. Using the fact that $\sum_{i=1}^{N} w_{i\,N} = 1$ and recalling that $N > \bar{N}$ we write

$$\sqrt{T}\left(\hat{\mu}(\boldsymbol{w}_N) - \boldsymbol{\mu}(\boldsymbol{\theta}_1)\right) = \sum_{i=1}^{N} w_{i\,N} Y_{i\,T} = \sum_{i=1}^{\bar{N}} w_{i\,N} Y_{i\,T} + \sum_{i=\bar{N}+1}^{N} w_{i\,N} Y_{i\,T} \ .$$

The first sum contains the units whose weights are allowed to be asymptotically non-negligible. By lemma OA.1.3, as $N, T \to \infty$ jointly, it holds that

$$\sum_{i=1}^{\bar{N}} w_{i\,N} Y_{i\,T} \Rightarrow \sum_{i=1}^{\bar{N}} w_i \Lambda_i \sim N\left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1), \sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_1' \boldsymbol{V}_i \boldsymbol{d}_1\right) \ .$$

The second sum contains the units whose weights satisfy $\sup_{i>\bar{N}} w_{i\,N} = o(N^{-1/2})$. By appealing to lemma OA.1.2, we show that $\sum_{i=\bar{N}+1}^{N} w_{i\,N} Y_{i\,T} \xrightarrow{p} -\left(1 - \sum_{i=1}^{\bar{N}} w_i\right) \boldsymbol{d}_0' \boldsymbol{\eta}_1$ as $N, T \to \infty$ jointly. We turn to verifying the conditions of lemma OA.1.2:

1. *Assumption 1* (large T step): follows from lemma 1 as

$$Y_{i\,T} \Rightarrow \Lambda_i \sim N\left(\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1), \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0\right) \ .$$

2. *Assumption 2* (large N step): by lemma OA.1.4, $\sum_{i=\bar{N}+1}^{N} w_{i\,N} \Lambda_i$ converges in probability to $-\left(1 - \sum_{i=1}^{\bar{N}} w_i\right) \boldsymbol{d}_0' \boldsymbol{\eta}_1$

3. *Assumptions 3-6* are verified by lemmas OA.1.5-OA.1.8, respectively.

Last, by Slutsky's theorem

$$\sum_{i=1}^{\bar{N}} w_{i\,N} Y_{i\,T} + \sum_{i=\bar{N}+1}^{N} w_{i\,N} Y_{i\,T}$$

$$\Rightarrow N\left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) - \left(1 - \sum_{i=1}^{\bar{N}} w_i\right) \boldsymbol{d}_0'\boldsymbol{\eta}_1, \sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0\right) ,$$

which establishes the claim. □

OA.2 Alternative Loss Functions

It may be of interest to measure the quality of the averaging estimator using criteria other than the mean squared error (MSE). Generically, let l be a loss function, and suppose that we measure estimation quality with the corresponding risk:

$$R(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N)) \coloneqq \mathbb{E}\left[l(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N))\right]. \tag{OA.10}$$

In this section, we extend the analysis of the main text to accommodate two different classes of loss functions l. First, in section OA.2.1 we show that $R(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N))$ behaves essentially like the MSE for a broad class of smooth loss functions. In this case the MSE-optimal weights of section 2 also serve as feasible risk-optimal weights under the risk R. Second, in section OA.2.2 we obtain an explicit local approximation to the risk if l is the absolute loss, in which case R is the mean absolute deviation (MAD). The local approximation to the MAD is different from the MSE, but still amenable to minimization over averaging weights. The proofs for this section are collected in subsection OA.2.3.

OA.2.1 Smooth Loss Functions

We consider the class of *locally quadratic* loss functions (Hansen, 2016). Intuitively, a loss function is locally quadratic if it is a smooth function of the estimator, and the corresponding second derivative is nonzero when the estimator is close to the target value; assumption OA.A.1 below provides a formal definition. This is a broad class of loss functions that includes the squared and linear-exponential losses, losses based on smooth non-linear utility functions, and various integrated losses such as the Hellinger distance; see section 2.2 in

Hansen (2016) for a list of examples.

For locally quadratic losses, the corresponding risk (OA.10) behaves like the MSE up to a negligible difference term in our framework, as theorem OA.2.1 below shows. The local approximations for the MSE of section 3 of the main text are then also valid local approximations for risk (OA.10). In practical terms, this result means that one may use the fixed-N and large-N weights of eqs. (4) and (6) as feasible minimal risk weights for risk (OA.10).

Formally, we assume that the loss function l in eq. (OA.10) satisfies the following assumption:

OA.A.1 (Locally quadratic loss). *(i) The function $l(\cdot,\cdot)$ is defined on $\mu(\Theta)\times\mu(\Theta)$ where Θ is the parameter space (as in A.3) and $\mu(\Theta)$ is the image of Θ under μ.*

(ii) l is a loss function: the function $l(\cdot,\cdot)$ satisfies $l(x,x)=0$ and $l(x,y)>0$ for $x,y\in\mu(\Theta)$, $x\neq y$.

(iii) l is smooth: for any $x\in\mu(\Theta)$ the function $l(x,y)$ is at least three times differentiable in y.

(iv) Bounded second and third derivative: let $\partial_2^k l(\cdot,\cdot)$ be the kth partial derivative of $l(\cdot,\cdot)$ with respect to its second argument. Then there exist finite constants C_k such that $\sup_{x,y\in\mu(\Theta)}\left|\partial_2^k l(x,y)\right|\leq C_k$ for $k=2,3$.

(v) Nonzero second derivative around target value: there exists a $\varepsilon>0$ such that $\|x-y\|<\varepsilon$ implies that $|\partial_2^2 l(x,y)|>\varepsilon$.

Assumption OA.A.1 formally defines the class of locally quadratic loss functions. It is fairly mild and covers a number of standard loss functions, as noted above. Condition (iv) holds if the parameter space Θ is compact and (i)-(iii) hold; alternatively, (iv) may be relaxed to an integrability condition by trimming the loss as in Hansen (2016).

We now show that risk OA.10 locally behaves similarly to (scaled) MSE, up to a negligible component. Let the weight sequence $\boldsymbol{w}_N=(w_{i\,N})$ be as defined before theorem 1 in the main text.

Theorem OA.2.1. *Let condition A.3 hold with $\delta > 2$. Let the conditions of theorem OA.1.1 hold. Let the loss function $l(\cdot, \cdot)$ satisfy OA.A.1.*

Then (i) for any N and any $T > T_0$ the risk (OA.10) of the unit averaging estimator is finite; (ii) as $N, T \to \infty$ jointly, it holds that

$$T \times R(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N)) = T \times \left(\frac{1}{2} \partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right) \times MSE(\hat{\mu}(\boldsymbol{w}_N)) \right) + o(1). \quad \text{(OA.11)}$$

Intuitively, for locally quadratic losses, the risk of a consistent estimator is dominated by its (suitably rescaled) bias and variance. Accordingly, minimizing the MSE of the unit averaging estimator also approximately minimizes its risk (OA.10). See also Hansen (2016) for a similar result for shrinkage estimators in parametric settings.

Remark OA.2.1 (Higher-order loss functions)**.** The assumption of a (locally) nonzero second derivative is crucial to the results of theorem OA.2.1. Eq. (OA.11) does not hold if the loss function $l(\cdot, \cdot)$ is a higher-order loss in the sense that $\partial_2^2 l(x, x) = 0$ for all x; the quartic loss $l(x, y) = (x - y)^4$ is an example of a such a function.

However, it is possible to obtain results in the spirit of theorem OA.2.1 for higher-order losses. Suppose that $l(\cdot, \cdot)$ is a kth order loss in sense that $\partial_2^j l(x, x) = 0$ for $j = 0, 1, \dots, k-1$ and $\partial_2^k l(x, x)$ is bounded away from zero for some even k. In this case $T^{k/2} R(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N)) = T^{k/2}\, \mathbb{E}\left[(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N))^k\right] + o(1)$. In turn, an explicit local approximation for $T^{k/2}\, \mathbb{E}[(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N))^k]$ may be obtained by suitably modifying the proof of theorem 1 in the main text.

OA.2.2 Absolute Loss

An important loss function that does not satisfy OA.A.1 is the absolute loss, which leads to the mean absolute error (MAD):

$$MAD(\hat{\mu}(\boldsymbol{w}_N)) := \mathbb{E}\left[|\hat{\mu}(\boldsymbol{w}_N) - \mu(\boldsymbol{\theta}_1)|\right].$$

The following theorem shows that an explicit local approximation to the MAD is possible, although the expression is notably different from that for smooth loss functions.

Theorem OA.2.2. *Let conditions of theorem OA.1.1 hold.*

Then (i) for all $N, T > T_0$, the MAD of the averaging estimator is finite; (ii) as $N, T \to \infty$ jointly, it holds that

$$\begin{aligned} &T^{1/2} \times MAD(\hat{\mu}(\boldsymbol{w}_N)) \\ &\to \sqrt{\sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{\left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{d}_0' \boldsymbol{\eta}_i - \boldsymbol{d}_0' \boldsymbol{\eta}_1\right)^2}{2\sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0} \right) \\ &\quad + \left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{d}_0' \boldsymbol{\eta}_i - \boldsymbol{d}_0' \boldsymbol{\eta}_1 \right) \left(1 - 2\Phi\left(-\frac{\sum_{i=1}^{\bar{N}} w_i \boldsymbol{d}_0' \boldsymbol{\eta}_i - \boldsymbol{d}_0' \boldsymbol{\eta}_1}{\sqrt{\sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}} \right) \right). \end{aligned} \tag{OA.12}$$

Feasible minimum MAD weights may now be obtained by replacing the population quantities in eq. (OA.12) with the estimators of section 3 and minimizing the resulting function.

OA.2.3 Proofs For Section OA.2

Proof of theorem OA.2.1. To begin, we expand $l\left(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N)\right)$ around $\mu(\boldsymbol{\theta}_1)$ to obtain

$$\begin{aligned} l\left(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N)\right) &= l(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)) + \partial_2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right)\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right) \\ &\quad + \frac{1}{2}\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \tilde{\mu}\right)\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right)^2 \\ &= \frac{1}{2}\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \tilde{\mu}\right)\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right)^2, \end{aligned} \tag{OA.13}$$

where $\partial_2 l(\cdot, \cdot)$ is the derivative with respect to the second argument of l; $\tilde{\mu}$ lies in the interval between $\mu(\boldsymbol{\theta}_1)$ and $\hat{\mu}(\boldsymbol{w}_N)$; and we use that $l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right) = \partial_2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right) = 0$ under OA.A.1.

By OA.A.1, $|\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \tilde{\mu}\right)| < C_2$ for some finite constant C_2. (i) then follows from eq. (OA.13) and theorem 1 as $\mathbb{E}\left[\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right)^2\right] < \infty$ for all $N, T > T_0$.

We now turn to establishing (ii). By equation (OA.13), the risk (OA.A.1) can be written

$$\begin{aligned}
&R(\mu(\boldsymbol{\theta}_1), \hat{\mu}(\boldsymbol{w}_N)) \\
&= \frac{1}{2}\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right) MSE(\hat{\mu}(\boldsymbol{w}_N)) \\
&+ \mathbb{E}\left[\left(\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \tilde{\mu}\right) - \partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right)\right)\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right)^2\right], \qquad \text{(OA.14)}
\end{aligned}$$

where the existence of the last moment in eq. (OA.14) is established below in eq. (OA.16).

To bound the moment term in eq (OA.14), first note that by OA.A.1 and the mean value theorem it holds that

$$\begin{aligned}
\left|\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \tilde{\mu}\right) - \partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right)\right| &= \left|\partial_2^3 l\left(\mu(\boldsymbol{\theta}_1), \dot{\mu}\right)\left(\mu(\boldsymbol{\theta}_1) - \tilde{\mu}\right)\right| \\
&\leq C_3|\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)| \\
&\xrightarrow{p} 0,
\end{aligned}$$

where $\dot{\mu}$ lies in the interval between $\mu(\boldsymbol{\theta}_1)$ and $\tilde{\mu}$; the penultimate line uses that $\tilde{\mu}$ lies in the interval between $\mu(\boldsymbol{\theta}_1)$ and $\hat{\mu}(\boldsymbol{w}_N)$; and where the last line follows from theorem OA.1.1. Applying theorem OA.1.1 again, we conclude that

$$\left(\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \tilde{\mu}\right) - \partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right)\right) T\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right)^2 \xrightarrow{p} 0. \qquad \text{(OA.15)}$$

Further, by applying lemma A.2.2 with $\delta > 2$ and suitably modifying the proof of theorem 1 it may be seen that

$$\sup_{N,T>T_0} \mathbb{E}\left[\left|\sqrt{T}\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right)\right|^4\right] < \infty.$$

Accordingly, the random variable in eq. (OA.15) is uniformly bounded in $L_{4/3}$ ($\boldsymbol{\eta}$-a.s.) as

$$\mathbb{E}\left[\left|\left(\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \tilde{\mu}\right) - \partial_2^2 l\left(\mu(\boldsymbol{\theta}_1), \mu(\boldsymbol{\theta}_1)\right)\right) T\left(\mu(\boldsymbol{\theta}_1) - \hat{\mu}(\boldsymbol{w}_N)\right)^2\right|^{4/3}\right]$$

$$\leq \sup_{N,T>T_0} C_3^{4/3}\,\mathbb{E}\left[\left|\sqrt{T}\left(\mu(\boldsymbol{\theta}_1)-\hat{\mu}(\boldsymbol{w}_N)\right)\right|^4\right]. \tag{OA.16}$$

By the dominated convergence theorem and eqs. (OA.15)-(OA.16) it now follows that

$$\mathbb{E}\left[\left(\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1),\tilde{\mu}\right)-\partial_2^2 l\left(\mu(\boldsymbol{\theta}_1),\mu(\boldsymbol{\theta}_1)\right)\right)T\left(\mu(\boldsymbol{\theta}_1)-\hat{\mu}(\boldsymbol{w}_N)\right)^2\right]\to 0 \tag{OA.17}$$

Eq. (OA.11) now follows directly from eqs. (OA.14) and (OA.17). □

The proof of theorem OA.2.2 on the local normality result of theorem OA.1.1. We establish that the moments of the estimator converge to the moments of the limit distribution.

Proof of theorem OA.2.2. First, the MAD of the averaging estimator is finite since by Hölder's inequality it holds that

$$\begin{aligned}\sqrt{T}\,\mathbb{E}\left[\left|\hat{\mu}(\boldsymbol{w}_N)-\mu(\boldsymbol{\theta}_1)\right|\right] &\leq \sqrt{T\times\mathbb{E}\left[\left(\hat{\mu}(\boldsymbol{w}_N)-\mu(\boldsymbol{\theta}_1)\right)^2\right]} \\ &\equiv \sqrt{T\times MSE\left(\hat{\mu}(\boldsymbol{w}_N)\right)}.\end{aligned} \tag{OA.18}$$

The proof of theorem 1 establishes that it holds that

$$\sup_{N,T>T_0} T\times MSE\left(\hat{\mu}(\boldsymbol{w}_N)\right)<\infty. \tag{OA.19}$$

(i) follows.

Second, by theorem OA.1.1

$$\sqrt{T}\left(\hat{\mu}(\boldsymbol{w}_N)-\mu(\boldsymbol{\theta}_1)\right)\Rightarrow N\left(\sum_{i=1}^{\bar{N}} w_i\boldsymbol{d}_0'\boldsymbol{\eta}_i-\boldsymbol{d}_0'\boldsymbol{\eta}_1,\sum_{i=1}^{\bar{N}} w_i^2\boldsymbol{d}_0'\boldsymbol{V}_i\boldsymbol{d}_0\right). \tag{OA.20}$$

The first absolute moment of the limiting random variable in eq. (OA.20) is given by the right hand expression in eq. (OA.12) (Elandt, 1961). The first absolute moment of $\sqrt{T}\left(\hat{\mu}(\boldsymbol{w}_N)-\mu(\boldsymbol{\theta}_1)\right)$ is exactly $T^{1/2}\times MAD(\hat{\mu}(\boldsymbol{w}_N))$. This random variable is uniformly bounded in L^2 by eqs. (OA.18)-(OA.19). Accordingly, this first absolute moment (scaled

MAD) converges to the first absolute moment of the limit by the dominated convergence theorem from from eqs. (OA.18)-(OA.20). □

OA.3 Confidence Intervals with the Minimum MSE Unit Averaging Estimator

Inference based on the minimum MSE estimator is challenging. A valid confidence interval (CI) must account for the variability of the individual estimators, the variability in the estimated weights, the bias of the averaging estimator, and the uncertainty about the bias.

In this section, we propose a practical simulation-based CI that tackles all four of the above challenges. In order to motivate its construction, we first propose a valid two-step asymptotic CI (subsection OA.3.1). This asymptotic CI may be challenging to compute in practice. Accordingly, we propose a one-step simulation-based CI based on the same principles as the asymptotic CI (subsection OA.3.2). This CI is straightforward to compute and shows favorable coverage and length properties in a Monte Carlo study (see subsection OA.4.4).

OA.3.1 Asymptotic Confidence Interval

To motivate the confidence intervals of algorithms 1-2, we first recall the result of theorem OA.1.1. Let $\bar{N}$ be a fixed positive integer, $\{\boldsymbol{w}_N\}$ an $\bar{N}$-vector of weights that does depend on data, and let $(w_{1N}, \dots, w_{\bar{N}N}) \to (w_1, \dots, w_{\bar{N}})$, where $\sum_{i=1}^{\bar{N}} w_i = 1$ (we consider the fixed-N case, the extension to the large-N case is immediate). Then the unit averaging estimator satisfies

$$\sqrt{T}\left(\hat{\mu}(\boldsymbol{w}_N) - \mu(\boldsymbol{\theta}_1)\right) \Rightarrow N\left(\sum_{i=1}^{\bar{N}} w_i \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1), \sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0\right).$$

For clarity, we assume that $\boldsymbol{d}_0$ and $\{\boldsymbol{V}_i\}_{i=1}^{\bar{N}}$ are known; these parameters can be consistently estimated as in the main text.

If $\{\boldsymbol{\eta}_i\}_{i=1}^{\bar{N}}$ were known, it would be possible to compute the true optimal weights

$$\boldsymbol{w}^o := \underset{\boldsymbol{w}^{\bar{N}} \in \Delta^{\bar{N}}}{\arg\min}\, \boldsymbol{w}' \boldsymbol{\Psi}_{\bar{N}} \boldsymbol{w}, \qquad \text{(OA.21)}$$

where $\boldsymbol{\Psi}_{\bar{N}}$ is an $\bar{N} \times \bar{N}$ matrix with elements $[\boldsymbol{\Psi}_{\bar{N}}]_{i\,i} = \boldsymbol{d}_0' \left((\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) (\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)' + \boldsymbol{V}_i \right) \boldsymbol{d}_0$ and $[\boldsymbol{\Psi}_{\bar{N}}]_{i\,j} = \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1) (\boldsymbol{\eta}_j - \boldsymbol{\eta}_1)' \boldsymbol{d}_0$.

As the optimal weights $\boldsymbol{w}^o$ do not depend on the observed sample, a valid $(1-\alpha) \times 100\%$ asymptotic confidence interval for $\mu(\boldsymbol{\theta}_1)$ would then be given by

$$\Bigg[\hat{\mu}(\boldsymbol{w}_N) - \frac{z_{1-\alpha/2}\sqrt{\sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}}{\sqrt{T}} - \frac{\sum_{i=2}^{\bar{N}} w_i \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)}{\sqrt{T}}, \\ \hat{\mu}(\boldsymbol{w}_N) + \frac{z_{1-\alpha/2}\sqrt{\sum_{i=1}^{\bar{N}} w_i^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}}{\sqrt{T}} - \frac{\sum_{i=2}^{\bar{N}} w_i \boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)}{\sqrt{T}} \Bigg], \quad \text{(OA.22)}$$

where z_α is the αth quantile of the standard normal distribution.

The key obstacle to forming interval (OA.22) is the unavailability of consistent estimators for $\boldsymbol{\eta}_i$ in the local framework (see the discussion before lemma 2). There is non-diminishing uncertainty around both the bias terms $\boldsymbol{d}_0(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)$ in (OA.22) and the optimal weights $\boldsymbol{w}^o$.

Algorithm 1: Asymptotic $(1-\alpha-\gamma) \times 100\%$ Confidence Interval for $\mu(\boldsymbol{\theta}_1)$

1 Let $L_{\bar{N},T}$ be an $(1-\gamma) \times 100\%$ asymptotic confidence set for $(\boldsymbol{d}_0'(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1), \ldots, \boldsymbol{d}_0'(\boldsymbol{\eta}_{\bar{N}} - \boldsymbol{\eta}_1))$;

2 **for** *each guess* $\boldsymbol{E}^\kappa \in L_{\bar{N},T}$ **do**

3 Form the $\bar{N} \times \bar{N}$ matrix $\boldsymbol{\Psi}_{\bar{N}}^\kappa$ as $[\boldsymbol{\Psi}_{\bar{N}}^\kappa]_{1,1} = \boldsymbol{d}_0' \boldsymbol{V}_1 \boldsymbol{d}_0$, $[\boldsymbol{\Psi}_{\bar{N}}^\kappa]_{1\,i} = [\boldsymbol{\Psi}_{\bar{N}}^\kappa]_{i\,1} = 0$, $[\boldsymbol{\Psi}_{\bar{N}}^\kappa]_{i\,i} = (E_i^\kappa)^2 + \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0$ and $[\boldsymbol{\Psi}_{\bar{N}}^\kappa]_{i\,j} = E_i^\kappa E_j^\kappa$ for $i, j = 2, \ldots, \bar{N}$, $i \neq j$.

4 Define

$$\boldsymbol{w}^{\boldsymbol{E}^\kappa} = \arg\min_{\boldsymbol{w}^{\bar{N}} \in \Delta^{\bar{N}}} \boldsymbol{w}^{\bar{N}\prime} \boldsymbol{\Psi}_{\bar{N}}^\kappa \boldsymbol{w}^{\bar{N}}.$$

5 Define

$$m_{\bar{N},T}(\boldsymbol{E}^\kappa) = \hat{\mu}(\boldsymbol{w}^{\boldsymbol{E}^\kappa}) - \frac{z_{1-\alpha/2}\sqrt{\sum_{i=1}^{\bar{N}} (w_i^{\boldsymbol{E}^\kappa})^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}}{\sqrt{T}} - \frac{\sum_{i=2}^{\bar{N}} w_i^{\boldsymbol{E}^\kappa} E_i^\kappa}{\sqrt{T}}$$

$$M_{\bar{N},T}(\boldsymbol{E}^\kappa) = \hat{\mu}(\boldsymbol{w}^{\boldsymbol{E}^\kappa}) + \frac{z_{1-\alpha/2}\sqrt{\sum_{i=1}^{\bar{N}} (w_i^{\boldsymbol{E}^\kappa})^2 \boldsymbol{d}_0' \boldsymbol{V}_i \boldsymbol{d}_0}}{\sqrt{T}} - \frac{\sum_{i=2}^{\bar{N}} w_i^{\boldsymbol{E}^\kappa} E_i^\kappa}{\sqrt{T}},$$

where z_α is the αth quantile of the standard normal distribution.

6 **end**

7 Define

$$\mathcal{I}_{\bar{N},T} = \left[\min_{\boldsymbol{E}^\kappa \in L_{\bar{N},T}} m_{\bar{N},T}(\boldsymbol{E}^\kappa), \max_{\boldsymbol{E}^\kappa \in L_{\bar{N},T}} M_{\bar{N},T}(\boldsymbol{E}^\kappa) \right]$$

To fully account for the above uncertainty, we propose a two-step interval $\mathcal{I}_{\bar{N},T}$, formally constructed in algorithm 1. First, we form a confidence region $L_{\bar{N},T}$ that asymptotically contains the true value $\{\boldsymbol{d}_0'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)\}_{i=1}^{\bar{N}}$ with probability at least $1-\gamma$. Each point in $L_{\bar{N},T}$ then forms a guess for the true bias parameters $(\boldsymbol{d}_0'(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1), \dots, \boldsymbol{d}_0'(\boldsymbol{\eta}_{\bar{N}} - \boldsymbol{\eta}_1))'$. Second, for each candidate bias vector $\boldsymbol{E}^\kappa$ in $L_{\bar{N},T}$, we form the corresponding optimal weights (OA.21). For these weights and the guess $\boldsymbol{E}^\kappa$ for the bias, we form an interval of the form (OA.22). The overall CI $\mathcal{I}_{\bar{N},T}$ is the union of such intervals for all of the values of $\boldsymbol{E}^\kappa$ considered.

A suitable asymptotic confidence region $L_{\bar{N},T}$ for the bias parameters can be formed using lemma 2. The lemma implies that

$$\begin{pmatrix} \sqrt{T}\left(\mu\left(\hat{\boldsymbol{\theta}}_2\right) - \mu\left(\hat{\boldsymbol{\theta}}_1\right)\right) \\ \vdots \\ \sqrt{T}\left(\mu\left(\hat{\boldsymbol{\theta}}_{\bar{N}}\right) - \mu\left(\hat{\boldsymbol{\theta}}_1\right)\right) \end{pmatrix} \Rightarrow \begin{pmatrix} \Lambda_2 - \Lambda_1 \\ \vdots \\ \Lambda_{\bar{N}} - \Lambda_1 \end{pmatrix}, \tag{OA.23}$$

$$\begin{pmatrix} \Lambda_2 - \Lambda_1 \\ \vdots \\ \Lambda_{\bar{N}} - \Lambda_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{d}_0'(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \\ \vdots \\ \boldsymbol{d}_0'(\boldsymbol{\eta}_{\bar{N}} - \boldsymbol{\eta}_1) \end{pmatrix}, \begin{pmatrix} \boldsymbol{d}_0'(\boldsymbol{V}_2 + \boldsymbol{V}_1)\boldsymbol{d}_0 & \cdots & \boldsymbol{d}_0'\boldsymbol{V}_1\boldsymbol{d}_0 \\ \vdots & \ddots & \vdots \\ \boldsymbol{d}_0'\boldsymbol{V}_1\boldsymbol{d}_0 & \cdots & \boldsymbol{d}_0'(\boldsymbol{V}_{\bar{N}} + \boldsymbol{V}_1)\boldsymbol{d}_0 \end{pmatrix}\right).$$

$L_{\bar{N},T}$ may then be formed as the highest density region of $\boldsymbol{\Lambda} = (\Lambda_2 - \Lambda_1, \dots, \Lambda_{\bar{N}} - \Lambda_1)$, replacing $\boldsymbol{d}_0$ and $\boldsymbol{V}_i$ with $\hat{\boldsymbol{d}}_1$ and $\hat{\boldsymbol{V}}_i$ if necessary (see remark OA.3.1 below for an example construction).

As the following theorem shows, the interval of algorithm 1 has correct coverage. The proof of theorem OA.3.1 can be found in subsection OA.3.3.

Theorem OA.3.1. *Let assumptions A.1-A.5 hold. Let interval $\mathcal{I}_{\bar{N},T}$ be defined as in algorithm 1. Then*

$$\liminf_{T\to\infty} P\left(\mu(\boldsymbol{\theta}_1) \in \mathcal{I}_{\bar{N},T}\right) \geq 1 - \alpha - \gamma.$$

Interval $\mathcal{I}_{\bar{N},T}$ tackles all four challenges set out at the beginning of section OA.3. It accounts for the variability of the individual estimators and the variability of the averaging estimator due to sampling uncertainty. It does so by including a suitable asymptotic variance term in $m_{\bar{N},T}$ and $M_{\bar{N},T}$ (see alg. 1). The last term in $m_{\bar{N},T}$ and $M_{\bar{N},T}$ accounts for the bias of the averaging estimator. Finally, searching through the confidence region $L_{\bar{N},T}$ quantifies the uncertainty in the bias parameters and the variability in the averaging weights due to $\boldsymbol{\eta}$. The coverage guarantee of $\mathcal{I}_{\bar{N},T}$ stems from the fact that $L_{\bar{N},T}$ includes the true bias vector with a high probability.

OA.3.2 One-Step Confidence Interval

In practice, it may be challenging to store $L_{\bar{N},T}$ in the memory of a computer. If $L_{\bar{N},T}$ is discretized into a grid of points, the number of potential candidate vectors $\boldsymbol{E}^{\kappa}$ scales exponentially with $\bar{N}$.

To overcome this challenge, we propose a one-step interval, formally defined in algorithm 2. In it, we replace an asymptotic confidence set $L_{\bar{N},T}$ with an approximation based on bootstrapping the individual time series. Intuitively, we use each bootstrap sample directly to form a possible guess $\boldsymbol{E}^{\kappa}$, instead of precomputing and storing the confidence region $L_{\bar{N},T}$. For each such bootstrap guess, we compute the corresponding averaging weights and the debiased averaging estimator. The interval $\mathcal{I}^{B}_{\bar{N},T}$ is formed using the quantiles of the bootstrap estimates. $\mathcal{I}^{B}_{\bar{N},T}$ captures both the variability of the estimator and the uncertainty about its bias.

There are two key parallels between intervals $\mathcal{I}^{B}_{\bar{N},T}$ and $\mathcal{I}_{\bar{N},T}$. First, algorithm 2 implicitly constructs a bootstrap version of $L_{\bar{N},T}$ for $\gamma = 0$ as $B \to \infty$. Second, algorithm 2 quantifies the variability of the averaging estimator by computing the bootstrap unit averaging estimators $\hat{\mu}^{(b)}(\boldsymbol{w})$. Interval $\mathcal{I}_{\bar{N},T}$ instead uses both the asymptotic distribution of $\hat{\mu}(\boldsymbol{w})$, and the guesses in $L_{\bar{N},T}$.

Given the above parallels, the asymptotic validity of the one-step interval $\mathcal{I}^{B}_{\bar{N},T}$ follows from the validity of the boostrap method applied to the individual time series and theorem OA.3.1.

In the simulation study of subsection OA.4.4, we find that $\mathcal{I}^{B}_{\bar{N},T}$ enjoys favorable coverage

Algorithm 2: One-step $(1-\alpha)\times 100\%$ Confidence Interval for $\mu(\boldsymbol{\theta}_1)$

Draw B bootstrap samples of the data, using the full cross-section and only resampling the time dimension observations (e.g. using the stationary bootstrap (Politis and Romano, 1994))

Set $b=1$ and **while** $b \leq B$ **do**

- Run individual estimation on the bth bootstrap sample to obtain $\{\hat{\boldsymbol{\theta}}_i^{(b)}\}_{i=1}^{\bar{N}}$ and the corresponding estimated variances $\{\hat{\boldsymbol{V}}_i^{(b)}\}_{i=1}^{\bar{N}}$.
- Set
$$\hat{\boldsymbol{d}}_1^{(b)} \coloneqq \nabla\mu(\hat{\boldsymbol{\theta}}_1^{(b)}).$$
- Form the matrix $\hat{\boldsymbol{\Psi}}_N^{(b)}$ using $\{\hat{\boldsymbol{\theta}}_i^{(b)}\}_{i=1}^{\bar{N}}$, $\{\hat{\boldsymbol{V}}_i^{(b)}\}_{i=1}^{\bar{N}}$, and $\hat{\boldsymbol{d}}_1^{(b)}$. according to the expression after eq. (3) in the main text.
- Compute the bootstrap minimum MSE averaging weights $\hat{\boldsymbol{w}}^{(b)}$ using $\hat{\boldsymbol{\Psi}}_N^{(b)}$ according to eq. (4) in the main text.
- Compute the debiased unit averaging estimator in the bth sample as
$$m_{\bar{N},T}^{(b)} \coloneqq \hat{\mu}^{(b)}(\hat{\boldsymbol{w}}^{(b)}) - \frac{\sum_{i=2}^{\bar{N}} \hat{w}_i^{(b)} \hat{\boldsymbol{d}}_1^{(b)} (\hat{\boldsymbol{\theta}}_i^{(b)} - \hat{\boldsymbol{\theta}}_1^{(b)})}{\sqrt{T}},$$
$$\hat{\mu}^{(b)}(\boldsymbol{w}) \coloneqq \sum_{i=1}^{\bar{N}} w_i \mu(\hat{\boldsymbol{\theta}}_i^{(b)}).$$
- **if** $b < B$ **then**
  - Set $b = b+1$.
- **end**

end

Define the one-step confidence interval $\mathcal{I}_{\bar{N},T}^B$ as
$$\mathcal{I}_{\bar{N},T}^B \coloneqq \left[q_{\bar{N},T}^{(b)}(\alpha/2), q_{\bar{N},T}^{(b)}(1-\alpha/2)\right],$$
where $q_{\bar{N},T}^{(b)}(\tau)$ be the τth quantile of $m_{\bar{N},T}^{(b)}$ (across b).

and length properties. Compared to the CI based on the individual estimator only, $\mathcal{I}_{\bar{N},T}^B$ generally has the same coverage, but is somewhat shorter. This reduction in length is stronger if the minimum MSE estimator is comparatively more efficient than the individual estimator.

OA.3.3 Proof of Theorem OA.3.1

Proof of theorem OA.3.1. Let $\boldsymbol{E}^{True} = (\boldsymbol{d}_0'(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1), \dots, \boldsymbol{d}_0'(\boldsymbol{\eta}_N - \boldsymbol{\eta}_1))$ be the true values of the bias parameters, and let P_0 be the probability when the true heterogeneity parameters are $[\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_{\bar{N}}]$. Define the event $A_{\bar{N},T} = \{\mu(\boldsymbol{\theta}_1) \in [m_{\bar{N},T}(\boldsymbol{E}^{True}), M_{\bar{N},T}(\boldsymbol{E}^{(True)})\}$.

Observe that since $\{\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{\bar{N}}\}$ are the asymptotic variances of individual estimators under P_0, it holds that $P_0(A_{\bar{N},T}) \to 1 - \alpha$.

Now also define the event that the true bias components are captured by the ellipse $L_{\bar{N},T}$: $B_{\bar{N},T} = \{\boldsymbol{E}^{True} \in L_{\bar{N},T}\}$. Then

$$\liminf_{T\to\infty} P_0(A_{\bar{N},T}) = \liminf_{T\to\infty} \left[P_0(A_{\bar{N},T} \cap B_{\bar{N},T}) + P_0(A_{\bar{N},T} \cap B^c_{\bar{N},T})\right] = 1 - \alpha.$$

By definition of $L_{\bar{N},T}$, $\limsup_{T\to\infty} P_0(B^c_{\bar{N},T}) \leq \gamma$, which implies that

$$\liminf_{T\to\infty} P_0(A_{\bar{N},T} \cap B_{\bar{N},T}) \geq 1 - \alpha - \gamma. \tag{OA.24}$$

When $B_{\bar{N},T}$ holds, $\boldsymbol{E}^{True} \in L_{\bar{N},T}$. Thus $\boldsymbol{E}^{True}$ is one of the values $\boldsymbol{E}^\kappa$ considered in the second step, which implies that under the event $B_{\bar{N},T}$

$$[m(\boldsymbol{E}^{True}), M(\boldsymbol{E}^{True})] \subset \left[\min_{\boldsymbol{E}^\kappa \in L_{\bar{N},T}} m_{\bar{N},T}(\boldsymbol{E}^\kappa), \max_{\boldsymbol{E}^\kappa \in L_{\bar{N},T}} M_{\bar{N},T}(\boldsymbol{E}^\kappa)\right] \equiv \mathcal{I}_{\bar{N},T}. \tag{OA.25}$$

Combining eqs. (OA.24) and (OA.25), we conclude that

$$\liminf_{T\to\infty} P_0\left(\left\{\mu(\boldsymbol{\theta}_1) \in \mathcal{I}_{\bar{N},T}\right\} \cap B_{\bar{N},T}\right) \geq 1 - \alpha - \gamma.$$

Last, trivially it holds that $P_0\left(\left\{\mu(\boldsymbol{\theta}_1) \in \mathcal{I}_{\bar{N},T}\right\}\right) \geq P_0\left(\left\{\mu(\boldsymbol{\theta}_1) \in \left\{\mu(\boldsymbol{\theta}_1) \in \mathcal{I}_{\bar{N},T}\right\}\right\} \cap B_{\bar{N},T}\right)$, which yields the desired statement about coverage of $\mathcal{I}_{\bar{N},T}$:

$$\liminf_{T\to\infty} P_0\left(\left\{\mu(\boldsymbol{\theta}_1) \in \mathcal{I}_{\bar{N},T}\right\}\right) \geq 1 - \alpha - \gamma.$$

□

Remark OA.3.1 (Forming $L_{\bar{N},T}$ for $\bar{N} = 2$)**.** The key ingredient of algorithm 1 is the confidence region $L_{\bar{N},T}$. Such an region may be based on the convergence relation (OA.23).

For example, if $\bar{N} = 2$, then

$$L_{2,T} = \left[\sqrt{T}\left(\mu\left(\hat{\boldsymbol{\theta}}_2\right) - \mu\left(\hat{\boldsymbol{\theta}}_1\right)\right) - z_{1-\gamma/2}\sqrt{\boldsymbol{d}_0'(\boldsymbol{V}_2 + \boldsymbol{V}_1)\boldsymbol{d}_0}, \right.$$
$$\left. \sqrt{T}\left(\mu\left(\hat{\boldsymbol{\theta}}_2\right) - \mu\left(\hat{\boldsymbol{\theta}}_1\right)\right) - z_{\gamma/2}\sqrt{\boldsymbol{d}_0'(\boldsymbol{V}_2 + \boldsymbol{V}_1)\boldsymbol{d}_0} \right]$$

forms a suitable $(1-\gamma)\times 100\%$ asymptotic CI for $\boldsymbol{d}_1'(\boldsymbol{\eta}_i - \boldsymbol{\eta}_1)$, where z_τ is the τth quantile of the standard normal distribution. $\boldsymbol{V}_1, \boldsymbol{V}_2$, and $\boldsymbol{d}_0$ may be replaced by consistent estimators. This logic generalizes to higher dimensions.

OA.4 Further Materials for the Simulation Study

In this section, we extend the analysis of the Monte Carlo study of section 4. First, we consider two further focus parameters and an additional sample size $T = 180$ for the estimators of section 4 (subsection OA.4.1). Second, we analyze how the performance of the large-N estimators depends on their tuning parameters (subsection OA.4.2). Third, we study the weights generated by the minimum MSE estimator (subsection OA.4.3). Finally, we analyze the coverage and length properties of the confidence interval of section OA.3.2 (subsection OA.4.4).

The overall practical conclusions are broadly in line with the results of sections 4-5. In the absence of prior information, we recommend using the fixed-N or the top units large-N estimator. Both offer gains in the MSE for all focus parameters almost everywhere in the parameter space, without need for prior information. Furthermore, the top units specification is generally insensitive to the number of top units, while the fixed-N estimator has no tuning parameters. However, if prior information is available, using it can yield stronger improvements in the MSE.

The design of the study is as in section 4, with three additions. First, we consider a new intermediate value of $T = 180$. This value lies between the moderate- and large-T settings, with an average t-statistics of 5 (see remark 1). Second, we consider two additional focus parameters — the coefficient β_1 and the MSE-optimal forecast for y_{1T+1} given $x_{1T+1} = 1$ and y_{1T} — the conditional expectation $\mathbb{E}[y_{1T+1}|y_{1T}, x_{1T+1} = 1] = \lambda_1 y_{1T} + \beta_1$. We evaluate the performance of the averaging estimators for the same grid of values of λ_1 as in the main text. Note that both new focus parameters are sample-dependent, even for a given value of λ_1. The key measure of interest — the MSE of the form $\mathbb{E}\left[(\hat{\mu}(\boldsymbol{w}) - \mu(\boldsymbol{\theta}_1))^2|\lambda_1 = c\right], \lambda_1 \in [0.2, 0.8]$ — averages over the distributions of these focus parameters. Third and last, we consider an additional "coefficient pre-clustering" data-driven large-N specification. For this approach, we first cluster the individual estimates into k clusters using k-means. The units allocated to the cluster of the target unit are left unrestricted.

OA.4.1 MSE for All Sample Sizes and Focus Parameters

In this section, we report the results of estimating all three focus parameters (λ_1, β_1, the optimal forecast $\mathbb{E}[y_{1T+1}|y_{1T}, x_{1T+1}]$) using the estimators considered in section 4. We also use the pre-clustering large-N estimator defined above with $k = 4$ coefficient clusters. The results are visually presented on figs. OA.1-OA.5. On figs. OA.1-OA.3, we plot the MSE of each averaging approach relative to the MSE of the individual estimators. Additionally, figs. OA.4-OA.5 depict the bias and the relative variance of the averaging estimators for $\mu(\boldsymbol{\theta}_1) = \lambda_1$ for all values of (N, T) considered. Note that in this case the focus parameter is held fixed for each given value of λ_1, permitting the bias and the variance to be computed.

A common pattern in figs. OA.1-OA.5 is that the fixed-N, the associated top units large-N, and the "most similar" large-N estimators offer gains in the MSE for (almost) all values of λ_1 considered. The results for these estimators align with those in the main text, to which we refer for discussion.

The other estimators generally perform worse for at least one parameter for a non-trivial share of λ_1. For example, pre-clustering coefficients performs favorably for estimating λ_1, but does not improve forecasting or estimation of β_1 relative to the individual estimator (fig. OA.3), paralleling the results of the empirical application (see section OA.5). The mean group and the AIC-weighted estimators for β_1 and the forecast perform worse than the individual estimator by an order of magnitude, and thus are not reported on the figures.

There are several other features of interest in the results. First, the gains in the MSE for $T = 180$ are stronger than for $T = 600$ and weaker than for $T = 60$, as expected. Second, recall that β_1 and λ_1 are independent. Consequently, the reported MSE for β_1 and the forecast averages over the unconditional distribution of β_1. Third, the MSE for β_1 and the forecast depends on the value of λ_1 through the value of the focus parameter (the forecast only) and the statistical properties of the individual estimator of unit 1. The dependence on λ_1 is somewhat complex for the forecast, as λ_1 affects both the actual value of the forecast and the distribution of y_{1T}.

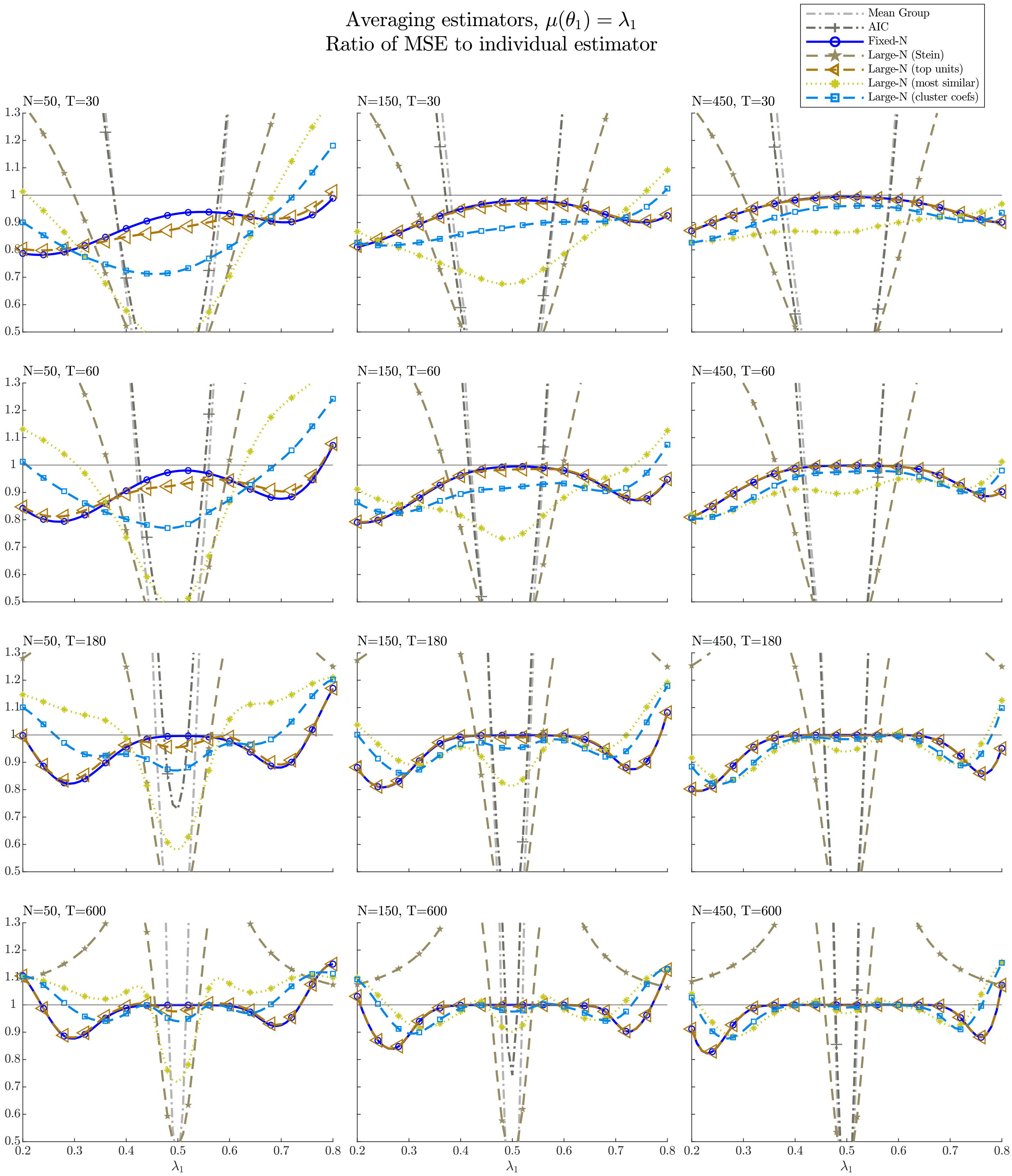

Figure OA.1: MSE of unit averaging estimators relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Note: part of this figure is reported as fig. 1 in the main text.

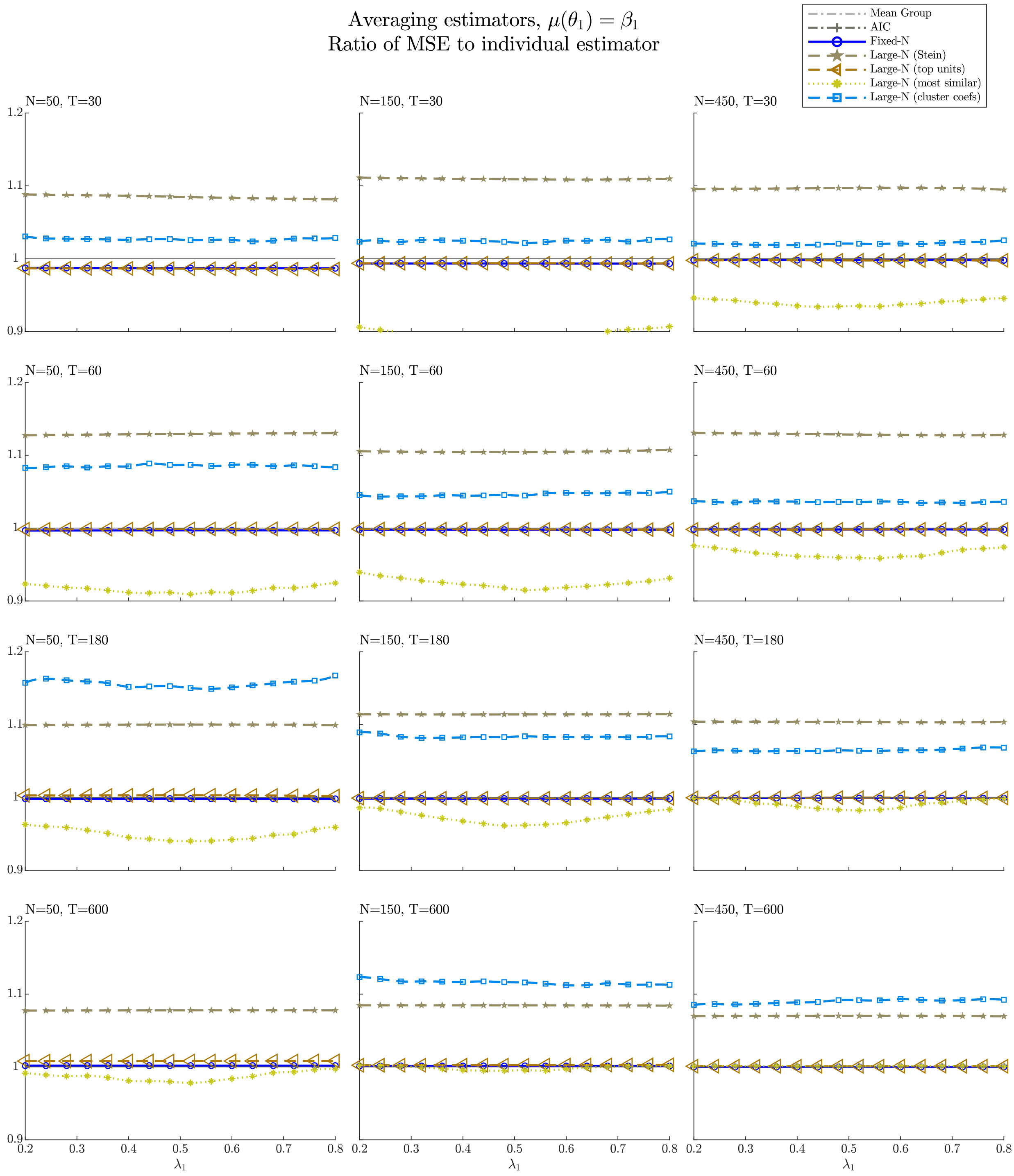

Figure OA.2: MSE of unit averaging estimators relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \beta_1$. Note: mean group and AIC estimators have MSE>1.5, and are not captured by the plot.

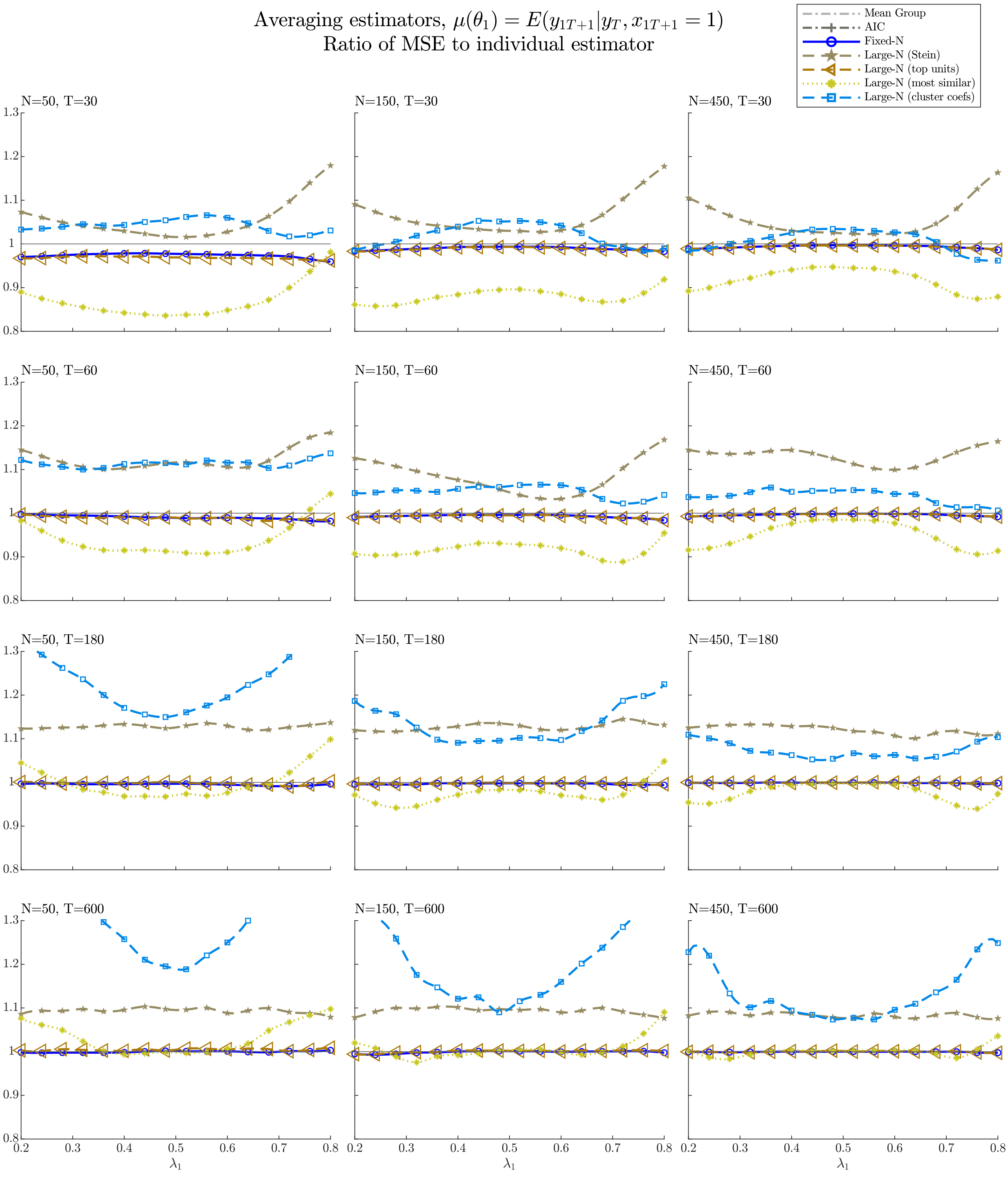

Figure OA.3: MSE of unit averaging estimators relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \mathbb{E}[y_{1T+1}|y_{1T}, x_{1T+1} = 1]$ (MSE-optimal forecast for y_{1T+1}). Note: mean group and AIC estimators have MSE>1.5, and are not captured by the plot.

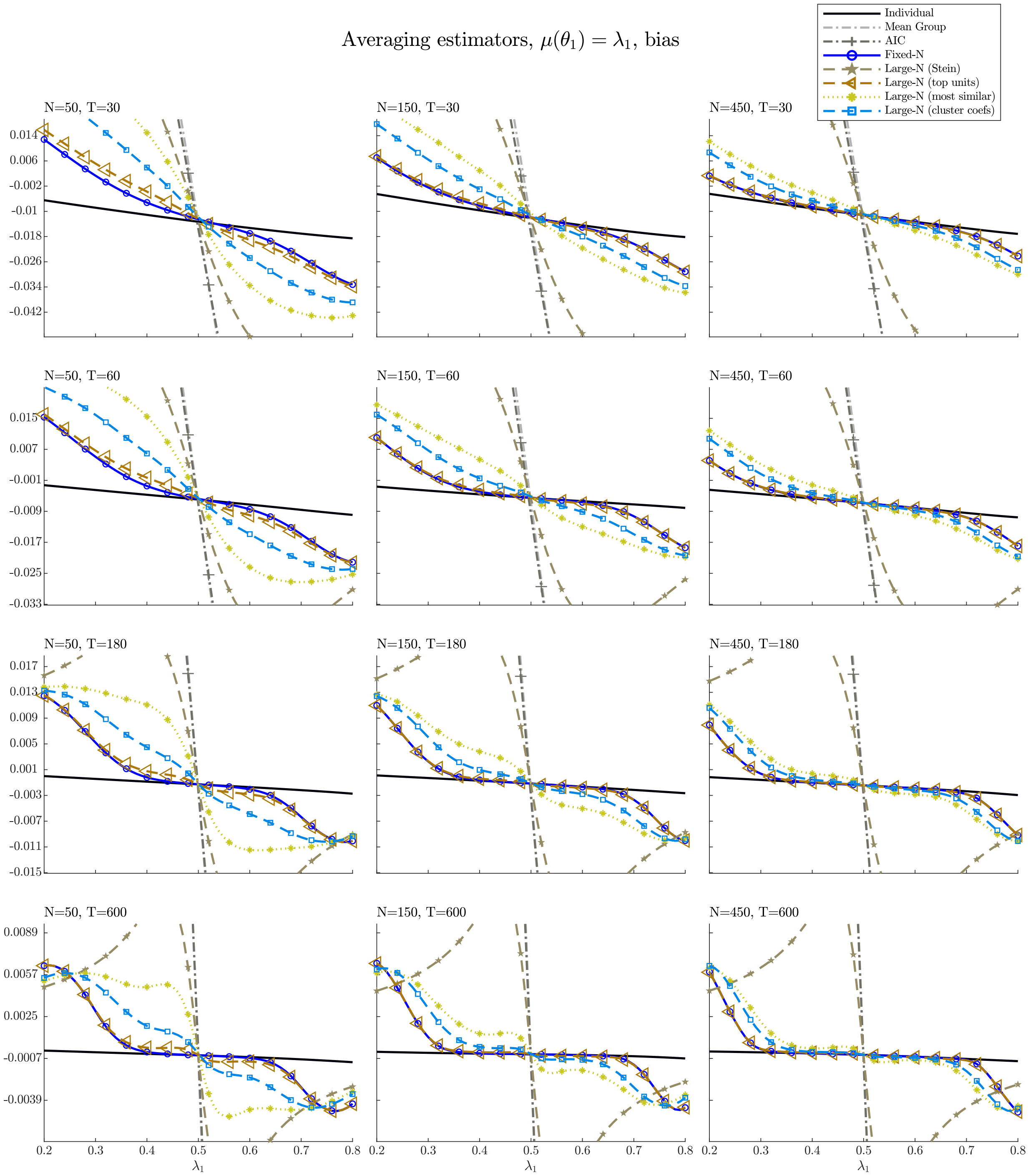

Figure OA.4: Bias of unit averaging estimators. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Note: part of this figure is reported as fig. 2 in the main text.

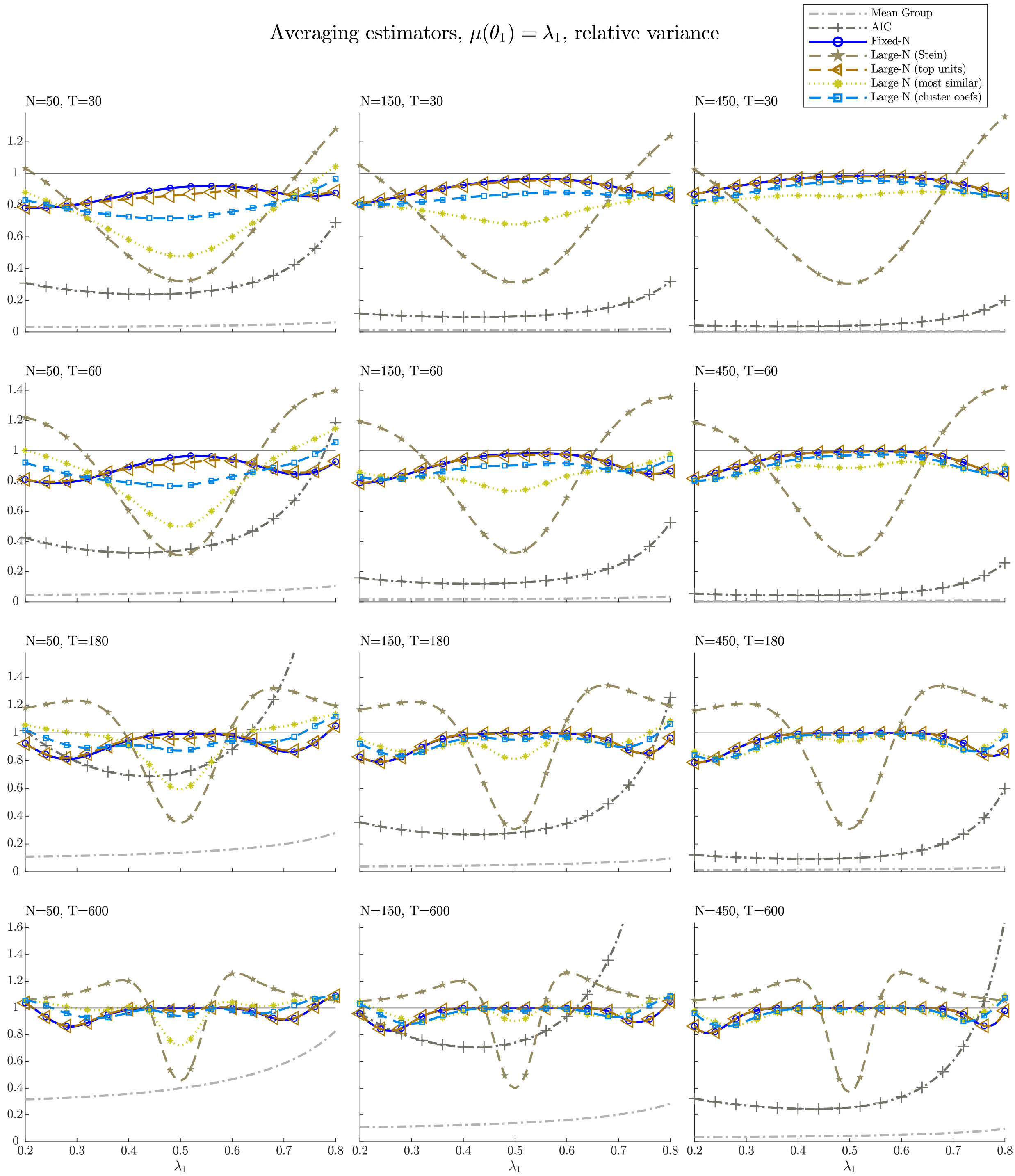

Figure OA.5: Variance of unit averaging estimators relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Note: part of this figure is reported as fig. 2 in the main text.

OA.4.2 Choice of Unrestricted Units

The set of unrestricted units acts as a tuning parameter for the large-N estimator. In this section, we examine how the choice of this set affects the performance of the large-N estimators of section 4, along with the pre-clustering large-N specification. For the estimators considered, the set of unrestricted units is fully determined by a scalar parameter (except for the Stein-like estimator that has no tuning parameters). Specifically,

- For the "most similar" estimator, the scalar parameter is the number k of units whose parameter vector $\boldsymbol{\theta}_i$ is closest to the parameter vector $\boldsymbol{\theta}_1$ of the target unit. We consider five specifications for k, two of which are independent of N ($k = 10, 25$) and three that depend on N ($k = 0.1N, 0.25N$ and $0.5N$).
- For the top units specification, the parameter is the number k of the units with the largest fixed-N weights. We consider the same values of k as for the "most similar" estimator: $k = 10, 25, 0.1N, 0.25N, 0.5N$.
- For the pre-clustering, the parameter is the number k of coefficient clusters. We consider $k = 2, 4$, and 8 clusters.

Figs. OA.6-OA.14 report the MSE, bias, and variance for the specifications considered. We report the results only for $\mu(\boldsymbol{\theta}_1) = \lambda_1$, as the ranking of the estimators is identical for the other two focus parameters.

As in the main text, the flexibility of the estimator controls a trade-off between stronger improvements for units close to the mean $\mathbb{E}[\lambda_1] = 0.5$ versus a stronger improvements for less typical units. The trade-off appears for all of the estimators, though it is less pronounced for the top units estimator (fig. OA.9). This trade-off is not identical to the bias-variance trade-off. More flexible estimators have uniformly lower bias (figs. OA.7, OA.10, OA.13). However, there is no uniform domination in terms of the variance: more flexible estimators have lower variance for more extreme values of λ_1; less flexible estimators have lower variance for λ_1 closer to the $\mathbb{E}[\lambda_1]$.

The performance of the top units estimator only weakly depends on the number of the top units chosen. The MSE profile of the estimator is close to that of the fixed-N estimator, although it is also somewhat affected by trade-off described above.

In contrast, the MSE profile of the "most similar" and the pre-clustering estimators varies more strongly with their tuning parameters. The variation follows the above trade-off as well, and no specification dominates any other. However, all the specifications yield an improvement over the unit-specific estimator, provided the cross-section is large enough (except potentially for $\lambda_1 \approx 0.8$, see the discussion in the main text).

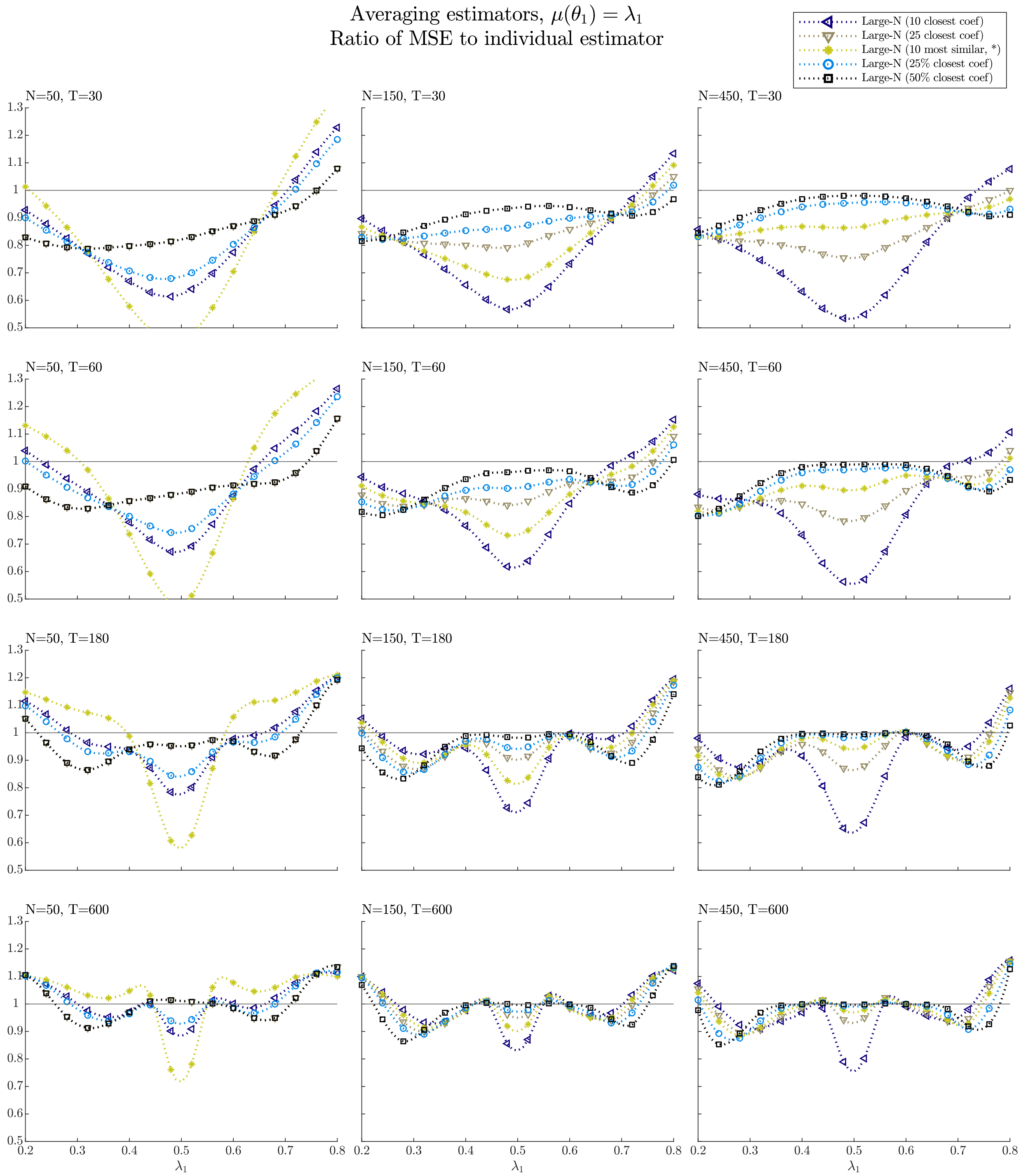

Figure OA.6: Relative MSE of large-N averaging estimators using coefficient similarity information. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. * – specification reported in the main text.

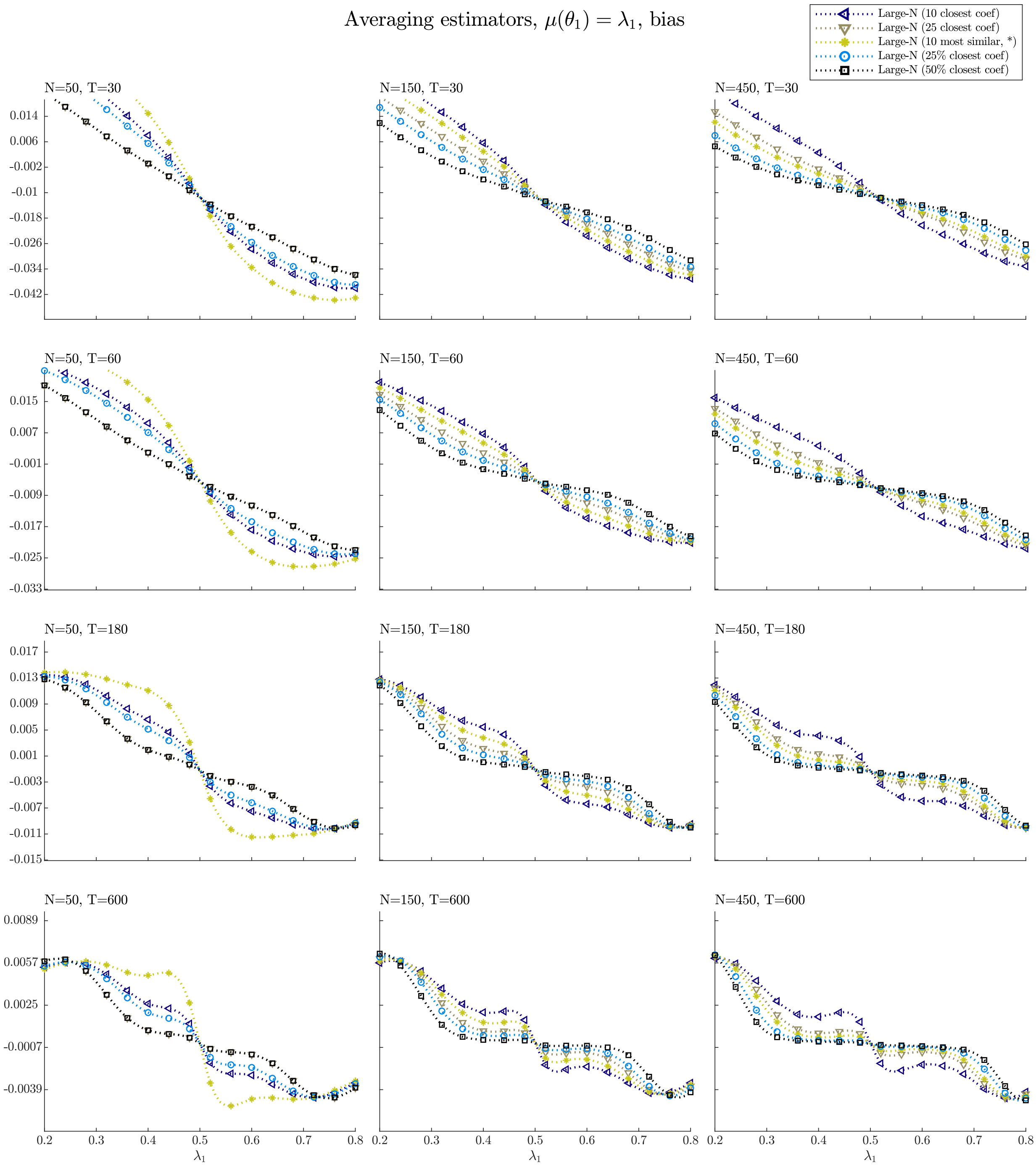

Figure OA.7: Bias of large-N averaging estimators using coefficient similarity information. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. * – specification reported in the main text.

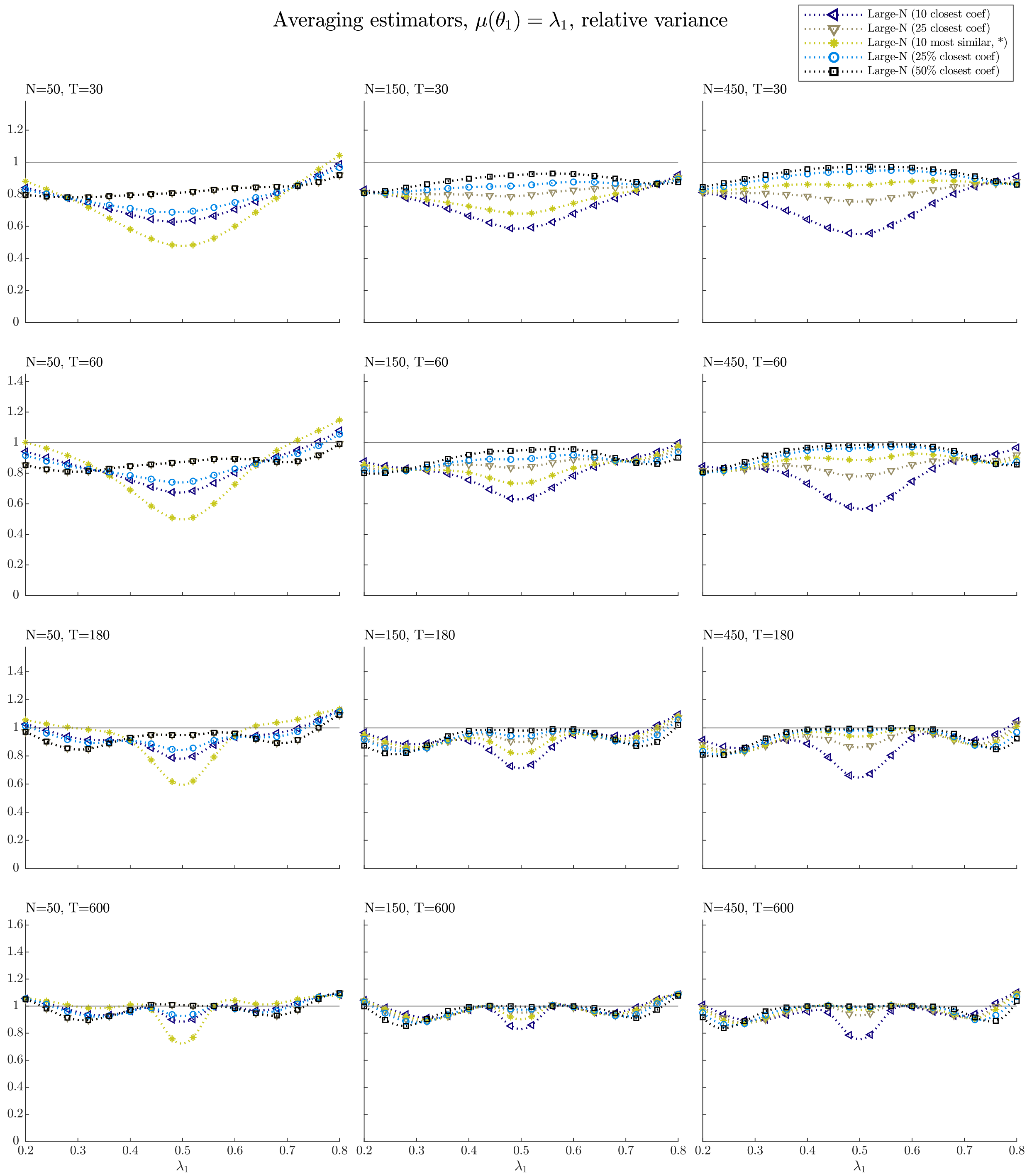

Figure OA.8: Relative variance of large-N averaging estimators using coefficient similarity information. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. * – specification reported in the main text.

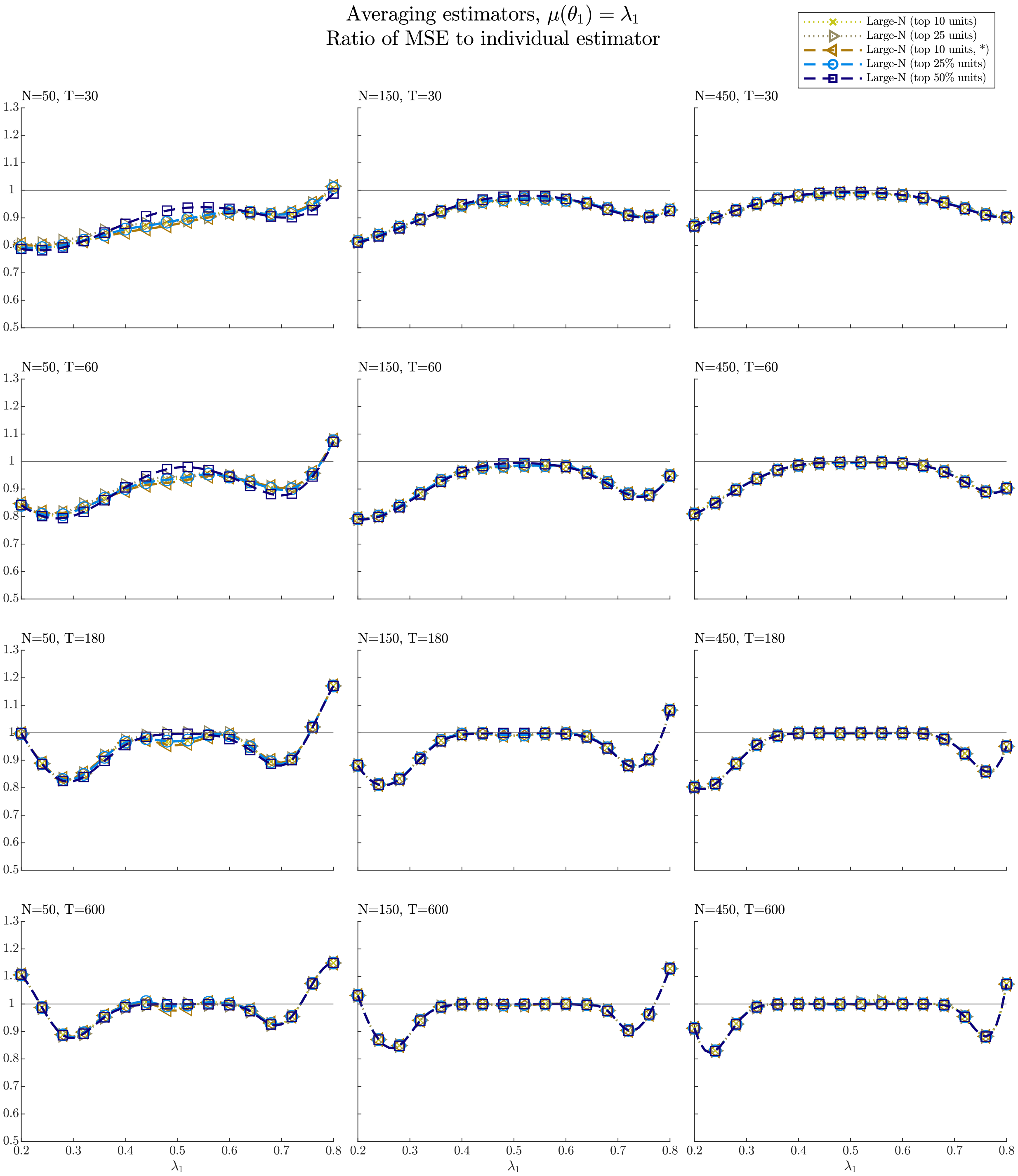

Figure OA.9: MSE of top unit large-N averaging estimators relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. * – specification reported in the main text.

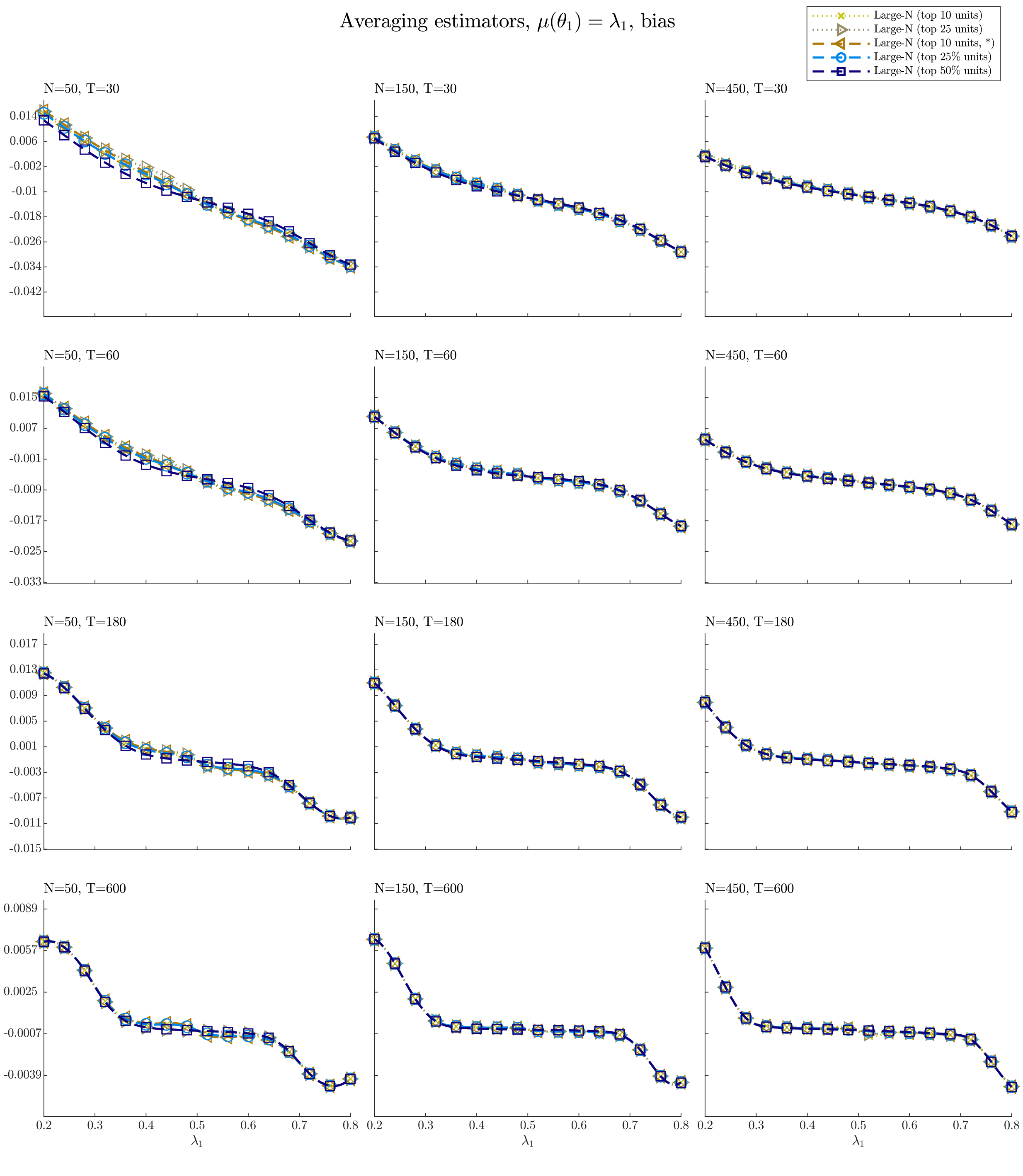

Figure OA.10: Bias of top unit large-N averaging estimators. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. * – specification reported in the main text.

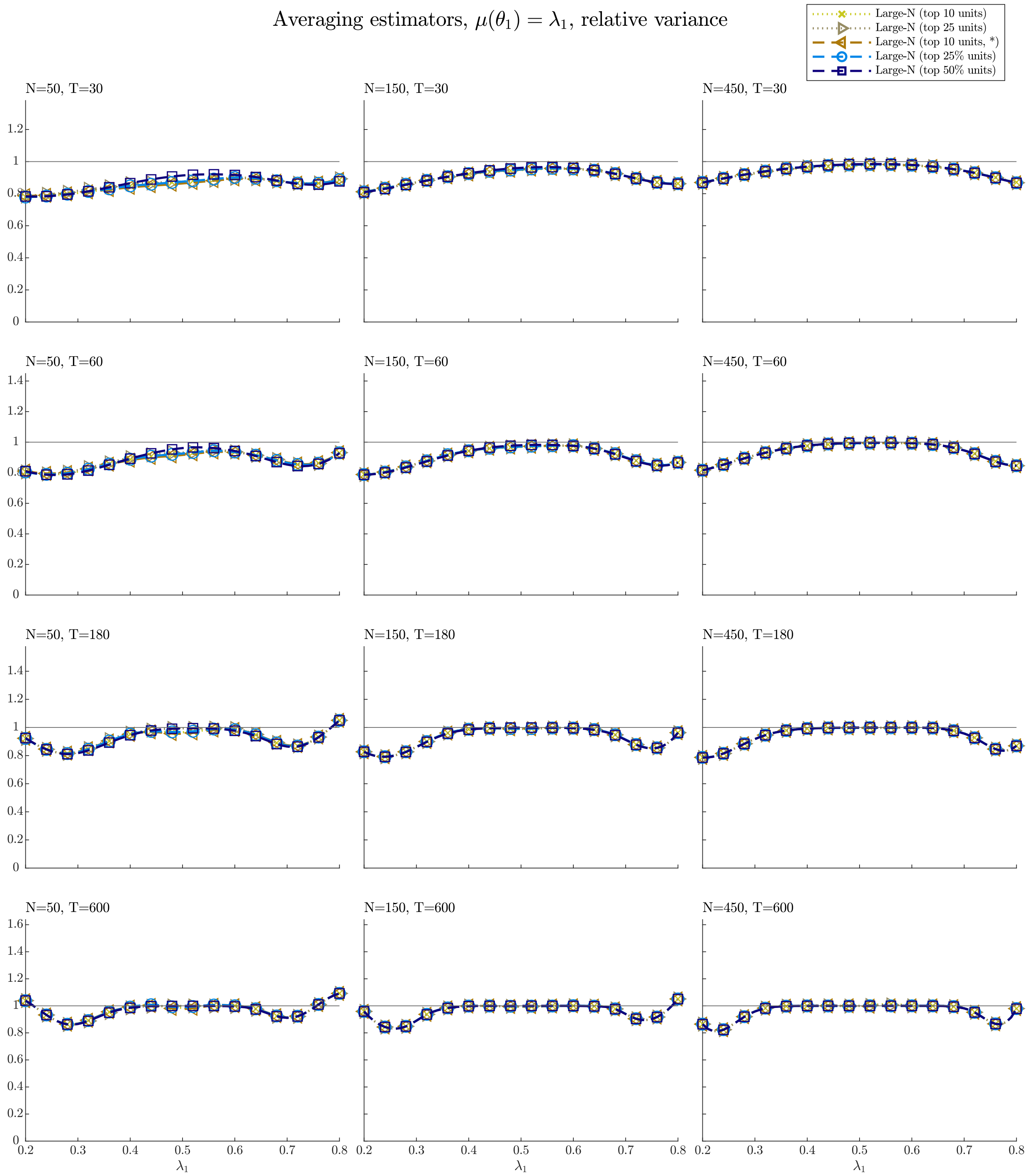

Figure OA.11: Relative variance of top unit large-N averaging estimators relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. * – specification reported in the main text.

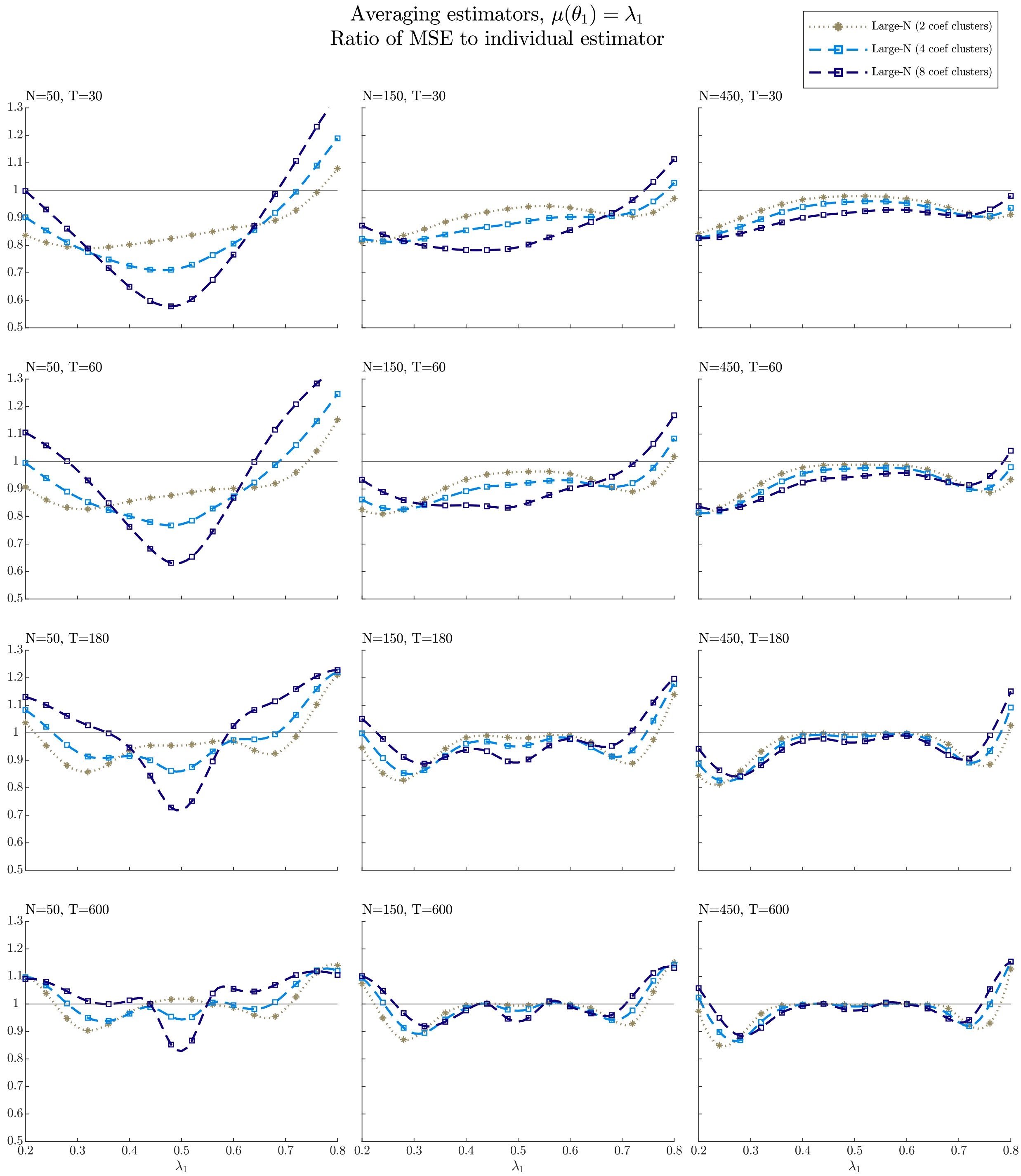

Figure OA.12: MSE of large-N averaging estimators with coefficient clustering relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$.

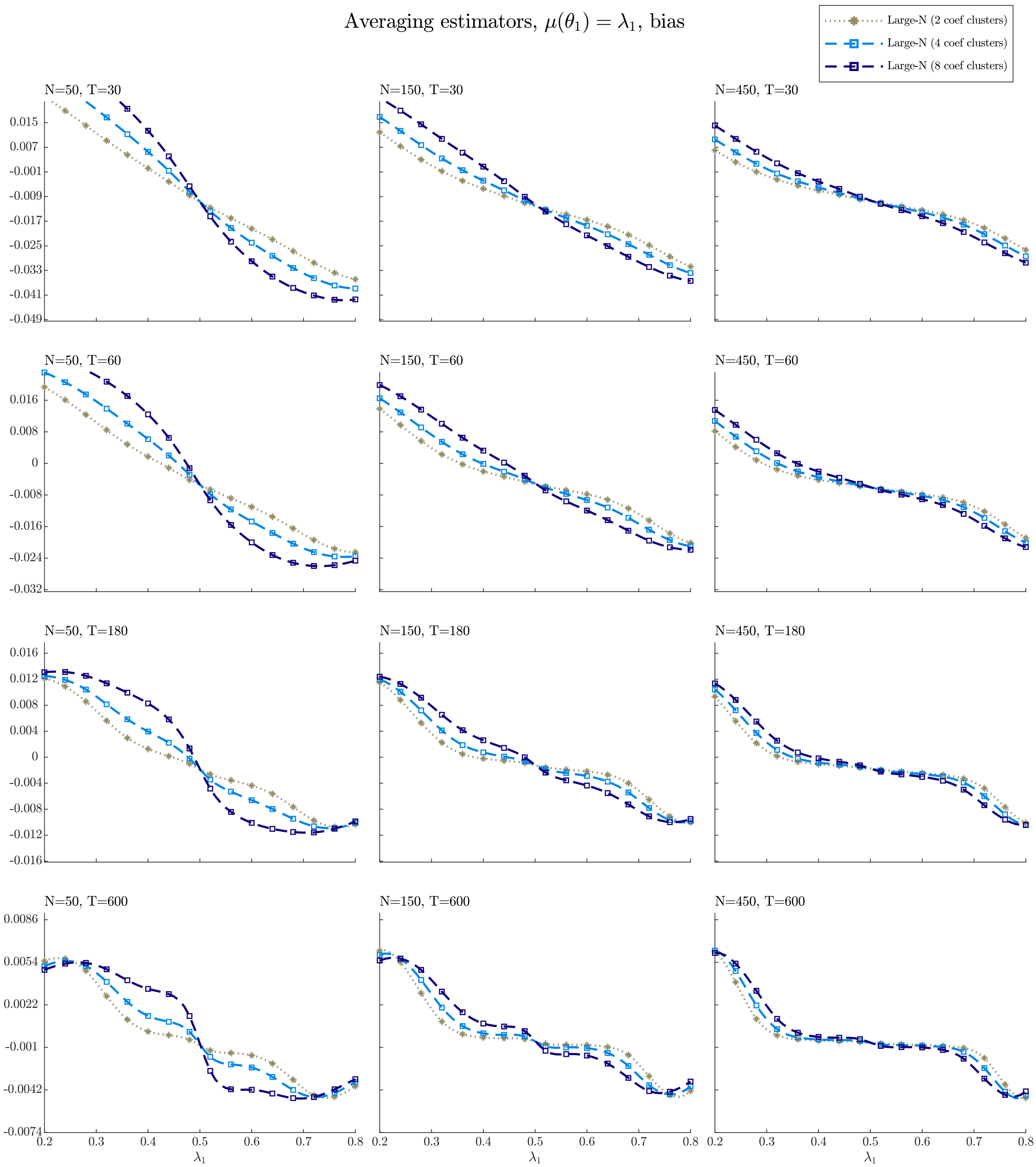

Figure OA.13: Bias of large-N averaging estimators with coefficient clustering. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$.

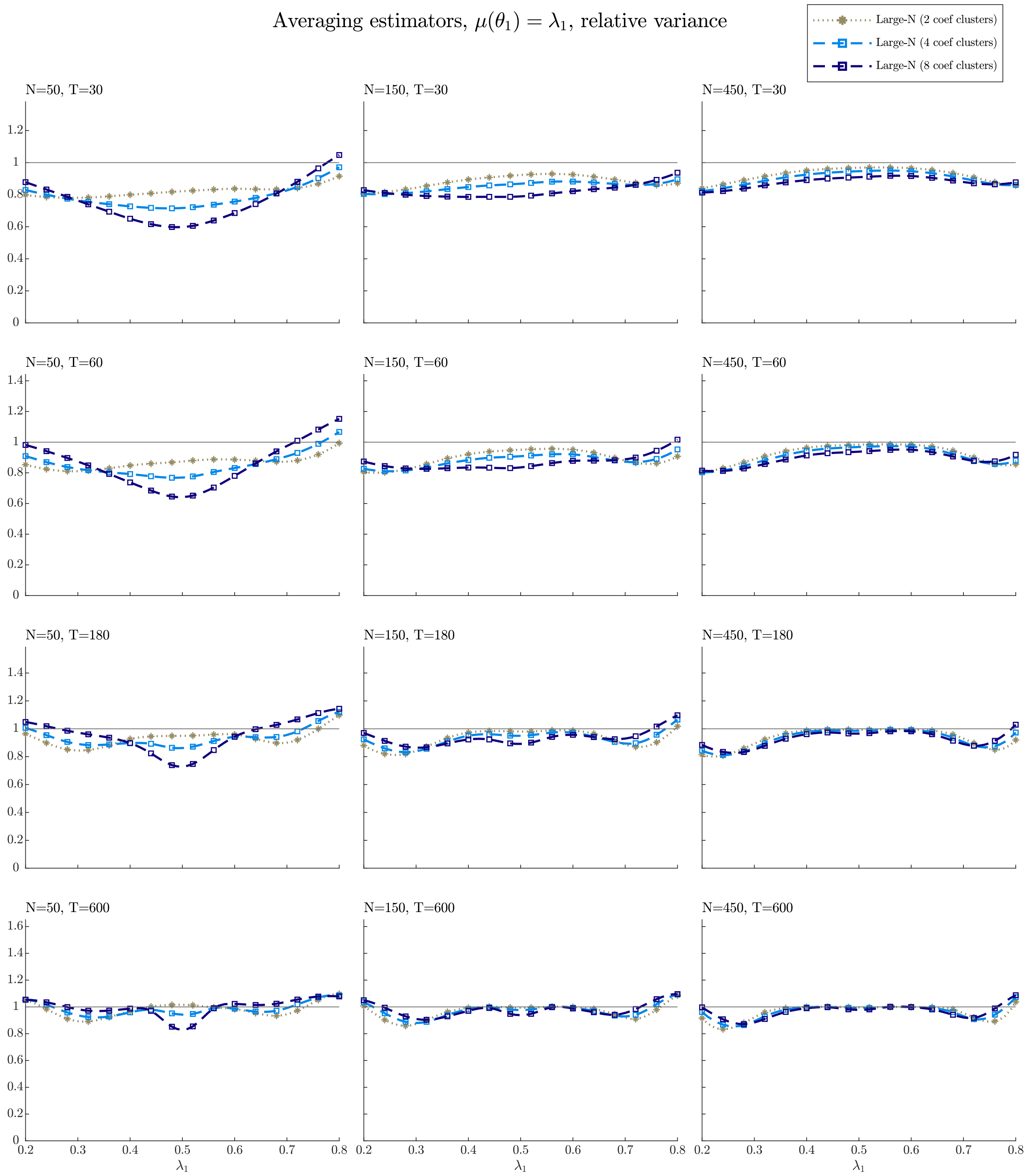

Figure OA.14: Relative variance of large-N averaging estimators with coefficient clustering relative to the individual estimator. Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$.

OA.4.3 Optimal Weights and Unrestricted Units

We now take a deeper look at the weights and the unrestricted unit chosen by the minimum MSE estimators of this simulation study. Specifically, we consider:

(i) The weight assigned to the target unit. On figs. OA.15-OA.17, we plot the average weight assigned to unit 1 for all of the focus parameters considered (λ_1, β_1, and the optimal forecast $\mathbb{E}[y_{1T+1}|y_{1T}, x_{1T+1} = 1]$). Note that fig. OA.15 expands on fig. 3 in the main text.

(ii) The weights assigned to non-target units, as a function of their own parameter value and the value of the target parameter. On figs. OA.18-OA.21, we report the expected weight assigned to a unit with parameter λ_{alt} when the focus parameter is λ_1, for all possible values of $(\lambda_1, \lambda_{alt})$.

(iii) The probability of a non-target unit being unrestricted, as a function of their own parameter value and the value of the target parameter. This probability is reported on figs. OA.22-OA.24 for the large-N estimators of subsection OA.4.1.

(iv) The average maximum difference between the weights of restricted units for several large-N estimators vs. the weights assigned to those units by the fixed-N estimator (fig. OA.25).

(v) The average difference in total mass assigned to the restricted set by several large-N estimators vs. the total mass assigned to those units by the fixed-N estimator (fig. OA.26).

Our key result is that the minimum MSE estimator is responsive to the target value in the following three senses. First, the estimator assigns larger weights to units with more similar values of the target parameter, regardless of the weighting scheme (figs. OA.18-OA.21). Second, the estimator detects whether the target unit is close to the mean or closer to the boundaries (figs. OA.15, OA.22-OA.24). In the former case, it assigns less weight to the target unit. Instead, it places more mass on the restricted set, which estimates the expected value of the target parameter with greater precision. In the latter case, more weight is assigned to the individual-specific estimator of the target unit. Third, the large-N estimators with data-driven unrestricted unit sets (top units and pre-clustering) select units with similar values of the target parameter into the unrestricted set (figs. OA.25-OA.26).

The flexibility of the estimator significantly influences the dispersion of the weights. More flexible estimators spread the weights more widely across non-target units, as can be seen by contrasting the weights of the fixed-N estimator (fig. OA.18) with the weights of the "most similar" large-N estimator (fig. OA.19). More flexible estimators also place less weight on the target unit for all focus parameters and sample sizes (OA.15-OA.17).

The value of λ_1 is another key driver of the weights and the unrestricted units. For λ_1 close to $\mathbb{E}[\lambda_1] = 0.5$, all of the weighting schemes spread their weights more widely and assign lower weights to each given unit (figs. OA.18-OA.21). Two factors drive this effect. First, there are generally more units with λ_i close to $\mathbb{E}[\lambda_i]$ under the DGP. Spreading the unrestricted weights across such units permits a greater reduction in variance without any increase in bias. Second, the restricted component, if present, estimates $\mathbb{E}[\lambda_1]$ with high precision (recall that the restricted units are assigned equal weights, see section 2). Accordingly the estimators also place a larger mass on the restricted set in this region (fig. OA.26). In contrast, if λ_1 is more extreme, units with similar λ_i are relatively scarce. In this case, similar units receive relatively higher weights, more weight is assigned to the individual estimator of unit 1, and less weight is given to the restricted set.

Finally, the two data-driven large-N procedures (top units and pre-clustering) select similar weights, but differ somewhat in their unrestricted sets (figs. OA.20-OA.21, OA.23-OA.24). The difference arises for λ_1 close to $\mathbb{E}[\lambda_1]$. For these values, each individual unit has a notably lower probability of being unrestricted for the top units estimator than for the pre-clustering estimator. This effect is driven by the fixed size of the unrestricted size of the top units estimator (10% of N) and the greater abundance of units with λ_i close to $\mathbb{E}[\lambda_i]$. In contrast, the unrestricted set of the pre-clustered estimator does not have a limited size, and so it consistently includes all units within a given distance of λ_1.

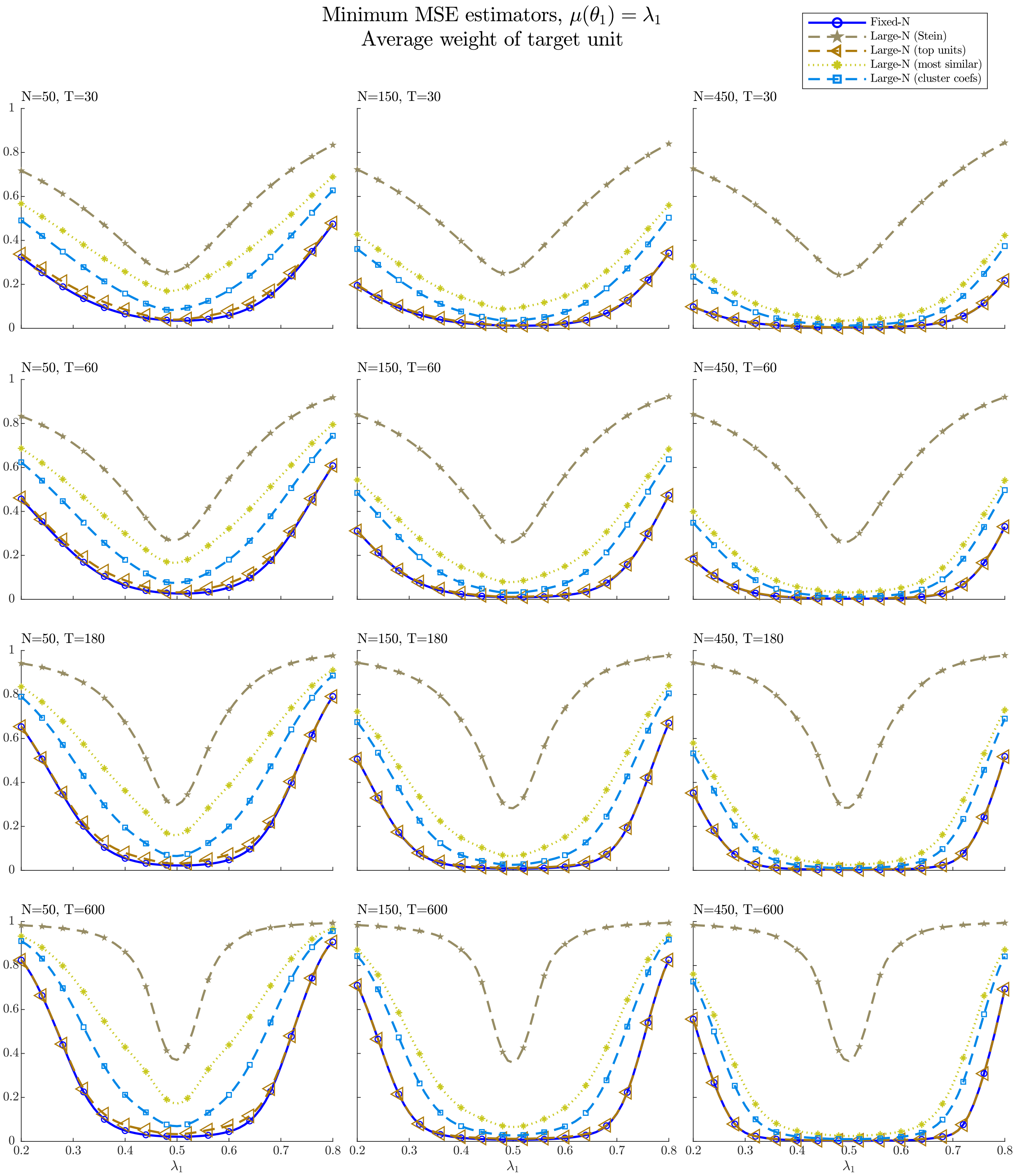

Figure OA.15: Average weight of target unit (unit 1). Focus parameter $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Note: part of this figure is reported as fig. OA.15 in the main text.

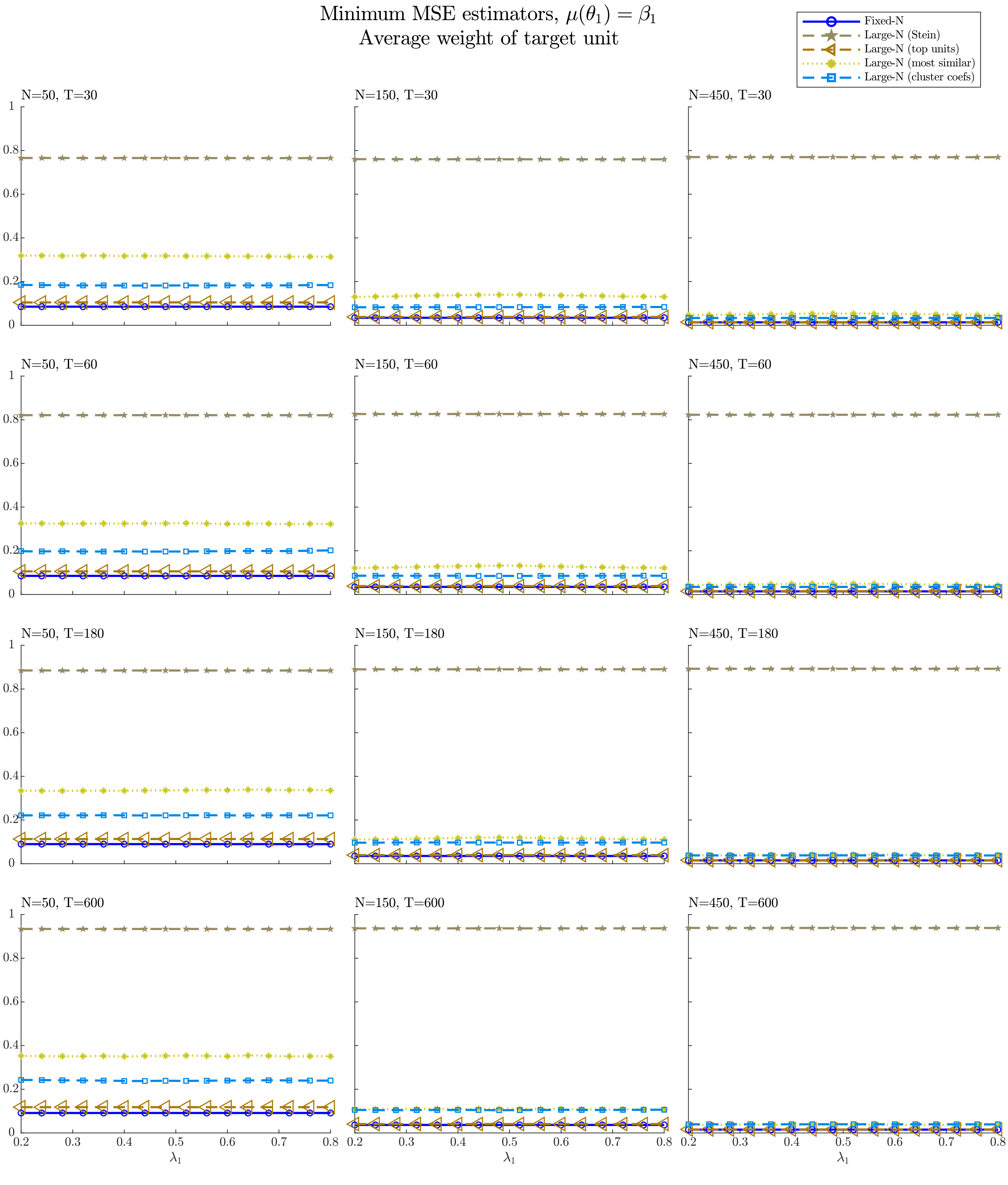

Figure OA.16: Average weight of target unit (unit 1). Focus parameter $\mu(\boldsymbol{\theta}_1) = \beta_1$.

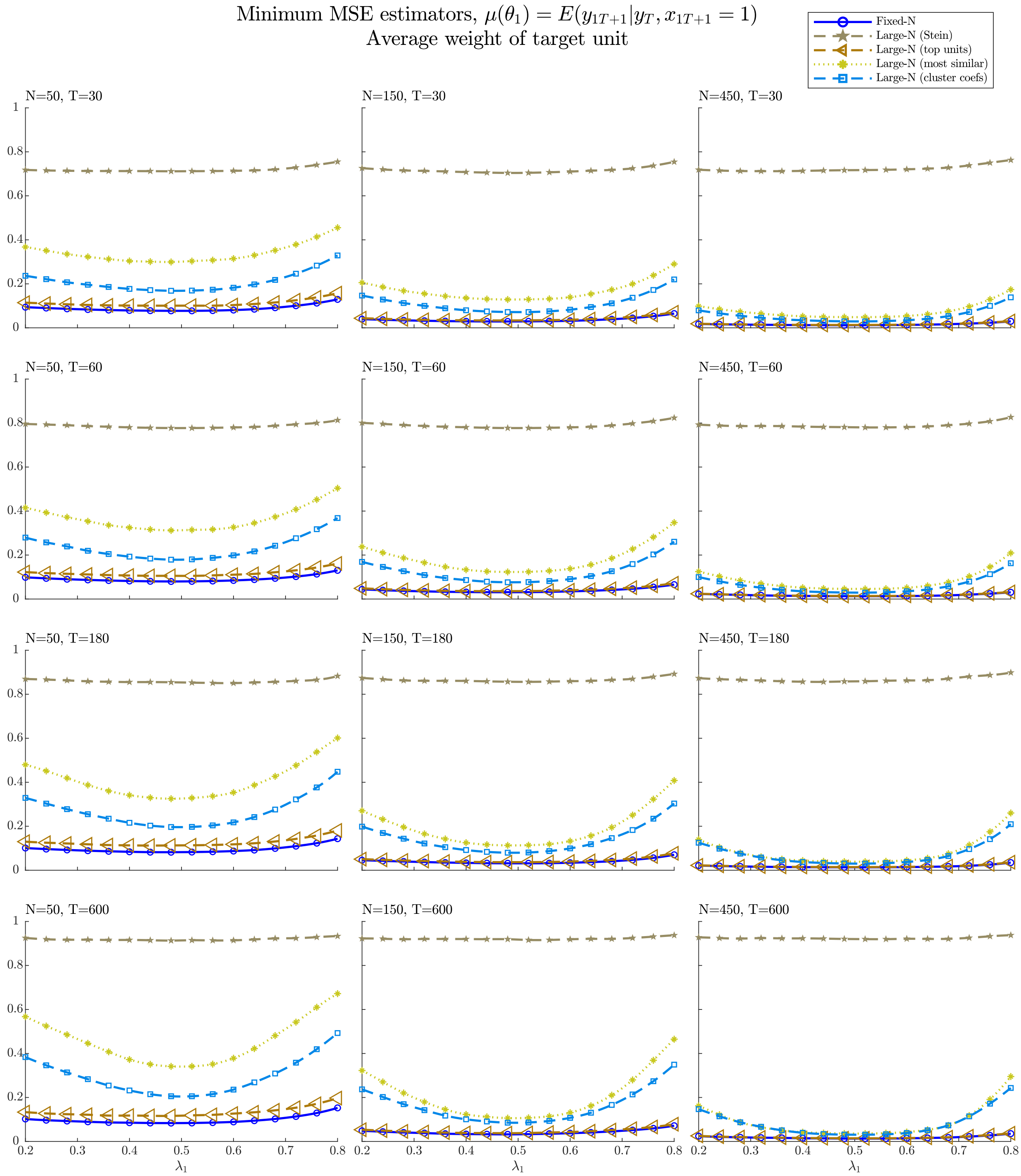

Figure OA.17: Average weight of target unit (unit 1). Focus parameter $\mu(\boldsymbol{\theta}_1) = \mathbb{E}[y_{1T+1}|y_{1T}, x_{1T+1} = 1]$.

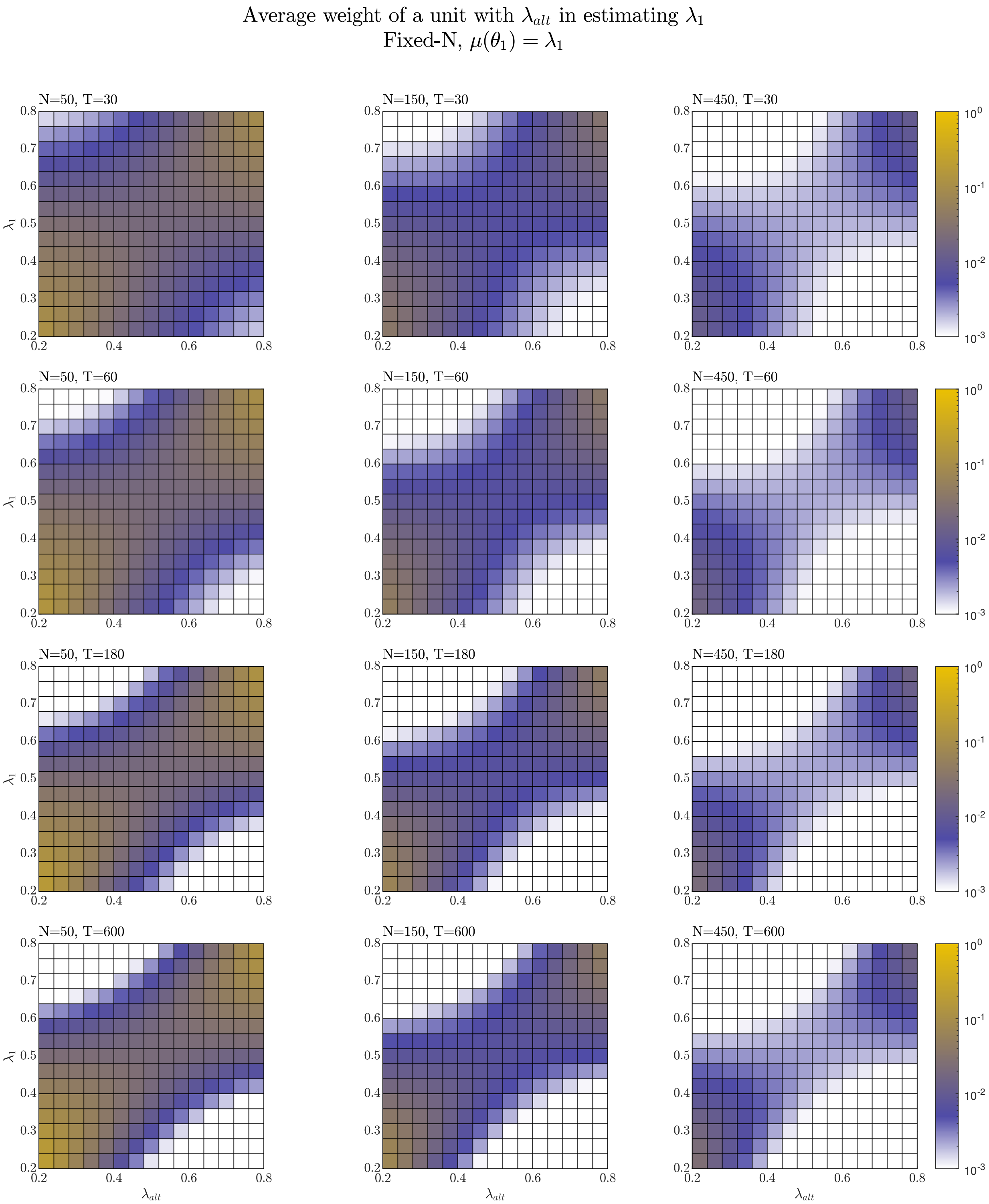

Figure OA.18: Average weight of a unit with $\lambda_i = \lambda_{alt}$ for estimating $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Fixed-$N$ estimator.

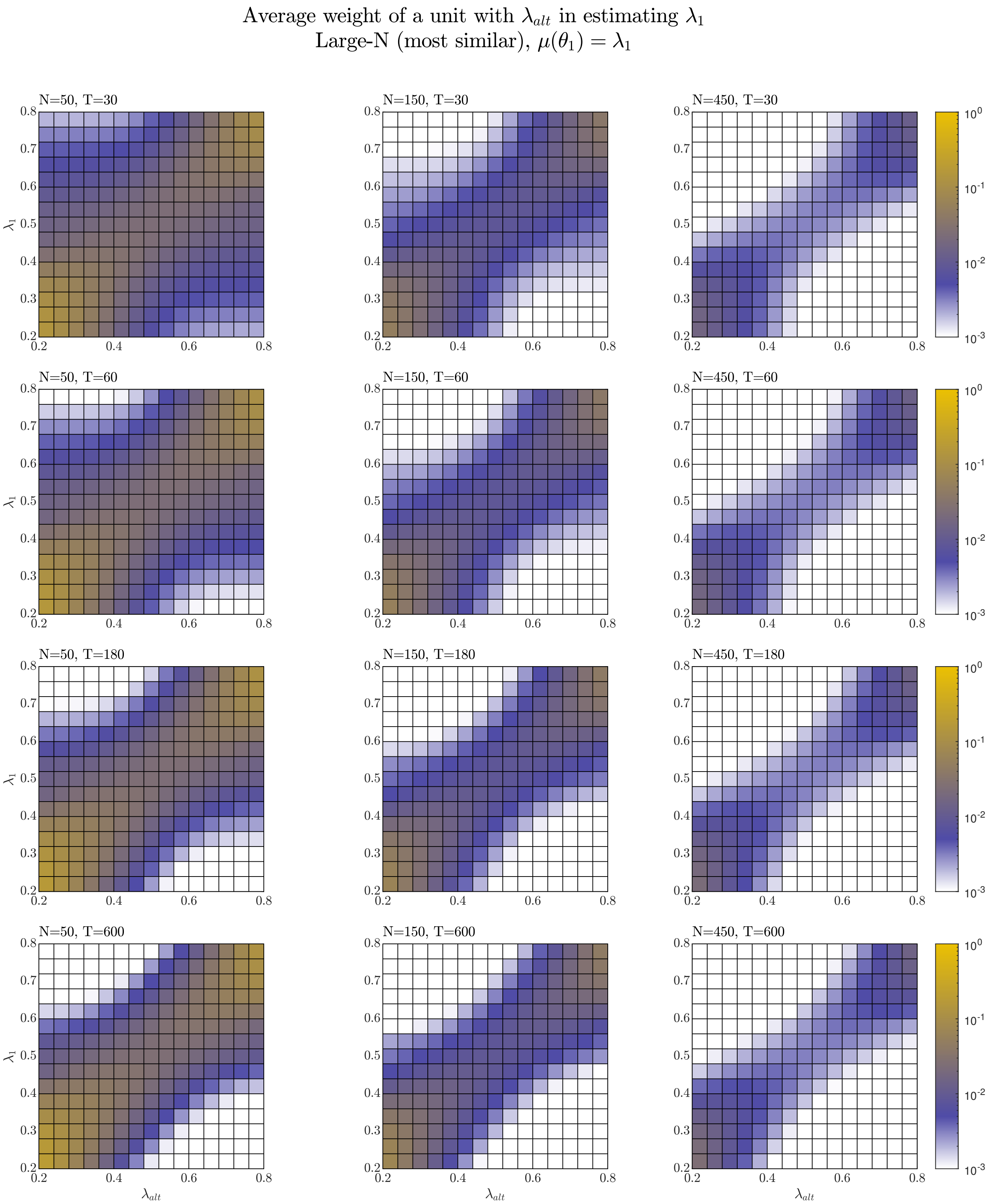

Figure OA.19: Average weight of a unit with $\lambda_i = \lambda_{alt}$ for estimating $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Large-$N$ estimator (10% most similar units are unrestricted)

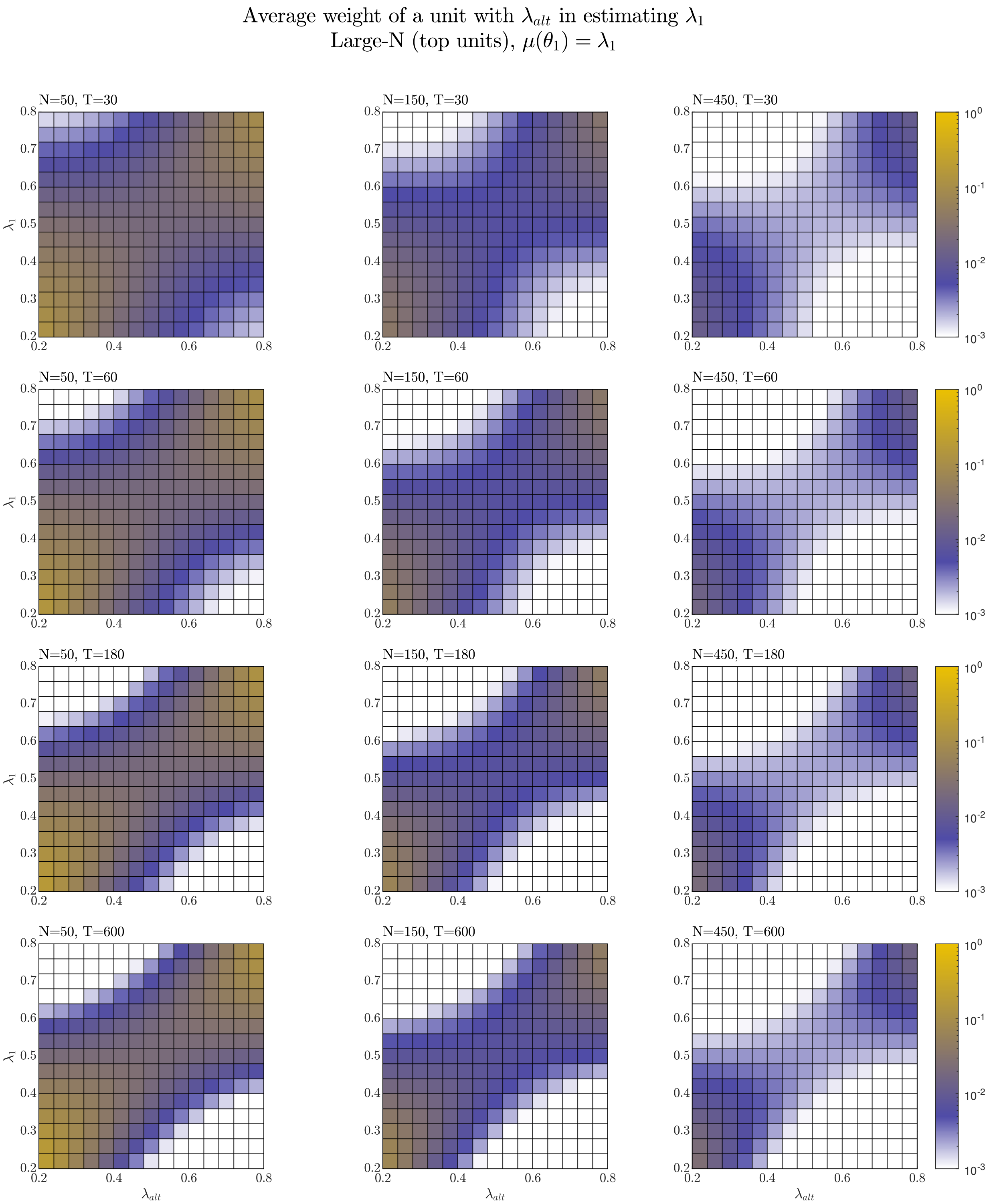

Figure OA.20: Average weight of a unit with $\lambda_i = \lambda_{alt}$ for estimating $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Large-$N$ estimator (10% units with largest fixed-N weights are unrestricted).

Average weight of a unit with λ_{alt} in estimating λ_1
Large-N (cluster coefs), $\mu(\theta_1) = \lambda_1$

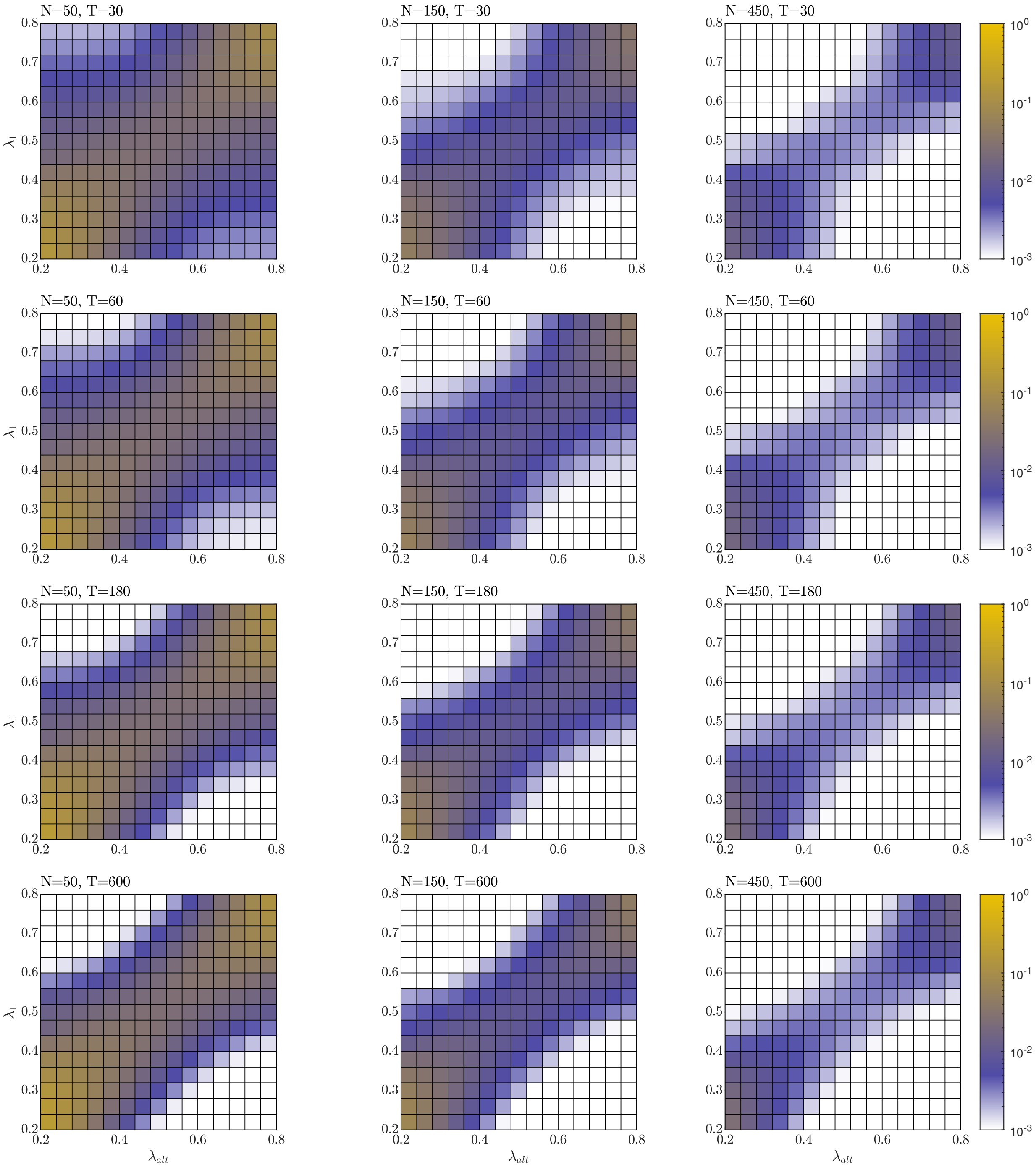

Figure OA.21: Average weight of a unit with $\lambda_i = \lambda_{alt}$ for estimating $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Large-$N$ estimator (estimated coefficients are preclustered in 4 clusters; units in the same cluster as the target unit are the unrestricted units).

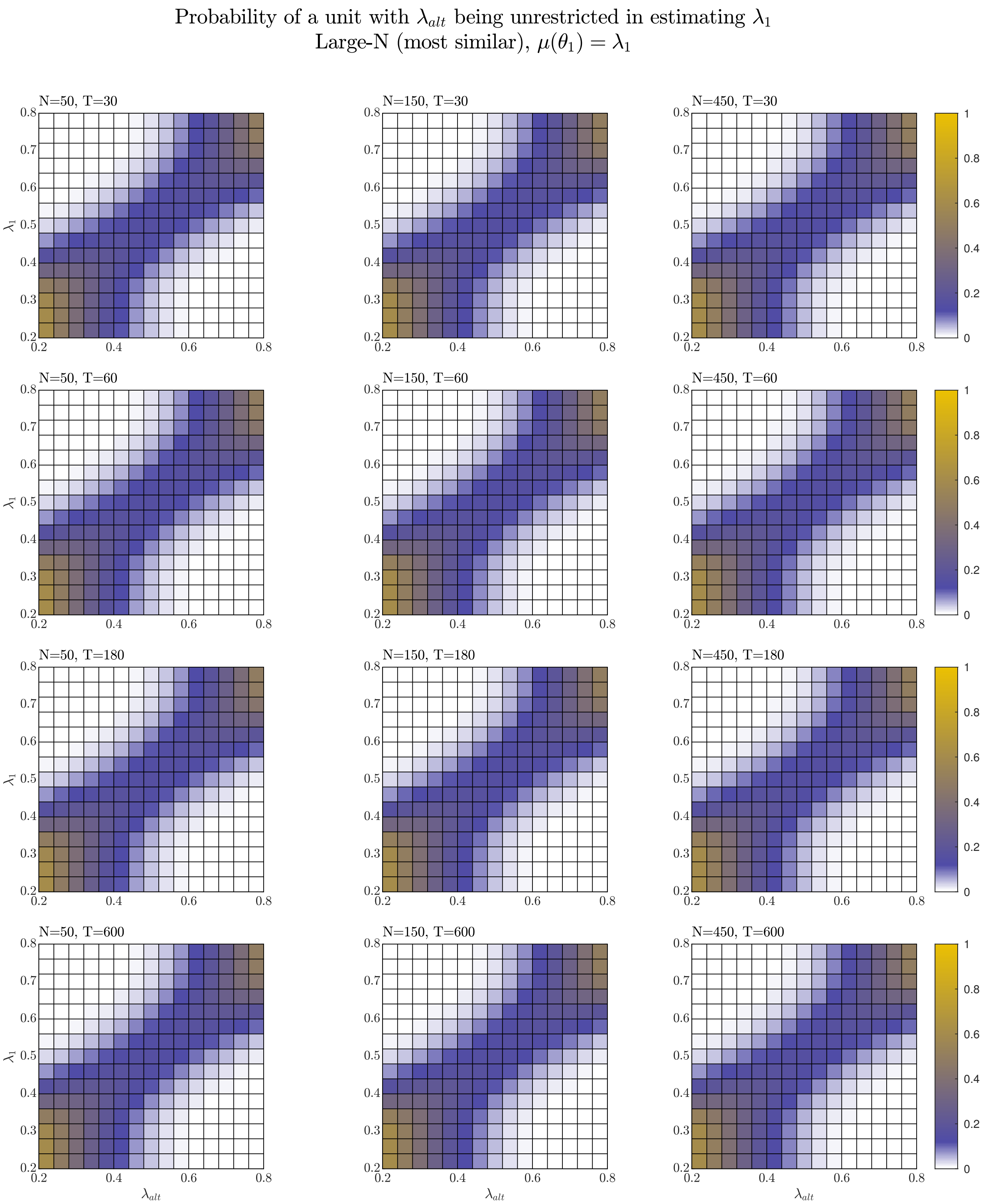

Figure OA.22: Probability of a unit with $\lambda_i = \lambda_{alt}$ being unrestricted for estimating $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Large-$N$ estimator (10% most similar units are unrestricted).

Probability of a unit with λ_{alt} being unrestricted in estimating λ_1
Large-N (top units), $\mu(\theta_1) = \lambda_1$

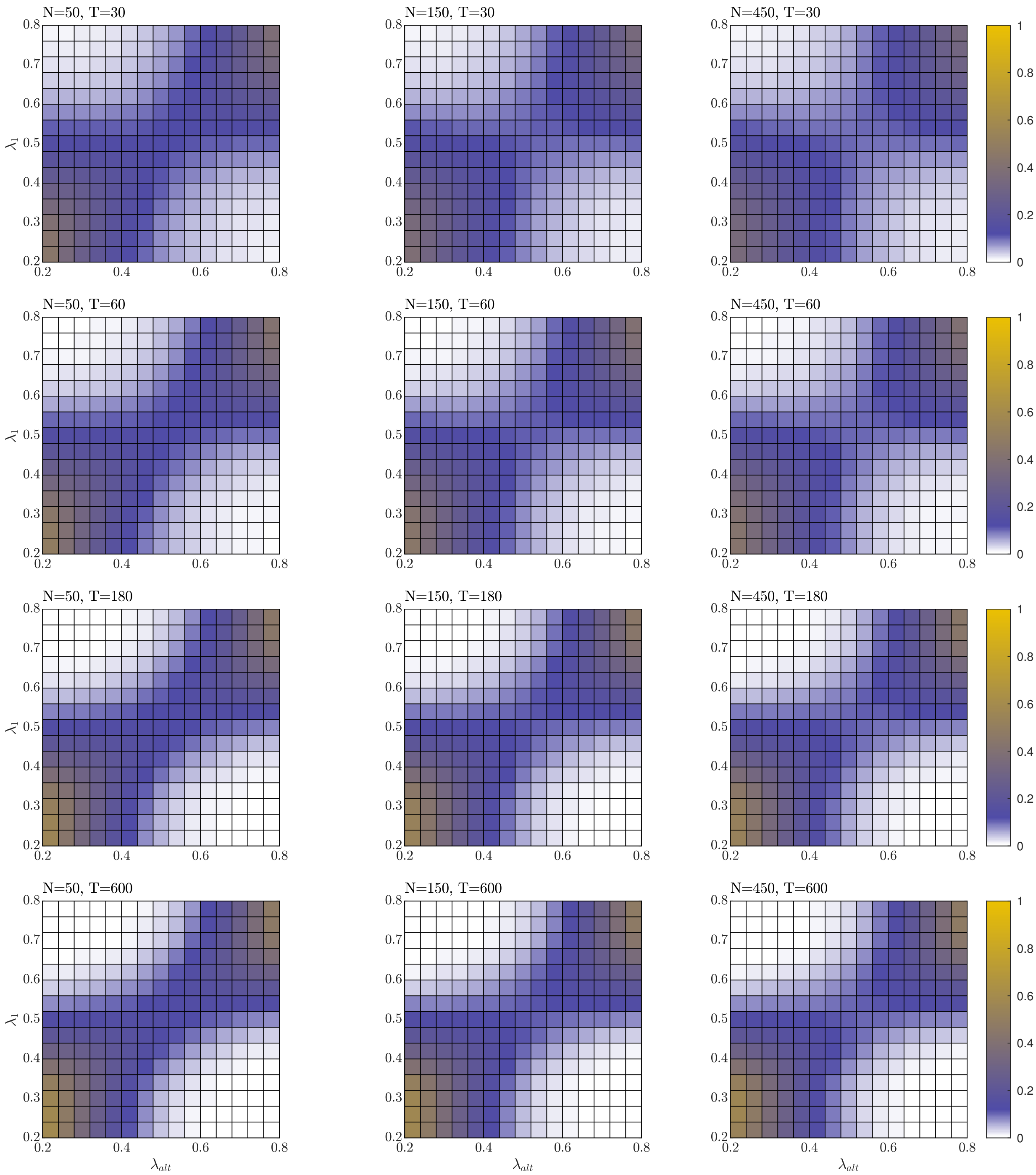

Figure OA.23: Probability of a unit with $\lambda_i = \lambda_{alt}$ being unrestricted for estimating $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Large-$N$ estimator (10% units with largest fixed-N weights are unrestricted).

Probability of a unit with λ_{alt} being unrestricted in estimating λ_1
Large-N (cluster coefs), $\mu(\theta_1) = \lambda_1$

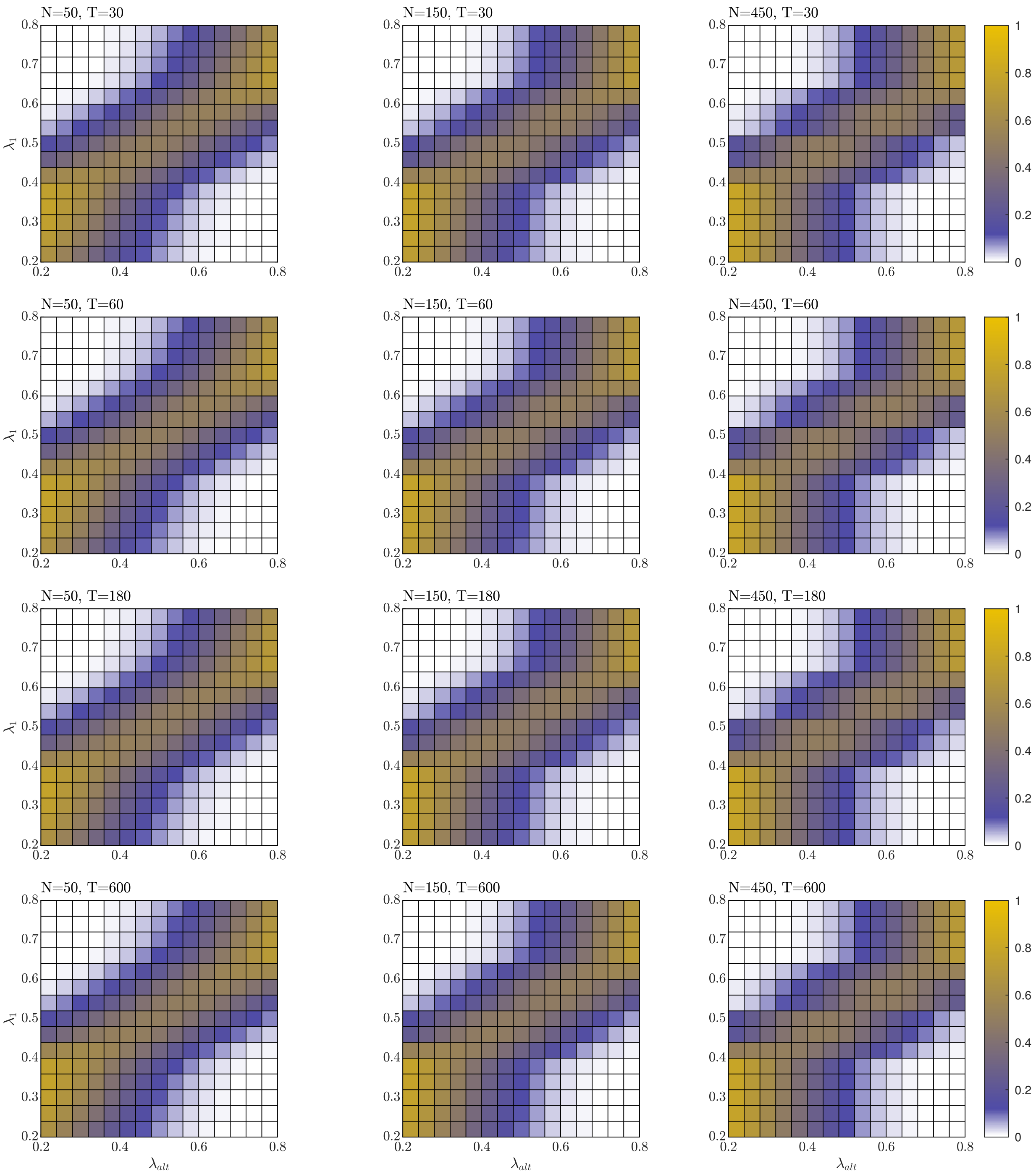

Figure OA.24: Probability of a unit with $\lambda_i = \lambda_{alt}$ being unrestricted for estimating $\mu(\boldsymbol{\theta}_1) = \lambda_1$. Large-$N$ estimator (estimated coefficients are preclustered in 4 clusters; units in the same cluster as the target unit are the unrestricted units).

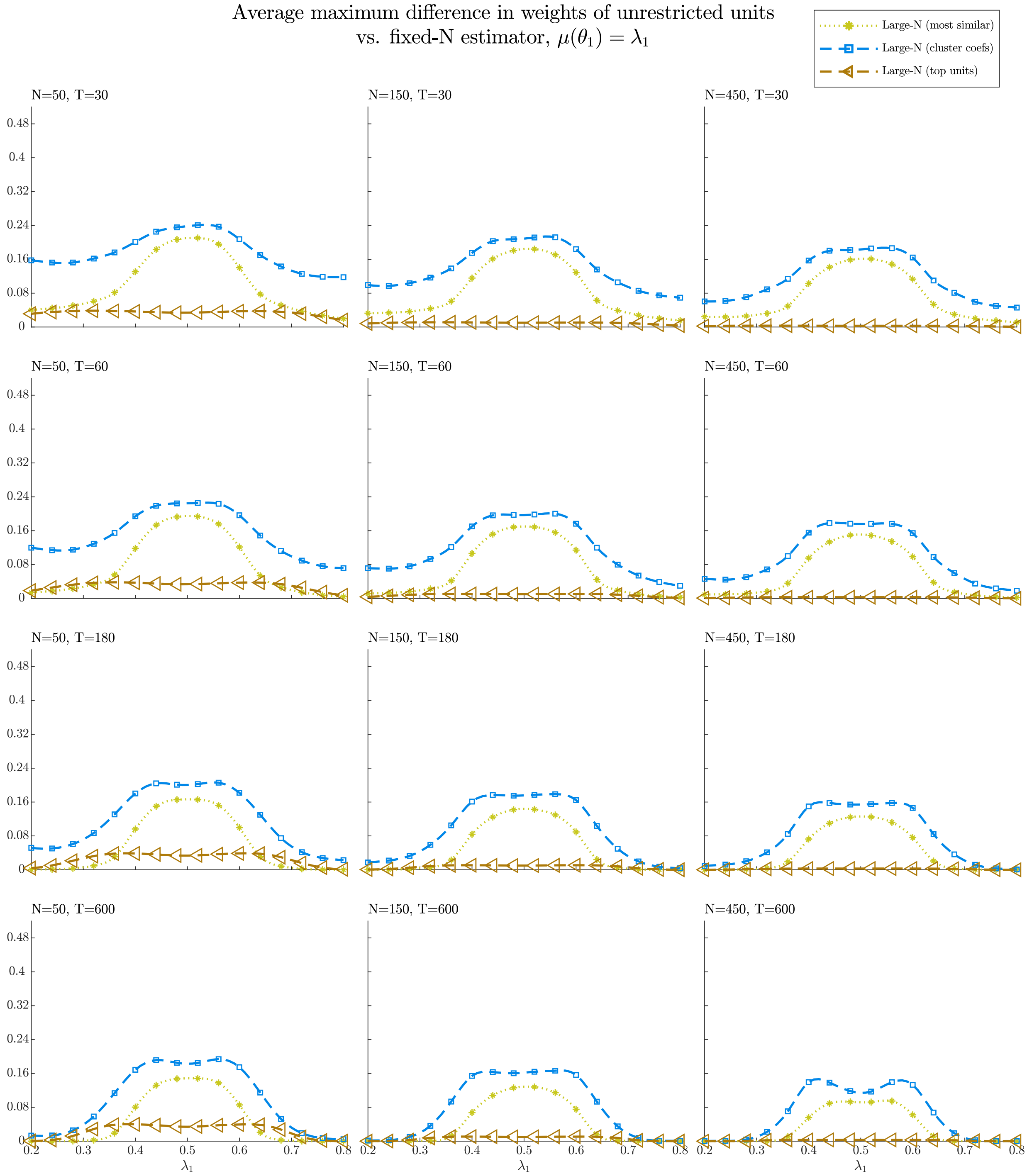

Figure OA.25: Average maximum difference between the weights of restricted units for several large-N estimators vs. the weights assigned to those units by the fixed-N estimators. Estimators of section 4.

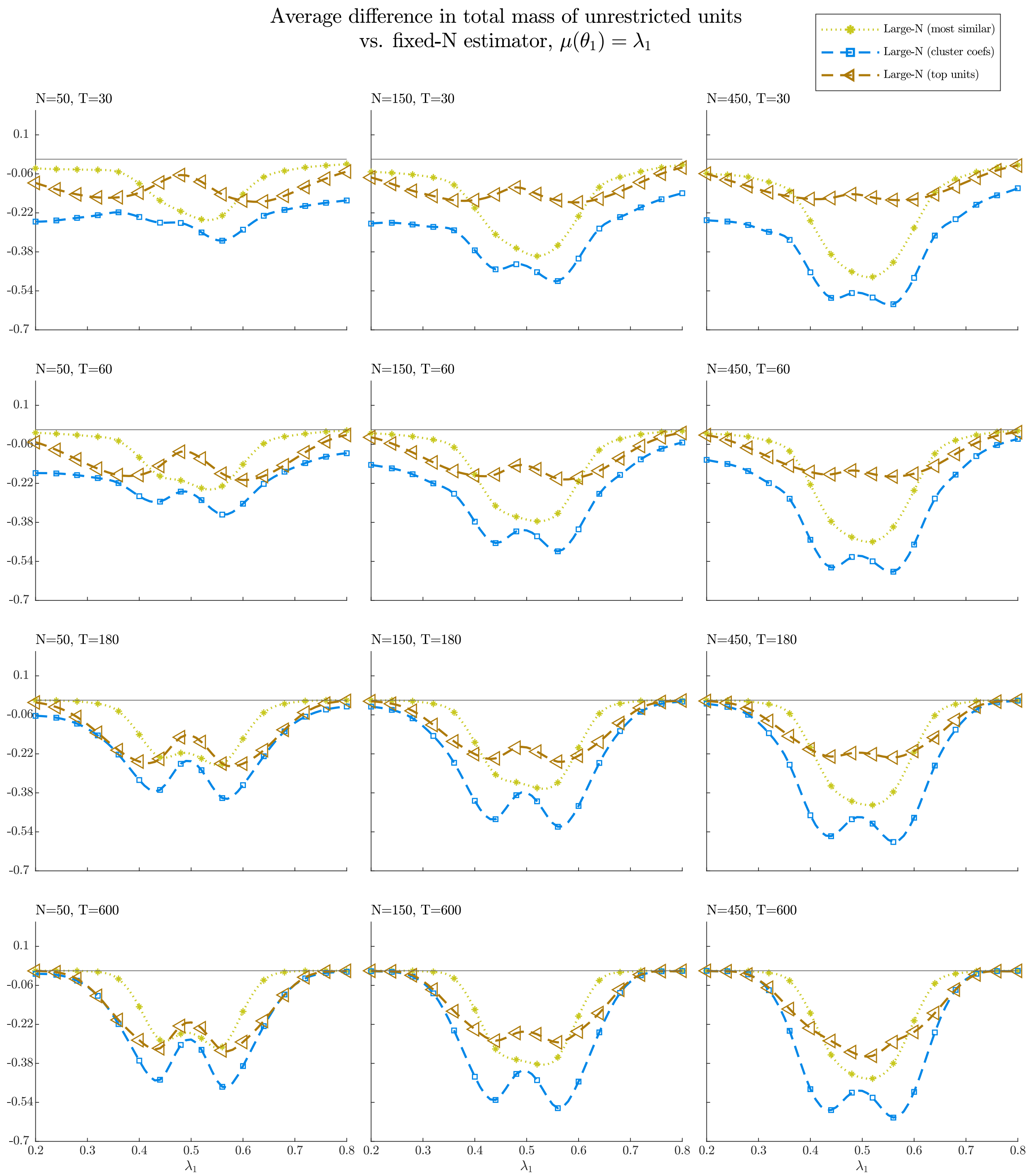

Figure OA.26: Average difference in total mass assigned to the restricted set by several large-N estimators vs. the total mass assigned to those units by the fixed-N estimator.

OA.4.4 Confidence Interval Coverage and Length

We now turn to the coverage and length properties of the one-step confidence interval $\mathcal{I}^B_{\bar{N},T}$ for the focus parameter (subsection OA.3.2). We consider 5 equally spaced values for the parameter $\lambda_1 \in \{0.2, 0.35, 0.5, 0.65, 0.8\}$. The sample sizes considered are $N = 50, 150$ and $T = 60, 180$. For each value of λ_1 and each pair (N, T), we draw 500 Monte Carlo samples. The focus parameters are λ_1, β_1, and $\mathbb{E}[y_{1T+1}|y_{1T}, x_{1T+1} = 1]$. In each sample, we compute $\mathcal{I}^B_{\bar{N},T}$ for each focus parameter, drawing $B = 500$ bootstrap samples. The bootstrap samples themselves are drawn using the stationary bootstrap (Politis and Romano, 1994) with expected block size $\lfloor T^{1/3} \rfloor$. The target coverage level is 95%. For comparison, we also compute a 95% bootstrap confidence interval based on the individual estimator of unit 1 only, using the same bootstrap samples. The results are reported in tables OA.1-OA.3.

We find that $\mathcal{I}^B_{\bar{N},T}$ compares favorably with the CI based on the individual estimator. In terms of coverage, $\mathcal{I}^B_{\bar{N},T}$ generally matches the performance of the individual-only confidence interval. The overage rates are close to nominal for all parameters, with some minor distortions for $\mu(\boldsymbol{\theta}_1) = \lambda_1$ for large values of λ_1. At the same time, the interval $\mathcal{I}^B_{\bar{N},T}$ is slightly shorter without loss of coverage. The reduction in length is more pronounced for

		$\lambda_1 = 0.2$		$\lambda_1 = 0.35$		$\lambda_1 = 0.5$		$\lambda_1 = 0.65$		$\lambda_1 = 0.8$	
N	T	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$
		Coverage									
50	60	0.90	0.91	0.89	0.89	0.90	0.90	0.90	0.90	0.90	0.87
	180	0.93	0.93	0.93	0.93	0.92	0.92	0.93	0.93	0.92	0.91
150	60	0.93	0.93	0.93	0.93	0.91	0.91	0.91	0.91	0.91	0.89
	180	0.95	0.96	0.94	0.94	0.94	0.94	0.94	0.93	0.93	0.92
		Length									
50	60	0.28	0.27	0.27	0.26	0.25	0.25	0.22	0.21	0.18	0.17
	180	0.17	0.16	0.16	0.15	0.15	0.15	0.13	0.13	0.11	0.10
150	60	0.29	0.27	0.27	0.27	0.26	0.25	0.23	0.23	0.19	0.18
	180	0.17	0.16	0.16	0.16	0.15	0.15	0.13	0.13	0.11	0.10

Table OA.1: Coverage and length of the one-step confidence interval based on the fixed-N estimator ($\mathcal{I}^B_{\bar{N},T}$) and the bootstrap confidence interval based on the individual estimator of the target unit (Ind). Focus parameter: $\mu(\boldsymbol{\theta}_1) = \lambda_1$

		$\lambda_1 = 0.2$		$\lambda_1 = 0.35$		$\lambda_1 = 0.5$		$\lambda_1 = 0.65$		$\lambda_1 = 0.8$	
N	T	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$
		Coverage									
50	60	0.93	0.93	0.91	0.92	0.93	0.93	0.91	0.91	0.92	0.93
	180	0.94	0.94	0.94	0.94	0.93	0.93	0.93	0.93	0.93	0.93
150	60	0.93	0.93	0.92	0.92	0.93	0.93	0.93	0.93	0.93	0.93
	180	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
		Length									
50	60	0.43	0.42	0.43	0.43	0.43	0.42	0.43	0.43	0.43	0.43
	180	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
150	60	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	180	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25

Table OA.2: Coverage and length of the one-step confidence interval based on the fixed-N estimator ($\mathcal{I}^B_{\bar{N},T}$) and the bootstrap confidence interval based on the individual estimator of the target unit (Ind). Focus parameter: $\mu(\boldsymbol{\theta}_1) = \beta_1$

		$\lambda_1 = 0.2$		$\lambda_1 = 0.35$		$\lambda_1 = 0.5$		$\lambda_1 = 0.65$		$\lambda_1 = 0.8$	
N	T	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$	Ind	$\mathcal{I}^B_{\bar{N},T}$
		Coverage									
50	60	0.95	0.96	0.95	0.95	0.95	0.95	0.94	0.94	0.94	0.94
	180	0.94	0.94	0.94	0.94	0.95	0.94	0.95	0.95	0.94	0.94
150	60	0.94	0.94	0.92	0.92	0.91	0.91	0.91	0.91	0.91	0.91
	180	0.93	0.93	0.94	0.94	0.96	0.96	0.95	0.95	0.95	0.95
		Length									
50	60	1.47	1.48	2.42	2.42	3.59	3.59	5.15	5.14	7.44	7.42
	180	1.34	1.35	2.34	2.35	3.59	3.59	5.23	5.23	8.01	8.00
150	60	1.44	1.44	2.32	2.32	3.43	3.43	4.89	4.88	7.12	7.11
	180	1.30	1.30	2.24	2.24	3.42	3.42	4.98	4.98	7.61	7.60

Table OA.3: Coverage and length of the one-step confidence interval based on the fixed-N estimator ($\mathcal{I}^B_{\bar{N},T}$) and the bootstrap confidence interval based on the individual estimator of the target unit (Ind). Focus parameter: $\mu(\boldsymbol{\theta}_1) = \mathbb{E}[y_{1T+1}|y_{1T}, x_{1T+1} = 1]$. Note: the estimated depends on T and thus length properties not directly comparable across T

points where the fixed-N minimum MSE estimator is relatively more efficient (see also the the results of subsection OA.4.1).

OA.5 Further Materials for the Empirical Application

Here we present further estimation results for our empirical application to forecasting regional unemployment for a panel of German labor market districts. The setting is as in section 5, to which we refer for details. In subsection OA.5.1, we provide estimation results that expand on the results in section 5. In subsection OA.5.2, we examine the weights chosen by our minimum MSE estimator.

OA.5.1 Full Estimation Results

In this section, we present the full results for the MSE of the averaging estimators considered in section 5. We also consider a coefficient pre-clustering large-N specification. For this approach, we first cluster the coefficient individual estimates into k clusters using k-means. The units allocated to the cluster of the target unit are left unrestricted (see also p. OA-30). Further, we consider the impact of the tuning parameters of the two data-driven large-N specifications (top units and pre-clustering).

Our results are visually presented on figs. OA.27-OA.28. First, fig. OA.27 plots the geographical distribution of the MSE for all of the approaches and rolling window sizes considered in the main text. Second, figs. OA.28-OA.29 expand on figs. 4 and 6 in the main text by also considering the pre-clustering large-N estimator. Fig. OA.28 provides a box plot for the MSE for all averaging approaches relative to the MSE of the individual estimator. For the data-driven large-N specifications, this figure plots the results for a range of tuning parameter values. Fig. OA.29 depicts the best-performing approach for each labor market district (AAB). In contrast to fig. 6, it includes the pre-clustering large-N estimator, but drops the Stein-like one for readability. This change does not affect the rankings, as the Stein-like estimator is best only for the same two districts as on fig. 6.

The results presented in this sections are fully in line with the results of section 5, to which we refer for a full discussion. Accordingly, here we limit ourselves to discussing two aspects of the results which do not appear in section 4. First, we find that the choice of the tuning parameter has a fairly minor impact on the performance of the top units and the pre-clustering large-N estimators. As fig. OA.28 shows, all of the specifications considered have broadly similar MSE profiles, although less flexible specifications (top 10% and 8 clusters) generally have more favorable worst-case performance. Second, the top units large-N estimator appears to be somewhat better than the pre-clustering specification, in line with the simulation results (section OA.4.1). Although the two approaches do not dominate each other, figs. OA.28-OA.29 show that the top units estimator has a more favorable MSE distribution and performs better for a larger number of labor market districts (AABs).

OA.5.2 Optimal Weights

We turn to the weights chosen by the minimum MSE estimators in the empirical application. Specifically, we consider:

- The weight assigned to the target AAB. Fig. OA.30 plots the average weight each AAB receives when it is the target unit, split by the estimator and the rolling window sizes.
- The average maximum difference between the weights of restricted units for the top units and pre-clustering large-N estimators vs. the weights assigned to those units by the fixed-N estimator. These results are plotted geographically for all AABs and rolling window sizes (figs. OA.31-OA.32).
- The average difference in total mass assigned to the restricted set by the large-N estimators vs. the mass assigned to those units by the fixed-N estimator (fig. OA.33).

We have also computed full geographic weight distributions for each AAB. The full set of maps (one per AAB, averaging approach, and rolling window size) is available on request.

Together, figures OA.30-OA.33 reveal two divides in the allocation of weights – a north-south and a urban-less urban one. Generally, AABs in southern regions (primarily Bavaria, Baden-Württemberg, and Hesse) and less urban AABs have considerably lower

MSE Relative to the Individual Estimator

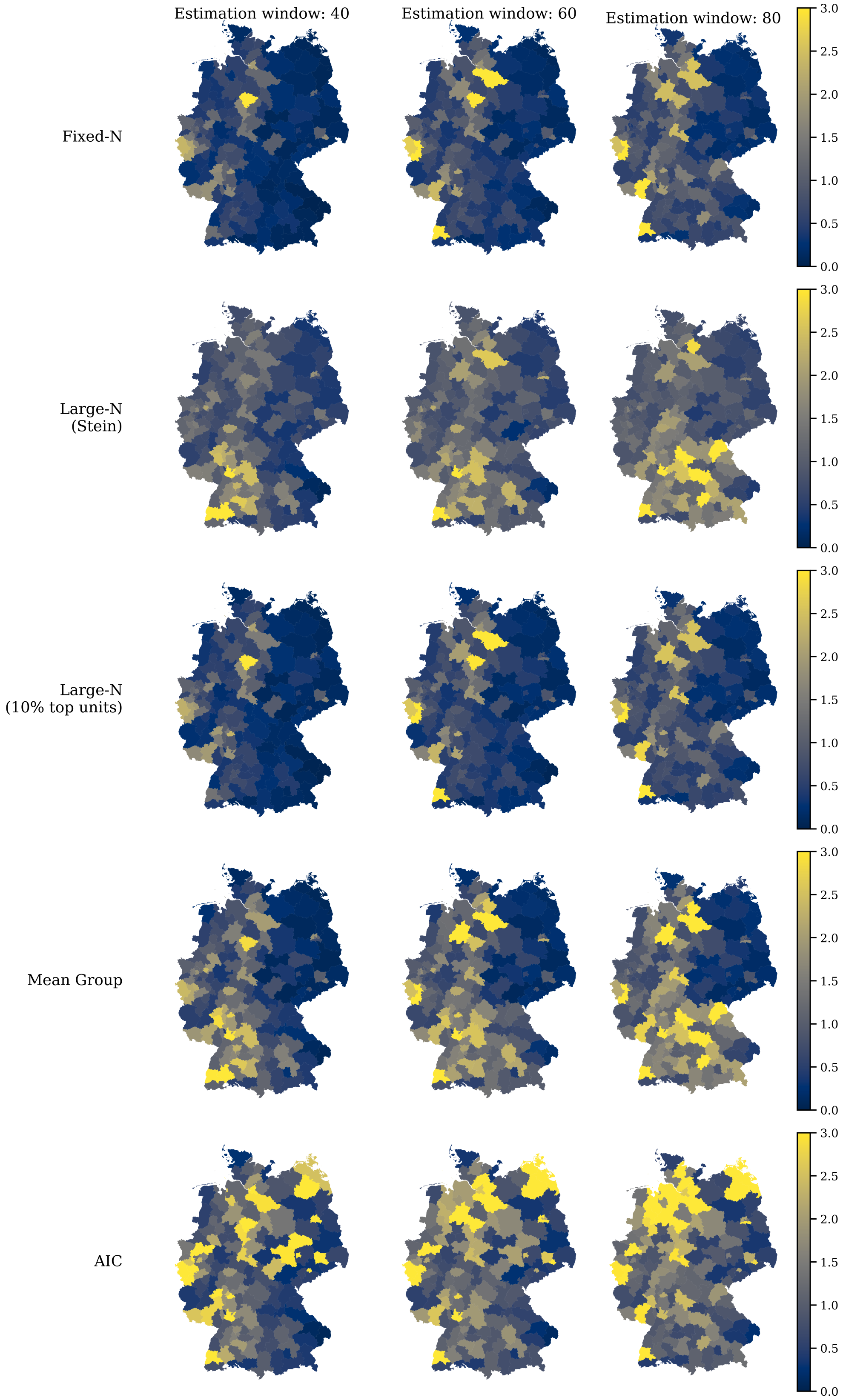

Figure OA.27: Geographic distribution of MSE relative to the individual estimator. Thin lines denote borders of labor market districts (AABs). Unit averaging approaches considered in the main text.

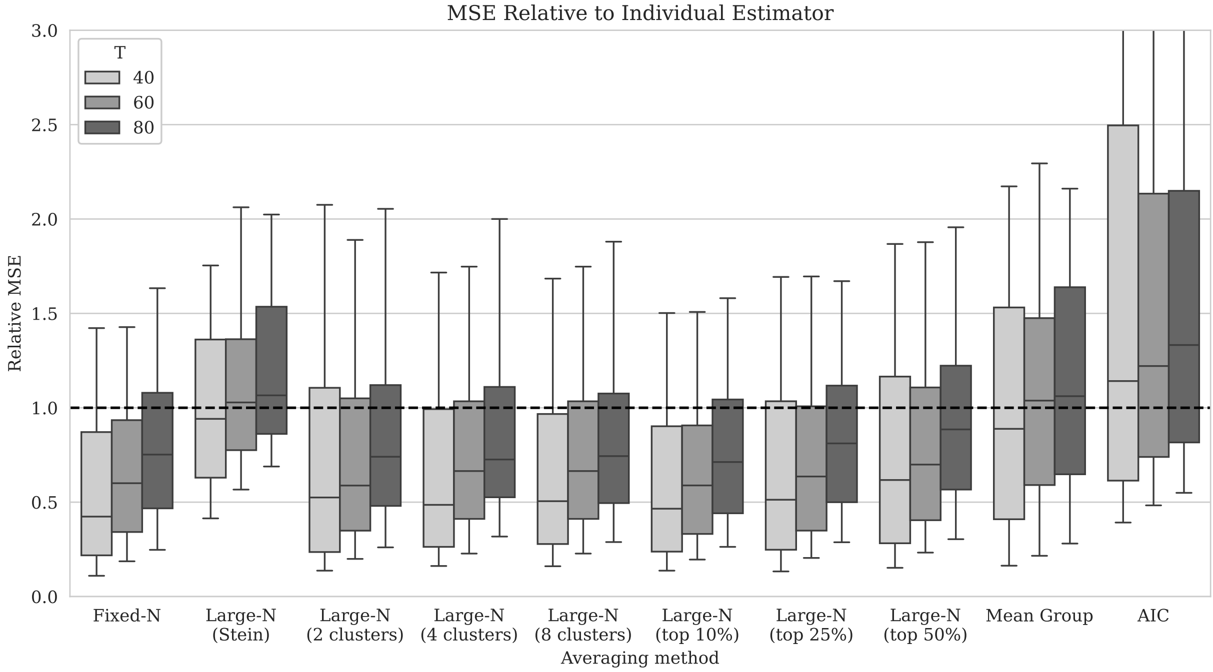

Figure OA.28: Distribution of relative MSEs across labor market districts (AABs). Split by different averaging strategies and estimation window size. Large-N (top x%) – top units large-N specification; Large-N (k clusters) – pre-clustering specification with k clusters. Whiskers – 10th and 90th percentiles; box boundaries – 25th and 75th percentiles; box crossbar – median. Note: the large-N (10%) specification is labeled "large-N (top units)" on fig. 4 in the main text

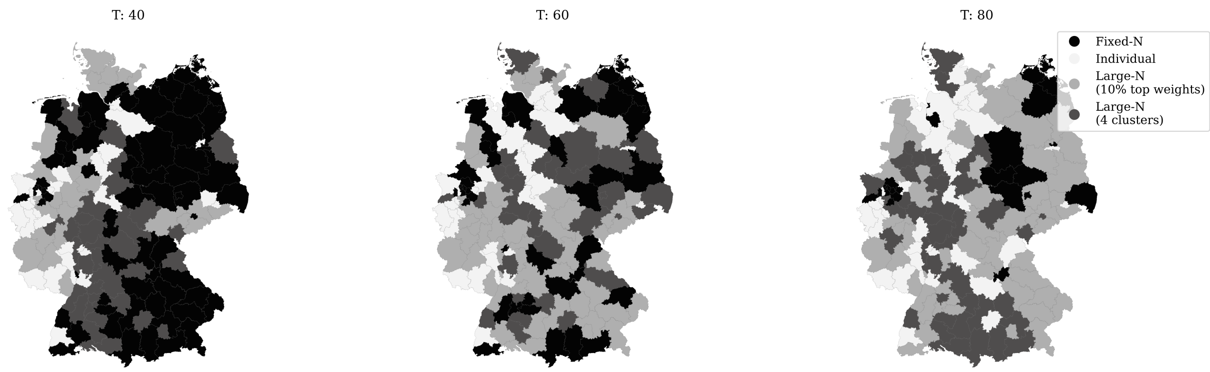

Figure OA.29: Best averaging approach for every labor market district (AAB), including the pre-clustering large-N estimator. Thin lines denote borders of AABs.

own weights, as fig. OA.30 shows. Further, there is a wider difference between the weights of the restricted units and the weights assigned to those units by the fixed-N estimators (though the difference is at most 0.025, as figs. OA.31-OA.32 show). The total mass of the restricted set is also higher for southern AABs (compared to the mass allocated to those units by the fixed-N estimator; fig. OA.33). The opposite pattern obtains for the northern regions and for more urban AABs.

The results of section OA.4.3 suggest a possible explanation for the divide. In our simulations results, units that lie closer to the mean of the focus parameter distribution have lower own weight, higher mass assigned to the restricted set, and larger maximal difference between the fixed-N and the large-N weights. A similar pattern hold for southern and non-urban AABs. Accordingly, these AABs may be interpreted as being "typical" in the sense of lying closer to the mean of the focus parameter distribution. In contrast, more urban AABs (Berlin, Munich, AABs in the Rhine-Ruhr region, etc.) and northern AABs may be viewed as being less similar to the average labor market district.

In other aspects, the results presented here agree with the simulation results. In particular, as fig. OA.30 shows, more flexible estimators generally place lower weights on the target unit for all regions and estimation window sizes. The same pattern appears in our simulation results.

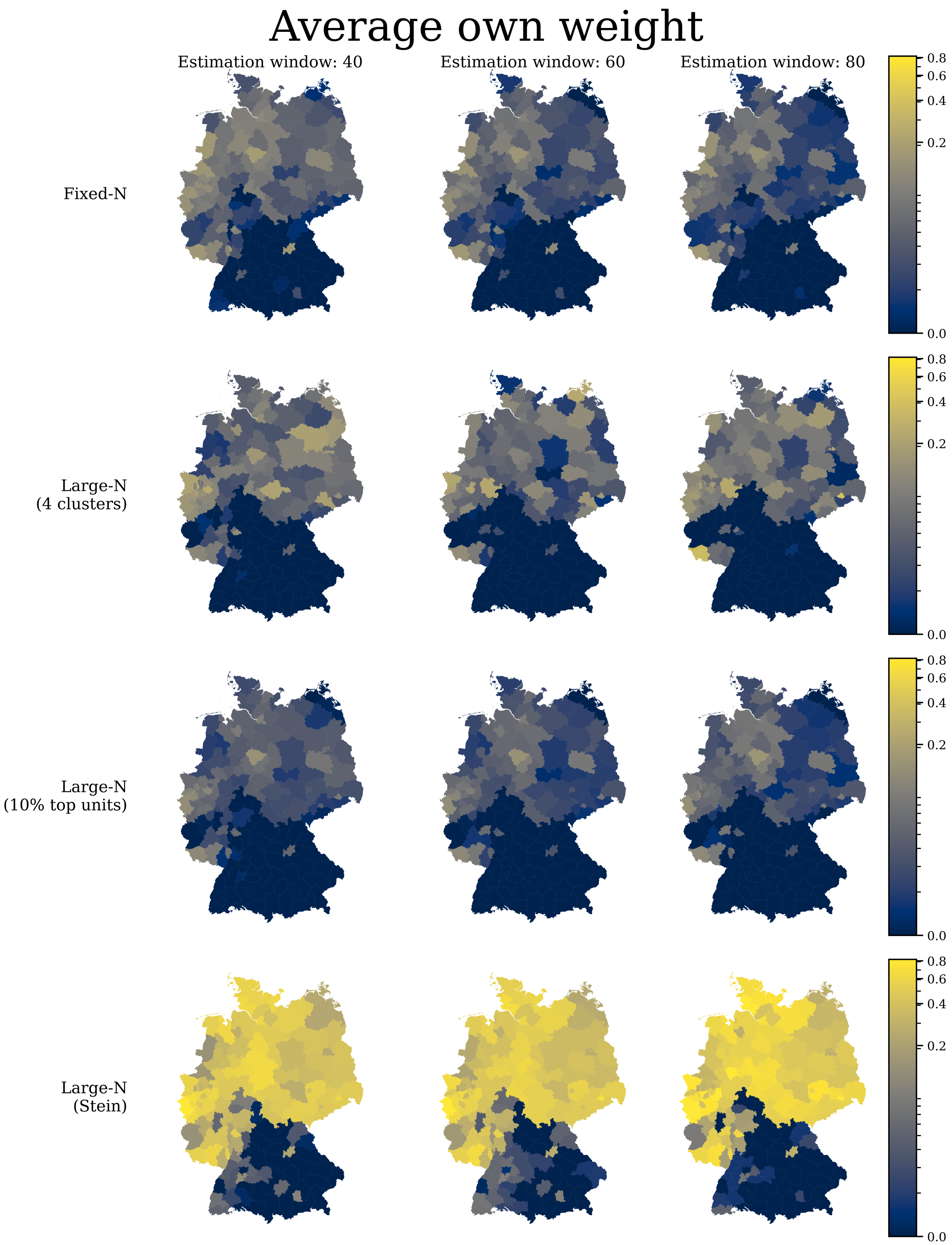

Figure OA.30: Average weight each AAB receives when it is the target unit. Split by estimator, AAB, and estimation window.

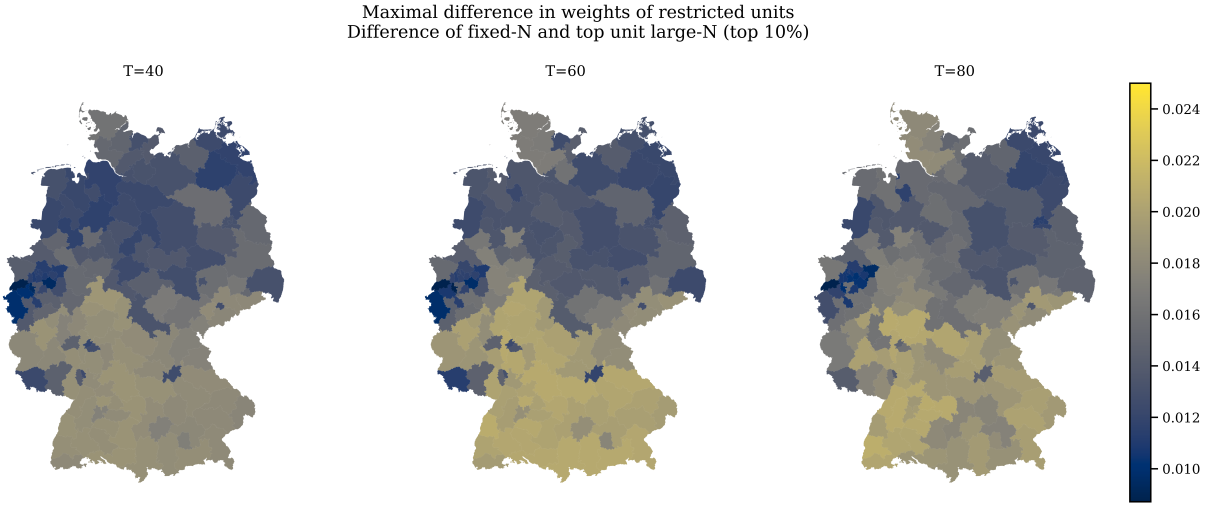

Figure OA.31: Average maximum difference between the weights of the restricted units for the top units large-N estimator vs. the weights assigned to those units by the fixed-N estimator (for each AAB).

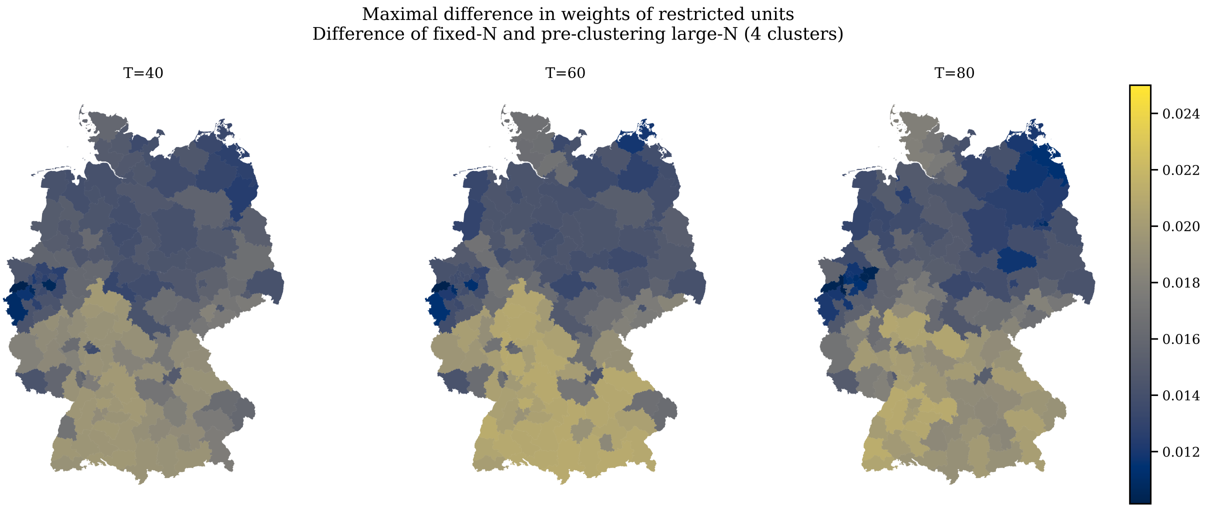

Figure OA.32: Average maximum difference between the weights of the restricted units for the pre-clustering large-N estimator vs. the weights assigned to those units by the fixed-N estimator (for each AAB).

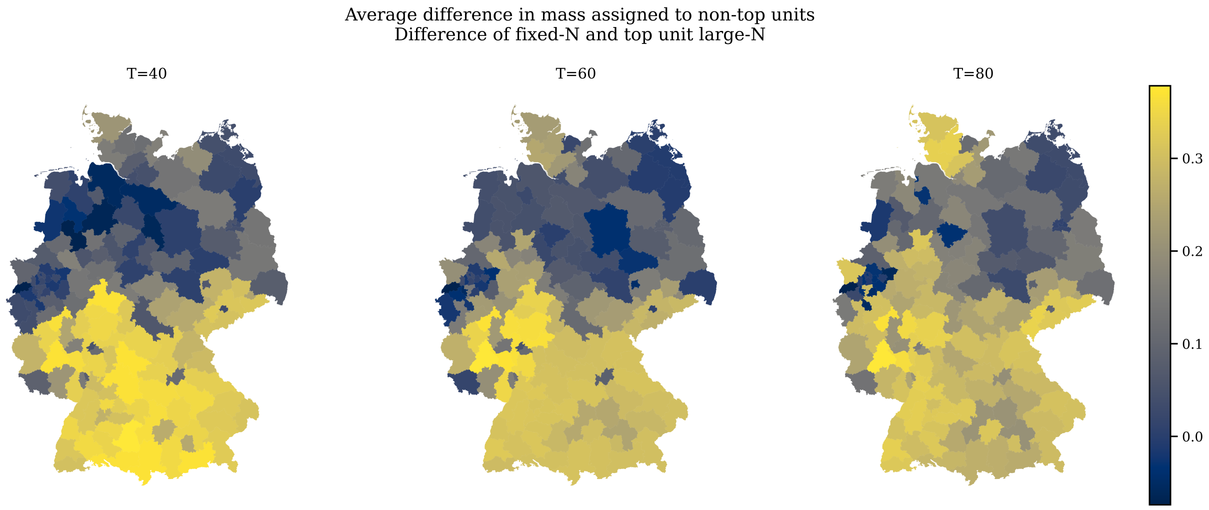

Figure OA.33: Average difference in mass assigned to the restricted set by the top units large-N estimator vs. the mass assigned to those units by the fixed-N estimator (for each AAB).

OA.6 Unit Averaging for GDP Nowcasting

OA.6.1 Setting and Methodology

In this section, we provide an additional application of our methodology to nowcasting quarterly GDP for a panel of European countries. GDP prediction provides another natural application of our unit averaging methodology. There is evidence of considerable heterogeneity between countries, yet at the same time pooling the data at least partially improves prediction accuracy (Garcia-Ferrer, Highfield, Palm, and Zellner, 1987; Hoogstrate, Palm, and Pfann, 2000; Marcellino, Stock, and Watson, 2003). The design of our application follows standard practices in the nowcasting literature (Marcellino and Schumacher, 2010; Schumacher, 2016). The literature on nowcasting is vast and we do not to cover it here. We refer to Bańbura, Giannone, Modugno, and Reichlin (2013) for a survey.

We use quarterly GDP data from 1995Q1 to 2019Q4 for 12 European countries: the 11 founding Eurozone economies and the UK. We enrich our dataset with a set of 162 monthly GDP predictors for each country. The set of predictors include both real, price, and survey data. Table OA.1 in the online appendix contains the complete list of variables and descriptions. All non-survey data is available from Eurostat whereas the survey data is available from the DG ECFIN. We use final data releases incorporating all revisions, making our study a pseudo-real time one.

Our empirical design takes into account both the delays in publication of monthly data ("ragged-edge problem") and the impact of timing on the information set available ("vintages" of data). First, the predictor variables are typically released with different delays after the end of the corresponding month, which is known as the "ragged-edge" problem (Wallis, 1986).[1] We adopt a stylized release calendar of bimonthly releases to account for this (table OA.1 in the online appendix lists the release delay for all variables). Second, as the quarter goes by, more data becomes available.[2] Each possible position in time determines

[1]For example, industrial production data is released 6 weeks after the end of the month, while survey data is released at the end of the month without delay.

[2]For example, nowcasting Q4 GDP can be done at any moment between October 1 when no data on Q4

a data "vintage". We assume that a month has 4 weeks; in accordance with our release calendar, we nowcast 6 weeks into the quarter (−6 weeks relative to quarter end), at quarter end (0 weeks), and +4 weeks after the end of the quarter (GDP is released at +6 weeks). Formally, let t index months. Then $v = -3/2, 0, 1$ is a fractional value that describes the monthly position (or vintage) relative to the end of the quarter.

We nowcast GDP in quarter $3t$ using all information available at time $3t + v$, separately for each value of $v \in \{-3/2, 0, +1\}$. As we have a large number of predictors available at monthly frequency, we opt for factor unrestricted MIDAS (U-MIDAS) (Foroni, Marcellino, and Schumacher, 2015). Given v, for each country we estimate monthly factors $f_{i\,t}$ with $\hat{f}_{i\,t|v}$ for all $t = 1, \ldots, \lfloor T + v \rfloor$ using the full dataset available at $T + v$.[3] The GDP is modeled as

$$y_{i\,3t} = \alpha_{i|v} + \sum_{k=0}^{11} \beta_{i\,k|v} \hat{f}_{i\,\lfloor 3t+v-k \rfloor|v} + \lambda_{i|v} y_{i\,3(t-1)} + \varepsilon_{i\,3t|v},$$

where $y_{i\,3t}$ is GDP of country i in quarter $3t$ and $\varepsilon_{i\,3t|v}$ is the prediction error. The country factors estimates $\hat{f}_{i\,t|v}$ are extracted from the large set of predictor variables using the EM-PCA method (Stock and Watson, 1999). We use only one factor for prediction following Marcellino and Schumacher (2010) and we include the lag of GDP following Clements and Galvão (2008). We nowcast GDP for each country using the conditional mean of GDP implied by the U-MIDAS specification. Parameter estimation is carried out using a rolling-windows of sizes 44, 60 and 76 quarters.[4] Factors are also re-estimated every two weeks using the all the data available at each point in time.

We estimate the conditional mean using only the fixed-N minimum MSE unit averaging estimator, since cross-sectional dimension is not large and each unit is potentially relevant. The performance of our minimum MSE unit averaging estimator is benchmarked against the individual, mean group, and averaging estimators using AIC/BIC weights.

is available yet up to the middle of the following February, when GDP data for Q4 is released. The amount of data available increases monotonically between these two dates.

[3] For example, suppose we wish to nowcast Q4 GDP. If $v = 1$, we estimate factors up to January of the following year using information available at the end of January. If $v = 1/2$, we estimate factors up to December using all the information available in the middle of January.

[4] Forecast evaluation begins in 2006Q1 for window size 44, 2010Q1 for T=60 and 2014Q1 for T=76.

OA.6.2 Results

In table OA.4 we provide a summary of forecasting performance results for GDP nowcasting. The table reports the MSE of the individual estimator as well as the MSE relative to the individual estimator for all other strategies. The table reports results for the five largest economies in our sample along with the GDP-weighted mean.[5]

		−6 weeks			0 weeks			+4 weeks		
	Averaging	44q	60q	76q	44q	60q	76q	44q	60q	76q
Mean	Individual	1.113	0.986	1.167	0.973	1.010	1.196	0.933	0.914	1.124
	minMSE	0.916	0.936	0.907	0.889	0.936	0.910	0.881	0.928	0.901
	AIC	0.933	0.962	0.980	0.908	0.960	0.974	0.878	0.949	0.955
	Mean group	1.417	1.570	1.524	1.635	1.696	1.505	1.879	1.704	1.489
DE	Individual	0.661	0.546	0.537	0.509	0.421	0.434	0.565	0.449	0.456
	minMSE	0.793*	0.822	0.815	0.787*	0.818	0.775	0.821*	0.809	0.794
	AIC	0.963*	0.977*	0.989	0.974	0.982*	0.973	0.978	0.989*	0.978
	Mean group	0.987	0.937	0.773	1.069	1.157	0.742	0.957	1.153	0.849
FR	Individual	0.194	0.154	0.129	0.143	0.100	0.086	0.155	0.121	0.098
	minMSE	0.988	1.067	1.037	0.971	1.059	1.159	0.916	0.978	1.069
	AIC	0.883*	0.934	0.975	0.833*	0.978	1.049	0.828*	0.935*	0.999
	Mean group	2.125*	2.068*	1.348	2.736*	2.942*	2.169*	2.473*	2.652*	2.156*
IT	Individual	0.591	0.253	0.156	0.279	0.178	0.116	0.232	0.131	0.082
	minMSE	0.893*	0.908*	0.852*	0.973	0.974	0.858*	1.046	1.025	0.857*
	AIC	0.945	0.972*	0.980	0.955	0.951*	0.976*	0.947	0.969*	0.975*
	Mean group	0.895	0.901	0.822	1.289	1.042	0.719	1.491*	1.595*	1.239
ES	Individual	0.288	0.198	0.147	0.233	0.121	0.106	0.253	0.114	0.102
	minMSE	0.919	0.909*	0.856*	0.955	0.951	0.927	0.957	0.919	0.889
	AIC	0.958	0.961*	0.974*	0.860	0.940	0.940*	0.813*	0.934	0.933*
	Mean group	1.237*	1.427*	1.225	1.011	1.561*	1.352	0.946	1.886*	1.248
UK	Individual	0.281	0.116	0.044	0.254	0.142	0.047	0.244	0.142	0.047
	minMSE	0.928	0.988	0.953	0.840	0.984	0.868	0.743	1.034	0.913
	AIC	0.871*	0.933	0.953	0.876	0.958	0.941	0.726*	0.917*	0.898*
	Mean group	1.714	2.444*	3.688*	2.530*	2.445*	3.121*	4.330*	2.214*	2.650*

Table OA.4: Nowcasting MSE. *For individual estimator*: absolute value. *For averaging estimators*: MSE relative to individual estimator. For different estimation window sizes (44, 60, 76); selected weekly horizons relative to quarter end ($-6, 0, +4$ weeks). * – forecasting performance difference significant at 10% in Diebold-Mariano test (Diebold and Mariano, 1995)

Our key finding is that using the minimum MSE estimator generally generally leads to improved nowcasting performance, mirroring the results of section 5. This is clear from table OA.4, as the vast majority of entries corresponding to those weights display relative MSE smaller than one, with improvements reaching up to 20%. The degree of improvement varies with the country in question. However, the average gain in performance is on the scale of about 9%.

[5]Weighing by GDP as in Marcellino et al. (2003) emulates forecasting the Eurozone GDP using individual forecasts.

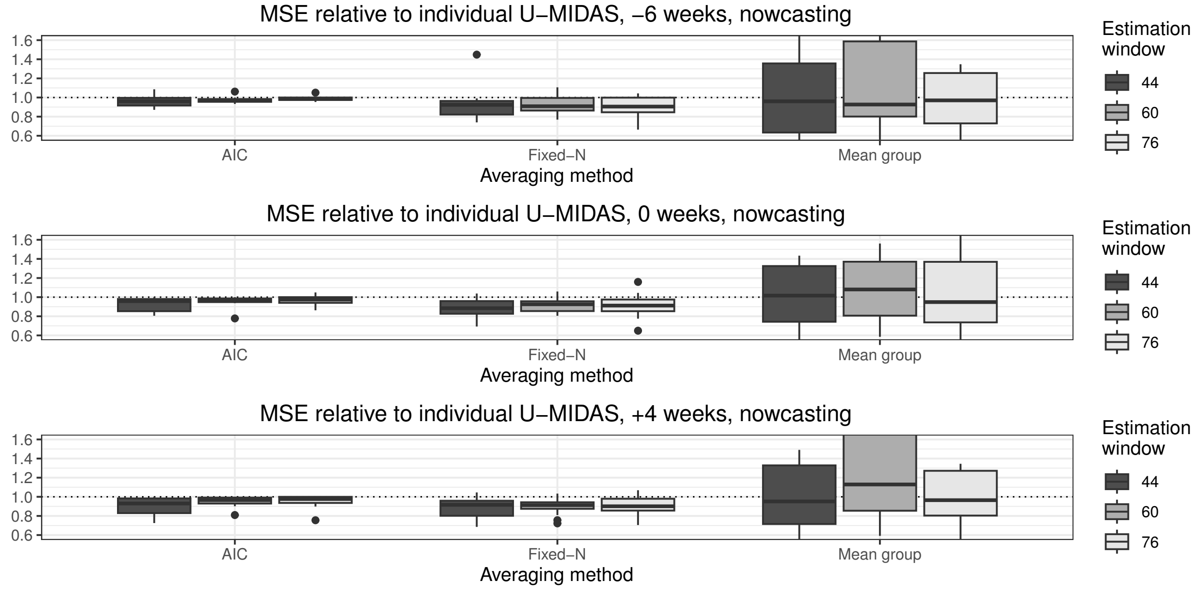

Figure OA.34: Distribution of relative MSEs across countries. Split by different averaging strategies and estimation window size. Same weekly positions as reported in table OA.4

Unlike in section 5, AIC weights offer fairly robust improvements for nowcasting. The MSE is on average 5% lower for AIC weights relative to the individual-specific estimator. We also observe that minimum MSE weights and AIC weights do not uniformly dominate each other.

Figure OA.34 provides a box plot for relative MSEs for nowcasting GDP for all the countries in the panel for the vintages considered in table OA.4. The figure illustrates that the favorable performance is robust across countries and not limited to the five biggest economies reported in table OA.4. Both minimum MSE and AIC weights generally lead to an improvement in performance, as both rarely have relative MSE above one. There is some evidence that the minimum MSE weights have a greater upside, at the price of potentially some more variability in the results, while AIC leads to smaller, but more tightly concentrated improvements. Further, we find that averaging is more attractive for the smallest sample size of $T = 44$, with relative MSE generally approaching one as T increases. This can be clearly seen in figure OA.34, as the improvement range for AIC and minimum MSE weights becomes more concentrated and to closer to one. As previously remarked, as T increases, the minimum MSE estimator converges to the individual estimator; a similar point applies to AIC weights if the log likelihood is not divided by samples size and allowed to diverge as sample size grows.

OA.6.3 Description of the Variables Used

In this section we describe the monthly predictors used for forecasting. The descriptions are contained in table OA.5. Table OA.5 is split in two parts. The first part broadly corresponds to "hard" economic activity data obtained from Eurostat. The second part lists the business and economic survey data available from the Directorate-General for Economic and Financial Affairs of the European Commission.

Description of the first part:

- Columns `Name`, `Description`, `Group` list the Eurostat name, description, and economic content of a given variable.
- `Code` provides the Eurostat code.
- `Delay weeks`: how many weeks after the end of the relevant month is the variable released. We simplify the calendar to two releases in a month: at the beginning of a month and in the middle. The full calendar is available from the Eurostat.
- `Tr.`: code of the transformation applied to the data to transform it to an $I(0)$ variable. Transformation codes are as follows:

  0 None, data used in levels

  1 First difference

  2 Log difference

  3 Translation upwards to ensure that variable is strictly positive + log difference

  4 Quarterly difference (quarterly GDP)

For survey variables we use their DG ECFIN survey codes. We omit the descriptions of each individual question; see the official description available from DG ECFIN.[6]

[6]https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys_en

Table OA.5: List of variables. Note: not all variables are available for all countries at a given time. This only impacts the precision in estimating country-specific factors.

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
1	PROD-B	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[B] Mining and quarrying
2	PROD-B-D	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[B-D] Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply
3	PROD-B-D_F	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[B-D_F] Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply; construction
4	PROD-B_C	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[B_C] Mining and quarrying; manufacturing
5	PROD-C	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[C] Manufacturing
6	PROD-C_HTC	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[C_HTC] High-technology manufacturing
7	PROD-C_LTC	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[C_LTC] Low-technology manufacturing
8	PROD-D	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[D] Electricity, gas, steam and air conditioning supply

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
9	PROD-MIG_CAG	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[MIG_CAG] MIG - capital goods
10	PROD-MIG_COG	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[MIG_COG] MIG - consumer goods
11	PROD-MIG_DCOG	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[MIG_DCOG] MIG - durable consumer goods
12	PROD-MIG_ING	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[MIG_ING] MIG - intermediate goods
13	PROD-MIG_NDCOG	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[MIG_NDCOG] MIG - non-durable consumer goods
14	PROD-MIG_NRG_X_E	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[MIG_NRG_X_E] MIG - energy (except section E)
15	PROD-MIG_ING_CAG	1	[I15] Index, 2015=100	6	Industrial Production	STS_INPR_M	[MIG_ING_CAG] MIG - intermediate and capital goods
16	PSQM-F_CC1	2	[PSQM] Building permits - m2 of useful floor area	6	Building Permits	STS_COBP_M	[F_CC1] Buildings
17	PSQM-F_CC11	2	[PSQM] Building permits - m2 of useful floor area	6	Building Permits	STS_COBP_M	[F_CC11] Residential buildings
18	TOVT-B	1	[I15] Index, 2015=100	6	Turnover in Industry	STS_INTV_M	[B] Mining and quarrying
19	TOVT-C	1	[I15] Index, 2015=100	6	Turnover in Industry	STS_INTV_M	[C] Manufacturing
20	TOVT-MIG_CAG	1	[I15] Index, 2015=100	6	Turnover in Industry	STS_INTV_M	[MIG_CAG] MIG - capital goods

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
21	TOVT-MIG_COG	1	[I15] Index, 2015=100	6	Turnover in Industry	STS_INTV_M	[MIG_COG] MIG - consumer goods
22	TOVT-MIG_ING	1	[I15] Index, 2015=100	6	Turnover in Industry	STS_INTV_M	[MIG_ING] MIG - intermediate goods
23	TOVV-G47	1	[I15] Index, 2015=100	6	Wholesale and retail, turnover	STS_TRTU_M	[G47] Retail trade, except of motor vehicles and motorcycles
24	TOVT-G-N_STS	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[G-N_STS] Services required by STS regulation
25	TOVT-H	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[H] Transportation and storage
26	TOVT-H51_I55_N79	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[H51_I55_N79] Air transport; accommodation; travel agency, tour operator and other reservation service and related activities
27	TOVT-I	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[I] Accommodation and food service activities
28	TOVT-J	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[J] Information and communication

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
29	TOVT-M69_M702	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[M69_M702] Legal, accounting and management consultancy activitie
30	TOVT-M_STS	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[M_STS] Professional, scientific and technical activities required by STS regulation
31	TOVT-N_STS	0	[PCH_PRE] Percentage change on previous period	6	Turnover in services	STS_SETU_M	[N_STS] Administrative and support service activities required by STS regulation
32	PCH_SM-I551-I553	0	[PCH_SM] Percentage change compared to same period in previous year	8	Nights spent at tourist accommodation	TOUR_OCC_NIM	[I551-I553] Hotels; holiday and other short-stay accommodation; camping grounds, recreational vehicle parks and trailer parks
33	PAS-PAS_BRD	2	[PAS] Passenger	8	Air transport of passengers	TTR00016	[PAS_BRD] Passengers on board
34	PAS-PAS_BRD_ARR	2	[PAS] Passenger	8	Air transport of passengers	TTR00016	[PAS_BRD_ARR] Passengers on board (arrivals)
35	PAS-PAS_BRD_DEP	2	[PAS] Passenger	8	Air transport of passengers	TTR00016	[PAS_BRD_DEP] Passengers on board (departures)

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
36	PAS-PAS_CRD	2	[PAS] Passenger	8	Air transport of passengers	TTR00016	[PAS_CRD] Passengers carried
37	PAS-PAS_CRD_ARR	2	[PAS] Passenger	8	Air transport of passengers	TTR00016	[PAS_CRD_ARR] Passengers carried (arrival)
38	PAS-PAS_CRD_DEP	2	[PAS] Passenger	8	Air transport of passengers	TTR00016	[PAS_CRD_DEP] Passengers carried (departures)
39	FA	3	[BAL] Balance, Million Euro	6	Balance of Payments	BOP_C6_M	[FA] Financial account
40	CA	3	[BAL] Balance, Million Euro	6	Balance of Payments	BOP_C6_M	[CA] Current account
41	CKA	3	[BAL] Balance, Million Euro	6	Balance of Payments	BOP_C6_M	[CKA] Current plus capital account (balance = Net lending (+) / net borrowing (-))
42	GS	3	[BAL] Balance, Million Euro	6	Balance of Payments	BOP_C6_M	[GS] Goods and services
43	KA	3	[BAL] Balance, Million Euro	6	Balance of Payments	BOP_C6_M	[KA] Capital account
44	THS_T-B_195500	3	[THS_T] Thousand tonnes	6	Crude oil	NRG_JODI	[B_195500] Demand
45	TJ_GCV-B_190400	3	[TJ_GCV] Terajoule (gross calorific value - GCV)	6	Natural gas supply	NRG_IND_343M	[B_190400] Stock Changes
46	TJ_GCV-B_190900	2	[TJ_GCV] Terajoule (gross calorific value - GCV)	6	Natural gas supply	NRG_IND_343M	[B_190900] Supply

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
47	GWH	2	[GWh] GWh	6	Electricity available	NRG_CB_EIM	Electricity available to internal market
48	GWH-B_190600	3	[GWh] GWh	6	Electricity supply	NRG_IND_342M	[B_190600] Net Imports
49	GWH-B_190900	2	[GWh] GWh	6	Electricity supply	NRG_IND_342M	[B_190900] Supply
50	GWH-B_197000	2	[GWh] GWh	6	Electricity supply	NRG_IND_342M	[B_197000] Total gross production
51	IRT_DTD	0	% rate	0	Money market interest rates	IRT_ST_M	[IRT_DTD] Day-to-day rate
52	IRT_M1	0	% rate	0	Money market interest rates	IRT_ST_M	[IRT_M12] 12-month rate
53	IRT_M12	0	% rate	0	Money market interest rates	IRT_ST_M	[IRT_M1] 1-month rate
54	IRT_M3	0	% rate	0	Money market interest rates	IRT_ST_M	[IRT_M3] 3-month rate
55	IRT_M6	0	% rate	0	Money market interest rates	IRT_ST_M	[IRT_M6] 6-month rate
56	NEER_IC42	1	[I10] Index, 2010=100	4	Effective exchange rate	ERT_EFF_IC_M	[NEER_IC42] Nominal effective exchange rate - 42 trading partners (industrial countries)
57	MCBY	0	% rate	4	EMU convergence criterion	IRT_LT_MCBY_M	[MCBY] EMU convergence criterion bond yields

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
58	RCH_M-CP00	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP00] All-items HICP
59	RCH_M-CP01	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP01] Food and non-alcoholic beverages
60	RCH_M-CP03	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP03] Clothing and footwear
61	RCH_M-CP04	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP04] Housing, water, electricity, gas and other fuels
62	RCH_M-CP045	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP045] Electricity, gas and other fuels
63	RCH_M-CP05	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP05] Furnishings, household equipment and routine household maintenance
64	RCH_M-CP06	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP06] Health
65	RCH_M-CP07	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP07] Transport
66	RCH_M-CP08	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP08] Communications
67	RCH_M-CP10	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[CP10] Education

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
68	RCH_M-FOOD	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[FOOD] Food including alcohol and tobacco
69	RCH_M-FUEL	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[FUEL] Liquid fuels and fuels and lubricants for personal transport equipment
70	RCH_M-GD	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[GD] Goods (overall index excluding services)
71	RCH_M-IGD	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[IGD] Industrial goods
72	RCH_M-NRG	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[NRG] Energy
73	RCH_M-SERV	0	% monthly change	2	Harmonised Index of Consumer Prices	PRC_HICP_MMOR	[SERV] Services (overall index excluding goods)
74	PRON-B	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[B] Mining and quarrying
75	PRON-B-E36	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[B-E36] Industry (except construction, sewerage, waste management and remediation activities)
76	PRON-C	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[C] Manufacturing
77	PRON-D	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[D] Electricity, gas, steam and air conditioning supply

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
78	PRON-MIG_CAG	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[MIG_CAG] MIG - capital goods
79	PRON-MIG_COG	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[MIG_COG] MIG - consumer goods
80	PRON-MIG_DCOG	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[MIG_DCOG] MIG - durable consumer goods
81	PRON-MIG_NDCOG	1	[I15] Index, 2015=100	2	Producer prices in industry, total	STS_INPP_M	[MIG_NDCOG] MIG - non-durable consumer goods
82	TOTAL-PC_ACT	1	[PC_ACT] Percentage of population in the labour force	4	Unemployment	UNE_RT_M	[TOTAL] Total
83	Y25-74-PC_ACT	1	[PC_ACT] Percentage of population in the labour force	4	Unemployment	UNE_RT_M	[Y_LT25] Less than 25 years
84	Y_LT25-PC_ACT	1	[PC_ACT] Percentage of population in the labour force	4	Unemployment	UNE_RT_M	[Y25-74] From 25 to 74 years
85	BUIL-TOT-COF-BS	1		0	Survey		
86	BUIL-TOT-1-BS	1		0	Survey		
87	BUIL-TOT-2-F1S	1		0	Survey		
88	BUIL-TOT-2-F2S	1		0	Survey		
89	BUIL-TOT-2-F3S	1		0	Survey		
90	BUIL-TOT-2-F4S	1		0	Survey		

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
91	BUIL-TOT-2-F5S	1		0	Survey		
92	BUIL-TOT-2-F6S	1		0	Survey		
93	BUIL-TOT-2-F7S	1		0	Survey		
94	BUIL-TOT-3-BS	1		0	Survey		
95	BUIL-TOT-4-BS	1		0	Survey		
96	BUIL-TOT-5-BS	1		0	Survey		
97	CONS-TOT-COF-BS	1		0	Survey		
98	CONS-TOT-1-BS	1		0	Survey		
99	CONS-TOT-2-BS	1		0	Survey		
100	CONS-TOT-3-BS	1		0	Survey		
101	CONS-TOT-4-BS	1		0	Survey		
102	CONS-TOT-5-BS	1		0	Survey		
103	CONS-TOT-6-BS	1		0	Survey		
104	CONS-TOT-7-BS	1		0	Survey		
105	CONS-TOT-8-BS	1		0	Survey		
106	CONS-TOT-9-BS	1		0	Survey		
107	CONS-TOT-10-BS	1		0	Survey		
108	CONS-TOT-11-BS	1		0	Survey		
109	CONS-TOT-12-BS	1		0	Survey		
110	INDU-TOT-COF-BS	1		0	Survey		
111	INDU-TOT-1-BS	1		0	Survey		
112	INDU-TOT-2-BS	1		0	Survey		
113	INDU-TOT-3-BS	1		0	Survey		

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
114	INDU-TOT-4-BS	1		0	Survey		
115	INDU-TOT-5-BS	1		0	Survey		
116	INDU-TOT-6-BS	1		0	Survey		
117	INDU-TOT-7-BS	1		0	Survey		
118	INDU	1		0	Survey		
119	SERV	1		0	Survey		
120	CONS	1		0	Survey		
121	RETA	1		0	Survey		
122	BUIL	1		0	Survey		
123	ESI	1		0	Survey		
124	EEI	1		0	Survey		
125	RETA-TOT-COF-BS	1		0	Survey		
126	RETA-TOT-1-BS	1		0	Survey		
127	RETA-TOT-2-BS	1		0	Survey		
128	RETA-TOT-3-BS	1		0	Survey		
129	RETA-TOT-4-BS	1		0	Survey		
130	RETA-TOT-5-BS	1		0	Survey		
131	RETA-TOT-6-BS	1		0	Survey		
132	SERV-TOT-COF-BS	1		0	Survey		
133	SERV-TOT-1-BS	1		0	Survey		
134	SERV-TOT-2-BS	1		0	Survey		
135	SERV-TOT-3-BS	1		0	Survey		
136	SERV-TOT-4-BS	1		0	Survey		

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
137	SERV-TOT-5-BS	1		0	Survey		
138	SERV-TOT-6-BS	1		0	Survey		
139	EMPL-B	1		0	Survey		
140	EMPL-B-E36	1		0	Survey		
141	EMPL-C	1		0	Survey		
142	EMPL-D	1		0	Survey		
143	EMPL-MIG_CAG	1		0	Survey		
144	EMPL-MIG_COG	1		0	Survey		
145	EMPL-MIG_DCOG	1		0	Survey		
146	EMPL-MIG_NDCOG	1		0	Survey		
147	HOWK-B	1		0	Survey		
148	HOWK-B-E36	1		0	Survey		
149	HOWK-C	1		0	Survey		
150	HOWK-D	1		0	Survey		
151	HOWK-MIG_CAG	1		0	Survey		
152	HOWK-MIG_COG	1		0	Survey		
153	HOWK-MIG_DCOG	1		0	Survey		
154	HOWK-MIG_NDCOG	1		0	Survey		
155	WAGE-B	1		0	Survey		
156	WAGE-B-E36	1		0	Survey		
157	WAGE-C	1		0	Survey		
158	WAGE-D	1		0	Survey		
159	WAGE-MIG_CAG	1		0	Survey		

Table OA.5 continued from previous page

n	Name	Tr.	Units	Delay weeks	Group	Code	Description
160	WAGE-MIG_COG	1		0	Survey		
161	WAGE-MIG_DCOG	1		0	Survey		
162	WAGE-MIG_NDCOG	1		0	Survey		
163	CLV_I15	4	[I15]	6	Quarterly GDP	CLV_I15	Quarterly GDP

References

M. Bańbura, D. Giannone, M. Modugno, and L. Reichlin. Now-Casting and the Real-Time Data Flow. In G. Elliott and A. Timmermann, editors, *Handbook of Economic Forecasting, Vol. 2 Part A*, chapter 4, pages 195–237. 2013.

M. P. Clements and A. B. Galvão. Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States. *Journal of Business and Economic Statistics*, 26(4):546–554, 2008.

F. X. Diebold and R. S. Mariano. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253, 1995.

R. C. Elandt. The Folded Normal Distribution: Two Methods of Estimating Parameters from Moments. *Technometrics*, 3(4):551–562, 1961.

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.

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.

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

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.

M. Marcellino and C. Schumacher. Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4):518–550, 2010.

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.

P. Phillips and H. R. Moon. Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111, 1999.

D. N. Politis and J. P. Romano. The Stationary Bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994.

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.