university of groningen
faculty of economics and business

2019010-EEF

Confidence Regions for Averaging Estimators

November 2019

Tom Boot

SOM is the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:

- Economics, Econometrics and Finance
- Global Economics & Management
- Innovation & Organization
- Marketing
- Operations Management & Operations Research
- Organizational Behaviour

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 9090/7068/3815

www.rug.nl/feb/research

university of groningen
faculty of economics and business

Confidence Regions for Averaging Estimators

Tom Boot
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance
t.boot@rug.nl

Confidence regions for averaging estimators

Tom Boot*

September 3, 2019

Abstract

In models with many parameters, averaging unrestricted estimators with estimators from restricted models can reduce estimation risk. We construct valid confidence regions centered at such averaging estimators. When the number of observations and imposed restrictions are sufficiently large, these regions have lower expected volume compared to the standard confidence region. Power gains over the standard F-test are found when the estimator from the restricted model is close to the true parameter vector and increases the distance to the parameter vector under the null.

1 Introduction

Estimation of high-dimensional parameter vectors can be inefficient even when the number of observations is sufficiently large to estimate the parameters. To increase efficiency, we can average estimators from an unrestricted model with estimators from one or more restricted models. When the number of restrictions on the parameters of interest is sufficiently large and averaging weights are of the type proposed by James and Stein (1961), averaging estimators dominate the risk of the unrestricted estimator and are locally minimax efficient (Hansen, 2016).

In this paper, we develop joint confidence regions centered at model averaging estimators with James-Stein-type weights. This enables valid inference after averaging estimators from models with and without control variables, random effects and fixed effects models, and when averaging instrumental variables estimators with least squares estimators. Instead of averaging with a single restricted estimator, we also consider averaging estimators from a sequence of nested models.

*University of Groningen, t.boot@rug.nl

The proposed confidence regions are based on the observation by Stein (1981) that the difference between the mean squared error of the averaging estimator and an unbiased risk estimate satisfies a central limit theorem in the number of parameters. Beran (1995) formalizes this in a set-up where a normally distributed vector is averaged with a fixed vector. We extend these results to allow for the construction of confidence regions after averaging an unrestricted and a restricted estimator as in Hansen (2016).

To leverage Stein's lemma in the proofs, results are derived under sequential limits in the sample size (n) and the effective number of restrictions on the parameters of interest (d). The structure of the proof is then similar to that of Beran (1995), with several modifications needed to establish the limiting distribution. In particular, averaging with a random vector changes the asymptotic variance needed to calculate the confidence regions. A well-known problem with sequential limit theory is that it can be misleading with regard to finite sample properties. We therefore study the averaging estimator in the linear regression model in more detail. Under the rate condition $d/n \to 0$, we find that the limit distribution under sequential limits coincides with that under joint limits.

In line with the risk reduction, the expected volume of the recentered confidence regions is smaller compared to the standard confidence region centered at the unrestricted estimator, a property already anticipated by Stein (1956). This reduction in volume affects power when the confidence regions are used as an alternative to a standard F-test. When the restricted estimator is expected to be close to the true parameter vector, power is increased over the standard confidence region. When the restricted estimator is expected to be close to the parameter vector under the null, power is reduced. This emphasizes that the restrictions should be selected with care. This is crucially different from the mean squared error perspective, where it is possible to uniformly dominate the unrestricted estimator regardless of the imposed restrictions.

We numerically analyze the confidence regions in a set of linear and instrumental variables models. We consider *indirect* restrictions, where the unrestricted estimator is averaged with a more efficient, but potentially biased estimator from a restricted model, as well as *direct* restrictions where the unrestricted estimator is averaged with a fixed vector. The coverage rate is close to nominal, even when the number of restrictions is small. Indirect restrictions improve power over the standard F-test in some parts of the parameter space, yet lose power in others. Direct restrictions where the signs of the fixed parameters correspond to the true signs, increase power over the whole parameter space. The results are further illustrated in a cross-country growth regression derived from Magnus et al. (2010).

Recentered confidence regions for multiple parameters have been discussed for the case where the restricted estimator is a fixed vector. Casella and Hwang (2012) provide an overview of the literature on recentered confidence regions. If the same radius is used as for the standard confidence region, Casella and Hwang (1982) prove that recentering increases the coverage rate. Confidence sets with reduced volume are developed for example by Casella and Hwang (1983) and Samworth (2005). In our numerical evaluation, we find these confidence regions to be conservative, especially when the number of parameters increases.

Confidence intervals for individual parameters after model averaging are proposed by Hjort and Claeskens (2003). Based on this suggestion, Liu (2015) develops confidence intervals for the Mallows model averaging estimator of Hansen (2007), and the jackknife model averaging estimator of Hansen and Racine (2012). Simulation-based approaches are considered by Claeskens and Hjort (2008), DiTraglia (2016) and Zhang and Liu (2019). These papers find substantial reductions in the length of the confidence intervals. Leeb and Kabaila (2017) show that for one-dimensional intervals, in contrast with the multidimensional regions as considered in this paper, such length reductions are not uniform over the parameter space.

This paper is structured as follows. Section 2 introduces the averaging estimator and the associated confidence regions, and provides geometric intuition for the results. The theoretical validity and the volume of the confidence regions is discussed in Section 3. Section 4 provides numerical evidence for the coverage rate and power properties of associated hypothesis tests. Section 5 concludes.

The following notation is used. The symbol $\Rightarrow$ denotes convergence in distribution, $\rightarrow_p$ is convergence in probability. Almost surely is abbreviated as $a.s.$ $||\boldsymbol{A}||$ denotes the largest eigenvalue of the square matrix $\boldsymbol{A}$. $\boldsymbol{P}_X = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, and $\boldsymbol{M}_X = \boldsymbol{I} - \boldsymbol{P}_X$. $\Phi(x)$ is the standard normal cumulative distribution function.

2 General set-up

We consider the set-up as in Hansen (2016). Suppose we have n observations from a model which depends on a parameter vector $\boldsymbol{\theta}_n \in \mathbb{R}^k$. We are interested in a parameter vector $\boldsymbol{\beta}_n = \boldsymbol{g}(\boldsymbol{\theta}_n) \in \mathbb{R}^p$ for some differentiable function $\boldsymbol{g} : \mathbb{R}^k \rightarrow \mathbb{R}^p$. For example, $\boldsymbol{\beta}_n$ might be a subset of $\boldsymbol{\theta}_n$. Define $\boldsymbol{G} = \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{\theta})' \in \mathbb{R}^{k\times p}$. Consider a set of restrictions on $\boldsymbol{\theta}_n$ as $\boldsymbol{r}(\boldsymbol{\theta}_n) = \boldsymbol{0}$ for some differentiable function $\boldsymbol{r} : \mathbb{R}^k \rightarrow \mathbb{R}^r$, and define $\boldsymbol{R} = \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{r}(\boldsymbol{\theta})' \in \mathbb{R}^{k\times r}$.

2.1 Averaging estimator

The estimator of the parameter vector of interest $\boldsymbol{\beta}_n$ from the unrestricted model is denoted as $\hat{\boldsymbol{\beta}}_n = \boldsymbol{g}(\hat{\boldsymbol{\theta}}_n)$, and from the restricted model as $\tilde{\boldsymbol{\beta}}_n$. The averaging estimator is given by the linear combination

$$\begin{aligned} \hat{\boldsymbol{\beta}}^a_n &= \hat{w}_{n,d_n}\tilde{\boldsymbol{\beta}}_n + (1-\hat{w}_{n,d_n})\hat{\boldsymbol{\beta}}_n \\ &= \tilde{\boldsymbol{\beta}}_n + (1-\hat{w}_{n,d_n})\hat{\boldsymbol{\delta}}_n. \end{aligned} \tag{1}$$

Here, we write for the difference between the unrestricted and restricted estimator

$$\hat{\boldsymbol{\delta}}_n = \hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n, \quad \boldsymbol{\delta}_n = \mathrm{E}\left[\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n\right]. \tag{2}$$

Let $\boldsymbol{\Sigma}_u$, $\boldsymbol{\Sigma}_r$, and $\boldsymbol{\Sigma}_\delta$ denote the asymptotic variance matrices of $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$, with the corresponding estimators $\hat{\boldsymbol{\Sigma}}_{u,n}$, $\hat{\boldsymbol{\Sigma}}_{r,n}$, and $\hat{\boldsymbol{\Sigma}}_{\delta,n}$.

The averaging weight $\hat{w}_{n,d_n}$ in (1) aims to minimize the risk $\rho(\hat{\boldsymbol{\beta}}^a_n, \boldsymbol{\beta}_n)$, which is defined as the expectation of a quadratic loss function, i.e.

$$\rho(\hat{\boldsymbol{\beta}}^a_n, \boldsymbol{\beta}_n) = \mathrm{E}[\ell(\hat{\boldsymbol{\beta}}^a_n, \boldsymbol{\beta}_n)], \quad \ell(\hat{\boldsymbol{\beta}}^a_n, \boldsymbol{\beta}_n) = n(\hat{\boldsymbol{\beta}}^a_n - \boldsymbol{\beta}_n)'\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}(\hat{\boldsymbol{\beta}}^a_n - \boldsymbol{\beta}_n). \tag{3}$$

We analyze averaging weights closely related to the shrinkage factor of James and Stein (1961), which Hansen (2016) shows to be locally asymptotically minimax efficient,

$$\hat{w}_{n,d_n} = \frac{\hat{\tau}_n}{n\hat{q}_n}, \quad \hat{\tau}_n = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\hat{\boldsymbol{\Sigma}}_{\delta,n}) - 2||\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\hat{\boldsymbol{\Sigma}}_{\delta,n}||, \quad \hat{q}_n = \hat{\boldsymbol{\delta}}'_n\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\hat{\boldsymbol{\delta}}_n. \tag{4}$$

The parameter $\hat{\tau}_n$ can be interpreted as a measure of the variance reduction achieved by the imposed restrictions. When averaging with a fixed vector, $\hat{\boldsymbol{\Sigma}}_{\delta,n} = \hat{\boldsymbol{\Sigma}}_{u,n}$, and $\hat{\tau}_n = p-2$ as proposed by James and Stein (1961). The denominator of $\hat{w}_{n,d_n}$ measures the misspecification bias induced by the restrictions. The weight placed on the restricted model is large when the difference between the estimates from the unrestricted and the restricted model as measured by $\hat{q}_n$ is small, and the variance reduction from the restrictions as measured by $\hat{\tau}_n$ is large. A geometric motivation for the averaging weights is provided in Section 2.3.

The weights are indexed by the sample size n and the effective number of restrictions on the parameter of interest d_n introduced by Hansen (2016) as

$$d_n = \frac{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\hat{\boldsymbol{\Sigma}}_{\delta,n})}{||\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\hat{\boldsymbol{\Sigma}}_{\delta,n}||}. \tag{5}$$

We will consider asymptotics where $d_n \to \infty$. Under the assumptions given in Section 3.2, we have $d_n \leq \min[p, r]$, so that $d_n \to \infty$ implies $(p, r) \to \infty$. In the special case where we average with a fixed vector, $\hat{\mathbf{\Sigma}}_{\delta,n} = \hat{\mathbf{\Sigma}}_{u,n}$ and hence, $d_n = p = r$, where p is the dimension of $\boldsymbol{\beta}_n$ and r is the number of restrictions.

2.2 Confidence regions

We consider spherical confidence regions, which are defined as follows.

Definition 1 *For any estimator $\bar{\boldsymbol{\beta}}_n$ of the parameter vector of interest $\boldsymbol{\beta}_n$, a weighting matrix $\hat{\boldsymbol{W}}$, and critical values $\hat{b}_n$, the confidence region is defined as*

$$C(\bar{\boldsymbol{\beta}}_n, \hat{b}_n) = \left\{ \boldsymbol{t} : n(\bar{\boldsymbol{\beta}}_n - \boldsymbol{t})' \hat{\boldsymbol{W}} (\bar{\boldsymbol{\beta}}_n - \boldsymbol{t}) \leq \hat{b}_n^2 \right\}. \tag{6}$$

Denote the inverse $\chi^2(p)$-distribution function as $F^{-1}_{\chi^2(p)}(x)$. The conventional confidence region for $\boldsymbol{\beta}_n$ centered at $\hat{\boldsymbol{\beta}}_n$ with coverage rate $1-\alpha$, is

$$C_n(\hat{\boldsymbol{\beta}}_n, d_\chi) = \left\{ \boldsymbol{t} : n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{t})' \hat{\mathbf{\Sigma}}^{-1}_{u,n} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{t}) \leq b_\chi^2 \right\}, \quad b_\chi^2 = F^{-1}_{\chi^2(p)}(1-\alpha). \tag{7}$$

We consider recentered confidence regions where $\bar{\boldsymbol{\beta}}_n$ in Definition 1 is the averaging estimator (1). To make the results comparable to the confidence region defined by (7), the weighting matrix is taken as $\hat{\boldsymbol{W}} = \hat{\mathbf{\Sigma}}^{-1}_{u,n}$ throughout. We then consider the following critical values.

$$\hat{b}_n^2 = \max(0, \hat{e}_n), \quad \hat{e}_n = \hat{\rho}\left(\hat{\boldsymbol{\beta}}^a_n, \boldsymbol{\beta}_n\right) + p^{1/2} \hat{\sigma}(\hat{c}_n) \Phi^{-1}(1-\alpha). \tag{8}$$

Here $\hat{\rho}$ is the following risk estimate,

$$\hat{\rho}\left(\hat{\boldsymbol{\beta}}^a_n, \boldsymbol{\beta}_n\right) = p - 2\hat{\tau}_n \left[\frac{\operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n})}{n\hat{q}_n} - 2 \frac{n\hat{\boldsymbol{\delta}}'_n \hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n} \hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\boldsymbol{\delta}}_n}{(n\hat{q}_n)^2} \right] + \hat{\tau}_n^2 \frac{1}{n\hat{q}_n}, \tag{9}$$

and the asymptotic variance $\hat{\sigma}^2(\hat{c}_n)$ is estimated as

$$\hat{\sigma}^2(\hat{c}_n) = 2 - 4\frac{\hat{\tau}_n^2}{p} \left[\frac{\operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n})^{-1}}{\hat{c}_n + 1} - \frac{\operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n} \hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n}) / \operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n})^2}{(\hat{c}_n + 1)^2} \right], \tag{10}$$

with

$$\hat{c}_n = n\tilde{\boldsymbol{\delta}}'_n \hat{\mathbf{\Sigma}}^{-1}_{u,n} \tilde{\boldsymbol{\delta}}_n / \operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n}), \quad \tilde{\boldsymbol{\delta}}_n = \max\left[0, 1 - \operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}_{u,n} \hat{\mathbf{\Sigma}}_{\delta,n}) / (n\hat{q}_n)\right]^{1/2} \hat{\boldsymbol{\delta}}_n. \tag{11}$$

In large samples, the risk estimator $\hat{\rho}(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$ appearing in (8) is an unbiased estimator of the risk $\rho(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$ defined in (3). Setting $\boldsymbol{t} = \boldsymbol{\beta}_n$ in (6), bringing $\hat{\rho}(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$ to the left-hand side, and rescaling by $p^{-1/2}$, we obtain the difference $D(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n) = p^{-1/2}[\ell(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n) - \hat{\rho}(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)]$. By the unbiasedness of $\hat{\rho}(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$, this difference has expectation zero when n is large. If in addition the number of effective restrictions d_n is sufficiently large, we find that $D(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$ is asymptotically normally distributed with an asymptotic variance that is consistently estimated by (10)–(11). This then results in asymptotically correct coverage of $\boldsymbol{\beta}_n$.

2.3 Geometric motivation in the linear regression model

To gain intuition for the weights (4) and the properties of confidence regions centered at the averaging estimator, consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{12}$$

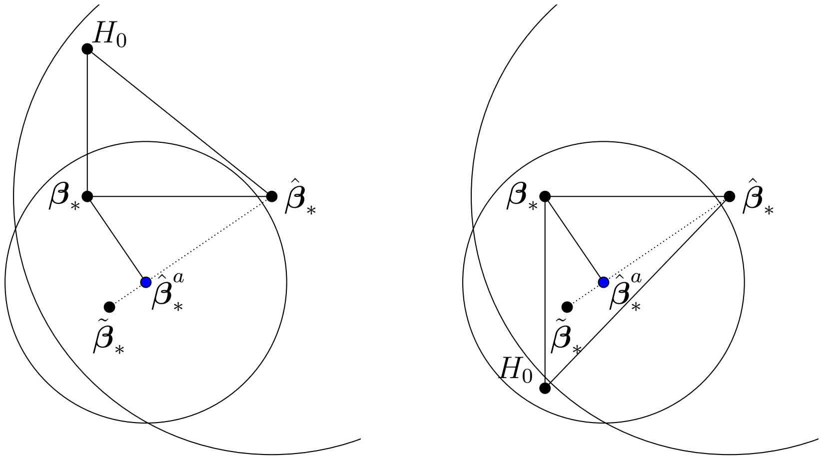
where $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]' \in \mathbb{R}^{n \times p}$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$. The errors are i.i.d. and satisfy $\mathrm{E}[\varepsilon_i|\boldsymbol{X}] = 0$, and $\mathrm{E}[\varepsilon_i^2|\boldsymbol{X}] = \sigma^2$. For simplicity, in this section we assume that σ^2 is known and we condition on $\boldsymbol{X}$. We also suppress the dependence of the various estimators on the sample size n.

The unrestricted estimator is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, with $\mathrm{var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = n^{-1}\hat{\boldsymbol{\Sigma}}_u = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Partition $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]$, and accordingly $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2']'$, where $\boldsymbol{\beta}_2 \in \mathbb{R}^r$. We consider a set of restrictions defined by $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{0}$, where $\boldsymbol{R} = [\boldsymbol{O}_{r \times p-r}, \boldsymbol{I}_r]$. This gives the restricted estimator $\tilde{\boldsymbol{\beta}} = [\boldsymbol{y}'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}, \boldsymbol{0}_r']'$, with $\mathrm{var}(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) = n^{-1}\hat{\boldsymbol{\Sigma}}_r = n^{-1}(\hat{\boldsymbol{\Sigma}}_u - \hat{\boldsymbol{\Sigma}}_u\boldsymbol{R}(\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u\boldsymbol{R})^{-1}\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u)$. The difference between the estimators is $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}$. As pointed out by Hausman (1978), $\mathrm{var}(\hat{\boldsymbol{\delta}}|\boldsymbol{X}) = n^{-1}\hat{\boldsymbol{\Sigma}}_\delta = n^{-1}(\hat{\boldsymbol{\Sigma}}_u - \hat{\boldsymbol{\Sigma}}_r)$. We also use below that $\mathrm{cov}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}|\boldsymbol{X}) = n^{-1}\hat{\boldsymbol{\Sigma}}_\delta$.

We are interested in averaging estimators with a low estimation risk $\rho(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = \mathrm{E}[\ell(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})] = \mathrm{E}[n(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})'\hat{\boldsymbol{\Sigma}}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})]$. Figure 1 displays the parameter vectors $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\beta}}$ rescaled by $(n^{-1}\hat{\boldsymbol{\Sigma}}_u)^{-1/2}$. The averaging estimator $\hat{\boldsymbol{\beta}}_*^a$ closest to $\boldsymbol{\beta}_*$ is given by the orthogonal projection of $\boldsymbol{\beta}_*$ on the line segment joining $\hat{\boldsymbol{\beta}}_*$ and $\tilde{\boldsymbol{\beta}}_*$. Defining $\hat{\boldsymbol{\delta}}_* = \hat{\boldsymbol{\beta}}_* - \tilde{\boldsymbol{\beta}}_*$, this suggests

$$\hat{\boldsymbol{\beta}}_*^a = \tilde{\boldsymbol{\beta}}_* + \frac{\hat{\boldsymbol{\delta}}_*'(\boldsymbol{\beta}_* - \tilde{\boldsymbol{\beta}}_*)}{\hat{\boldsymbol{\delta}}_*'\hat{\boldsymbol{\delta}}_*}\hat{\boldsymbol{\delta}}_* = \tilde{\boldsymbol{\beta}}_* + \left(1 - \frac{\hat{\boldsymbol{\delta}}_*'(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_*)}{\hat{\boldsymbol{\delta}}_*'\hat{\boldsymbol{\delta}}_*}\right)\hat{\boldsymbol{\delta}}_* \tag{13}$$

Figure 1: Power resulting from recentered confidence regions



Note: we write $\boldsymbol{\beta}_* = (n^{-1}\hat{\boldsymbol{\Sigma}}_u)^{-\frac{1}{2}}\boldsymbol{\beta}$, and similar for the other vectors. H_0 denotes the parameter vector under the null hypothesis.

Multiplying from the left with $(n^{-1}\hat{\boldsymbol{\Sigma}}_u)^{\frac{1}{2}}$, we get the averaging estimator

$$\hat{\boldsymbol{\beta}}^a = \tilde{\boldsymbol{\beta}} + \frac{n\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_u^{-1}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})}{n\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_u^{-1}\hat{\boldsymbol{\delta}}}\hat{\boldsymbol{\delta}} = \tilde{\boldsymbol{\beta}} + \left(1 - \frac{n\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_u^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{n\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_u^{-1}\hat{\boldsymbol{\delta}}}\right)\hat{\boldsymbol{\delta}}. \tag{14}$$

The denominator $n\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_u^{-1}\hat{\boldsymbol{\delta}}$ equals $n\hat{q}_n$ in the averaging weights (4). Also, for the numerator we have $\mathrm{E}[n\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_u^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|\boldsymbol{X}] = \mathrm{tr}(n\hat{\boldsymbol{\Sigma}}_u^{-1}\mathrm{cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}|\boldsymbol{X})) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_u^{-1}\hat{\boldsymbol{\Sigma}}_\delta) = r$, corresponding to the first term of $\hat{\tau}_n$ in (4). The second term in $\hat{\tau}_n$ is of lower order in r, and does not appear in the geometric picture sketched here. We see that the weights (4) achieve a low estimation risk by estimating the projection that minimizes the loss $\ell(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$.

Figure 1 also shows a particular realization of a confidence region centered at the unrestricted estimator $\hat{\boldsymbol{\beta}}_*$ given by the large circle, and one centered at the averaging estimator $\hat{\boldsymbol{\beta}}^a_*$ given by the small circle. The volume of the confidence regions centered at the averaging estimator can be reduced without sacrificing coverage since its distance to the true parameter vector is smaller.

Efron (2006) points out that smaller confidence regions do not necessarily improve the power of corresponding tests. This can be seen by comparing both panels of Figure 1. On the left, the restricted estimator $\tilde{\boldsymbol{\beta}}$ is further away from the null hypothesis than the true parameter vector $\boldsymbol{\beta}$. In this case, recentering shifts the confidence region away from the parameter vector under the null and we expect to gain power against H_0. On the right however, the restricted estimator is close to the parameter vector under the null. The recentered confidence region now does not reject the null, while the standard confidence region would.

The leading case in practice is to consider $H_0 : \boldsymbol{\beta} = \mathbf{0}$. From the discussion above, we expect to gain power if the restricted estimator has the same signs as the unknown parameter vector $\boldsymbol{\beta}$, and the magnitude of the parameters is not too small. Section 4 provides a suggestion to obtain the appropriate restrictions.

3 Theoretical results

3.1 Preliminaries

We defined the effective number of restrictions on the parameters of interest in (5). The main results in this paper are based on sequential asymptotic limits, where first the number of observations n goes to infinity, which we refer to as the (n)-asymptotic limit. Then, we consider the limit where the effective number of restrictions d goes to infinity, which we refer to as the (d, n)-asymptotic limit. Following Phillips and Moon (1999), this is also written as $(d, n \to \infty)_{\text{seq}}$. We study joint limits, written as $(d, n \to \infty)$, in the linear regression model in Section 3.6.

The theoretical results will show the validity and expected volume of the confidence regions centered at the averaging estimator (1) with critical values given by (8)–(11). The confidence regions are said to be valid under the following definition.

Definition 2 *Let $\hat{b}_n$ be the critical value for the confidence region for the estimator $\hat{\boldsymbol{\beta}}_n^a$ with weighting matrix $\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}$. The confidence region $C_n\left(\hat{\boldsymbol{\beta}}_n^a, \hat{b}_n, \hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\right)$ is (d, n)-asymptotically valid if*

$$\lim_{d\to\infty} \lim_{n\to\infty} \left| \mathrm{P}\left(\boldsymbol{\beta} \in C_n\left(\hat{\boldsymbol{\beta}}_n^a, \hat{b}_n\right)\right) - (1-\alpha) \right| = 0. \tag{15}$$

We measure the volume by the geometrical risk, see for example Beran (1995). This geometric risk is trimmed to ensure that it is well defined for all values of n. This trimming does not affect the expressions for the geometrical risk that we derive below.

Definition 3 *Suppose ξ is a finite positive constant. The confidence region $C_n = C_n(\hat{\boldsymbol{\beta}}_n, \hat{b}_n)$ has trimmed (n)-asymptotic geometrical risk*

$$\begin{aligned} GR(C) &= \lim_{n\to\infty} \mathrm{E}\left[\min\left\{\sup_{\boldsymbol{t}\in C_n} \sqrt{p^{-1} n(\boldsymbol{\beta}_n - \boldsymbol{t})' \hat{\boldsymbol{\Sigma}}_{u,n}^{-1} (\boldsymbol{\beta}_n - \boldsymbol{t})}, \xi\right\}\right] \\ &= \lim_{n\to\infty} \mathrm{E}\left[\min\left\{\sqrt{p^{-1} \ell(\hat{\boldsymbol{\beta}}_n, \boldsymbol{\beta}_n)} + p^{-1/2} \hat{b}_n, \xi\right\}\right]. \end{aligned} \tag{16}$$

If $\lim_{d\to\infty} GR(C) = g$, *where* g *does not depend on* ξ, *then the* (d,n)*-asymptotic geometrical risk of* C_n *equals* g.

The geometrical risk measures the expected distance between the most distant vector in the confidence region and the true parameter vector of interest. As indicated by the second line of (16), this equals the distance between the estimator and the true parameter vector plus the radius of the confidence sphere.

Finally, we will need a measure of the (n)-asymptotic risk of an estimator. Subtracting an (n)-asympotically unbiased estimator of this risk from the mean squared error will yield a quantity that asymptotically has mean zero.

Definition 4 *Suppose that as* $n \to \infty$, $\ell(\hat{\boldsymbol{\beta}}^a_n, \boldsymbol{\beta}_n) \Rightarrow \ell(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})$. *The* (n)*-asymptotic risk is defined as*

$$\rho(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = \mathrm{E}\left[\ell(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})\right]. \tag{17}$$

3.2 Assumptions

Assumption A1 *Define the restricted set* $\boldsymbol{\Theta}_r = \{\boldsymbol{\theta} : \boldsymbol{r}(\boldsymbol{\theta}) = \mathbf{0}\}$. *Let* $\boldsymbol{\theta}_r \in \boldsymbol{\Theta}_r$. *The true parameter vector* $\boldsymbol{\theta}_n$ *is close to the restricted parameter vector* $\boldsymbol{\theta}_r$, *in the sense that* $\boldsymbol{\theta}_n = \boldsymbol{\theta}_r + n^{-1/2}\boldsymbol{h}$

Assumption A2 *Let the* k*-dimensional random vector* $\boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{V})$, *and define* $\boldsymbol{V} = \boldsymbol{L}'\boldsymbol{L}$. *Along sequences* $\boldsymbol{\theta}_n$ *defined in Assumption A1, as* $n \to \infty$,

1. *The parameter estimates converge in distribution to*

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) &\Rightarrow \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{G}'\boldsymbol{z} \\ \sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) &\Rightarrow \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{G}'\left[\boldsymbol{z} - \boldsymbol{V}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{V}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{z}+\boldsymbol{h})\right]. \end{aligned} \tag{18}$$

2. *The covariance matrix estimates converge in probability to*

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{u,n} &\to_p \boldsymbol{\Sigma}_u = \boldsymbol{G}'\boldsymbol{V}\boldsymbol{G}, \\ \hat{\boldsymbol{\Sigma}}_{r,n} &\to_p \boldsymbol{\Sigma}_r = \boldsymbol{G}'\boldsymbol{L}'(\boldsymbol{I} - \boldsymbol{L}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{L}')\boldsymbol{L}\boldsymbol{G}, \\ \hat{\boldsymbol{\Sigma}}_{\delta,n} &\to_p \boldsymbol{\Sigma}_\delta = \boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_r. \end{aligned} \tag{19}$$

3. *The averaging weights converge in distribution to*

$$\hat{w}_{n,d_n} \Rightarrow w_d = \tau/\hat{q}. \tag{20}$$

with $\hat{q} = (\boldsymbol{z}+\boldsymbol{h})'\boldsymbol{B}(\boldsymbol{z}+\boldsymbol{h})$, $\boldsymbol{B} = \boldsymbol{R}(\boldsymbol{R}'\boldsymbol{V}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{V}\boldsymbol{G}\boldsymbol{V}^{-1}\boldsymbol{G}'\boldsymbol{V}\boldsymbol{R}'(\boldsymbol{R}'\boldsymbol{V}\boldsymbol{R})^{-1}\boldsymbol{R}'$, $\tau = \mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta) - 2||\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta||$.

Furthermore, the convergence in 1. and 3. occurs jointly.

Assumption A3 *Define* $q = \boldsymbol{h}'\boldsymbol{B}\boldsymbol{h}$ *and* $c_d = \frac{q}{\operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}$. *For all* d, $c_d < \infty$, *and* $\lim_{d\to\infty} c_d = c < \infty$. *Moreover,* $\lim_{d\to\infty} \tau/p = a_1$, $\lim_{d\to\infty} \operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)/p = a_2$.

Assumption A1 prevents the restrictions to cause an (n)-asymptotically infinite bias in the restricted estimator $\tilde{\boldsymbol{\beta}}_n$. Assumption A2 regards the (n)-asymptotic behavior of the estimators and their covariance matrices. The vector $\boldsymbol{h}$ captures the misspecification bias that arises from imposing invalid restrictions. We see that the difference between the asymptotic covariance matrices of the unrestricted and restricted estimator is positive definite, so that a bias-variance trade-off is apparent in imposing the restrictions. A consequence of Assumption A2 is that the restricted estimator is (n)-asymptotically independent of its difference with the unrestricted estimator, the same principle that underlies the specification tests by Hausman (1978). Assumption A3 ensures that a law of large numbers in the effective number of restrictions d applies to the averaging weights.

3.3 Confidence regions centered at unrestricted estimators

To highlight the ideas underlying the construction of confidence regions for the averaging estimator, we can construct a valid confidence region under sequential limits for the unrestricted estimator $\hat{\boldsymbol{\beta}}_n$. Since the unrestricted estimator only depends on the dimension of the parameter vector of interest p, we consider here (p,n)-sequential asymptotics. The following lemma provides a (p,n)-asymptotically valid confidence region.

Lemma 1 *Let* $C_n(\hat{\boldsymbol{\beta}}_n, b)$ *be a confidence region for the unrestricted estimator with*

$$b^2 = p + \sqrt{p}\sigma\Phi^{-1}(1-\alpha), \tag{21}$$

where $\sigma = \sqrt{2}$. *Then,* $C_n(\hat{\boldsymbol{\beta}}_n, b)$ *is* (p,n)*-asymptotically valid.*

Proof: By Assumption A2, as $n \to \infty$,

$$p^{-1/2}\left[n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) - p\right] \Rightarrow p^{-1/2}\sum_{i=1}^{p}(z_i^2 - 1), \tag{22}$$

where $\{z_1, z_2, \ldots, z_p\}$ is a sequence of mean zero independent random variables with variance 1. Then, as $(p, n \to \infty)_{\text{seq}}$,

$$p^{-1/2}\left[n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) - p\right] \Rightarrow N(0, \sigma^2), \tag{23}$$

where $\sigma^2 = 2$. ■

Reasoning similar to that in the proof of Lemma 1 is used to develop confidence intervals for the averaging estimator (1). We interpret the left-hand side of (23) as the difference between the mean squared error of $\hat{\boldsymbol{\beta}}_n$, and an (n)-asymptotically unbiased estimator for its (n)-asymptotic risk. We therefore first derive such a risk estimate for the averaging estimator (1). We then show that the difference between the mean squared error and risk estimate converges in distribution to a normal with an asymptotic variance that can be (d, n)-consistently estimated. This is then used to construct (d, n)-asymptotically valid confidence regions for the averaging estimator. For normally distributed estimators averaged with the zero vector, the results reduce to those by Beran (1995).

Having established validity of the confidence region defined by Lemma 1, we turn to the associated geometrical risk.

Lemma 2 *The (p, n)-asymptotic geometrical risk for the confidence region defined in Lemma 1 equals 2.*

Proof: In Definition 1, take $\zeta = 3$. Define $\hat{t}_n^2 = p^{-1}n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\hat{\boldsymbol{\Sigma}}_u^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$. Under Assumption A2, as $n \to \infty$, $\hat{t}_n^2 \Rightarrow \hat{t}^2 = p^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Also the critical value in (21) is such that $p^{-1/2}\hat{b}_n = p^{-1/2}b$. Then, $\lim_{n\to\infty} \mathrm{E}\left[\min\left\{\hat{t}_n + p^{-1/2}\hat{b}_n, \zeta\right\}\right] = \mathrm{E}\left[\min\{\hat{t} + p^{-1/2}b, \zeta\}\right]$ by the bounded convergence theorem. By Lemma A.1 in Appendix A.3, $\operatorname{plim}_{p\to\infty} \hat{t} = 1$. From (21), $\operatorname{plim}_{p\to\infty} p^{-1/2}b = 1$. Then, since $\zeta = 3$, $\lim_{p\to\infty} \mathrm{E}\left[\min\{\hat{t} + p^{-1/2}b, \zeta\}\right] = 2$. ■

3.4 Confidence regions centered at averaging estimators

To apply the reasoning leading to the confidence region (23), we first need an (n)-asymptotically unbiased estimator for the (n)-asymptotic risk of $\hat{\boldsymbol{\beta}}_n^a$ given in (17). This is provided in the following lemma.

Lemma 3 *Suppose Assumption A1–A3 hold. Consider the risk estimator (9). Then, as $n \to \infty$,*

$$\hat{\rho}\left(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n\right) \Rightarrow \hat{\rho}\left(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}\right) = p - 2\tau\left[\frac{\operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}{\hat{q}} - 2\frac{\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}}{\hat{q}^2}\right] + \tau^2\frac{1}{\hat{q}}, \tag{24}$$

and $\mathrm{E}\left[\hat{\rho}(\hat{\boldsymbol{\beta}}^{a}, \boldsymbol{\beta})\right] = \rho\left(\hat{\boldsymbol{\beta}}^{a}, \boldsymbol{\beta}\right)$*, with* $\rho(\hat{\boldsymbol{\beta}}^{a}, \boldsymbol{\beta})$ *as in* (17)*.*

Proof: Appendix A.2 shows that this follows from an application of Stein's lemma.

In line with the approach to obtain confidence regions centered at the unrestricted estimator in (23), consider the difference between the quadratic loss (3) and the unbiased estimator for its risk from Lemma 3,

$$D_n\left(\hat{\boldsymbol{\beta}}^{a}_n, \boldsymbol{\beta}_n\right) = p^{-1/2}\left[\ell(\hat{\boldsymbol{\beta}}^{a}_n, \boldsymbol{\beta}_n) - \hat{\rho}\left(\hat{\boldsymbol{\beta}}^{a}_n, \boldsymbol{\beta}_n\right)\right]. \tag{25}$$

The following theorem gives the (d, n)-asymptotic distribution of $D_n\left(\hat{\boldsymbol{\beta}}^{a}_n, \boldsymbol{\beta}_n\right)$.

Theorem 1 *Suppose Assumption A1–A3 hold. Then, as* $(d, n \to \infty)_{\mathrm{seq}}$*,*

$$D_n\left(\hat{\boldsymbol{\beta}}^{a}_n, \boldsymbol{\beta}_n\right) \Rightarrow N\left(0, \sigma^2(c)\right), \quad \sigma^2(c) = 2 - 4\left[\frac{a_1}{c+1} - \frac{a_2}{(c+1)^2}\right], \tag{26}$$

where (c, a_1, a_2) *defined in Assumption A3.*

A proof is provided in Appendix A.3. This theorem generalizes Theorem 2.1 of Beran (1995), which was derived for the case where the estimators are exactly normal, the restricted estimator is the zero vector, and $\hat{\boldsymbol{\Sigma}}^{-1}_{u,n} = \hat{\boldsymbol{\Sigma}}_{\delta,n} = \boldsymbol{I}_p$. In this case $a_1 = a_2 = 1$, and $c = \lim_{p\to\infty} \boldsymbol{h}'\boldsymbol{h}/p$ with $\boldsymbol{h}$ as in Assumption A1.

The parameter c in Theorem 1 measures the strength of the misspecification bias induced by the model restrictions relative to the efficiency gain. To construct a valid confidence region, we need a (d, n)-consistent estimator of c, which is provided in the following corollary.

Corollary 1 *Suppose Assumption A2–A3 hold, and* $\hat{c}_n = n\tilde{\boldsymbol{\delta}}'_n\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\tilde{\boldsymbol{\delta}}_n/\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\hat{\boldsymbol{\Sigma}}_{\delta,n})$*, where* $\tilde{\boldsymbol{\delta}}_n = \left[\max\left(0, 1 - \frac{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}_{u,n}\hat{\boldsymbol{\Sigma}}_{\delta,n})}{n\hat{q}_n}\right)\right]^{1/2}\hat{\boldsymbol{\delta}}_n$*. Then, as* $(d, n \to \infty)_{\mathrm{seq}}$*,* $\hat{c}_n \to_p c$*.*

The proof follows from Lemma A.1 in the Appendix A.1.

Corollary 1 leads to the main theorem.

Theorem 2 *The confidence region* $C_n\left(\hat{\boldsymbol{\beta}}^{a}_n, \hat{b}_n\right)$ *with critical values* $\hat{b}_n$ *as in* (8) *is* (d, n)*-asymptotically valid with* (d, n)*-asymptotic geometrical risk* $= 2\left(\frac{c+1-a_1}{c+1}\right)^{1/2}$*, with* a_1 *and* c *as in Theorem 1.*

A proof is provided in Appendix A.4.

Since $a_1 \geq 0$, Theorem 2 states that the geometrical risk is at least as low as when centering the confidence region centered at the unrestricted estimator. We expect the largest improvements when the misspecification bias relative to the variance improvements, as measured by the parameter c, is small.

3.5 Sequences of nested models

Instead of a single set of restrictions, we can also consider a sequence of m restricted models. Here, we have sets of restrictions $\boldsymbol{r}^i(\boldsymbol{\theta}_n) = 0$ for $i = 1, \ldots, m$, with $\boldsymbol{r}^i : \mathbb{R}^k \to \mathbb{R}^{r^i}$ differentiable with respect to $\boldsymbol{\theta}_n$. As for a single set of restrictions, define $\boldsymbol{R}_i = \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{r}^i(\boldsymbol{\theta})' \in \mathbb{R}^{k \times r_i}$. By nested models, we mean

$$\boldsymbol{R}_{i+1} = [\boldsymbol{R}_i, \tilde{\boldsymbol{R}}_{i+1}]. \tag{27}$$

Denote by $\hat{\boldsymbol{\beta}}_n$ the unrestricted estimator, and by $\tilde{\boldsymbol{\beta}}_n^{(i)}$ the estimator under the i-th set of restrictions. For $i = 1, \ldots, m$, define

$$\hat{\boldsymbol{\delta}}_n^{(i)} = \hat{\boldsymbol{\beta}}_n^{(i-1)} - \tilde{\boldsymbol{\beta}}_n^{(i)}, \tag{28}$$

where $\hat{\boldsymbol{\beta}}_n^{(0)} = \hat{\boldsymbol{\beta}}_n$. The covariance matrices of $\hat{\boldsymbol{\beta}}_n$, $\tilde{\boldsymbol{\beta}}_n^{(i)}$ and $\hat{\boldsymbol{\delta}}_n^{(i)}$ are denoted by

$$\text{var}[\hat{\boldsymbol{\beta}}_n] = n^{-1}\boldsymbol{\Sigma}_{u,n}, \quad \text{var}[\tilde{\boldsymbol{\beta}}_n^{(i)}] = n^{-1}\boldsymbol{\Sigma}_{r,n}^{(i)}, \quad \text{var}[\hat{\boldsymbol{\delta}}_n^{(i)}] = n^{-1}\boldsymbol{\Sigma}_{\delta^{(i)},n}. \tag{29}$$

with corresponding estimators $n^{-1}\hat{\boldsymbol{\Sigma}}_{u,n}$, $n^{-1}\hat{\boldsymbol{\Sigma}}_{r,n}^{(i)}$, and $n^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n}$.

Extending the averaging estimator (1) to the case with multiple nested models

$$\hat{\boldsymbol{\beta}}_n^a = \hat{\boldsymbol{\beta}}_n^{(m)} + \sum_{i=1}^m (1 - \hat{w}_{n,d_n^i}^{(i)}) \hat{\boldsymbol{\delta}}_n^{(i)}, \quad \hat{w}_{n,d_n^i}^{(i)} = \frac{\text{tr}(\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n}) - 2||\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n}||}{n\, \hat{\boldsymbol{\delta}}_n^{(i)\prime} \hat{\boldsymbol{\Sigma}}_{u,n}^{-1} \hat{\boldsymbol{\delta}}_n^{(i)}} \tag{30}$$

Define the following quantities,

$$\begin{aligned} \hat{\tau}_n^{(i)} &= \hat{s}_n^{(i)} - 2\hat{\lambda}_n^{(i)}, \quad \hat{s}_n^{(i)} = \text{tr}(\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n}), \quad \hat{\lambda}_n^{(i)} = ||\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n}||, \\ \hat{q}_n^{(i)} &= \hat{\boldsymbol{\delta}}_n^{(i)\prime} \hat{\boldsymbol{\Sigma}}_{u,n}^{-1} \hat{\boldsymbol{\delta}}_n^{(i)}, \quad q_n^{(i)} = \boldsymbol{\delta}_n^{(i)\prime} \boldsymbol{\Sigma}_{u,n}^{-1} \boldsymbol{\delta}_n^{(i)}. \end{aligned} \tag{31}$$

The effective number of restrictions imposed by the i-th set of restrictions equals

$$d_n^{(i)} = \hat{s}_n^{(i)} / \hat{\lambda}_n^{(i)}, \qquad \operatorname*{plim}_{n\to\infty} d_n^{(i)} = d_i. \tag{32}$$

The following extensions of Assumption A1–A3 are made.

Assumption M1 *For $i = 1, \ldots, m$, define restricted sets $\boldsymbol{\Theta}_{r^{(i)}} = \{\boldsymbol{\theta} : \boldsymbol{r}^{(i)}(\boldsymbol{\theta}) = \boldsymbol{0}\}$. Let $\boldsymbol{\theta}_{r^{(i)}} \in \boldsymbol{\Theta}_{r^{(i)}}$. The true parameter vector $\boldsymbol{\theta}_n$ is close to the restricted parameter vectors $\boldsymbol{\theta}_{r^{(i)}}$, in the sense that $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{r^{(i)}} + n^{-1/2}\boldsymbol{h}^{(i)}$*

Assumption M2 *Let the $k \times 1$ dimensional vector $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{V})$. As $n \to \infty$,*

1. *The parameter estimates converge in distribution to*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) \Rightarrow \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{G}'\boldsymbol{z},$$
$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^{(i)} - \boldsymbol{\beta}_n) \Rightarrow \tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta} = \boldsymbol{G}' \left[\boldsymbol{z} - \boldsymbol{V}\boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{V}\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'(\boldsymbol{z} + \boldsymbol{h}_i)\right], \tag{33}$$

for $i = 1, \ldots, m$, along sequences defined in Assumption M1.

2. *The covariance matrix estimates converge in probability to*

$$\hat{\boldsymbol{\Sigma}}_{u,n} \to_p \boldsymbol{\Sigma}_u = \boldsymbol{G}'\boldsymbol{V}\boldsymbol{G},$$
$$\hat{\boldsymbol{\Sigma}}_{r^{(i)},n} \to_p \boldsymbol{\Sigma}_{r^{(i)}} = \boldsymbol{G}'\boldsymbol{L}'(\boldsymbol{I} - \boldsymbol{L}\boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'\boldsymbol{L}')\boldsymbol{L}\boldsymbol{G}, \tag{34}$$
$$\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n} \to_p \boldsymbol{\Sigma}_{\delta^{(i)}} = \boldsymbol{\Sigma}_{r^{(i-1)}} - \boldsymbol{\Sigma}_{r^{(i)}}.$$

where $\boldsymbol{\Sigma}_u$ is invertible, and $\boldsymbol{\Sigma}_{r^{(0)}} = \boldsymbol{\Sigma}_u$.

3. *The averaging weights converge in distribution to*

$$\hat{w}_n^{(i)} \Rightarrow w_{d_i,i} = \frac{\tau_i}{\hat{q}_i}. \tag{35}$$

where $\hat{q}_i = (\tilde{\boldsymbol{\beta}}_{i-1} - \tilde{\boldsymbol{\beta}}_i)'\boldsymbol{\Sigma}_u^{-1}(\tilde{\boldsymbol{\beta}}_{i-1} - \tilde{\boldsymbol{\beta}}_i)$, $\tilde{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{\beta}}$, $\tau_i = s_i - 2\lambda_i$, $s_i = \mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i})$, $\lambda_i = ||\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i}||$.

Assumption M3 *Let $\boldsymbol{\delta}_1 = \boldsymbol{G}'\boldsymbol{V}\boldsymbol{R}_1(\boldsymbol{R}_1'\boldsymbol{V}\boldsymbol{R}_1)^{-1}\boldsymbol{R}_1'\boldsymbol{h}_1$, and for $i = 2, \ldots, m$ let $\boldsymbol{\delta}_i = \boldsymbol{G}'\boldsymbol{V}\left[\boldsymbol{R}_{i-1}(\boldsymbol{R}_{i-1}'\boldsymbol{V}\boldsymbol{R}_{i-1})^{-1}\boldsymbol{R}_{i-1}'\boldsymbol{h}_{i-1} - \boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{V}\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'\boldsymbol{h}_i\right]$. Then, $c_{d_i} = \boldsymbol{\delta}_i\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\delta}_i/\mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i}) < \infty$ and, as $d_i \to \infty$, $c_{d_i} \to c_i < \infty$ for all i. Moreover, $\lim_{d_i\to\infty} \tau_i/p = a_{1,i}$, $\lim_{d_i\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i})/p = a_{2,i}$.*

Assumption M4 *For $i = 2, \ldots, m$, define $\boldsymbol{\Delta}_i = \boldsymbol{L}\boldsymbol{R}_{i-1}(\boldsymbol{R}_{i-1}'\boldsymbol{V}\boldsymbol{R}_{i-1})^{-1}\boldsymbol{R}_{i-1}' - \boldsymbol{L}\boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{V}\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'$, $\boldsymbol{\Delta}_1 = -\boldsymbol{L}\boldsymbol{R}_1(\boldsymbol{R}_1'\boldsymbol{V}\boldsymbol{R}_1)^{-1}\boldsymbol{R}_1'$, and $\boldsymbol{P}_{LG} = \boldsymbol{L}\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{V}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{L}'$. Then, $\boldsymbol{P}_{LG}\boldsymbol{\Delta}_i = \boldsymbol{\Delta}_i$.*

Assumption M1–M3 parallel Assumption A1–A3 stated before. Combining (27) with (33) ensures that a restricted estimator has zero covariance with its difference from an estimator under fewer restrictions. Assumption M4 is new. Technically, it ensures that in the (n)-asymptotic limit, the cross terms $\hat{\boldsymbol{\delta}}_{i,n}\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\delta}}_{j,n}$ vanish for $i \neq j$. This is sufficient to prove a reduction in the both the risk and the geometrical risk of the averaging estimator over the unrestricted estimator in the (d, n)-asymptotic limit. The leading case where Assumption M4 is satisfied is when $\boldsymbol{G} = \boldsymbol{I}$, i.e. when restrictions are directly imposed on parameters of interest.

The theorems below follow by applying the techniques developed to establish Theorem 1 and Theorem 2. First Lemma 3 is extended,

Lemma 4 *Suppose Assumption M1–M4 hold. To estimate* (17), *consider the following risk estimate and its* (n)*-asymptotic analogue,*

$$\hat{\rho}\left(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n\right) = p - \sum_{i=1}^{m} \left\{ 2\hat{\tau}_n^{(i)} \left[\frac{\hat{s}_n^{(i)}}{n\hat{q}_n^{(i)}} - 2\frac{n\hat{\boldsymbol{\delta}}_n^{(i)\prime}\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n}\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\delta}}_n^{(i)}}{\left(n\hat{q}_n^{(i)}\right)^2} \right] - \frac{\left(\hat{\tau}_n^{(i)}\right)^2}{n\hat{q}_n^{(i)}} \right\},$$

$$\hat{\rho}\left(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}\right) = p - \sum_{i=1}^{m} \left\{ 2\tau_i \left[\frac{s_i}{\hat{q}_i} - 2\frac{\hat{\boldsymbol{\delta}}_i\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i}\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}_i}{\hat{q}_i^2} \right] - \frac{\tau_i^2}{\hat{q}_i} \right\}. \tag{36}$$

Then, as $n \to \infty$, $\hat{\rho}\left(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n\right) \Rightarrow \hat{\rho}\left(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}\right)$, *and* $\mathrm{E}\left[\hat{\rho}(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})\right] = \rho\left(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}\right)$.

A proof is provided in Appendix A.5.

The following theorem provides the distribution of (25), the difference between the mean squared error and the unbiased estimator of the risk of the averaging estimator given in Lemma 4. We denote by $(d, n \to \infty)_{\text{seq}}$ sequential limits where first $n \to \infty$, and then $d_i \to \infty$ for all i.

Theorem 3 *Suppose that Assumption M1–M4 hold. Define* $\boldsymbol{c} = (c_1, \ldots, c_m)$. *Then, as* $(d, n \to \infty)_{\text{seq}}$,

$$D_n(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n) \Rightarrow N\left(0, \sigma^2(\boldsymbol{c})\right), \quad \sigma^2(\boldsymbol{c}) = 2 - 4\sum_{i=1}^{m}\left[\frac{a_{1,i}}{c_i + 1} - \frac{a_{2,i}}{(c_i+1)^2}\right]. \tag{37}$$

with $(c_i, a_{1,i}, a_{2,i})$ *defined in Assumption M3.*

A consistent estimator for the parameters c_i is given by the following corollary.

Corollary 2 *Let* $\tilde{\boldsymbol{\delta}}_{i,n} = \max\left(0, 1 - \hat{s}_n^{(i)} / \left[n\hat{q}_n^{(i)}\right]\right)^{1/2} \hat{\boldsymbol{\delta}}_{i,n}$ *and* $\hat{c}_{i,n} = \tilde{\boldsymbol{\delta}}_{i,n}'\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\tilde{\boldsymbol{\delta}}_{i,n}/\hat{s}_n^{(i)}$. *By Assumption M2–M3, as* $(d, n \to \infty)_{\text{seq}}$, $\hat{c}_{i,n} \to_p c_i$.

The proof follows directly from the proof for the case with a single restricted estimator presented in Appendix A.1.

We consider the following estimator for $\hat{\sigma}^2(\boldsymbol{c})$,

$$\hat{\sigma}^2(\hat{\boldsymbol{c}}) = 2 - 4\sum_{i=1}^{m}\frac{(\hat{\tau}_n^{(i)})^2}{p}\left[\frac{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n})^{-1}}{\hat{c}_{i,n}+1} - \frac{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n}\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n})}{(\hat{c}_{i,n}+1)^2\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{u,n}^{-1}\hat{\boldsymbol{\Sigma}}_{\delta^{(i)},n})^2}\right]. \tag{38}$$

We then obtain the following theorem.

Theorem 4 *Suppose that Assumption M1–M4 hold. Consider the confidence region* $C_n\left(\hat{\boldsymbol{\beta}}_n^a, \hat{b}_n\right)$ *with* $\hat{b}_n^2 = \max(0, \hat{e}_n)$, $\hat{e}_n = \hat{\rho}(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n) + p^{1/2}\hat{\sigma}(\hat{\boldsymbol{c}})\Phi^{-1}(1-\alpha)$, *where* $\hat{\rho}(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$ *as in* (36)*, and* $\hat{\sigma}(\hat{\boldsymbol{c}})$ *from* (37)*. Then,* $C_n\left(\hat{\boldsymbol{\beta}}_n^a, \hat{b}_n\right)$ *is* (d,n)*-asymptotically valid with* (d,n)*-asymptotic geometrical risk* $2\left(1-\sum_{i=1}^m \frac{a_{1,i}}{c_i+1}\right)^{1/2}$.

Appendix A.6 gives the proof of Theorem 3–4, which is a component-wise application of the proofs for Theorem 1–2 facilitated by Assumption M4.

3.6 Joint limits in the linear regression model

In this section we establish under what conditions the developed sequential limit theory remains valid under joint limits, denoted by $(d, n \to \infty)$, in the linear regression model (12). Throughout this section, we suppress the dependence of the (estimated) parameter vectors on the sample size n. We consider restrictions $\boldsymbol{R}'\tilde{\boldsymbol{\beta}} = \boldsymbol{c}$, where $\boldsymbol{R} \in \mathbb{R}^{p\times r}$ and $\text{rank}(\boldsymbol{R}) = r$.

We make the following assumptions, where $M > 0$ denotes a generic finite constant that can differ across equations. Here, *a.s.* denotes almost surely, and *a.s.n.* almost surely for n large enough (Chao et al., 2012).

Assumption LR1 *The regressors and error terms satisfy the following:*
(a) $\{\boldsymbol{x}_i\}$ *is an i.i.d. sequence with* $\mathrm{E}[\boldsymbol{x}_i\boldsymbol{x}_i'] = \boldsymbol{Q}_X$ *and* $\boldsymbol{Q}_X$ *positive definite. Moreover,* $p^{-1}\text{var}(\boldsymbol{x}_i'\boldsymbol{x}_i) \leq M$*, and* $\mathrm{E}[x_{ij}^4] \leq M < \infty$ *for all* $j = 1, \ldots, p$.
(b) Let $\lambda_1, \ldots, \lambda_p$ *be the eigenvalues of* $n^{-1}\boldsymbol{X}'\boldsymbol{X}$ *sorted in decreasing order. There exist finite positive constants* b *and* B *such that* $b \leq \lambda_p \leq \lambda_1 \leq B$ *a.s.n.*
(c) Conditional on $\boldsymbol{X}$*,* $\{\varepsilon_i\}$ *is an i.i.d. sequence with* $\mathrm{E}[\varepsilon_i|\boldsymbol{X}] = 0$*,* $\mathrm{E}[\varepsilon_i^2|\boldsymbol{X}] = \sigma^2$*,* $\mathrm{E}[\varepsilon_i^4|\boldsymbol{X}] = \mathrm{E}[\varepsilon_i^4] \leq M < \infty$.

Assumption LR2 *Let* $\boldsymbol{h} \in \mathbb{R}^p$*. The restrictions satisfy the following:*
(a) $\boldsymbol{R}'\boldsymbol{\beta} - \boldsymbol{c} = n^{-1/2}\boldsymbol{R}'\boldsymbol{h}$.
(b) $d^{-1}\sigma^{-2}\boldsymbol{h}'\boldsymbol{h} < \infty$.
(c) Define $c_{d,n} = d^{-1}\sigma^{-2}\boldsymbol{h}'\boldsymbol{R}(\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{h}$*. Then, as* $(d, n \to \infty)$*,* $c_{d,n} \to_p c$ *for some constant* $c \geq 0$.

Assumption LR3 *As* $(d,n) \to \infty$*, (a)* $\frac{d}{n} \to 0$*, and (b)* $\frac{d}{p} \to a$ *with* $a \in (0,1]$.

Assumption LR1 replaces Assumption A2. Assumption LR2 combines Assumption A1 and Assumption A3. Note that in part (c) $c_{d,n} < \infty$ *a.s.n.* by part (b) of Assumption LR1 and part (b) of Assumption LR2. When $\boldsymbol{R} = \boldsymbol{I}$, part (c) follows from Assumption LR1 and part (b) of Assumption LR2. Finally, Assumption LR3 is the rate condition needed for the sequential limit distribution to coincide with the joint limit distribution. Since $d \leq p$, the fact that $d \to \infty$ implies $p \to \infty$.

Part (b) rules out the case where the number of restrictions is negligible compared to the number of parameters, in which case the effect from averaging is negligible.

Estimators and averaging weights The unrestricted estimator of $\boldsymbol{\beta}$ in (12) is the least squares estimator, which is also used to estimate the noise level σ^2, i.e.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \quad \hat{\boldsymbol{\Sigma}}_u = \hat{\sigma}^2 \left(n^{-1}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}, \quad \hat{\sigma}^2 = (n-p)^{-1}\boldsymbol{\varepsilon}'\boldsymbol{M}_X\boldsymbol{\varepsilon}. \tag{39}$$

In Appendix A.7, we show that $\hat{\sigma}^2$ is consistent for σ^2. The results do not require $\hat{\boldsymbol{\Sigma}}_u$ to converge in probability. Imposing $\boldsymbol{R}'\tilde{\boldsymbol{\beta}} = \boldsymbol{c}$, leads to the restricted estimator

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{c}), \\ \hat{\boldsymbol{\Sigma}}_r &= \hat{\boldsymbol{\Sigma}}_u - \hat{\boldsymbol{\Sigma}}_u\boldsymbol{R}(\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u\boldsymbol{R})^{-1}\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u. \end{aligned} \tag{40}$$

The difference $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Sigma}}_\delta = \hat{\boldsymbol{\Sigma}}_u\boldsymbol{R}(\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u\boldsymbol{R})^{-1}\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u$. The effective number of restrictions in this set-up is equal to the number of restrictions, as

$$d = \frac{\operatorname{tr}(\boldsymbol{R}(\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u\boldsymbol{R})^{-1}\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_u)}{||\boldsymbol{R}(\boldsymbol{R}'\hat{\boldsymbol{\Sigma}}_u\boldsymbol{R})^{-1}\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_u||} = r. \tag{41}$$

The averaging estimator is as in (1), i.e. $\hat{\boldsymbol{\beta}}^a = \hat{\omega}\tilde{\boldsymbol{\beta}} + (1-\hat{\omega})\hat{\boldsymbol{\beta}}$. Using the expressions for the restricted and unrestricted estimator above, the averaging weights are a function of the inverse F-statistic associated with the imposed restrictions,

$$\hat{w} = \frac{r-2}{r \cdot \hat{F}}, \qquad \hat{F} = \frac{\hat{\boldsymbol{\delta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\delta}}/r}{\hat{\sigma}^2}.$$

Appendix A.7 shows that $\operatorname{plim}_{(d,n\to\infty)} \hat{F} = c + 1$, with c as in Assumption LR2. This is also found by Calhoun (2011) and Anatolyev (2012), who consider testing many restrictions in the linear regression model.

Using the above expressions and Lemma 3, we find $\hat{\rho}(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = p - (r-2)\hat{\omega}$, and subsequently

$$D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = p^{-1/2}\left[n(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})\hat{\boldsymbol{\Sigma}}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta}) - p + (r-2)\hat{\omega}\right]. \tag{42}$$

The following lemma that states that the distribution of $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$ as given in Theorem 1 holds under joint limits.

Lemma 5 *Under Assumption LR1–LR3, as* $(d, n \to \infty)$,

$$D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) \Rightarrow N(0, \sigma^2(c)), \quad \sigma^2(c) = 2 - 4a \left[\frac{1}{c+1} - \frac{1}{(c+1)^2} \right],$$

with c *defined in Assumption LR2 and* a *in Assumption LR3.*

The proof is provided in Appendix A.7. Key underlying results are Theorem 2 from Phillips and Moon (1999) and Lemma A2 from Chao et al. (2012).

4 Numerical analysis

4.1 Implementation

The geometrical argument in Section 2.3 highlights the importance of the choice for the restricted estimator when using the confidence regions for hypothesis testing. To increase power against $H_0 : \boldsymbol{\beta} = \mathbf{0}$, we need to control the sign and magnitude of the restricted estimator. The most convenient way to control the sign of the restricted estimates is by using direct restrictions that set the signs in accordance with prior knowledge and/or economic theory. We propose to set the restricted estimator as

$$\tilde{\boldsymbol{\beta}}_n = \boldsymbol{L}\boldsymbol{c} \cdot \frac{m}{p^{1/4} n^{1/2}}, \qquad \boldsymbol{L}\boldsymbol{L}' = \hat{\boldsymbol{\Sigma}}_{u,n} \tag{43}$$

where $\boldsymbol{c}$ is a vector with elements in $\{-1, 1\}$ that ensure that $\tilde{\boldsymbol{\beta}}_n$ has the expected coefficient signs. To obtain $\boldsymbol{L}$, we use a Cholesky decomposition. The scaling of the estimator is such that the corresponding Wald statistic is local-to-zero, which is reasonable in empirical settings. The parameter m determines how far away from zero the restrictions are. We investigate the choice of m below.

A second practical consideration is the following. When the difference between the restricted and unrestricted estimator is large, the weight on the restricted estimator goes to zero, and the averaging estimator equals the unrestricted estimator. To get (n)-asymptotically correct coverage, the critical values should be equal to that of the $\chi^2(p)$-distribution, denoted by b_χ^2. However, we approximate b_χ^2 by $b_N = p + \sqrt{2p}\Phi^{-1}(\alpha)$. Although valid for large p, for practically relevant values of p, this will lead to undercoverage. Following the suggestion of Stein (1981), this can be corrected by choosing a higher value for α to achieve the desired nominal coverage rate. Setting $b_N = b_\chi^2$, and solving for α, we find $\alpha = \Phi((b_\chi^2 - p)/\sqrt{2p})$, with $\Phi(\cdot)$ the standard normal CDF. These adjusted levels are used throughout. This prevents that power differences result from incorrect size of the test.

4.2 Low- and high-dimensional models

We consider the case where we are interested in a parameter vector $\boldsymbol{\beta}$, and we need to include a large set of control variables to ensure that our estimates for $\boldsymbol{\beta}$ are unbiased. An application to instrumental variables regression is given in the appendix. The data generating process is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}),$$
$$\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{z}_i \end{pmatrix} \sim N\left[\boldsymbol{0}, \begin{pmatrix} \boldsymbol{I}_p & \rho\boldsymbol{I}_p & \boldsymbol{O}_{p\times k-2p} \\ \rho\boldsymbol{I}_p & \boldsymbol{I}_p & \boldsymbol{O}_{p\times k-2p} \\ \boldsymbol{O}_{k-2p\times p} & \boldsymbol{O}_{k-2p\times p} & \boldsymbol{I}_{k-2p} \end{pmatrix}\right]. \tag{44}$$

The parameter vector $\boldsymbol{\beta}$ is of interest, while the parameters $\boldsymbol{\gamma}$ are nuisance parameters. The number of parameters of interest is $p = \{6, 12, 24\}$ and the number of nuisance parameters $k-p = 24$. The sample size equals $n = \{150, 500\}$. The correlation ρ is varied as $\rho = \{0.2, 0.9\}$. For $j = 1, \ldots, p$,

$$\beta_j = \left[\frac{c_\beta}{np^{1/2}(1-\rho^2)}\right]^{1/2} \frac{j^{-1}}{(\sum_{i=1}^{p} i^{-2})^{1/2}}, \quad c_\beta = \{-12, \ldots, 12\} \tag{45}$$
$$\gamma_j = \left[\frac{c_\gamma}{n(k-p)^{1/2}(1-\rho^2)}\right]^{1/2} \frac{j^{-1}}{\left(\sum_{i=1}^{k-p} i^{-2}\right)^{1/2}}, \quad c_\gamma = 10.$$

The unrestricted estimator is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{M}_Z\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_Z\boldsymbol{y}$. We consider the indirectly restricted estimator $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ as well as the direct restricted vector (43). For the latter, we set $c_i = 1$ for $i = 1, \ldots, p$. We vary $m = \{-3, 0, 3\}$. Note that when $m > 0$ and $c_\beta > 0$, the restricted vector has the correct sign, as well as when $m < 0$ and $c_\beta < 0$. The choice for $m = \pm 3$ is motivated in Appendix B.2, where we find that this choice yields the highest power. All results are averaged over 100,000 draws of the set $\{\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\varepsilon}\}$.

In Table 1, we show the coverage rate for the proposed confidence regions. Throughout the coverage rate is close to the nominal level of 0.95. When $n = 150$, choosing the fixed restricted vector with $m = \pm 3$ yields in slight undercoverage that largely disappears when increasing the sample size to $n = 500$. The correlation parameter ρ only affects the coverage under indirect restrictions, although this effect disappears when the sample size increases to $n = 500$.

Figure 2 shows the power compared to the power of the standard F-test on the parameters of interest. In the left upper panel, we consider the case with $p = 12$ variables of interest, and we have weak correlation between the regressors

Table 1: Linear regression model: coverage rate.

n		$p=6$		$p=12$		$p=24$	
		$\rho=0.2$	$\rho=0.9$	$\rho=0.2$	$\rho=0.9$	$\rho=0.2$	$\rho=0.9$
150	$m=-3$	0.942	0.943	0.941	0.941	0.938	0.938
	$m=0$	0.964	0.964	0.961	0.960	0.951	0.951
	$m=3$	0.942	0.943	0.940	0.942	0.938	0.939
	OLS	0.946	0.946	0.942	0.947	0.931	0.948
500	$m=-3$	0.949	0.948	0.948	0.948	0.948	0.946
	$m=0$	0.966	0.965	0.963	0.964	0.959	0.959
	$m=3$	0.949	0.948	0.948	0.949	0.949	0.948
	OLS	0.950	0.949	0.949	0.949	0.946	0.951

Note: coverage rate under (44) at $\boldsymbol{\beta}=\mathbf{0}$, sample size $n=\{150,500\}$, number of parameters of interest $p=\{6,12,24\}$, and correlation between nuisance variables and variables of interest $\rho=\{0.2,0.9\}$. Coverage rates are reported for averaging with (43) choosing $m=\{-3,0,3\}$, and averaging with the OLS estimator that ignores the control variables. Nominal coverage equals 0.95.

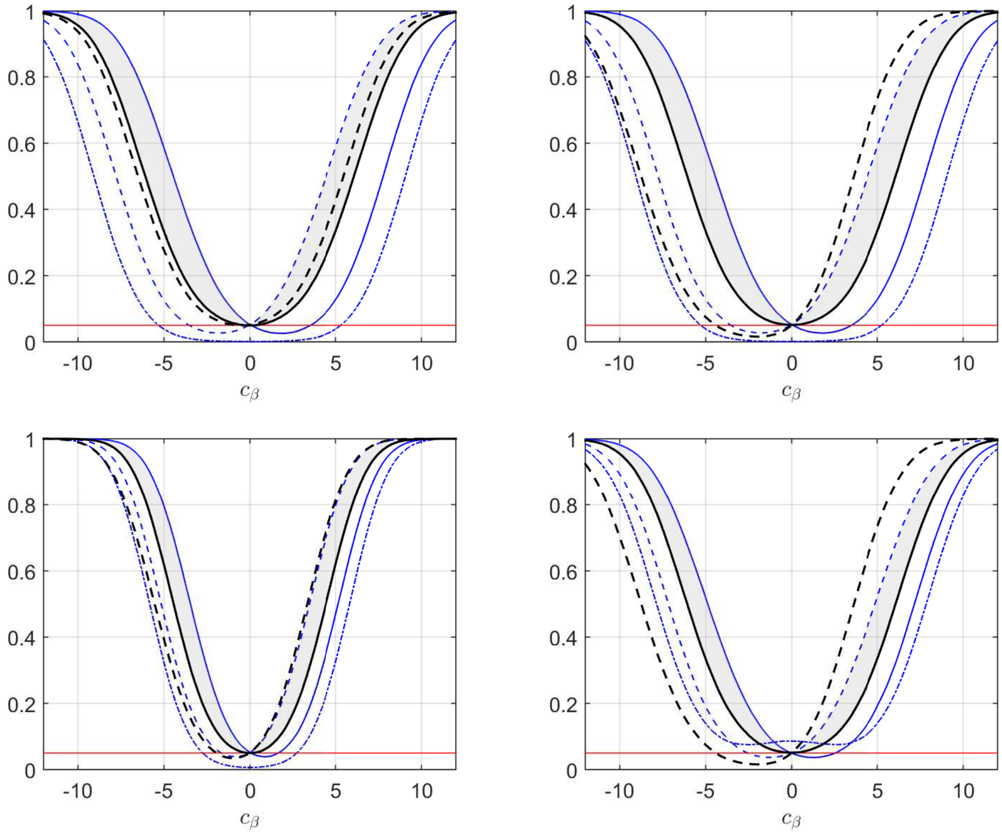
in $\boldsymbol{X}$ and $\boldsymbol{Z}$ ($\rho=0.2$). Power under the standard F-test is depicted by the black solid line. The restricted estimator (43) with $m=\{-3,0,3\}$ is displayed by the blue solid, dash-dotted and dashed line. We see that for power improvements, visualized by the gray area, it is essential to get the sign of the coefficient vector right. Setting $m=0$, a common choice when the interest is in risk reduction, substantially lowers power. When we use the indirectly restricted estimator, we see a power improvement when $c_\beta>0$, and a slight power loss when $c_\beta<0$. The reason is that because of the positive correlation, omitting the control variables leads tp an upward bias in the coefficients. When $c_\beta>0$ this results in a power increase, but the upward bias similarly reduces power when $c_\beta<0$.

In the right upper panel, we increase correlation between the regressors to $\rho=0.9$. This does not affect the averaging estimator when a fixed restricted estimator is used. However, when the restricted least squares estimator is used, a larger power gain is observed when $c_\beta>0$ and a larger loss when $c_\beta<0$.

In the left lower panel, we decrease the number of parameters of interest to $p=6$. The blue lines are again the power curves using the fixed restricted vector. Decreasing the number of parameters of interest also decreases both the power gains when the correct sign is used, and losses when the wrong sign is used. The positive correlation between the regressors again makes using the restricted least squares estimator useful only when $c_\beta>0$.

Finally, in the right lower panel, we consider the same setting as in the upper

Figure 2: Linear regression model: power.

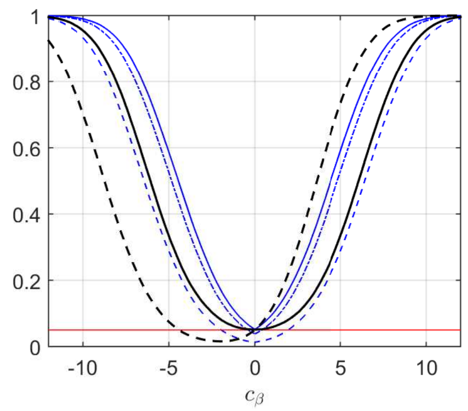


Note: the figure shows power against $H_0 : \boldsymbol{\beta} = \mathbf{0}$ at a sample size of $n = 500$. The black solid line corresponds to the usual F-test, the black dashed line to averaging with the least squares estimator that ignores the control variables. The blue lines correspond to averaging with the restricted estimator (43) with $c_i = 1$ for $i = 1, \ldots, p$, $m = -3$ (solid), $m = 0$ (dash-dotted), $m = 3$ (dashed). In the left upper panel, the correlation $\rho = 0.2$ and there are $p = 12$ parameters of interest. In the right upper panel, $\rho = 0.9$. The left lower panel is the same as the right upper panel, but now $p = 6$. The right lower panel again has $p = 12$, but the blue lines correspond to the multiple restricted estimator where the first set of restrictions sets $\beta_7, \ldots \beta_{12}$ equal to (43), and the second set of restrictions sets $\beta_1, \ldots, \beta_6$ equal to (43).

right panel, but now we use the multiple averaging estimator from Section 3.5. We choose a directly restricted estimator that sets only the final $p/2$ parameters equal to according to (43), but leaves the others unrestricted, as well as one that sets all parameters according to (43). We see that the both power gains and losses are smaller compared to using a single restricted estimator.

Alternative confidence regions In Figure 3, we compare the power under the critical values derived here to the confidence regions by Casella and Hwang (1983) and Samworth (2005). All average with the fixed restricted vector with $m = 3$ when $c_\beta > 0$ and $m = -3$ when $c_\beta < 0$. That is, we assume that the correct sign of

Figure 3: Linear regression model: comparison to alternative procedures.



Note: the figure shows power against $H_0 : \boldsymbol{\beta} = \mathbf{0}$ at a sample size of $n = 500$. In the upper panels, the black solid line is the power from the usual F-test, the black dashed line when averaging with the restricted least squares estimator. The solid blue line corresponds to the restricted estimator (43) with $m = 3$ and the correct sign. The dash-dotted blue line is power under the procedure by Casella and Hwang (1983), the dashed blue line using Samworth (2005). Both panels have $\rho = 0.9$. The left panel has $p = 12$, the right panel $p = 24$.

the coefficients is chosen. The construction of the confidence regions is discussed in Appendix B.1. The solid blue line corresponds to the regions developed in this paper, the dash-dotted blue line by those of Casella and Hwang (1983), and the dashed blue line by Samworth (2005). The black solid line is again the power from the standard F-test, and the black dashed line from using the indirectly restricted estimator. We find that the confidence regions developed in this paper offer higher power, especially when the number of parameters is large. From the numerical results in Casella and Hwang (1983) and Samworth (2005) this can be expected, as these regions generally lead to substantial overcoverage when p is large.

Skewed, heavy-tailed distributions and joint limits In Section 3.6, we studied the asymptotic theory under joint limits in the number of restrictions and the sample size. To test this theory empirically, we consider the same model as above with $\boldsymbol{\beta} = \mathbf{0}$. We only consider averaging with the fixed restricted vector (43), so that $p = r = d$. We set $m = 3$. We now consider regressors $\boldsymbol{X} = \tilde{\boldsymbol{X}}\boldsymbol{\Sigma}^{1/2}$. Here the covariance matrix $\boldsymbol{\Sigma}$ is as before, but the elements from $\tilde{\boldsymbol{X}}$ are generated by squaring independent $t(10)$ random variables and standardizing the columns. The elements of $\boldsymbol{\varepsilon}$ are also standardized squared $t(10)$ random variables. The number of degrees of freedom is chosen according to the requirements in Assumption LR1. Squaring induces skewness in both the regressors and the errors. We consider $p = \{6, 12, 24\}$ and $n = \{150, 500, 1500, 5000\}$. In this way, n grows faster than

Table 2: Linear regression model: coverage rate sensitivity

		$\rho = 0.2$			$\rho = 0.9$		
$\{\boldsymbol{X}, \boldsymbol{\varepsilon}\}$	n	$p = 6$	$p = 12$	$p = 24$	$p = 6$	$p = 12$	$p = 24$
$t^2(10)$	150	0.916	0.910	0.906	0.926	0.920	0.914
	500	0.928	0.921	0.923	0.937	0.933	0.930
	1500	0.935	0.931	0.932	0.943	0.940	0.938
	5000	0.940	0.937	0.940	0.947	0.946	0.944
N	150	0.942	0.940	0.938	0.943	0.942	0.939
	500	0.949	0.948	0.949	0.948	0.949	0.948
	1500	0.950	0.950	0.949	0.950	0.948	0.949
	5000	0.950	0.949	0.950	0.951	0.949	0.950

Note: coverage rate under (44) at $\boldsymbol{\beta} = \mathbf{0}$, sample size $n = \{150, 500, 1500, 5000\}$, number of parameters of interest $p = \{6, 12, 24\}$, and correlation between nuisance variables and variables of interest $\rho = \{0.2, 0.9\}$. Coverage rates are reported for averaging with (43) choosing $m = 3$. Regressors and errors are standardized squared $t(10)$ random variables (upper panel), or normal random variables (lower panel). Nominal coverage equals 0.95.

p, in line with Assumption LR3. For comparison, we also show the results for normally distributed regressors and errors.

The results are displayed in Table 2. For small n and large p, coverage drops slightly as a result of changing the distribution of the regressors and errors. Nevertheless, by moving diagonally across the table, we see that the coverage under skewed, heavy-tailed regressors and errors increases towards the nominal coverage as n increases faster than p.

4.3 Empirical illustration

As an illustration we consider the growth regression comparison of Magnus et al. (2010). Following the sets of auxiliary regressors in their Models 1 and 2, we divide twelve available regressors in three groups. The first group contains variables that approximate the Solow determinants: the log of GDP per capita in 1960 (abbreviation: GDP60, expected sign: –), the equipment investment share of GDP between 1960-1985 (EQUIPINV, +), total gross enrollment in primary school in 1960 (SCHOOL60, +), life expectancy at age zero in 1960 (LIFE60, +).

The second group is a set of variables that aims to capture the fundamentals of different countries: a rule of law index (LAW, +), the fraction of tropical area (TROPICS, –), ethnolinguistic fractionalization index (AVELF, –), and the fraction of Confucian population (CONFUCIAN, +).

Table 3: Empirical illustration: averaging estimates and test statistics

	Unrestricted	Restricted	$\hat{w}$	Average
GDP60	-0.0173 (0.0033)	-0.0071	0.1746	-0.0155
EQUIPINV	0.1324 (0.0579)	0.0996		0.1267
SCHOOL60	0.0144 (0.0096)	0.0215		0.0156
LIFE60	0.0006 (0.0004)	0.0002		0.0005
W	42.5347			36.5996
W_c	9.4877			7.7037
LAW	0.0200 (0.0068)	0.0144	0.2614	0.0191
TROPICS	-0.0055 (0.0041)	-0.0086		-0.0063
AVELF	-0.0040 (0.0060)	-0.0120		-0.0061
CONFUC	0.0538 (0.0169)	0.0221		0.0455
W	29.0063			24.6519
W_c	9.4877			7.3616
MINING	-0.0090 (0.0192)	-0.0407	0.1518	-0.0138
PRIGHTS	-0.0013 (0.0012)	0.0021		-0.0007
MALFAL	-0.0104 (0.0052)	-0.0121		-0.0107
DPOP	0.3352 (0.2542)	0.5212		0.3635
W	8.2393			8.0247
W_c	9.4877			7.8654

Note: the table reports the estimated coefficients and standard errors using the unrestricted estimator, the restricted estimator (43), the averaging weight ($\hat{w}$), and the averaging estimator. For each variable group, we report the Wald statistic (W) with the 5% critical value (W_c), and the corresponding analogues based on the averaging estimator.

The third group is a set of additional control variables whose relevance is unclear: population growth between 1960 and 1990 (DPOP, +), a political rights index (PRIGHTS, +), malaria prevalence in 1966 (MALARIA, –), and the fraction of GDP produced in mining (MINING, –).

For each group we construct a restricted estimator using (43) and the signs as indicated above. Following the simulation results from Section 4, we set $m = 3$. We then calculate the averaging estimator (1) together with the critical values (8)–(11) to determine whether each of the three groups is jointly significant.

The coefficient estimates are provided in Table 3. We report the unrestricted estimator, with the corresponding standard errors, the restricted estimator, and the averaging estimator. Except for the variable PRIGHTS, the unrestricted and restricted estimates agree on the sign of the coefficients. The weight w assigned to the restricted model is substantial for all groups. This implies that the restricted estimator (43) is a reasonable choice.

In terms of significance, we see that the first two groups are highly jointly significant according to a standard Wald test ($W \gg W_c$) at a 5% level. The test statistic based on the averaging estimator is slightly smaller that the standard test statistic, but this also holds for the relevant critical values. For the third group of variables, we see that the standard Wald test is insignificant at a 5% level, while test based on the recentered confidence region exceeds the critical value.

5 Conclusion

We construct confidence regions centered at averaging estimators. The regions yield correct coverage under sequential limits in the number of observations (n) and the number of effective restrictions (d). Specializing to the linear regression model, we find that the limit distribution is valid under joint limits in d and n as long as $d/n \to 0$. When using the confidence regions for hypothesis testing, the model restrictions play a crucial role. Power gains are observed when using a fixed restricted vector where the sign of the coefficients corresponds to that of the true parameter vector. In this case, the confidence regions can be used to increase power over standard F-tests.

References

Anatolyev, S. (2012). Inference in regression models with many regressors. *Journal of Econometrics*, 170(2):368–382.

Beran, R. (1995). Stein confidence sets and the bootstrap. *Statistica Sinica*, 5:109–127.

Calhoun, G. (2011). Hypothesis testing in linear regression when k/n is large. *Journal of Econometrics*, 165(2):163–174.

Casella, G. and Hwang, J. G. (2012). Shrinkage confidence procedures. *Statistical Science*, 27(1):51–60.

Casella, G. and Hwang, J. T. (1982). Limit expressions for the risk of James-Stein estimators. *Canadian Journal of Statistics*, 10(4):305–309.

Casella, G. and Hwang, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *Journal of the American Statistical Association*, 78(383):688–698.

Chao, J. C., Swanson, N. R., Hausman, J. A., Newey, W. K., and Woutersen, T.

(2012). Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory*, 28(1):42–86.

Claeskens, G. and Hjort, N. L. (2008). Model selection and model averaging. *Cambridge Books*.

DiTraglia, F. J. (2016). Using invalid instruments on purpose: focused moment selection and averaging for GMM. *Journal of Econometrics*, 195(2):187–208.

Efron, B. (2006). Minimum volume confidence regions for a multivariate normal mean vector. *Journal of the Royal Statistical Society: Series B*, 68(4):655–670.

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

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

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

Hansen, B. E. and Racine, J. S. (2012). Jackknife model averaging. *Journal of Econometrics*, 167(1):38–46.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

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

Huber, P. J. et al. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1(5):799–821.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 361–379.

Leeb, H. and Kabaila, P. (2017). Admissibility of the usual confidence set for the mean of a univariate or bivariate normal population: the unknown variance case. *Journal of the Royal Statistical Society: Series B*, 79(3):801–813.

Liu, C.-A. (2015). Distribution theory of the least squares averaging estimator. *Journal of Econometrics*, 186(1):142–159.

Magnus, J. R., Powell, O., and Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154(2):139–153.

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

Samworth, R. (2005). Small confidence sets for the mean of a spherically symmetric distribution. *Journal of the Royal Statistical Society: Series B*, 67(3):343–361.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on mathematical statistics and probability*, volume 1, pages 197–206.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9(6):1135–1151.

Stock, J. and Yogo, M. (2005). Asymptotic distributions of instrumental variables statistics with many instruments. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter 6. Cambridge University Press.

Ullah, A. (2004). *Finite sample econometrics*. Oxford University Press.

Zhang, X. and Liu, C.-A. (2019). Inference after model averaging in linear regression models. *Econometric Theory*, 35(4):816–841.

Appendix A Mathematical details

A.1 Preliminary lemma's

Lemma A.1 *Suppose Assumption A2 and Assumption A3 hold. Then,*

$$\operatorname*{plim}_{d\to\infty} \frac{\hat{q}}{\text{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)} = c + 1. \tag{A.1}$$

Proof: By Assumption A3

$$c_d = \frac{q}{\text{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)} < \infty. \tag{A.2}$$

Under Assumption A2, using standard results on quadratic forms in normally distributed random vectors,

$$\begin{aligned} \text{E}\left[\frac{\hat{q}}{\text{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}\right] &= c_d + 1, \\ \text{Var}\left[\frac{\hat{q}}{\text{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}\right] &= 2\frac{\text{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}{\text{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)^2} + 4\frac{\boldsymbol{\delta}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\delta}}{\text{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)^2} \leq \frac{2}{d} + \frac{4c}{d}. \end{aligned} \tag{A.3}$$

When $d \to \infty$, Chebyshev's inequality implies (A.1). ■

Lemma A.2 (Special case of Chao et al. (2012), Lemma A2) *Suppose that the following conditions hold a.s.*

(i) $\boldsymbol{P}$ *is a symmetric idempotent matrix with* $[\boldsymbol{P}]_{ii} < 1$*,*

(ii) Conditional on $\boldsymbol{X}$, $\{\varepsilon_i\}$ *is an i.i.d. sequence,*
(iii) $\mathrm{E}[\varepsilon_i|\boldsymbol{X}] = 0$, $\mathrm{E}[\varepsilon_i^2|\boldsymbol{X}] = E[\varepsilon_i^2] = \sigma^2$, *and* $\mathrm{E}[\varepsilon_i^4|\boldsymbol{X}] \leq M$,
(iv) $\mathrm{rk}(\boldsymbol{P}) \to \infty$ *as* $n \to \infty$.
Then for

$$V_n = \frac{2\sigma^4}{\mathrm{rk}(\boldsymbol{P})} \sum_{i \neq j} P_{ij}^2, \tag{A.4}$$

with $V_n > M$ *a.s.n., it follows that*

$$V_n^{-1/2} \frac{1}{\sqrt{\mathrm{rk}(\boldsymbol{P})}} \sum_{i \neq j} \varepsilon_i \varepsilon_j P_{ij} \Rightarrow N(0,1), \quad a.s. \tag{A.5}$$

A.2 Proof of Lemma 3

The first part follows from Assumption A2 and the continuous mapping theorem. What remains to be shown is that $\mathrm{E}\left[\hat{\rho}\left(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}\right)\right] = \rho\left(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}\right)$. Note first that by Assumption A2 the following weak convergence holds for $\hat{\delta}_n$ defined in (2)

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n) \Rightarrow_n \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} &= \boldsymbol{G}'\boldsymbol{V}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{V}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{z}, \\ \boldsymbol{\delta} &= \boldsymbol{G}'\boldsymbol{V}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{V}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{h}. \end{aligned} \tag{A.6}$$

The (n)-asymptotic representation of the averaging estimator (1) is then given by

$$\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta} = \tilde{\boldsymbol{\beta}} - (\boldsymbol{\beta} - \boldsymbol{\delta}) + (1 - \hat{w}_d)\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}, \tag{A.7}$$

By Assumption A2, we have

$$\tilde{\boldsymbol{\beta}} - (\boldsymbol{\beta} - \boldsymbol{\delta}) = \boldsymbol{G}'\boldsymbol{L}'\boldsymbol{M}_{LR}\boldsymbol{u}, \quad \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} = \boldsymbol{G}'\boldsymbol{L}'\boldsymbol{P}_{LR}\boldsymbol{u}, \tag{A.8}$$

where $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{P}_{LR} = \boldsymbol{L}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{V}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{L}'$, and $\boldsymbol{M}_{LR} = \boldsymbol{I} - \boldsymbol{P}_{LR}$.

From (A.8), it is clear that the covariance between $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ is zero, and since they are normal, this implies independence. As the weight $\hat{w}_d$ only depends on $\hat{\boldsymbol{\delta}}$, the (n)-asymptotic risk of the averaging estimator consists of two terms

$$\rho(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = \mathrm{E}\left[(\tilde{\boldsymbol{\beta}} - (\boldsymbol{\beta} - \boldsymbol{\delta}))'\boldsymbol{\Sigma}_u^{-1}(\tilde{\boldsymbol{\beta}} - (\boldsymbol{\beta} - \boldsymbol{\delta}))\right] + \rho(\hat{\boldsymbol{\delta}}^{JS}, \boldsymbol{\delta}), \tag{A.9}$$

where

$$\rho(\hat{\boldsymbol{\delta}}^{JS}, \boldsymbol{\delta}) = \mathrm{E}\left[(\hat{\boldsymbol{\delta}}^{JS} - \boldsymbol{\delta})'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\delta}}^{JS} - \boldsymbol{\delta})\right], \quad \hat{\boldsymbol{\delta}}^{JS} = (1 - \hat{w}_d)\hat{\boldsymbol{\delta}}. \tag{A.10}$$

To apply Stein's lemma to (A.10), we introduce the notation

$$\hat{\boldsymbol{\delta}} = \boldsymbol{\Sigma}_\delta^{1/2}\boldsymbol{m}, \quad \boldsymbol{m} \sim N(\boldsymbol{\mu}, \boldsymbol{I}_p), \quad \boldsymbol{\mu} = \boldsymbol{\Sigma}_\delta^{-1/2}\boldsymbol{\delta}. \tag{A.11}$$

The following quantities are helpful in the derivations below

$$\begin{aligned} \boldsymbol{S} &= \boldsymbol{\Sigma}_\delta^{1/2}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta^{1/2}, \quad \hat{w}_d = \frac{\tau}{\hat{q}} = \frac{\tau}{\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m}}, \\ \boldsymbol{g}(\boldsymbol{m}) &= -\hat{w}_d\hat{\boldsymbol{\delta}} = -\frac{\tau}{\hat{q}}\hat{\boldsymbol{\delta}} = -\frac{\tau}{\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m}}\boldsymbol{m}, \quad \boldsymbol{h}(\boldsymbol{m}) = -\frac{\tau}{\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m}}\boldsymbol{S}\boldsymbol{m}. \end{aligned} \tag{A.12}$$

In terms of the quantities in (A.12), the risk (A.10) is

$$\begin{aligned} \rho(\hat{\boldsymbol{\delta}}^{JS}, \boldsymbol{\delta}) &= \mathrm{E}\left[(\boldsymbol{m} + \boldsymbol{g}(\boldsymbol{m}) - \boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m} + \boldsymbol{g}(\boldsymbol{m}) - \boldsymbol{\mu})\right] \\ &= \mathrm{E}\left[(\boldsymbol{m} - \boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m} - \boldsymbol{\mu}) + 2\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m} - \boldsymbol{\mu}) + \boldsymbol{h}(\boldsymbol{m})'\boldsymbol{S}^{-1}\boldsymbol{h}(\boldsymbol{m})\right] \\ &= \mathrm{tr}(\boldsymbol{S}) + 2\mathrm{E}[\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})] + \mathrm{E}[\boldsymbol{h}(\boldsymbol{m})'\boldsymbol{S}^{-1}\boldsymbol{h}(\boldsymbol{m})], \end{aligned} \tag{A.13}$$

where the second term in the last line is obtained by applying Stein's lemma to the second term on the second line.

From (A.12),

$$\frac{\partial h_i(\boldsymbol{m})}{\partial \boldsymbol{m}_k} = -\tau\left[\frac{S_{ik}}{\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m}} - 2\frac{\sum_{l,n} S_{il}m_lS_{km}m_n}{(\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m})^2}\right], \tag{A.14}$$

such that

$$\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m}) = -\tau\left[\frac{\mathrm{tr}(\boldsymbol{S})}{\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m}} - 2\frac{\boldsymbol{m}'\boldsymbol{S}^2\boldsymbol{m}}{(\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m})^2}\right]. \tag{A.15}$$

The risk of the averaging estimator is then given by

$$\rho(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = \mathrm{tr}(\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_u^{-1} + \boldsymbol{S}) - 2\tau\mathrm{E}\left[\frac{\mathrm{tr}(\boldsymbol{S})}{\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m}} - 2\frac{\boldsymbol{m}'\boldsymbol{S}^2\boldsymbol{m}}{(\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m})^2}\right] + \tau^2\mathrm{E}\left[\frac{1}{\boldsymbol{m}'\boldsymbol{S}\boldsymbol{m}}\right]. \tag{A.16}$$

Using the definitions in (A.12), yields Lemma 3.

A.3 Proof of Theorem 1

By Assumption A2 and with $\hat{\rho}(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$ given by (24)

$$D_n(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n) \Rightarrow_n D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = p^{-\frac{1}{2}}\left\{(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta}) - \hat{\rho}\left(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}\right)\right\}. \tag{A.17}$$

It is immediately clear that $\mathrm{E}[D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})] = 0$, since $\hat{\rho}(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$ is an unbiased estimator of the (n)-asymptotic risk. For the variance, first use (A.9) and (A.13) to write

$$D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = A_{rr} + 2A_{r\delta} + A_{\delta\delta}, \tag{A.18}$$

where

$$\begin{aligned}
A_{rr} &= p^{-\frac{1}{2}} \left[(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])' \boldsymbol{\Sigma}_u^{-1} (\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]) - \mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_r) \right], \\
A_{r\delta} &= p^{-\frac{1}{2}} (\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])' \boldsymbol{\Sigma}_u^{-1} ((1-\hat{w}_d)\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \\
&= p^{-\frac{1}{2}} \left[(1-\hat{w}_d)(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])' \boldsymbol{\Sigma}_u^{-1} (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - \hat{w}_d (\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])' \boldsymbol{\Sigma}_u^{-1} \boldsymbol{\delta} \right], \qquad \text{(A.19)} \\
A_{\delta\delta} &= p^{-\frac{1}{2}} \left[(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})' \boldsymbol{\Sigma}_u^{-1} (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - s - 2\hat{w}_d \left(\hat{\boldsymbol{\delta}}' \boldsymbol{\Sigma}_u^{-1} (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - s + 2 \frac{\hat{\boldsymbol{\delta}}' \boldsymbol{\Sigma}_u^{-1} \boldsymbol{\Sigma}_\delta \boldsymbol{\Sigma}_u^{-1} \hat{\boldsymbol{\delta}}}{\hat{q}} \right) \right],
\end{aligned}$$

and $\hat{w}_d = \frac{\tau}{\hat{q}}$, $\tau = s - 2\lambda$, $s = \mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)$, and $\lambda = ||\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta||$.

Under Assumption A2 $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ are independent and asymptotically normal, and hence, $A_{rr}, A_{r\delta}, A_{\delta\delta}$ have zero covariance. It is therefore sufficient to determine the variance of the individual terms. Since each of the terms in (A.19) has expectation zero, we need to calculate $\mathrm{E}[A_{rr}^2]$, $\mathrm{E}[A_{r\delta}^2]$, $\mathrm{E}[A_{\delta\delta}^2]$.

The variance of A_{rr} follows from results on quadratic forms in normal vectors.

$$\mathrm{E}[A_{rr}^2] = 2p^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_r). \tag{A.20}$$

To calculate the variance of $A_{r\delta}$, define the matrix $\boldsymbol{A} = \boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_u^{-1}$. Then,

$$\begin{aligned}
\mathrm{E}(A_{r\delta}^2) &= p^{-1}\mathrm{E}\left\{ \left[(1-\hat{\omega})\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right]' \boldsymbol{\Sigma}_u^{-1} (\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])'(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]) \boldsymbol{\Sigma}_u^{-1} \left[(1-\hat{\omega})\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right] \right\} \\
&= p^{-1}\mathrm{E}\left\{ \left[(1-\hat{w}_d)\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right]' \boldsymbol{A} \left[(1-\hat{w}_d)\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right] \right\} \qquad \text{(A.21)} \\
&= p^{-1}\mathrm{E}\left[(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\boldsymbol{A}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right] - 2p^{-1}\mathrm{E}\left[\hat{w}_d\hat{\boldsymbol{\delta}}'\boldsymbol{A}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right] + p^{-1}\mathrm{E}\left[\hat{w}_d^2\hat{\boldsymbol{\delta}}'\boldsymbol{A}\hat{\boldsymbol{\delta}}\right] \\
&= \frac{\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_\delta)}{p} - \frac{2\tau}{p}\mathrm{E}\left[\frac{\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_\delta)}{\hat{q}} - 2\frac{\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{A}\hat{\boldsymbol{\delta}}}{\hat{q}^2}\right] + \frac{\tau^2}{p}\mathrm{E}\left[\frac{\hat{\boldsymbol{\delta}}\boldsymbol{A}\hat{\boldsymbol{\delta}}}{\hat{q}^2}\right],
\end{aligned}$$

where we applied Stein's lemma to the second term on the second to last line.

Finally, for the variance of $A_{\delta\delta}$, we use definitions (A.11) and (A.12) to write

$$\begin{aligned}\mathrm{E}[A_{\delta\delta}^2] &= p^{-1}\mathrm{E}\Big\{ [(\boldsymbol{m}+\boldsymbol{g}(\boldsymbol{m})-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}+\boldsymbol{g}(\boldsymbol{m})-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{S}) \\ &\qquad -2\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m}) - \boldsymbol{h}(\boldsymbol{m})'\boldsymbol{S}^{-1}\boldsymbol{h}(\boldsymbol{m})]^2\Big\} \\ &= p^{-1}\mathrm{E}\left\{[(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{S}) + 2\left(\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu}) - \boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})\right)]^2\right\} \\ &= p^{-1}\mathrm{E}\left\{[(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{S})]^2 + 4\left(\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu}) - \boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})\right)^2\right. \\ &\qquad \left. + 4\left(\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu}) - \boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})\right)[(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{S})]\right\} \\ &= 2p^{-1}\mathrm{tr}(\boldsymbol{S}^2) + 4p^{-1}\mathrm{E}\left\{\left(\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu})\right)^2 + \left(\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})\right)^2\right. \\ &\qquad - 2(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{h}(\boldsymbol{m})\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m}) + \boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu})(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu}) \\ &\qquad \left. - (\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu})\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})\right\}.\end{aligned} \tag{A.22}$$

To proceed, we use the following result derived in Theorem 3 of Stein (1981) by repeatedly applying Stein's lemma.

$$\begin{aligned}\mathrm{E}\left[(\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu}))^2\right] &= \mathrm{E}\Big[\boldsymbol{h}(\boldsymbol{m})'\boldsymbol{h}(\boldsymbol{m}) + (\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m}))^2 \\ &\qquad + \mathrm{tr}[(\boldsymbol{\nabla}\boldsymbol{h}(\boldsymbol{m})')^2] + 2\sum_{i=1}^{p}\sum_{j=1}^{p} h_i(\boldsymbol{m})\boldsymbol{\nabla}_j\boldsymbol{\nabla}_i h_j(\boldsymbol{m})\Big], \\ \mathrm{E}\left[\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu})\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})\right] &= \mathrm{E}\left[(\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m}))^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} h_i(\boldsymbol{m})\boldsymbol{\nabla}_j\boldsymbol{\nabla}_i h_j(\boldsymbol{m})\right]\end{aligned} \tag{A.23}$$

The final two terms of (A.22) require an extension to the results presented by Stein (1981). Applying Stein's lemma twice, we have

$$\begin{aligned}&\mathrm{E}[(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu})\boldsymbol{h}(\boldsymbol{m})'(\boldsymbol{m}-\boldsymbol{\mu})] \\ &\qquad = \mathrm{E}\left[(\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu}) + 2\boldsymbol{h}(\boldsymbol{m})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu})\right] \\ &\qquad = \mathrm{E}\left[\boldsymbol{\nabla}'\boldsymbol{h}(\boldsymbol{m})(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{S}(\boldsymbol{m}-\boldsymbol{\mu}) + 2\boldsymbol{\nabla}'\boldsymbol{S}\boldsymbol{h}(\boldsymbol{m})\right],\end{aligned} \tag{A.24}$$

where the first term will cancel against the last term of (A.22).

In total, we now have

$$\mathrm{E}[A_{\delta\delta}^2] = \frac{2\mathrm{tr}(\boldsymbol{S}^2)}{p} + \frac{4}{p}\mathrm{E}\left[\boldsymbol{h}(\boldsymbol{m})'\boldsymbol{h}(\boldsymbol{m}) + \mathrm{tr}\left[(\boldsymbol{\nabla}\boldsymbol{h}(\boldsymbol{m})')^2\right] + 2\boldsymbol{\nabla}'\boldsymbol{S}\boldsymbol{h}(\boldsymbol{m})\right] \tag{A.25}$$

We can work out the final two terms explicitly,

$$\begin{aligned}\operatorname{tr}\left[(\boldsymbol{\nabla h}(\boldsymbol{m})')^2\right] &= \tau^2\left[\frac{\operatorname{tr}(\boldsymbol{S}^2)}{(\boldsymbol{mSm})^2} + 4\frac{(\boldsymbol{mS}^2\boldsymbol{m})^2}{(\boldsymbol{m}'\boldsymbol{Sm})^4} - 4\frac{\boldsymbol{m}'\boldsymbol{S}^3\boldsymbol{m}}{(\boldsymbol{m}'\boldsymbol{Sm})^3}\right] \\ \boldsymbol{\nabla}'\boldsymbol{Sh}(\boldsymbol{m}) &= -\tau\left[\frac{\operatorname{tr}(\boldsymbol{S}^2)}{\boldsymbol{m}'\boldsymbol{Sm}} - 2\frac{\boldsymbol{m}'\boldsymbol{S}^3\boldsymbol{m}}{(\boldsymbol{m}'\boldsymbol{Sm})^2}\right]\end{aligned} \tag{A.26}$$

Substituting this into (A.25) and using the definitions (A.11) and (A.12) gives

$$\begin{aligned}\mathrm{E}[A_{\delta\delta}^2] &= \frac{2\operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}{p} \\ &+ \frac{4}{p}\mathrm{E}\frac{1}{\hat{q}}\left[\tau^2\frac{\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}}{\hat{q}} - 2\tau\left[\operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta) - 2\frac{\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}}{\hat{q}}\right]\right] \\ &+ \frac{4}{p}\tau^2\mathrm{E}\frac{\operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}{\hat{q}^2} + \mathrm{E}\frac{16\tau^2}{p\hat{q}^3}\left[\frac{(\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}})^2}{\hat{q}} - \hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}\right]\end{aligned} \tag{A.27}$$

Adding the variances of A_{rr}, $2A_{r\delta}$, and $A_{\delta\delta}$, we obtain

$$\begin{aligned}\mathrm{V}[D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})] = \frac{1}{p}\Bigg\{& 2\operatorname{tr}\left[\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\Sigma}_r + \boldsymbol{\Sigma}_\delta)\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\Sigma}_r + \boldsymbol{\Sigma}_\delta)\right] \\ &- 8\tau\mathrm{E}\left[\frac{\operatorname{tr}\left[\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\Sigma}_r + \boldsymbol{\Sigma}_\delta)\right]}{\hat{q}} - 2\frac{\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}}{\hat{q}^2}\right] \\ &+ 4\tau^2\mathrm{E}\left[\frac{\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\Sigma}_\delta + \boldsymbol{\Sigma}_r)\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}}{\hat{q}^2} + \frac{\operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta)}{\hat{q}^2}\right] \\ &+ 16\tau^2\mathrm{E}\left[\frac{(\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}})^2}{\hat{q}^4} - \frac{\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}}{\hat{q}^3}\right]\Bigg\}\end{aligned} \tag{A.28}$$

Choosing $\tau = \operatorname{tr}(\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta) - 2||\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_\delta||$, and using Lemma A.1, we have

$$\operatorname*{plim}_{d\to\infty} \mathrm{V}[D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})] = 2 - 4\left[\frac{a_1}{c+1} - \frac{a_2}{(c+1)^2}\right] \tag{A.29}$$

with (c, a_1, a_2) defined in Assumption A3. ∎

Normality of $D_n(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$ What remains to prove Theorem 1 is the (d, n)-asymptotic normality of $D_n(\hat{\boldsymbol{\beta}}_n^a, \boldsymbol{\beta}_n)$. Following (A.17), it suffices to show the

(d)-asymptotic normality of $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$. We start by noting that

$$\begin{aligned}\boldsymbol{\Sigma}_u^{-1/2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) &\sim N(\mathbf{0}, \boldsymbol{I} - \boldsymbol{\Sigma}_u^{-1/2}\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_u^{-1/2}) \\ \boldsymbol{\Sigma}_u^{-1/2}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]) &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_u^{-1/2}\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_u^{-1/2}).\end{aligned} \tag{A.30}$$

Using the eigenvalue decomposition $\boldsymbol{I} - \boldsymbol{\Sigma}_u^{-1/2}\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_u^{-1/2} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$,

$$\boldsymbol{U}'\boldsymbol{\Sigma}_u^{-1/2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \sim N(\mathbf{0}, \boldsymbol{D}), \quad \boldsymbol{U}'\boldsymbol{\Sigma}_u^{-1/2}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]) \sim N(\mathbf{0}, \boldsymbol{I} - \boldsymbol{D}). \tag{A.31}$$

Define the vectors

$$\boldsymbol{z}_\beta = \boldsymbol{U}'\boldsymbol{\Sigma}_u^{-1/2}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]), \quad \boldsymbol{z}_\delta = \boldsymbol{U}'\boldsymbol{\Sigma}_u^{-1/2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}), \quad \boldsymbol{\nu}_\delta = \boldsymbol{U}'\boldsymbol{\Sigma}_u^{-1/2}\boldsymbol{\delta}. \tag{A.32}$$

The random vectors $\boldsymbol{z}_\beta$ and $\boldsymbol{z}_\delta$ are independent by the independence between $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$. Also the elements within the random vectors are independent by (A.31). In terms of the quantities defined in (A.32),

$$D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = p^{-1/2}\sum_{i=1}^{p} X_i,$$

where

$$X_i = z_{i,\beta}^2 - \mathrm{E}[z_{i,\beta}^2] + 2(1-\hat{w}_d)z_{i,\beta}z_{i,\delta} - 2\hat{w}_d\nu_{i,\delta}z_{i,\beta} + (1-2\hat{w}_d)(z_{i,\delta}^2 - \mathrm{E}[z_{i,\delta}^2]) - 2\hat{w}_d\nu_{i,\delta}z_{i,\delta}.$$

Lemma A.1 implies $\mathrm{plim}_{d\to\infty}\hat{w}_d = \frac{1}{c+1}$, so that by Slutsky's theorem, we can replace $\hat{w}_d$ by its probability limit $\frac{1}{c+1}$. Then $\{X_1, \dots X_p\}$ is a sequence of asymptotically independent random variables, with mean zero and finite variance. Lyapunov's central limit theorem applies if $\mathrm{E}[|X_i|^{2+\epsilon}] = \Delta < \infty$ for some $\epsilon > 0$. Using $|a+b|^{2+\epsilon} \leq 2^{1+\epsilon}(|a|^{2+\epsilon} + |b|^{2+\epsilon})$, and the fact that z_β and z_δ are normally distributed, this condition is indeed satisfied. ■

A.4 Proof of Theorem 2

Since $\mathrm{plim}_{d\to\infty}\hat{c} = c$, as $(d, n \to \infty)_{\mathrm{seq}}$, $\sigma(\hat{c})^{-1}D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}, W) \Rightarrow N(0,1)$.

Calculating the geometric risk is similar to the approach used in Lemma 2. Choose $\xi = 3$. Since we trim the geometric risk by ξ, we can immediately analyze the (n)-asymptotic limit. Define $\hat{t}^2 = \frac{1}{p}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})$.

Rescaling the (n)-asymptotic unbiased risk estimate given in Lemma 3 by p^{-1} and using Lemma A.1, we know that $\lim_{d\to\infty}\hat{t} = \frac{c+1-a_1}{c+1}$. Since the leading term of $p^{-1/2}\hat{b}$ is the (n)-asymptotic risk, we also know that

$$\operatorname*{plim}_{d\to\infty} p^{-1/2}\hat{b} = \operatorname*{plim}_{d\to\infty} \sqrt{\max(0,\hat{e}/p)} = \frac{c+1-a_1}{c+1}. \tag{A.33}$$

We then have,

$$\lim_{d\to\infty} \mathrm{E}[\min(\hat{t}+p^{-1/2}\hat{b},\xi)] = 2\frac{c+1-a_1}{c+1}. \tag{A.34}$$

This completes the proof of Theorem 2. ■

A.5 Proof of Lemma 4

For any matrix $\boldsymbol{A}$, define throughout $\boldsymbol{P}_A = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, and $\boldsymbol{M}_A = \boldsymbol{I} - \boldsymbol{P}_A$. We first derive the covariance of two restricted estimators.

$$\begin{aligned}\operatorname{cov}(\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\beta}}_j) &= \boldsymbol{G}'\left[\boldsymbol{I}-\boldsymbol{V}\boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{V}\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'\right]\boldsymbol{V}\left[\boldsymbol{I}-\boldsymbol{R}_j(\boldsymbol{R}_j'\boldsymbol{V}\boldsymbol{R}_j)^{-1}\boldsymbol{R}_j'\boldsymbol{V}\right]\boldsymbol{G}\\ &= \boldsymbol{G}'\boldsymbol{L}'\left(\boldsymbol{I}-\boldsymbol{P}_{LR_i}\right)\left(\boldsymbol{I}-\boldsymbol{P}_{LR_j}\right)\boldsymbol{L}\boldsymbol{G},\end{aligned} \tag{A.35}$$

where we used $\boldsymbol{V} = \boldsymbol{L}'\boldsymbol{L}$.

Write $\boldsymbol{R}_j = [\boldsymbol{R}_i, \boldsymbol{S}]$, and define $\boldsymbol{A}_i = \boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{V}\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'$. Then,

$$\boldsymbol{A}_j = \boldsymbol{A}_i + (\boldsymbol{A}_i\boldsymbol{V}-\boldsymbol{I})\,\boldsymbol{S}\left[\boldsymbol{S}'(\boldsymbol{V}-\boldsymbol{V}\boldsymbol{A}_i\boldsymbol{V})\boldsymbol{S}\right]^{-1}\boldsymbol{S}'\left(\boldsymbol{V}\boldsymbol{A}_i-\boldsymbol{I}\right), \tag{A.36}$$

and hence $\boldsymbol{P}_{LR_j} = \boldsymbol{P}_{LR_i} + \boldsymbol{M}_{LR_i}\boldsymbol{L}\boldsymbol{S}\left[\boldsymbol{S}'\boldsymbol{L}'\boldsymbol{M}_{LR_i}\boldsymbol{L}\boldsymbol{S}\right]^{-1}\boldsymbol{S}'\boldsymbol{L}\boldsymbol{M}_{LR_i}$. For the covariance between two restricted estimators, this gives

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\beta}}_j) = \boldsymbol{G}'\boldsymbol{L}'(\boldsymbol{I}-2\boldsymbol{P}_{LR_i}-\boldsymbol{P}_{M_{LR_i}LS}+\boldsymbol{P}_{LR_i})\boldsymbol{L}\boldsymbol{G} = \mathrm{V}(\hat{\boldsymbol{\beta}}_j). \tag{A.37}$$

Hence, for $\hat{\boldsymbol{\delta}}_j$ defined in Assumption M3,

$$\operatorname{cov}(\hat{\boldsymbol{\delta}}_i,\hat{\boldsymbol{\delta}}_j) = 0, \qquad \operatorname{cov}(\tilde{\boldsymbol{\beta}}^m,\hat{\boldsymbol{\delta}}_j) = 0. \tag{A.38}$$

Combining this with the (n)-asymptotic normality of $\hat{\boldsymbol{\delta}}_i$ and $\tilde{\boldsymbol{\beta}}^m$ shows that the averaging estimator (30) is a sum of independent components. Similar to the case of a single restricted estimator, the risk factorizes as

$$\rho(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = p^{-1}\mathrm{E}\left[(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})\right]$$

$$= p^{-1}\mathrm{E}\left[(\hat{\boldsymbol{\beta}}_m - \mathrm{E}[\hat{\boldsymbol{\beta}}_m])\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}}_m - \mathrm{E}[\hat{\boldsymbol{\beta}}_m])\right] + p^{-1}\sum_{i\neq j}^{m}\mathrm{E}\left[\hat{w}_{d_i,i}\hat{w}_{d_j,j}\hat{\boldsymbol{\delta}}_i'\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}_j\right]$$

$$+ p^{-1}\sum_{i=1}^{m}\mathrm{E}\left[((1-\hat{w}_{d_i,i})\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i)'\boldsymbol{\Sigma}_u^{-1}((1-\hat{w}_{d_i,i})\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i)\right]. \tag{A.39}$$

By Assumption M4, the second term is equal to zero. To the third term, we can apply Stein's lemma to each summand as before. This yields

$$\rho(\hat{\boldsymbol{\beta}}^a) = 1 - \frac{1}{p}\sum_{i=1}^{m}\mathrm{E}\left\{2\tau_i\left[\frac{s_i}{\hat{q}_i} - 2\frac{\hat{\boldsymbol{\delta}}_i\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i}\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}_i}{\hat{q}_i^2}\right] - \tau_i^2\frac{1}{\hat{q}_i}\right\}. \tag{A.40}$$

A.6 Proof of Theorem 3 and Theorem 4

Define the difference between the mean squared error and the unbiased estimator for the risk of the averaging estimator (30) derived in (A.40) as

$$D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = p^{-1/2}\left\{(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta}) - p\right. \tag{A.41}$$

$$\left. + \sum_{i=1}^{m}\mathrm{E}\left\{2\tau_i\left[\frac{s_i}{\hat{q}_i} - 2\frac{\hat{\boldsymbol{\delta}}_i\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Sigma}_{\delta_i}\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}_i}{\hat{q}_i^2}\right] - \tau_i^2\frac{1}{\hat{q}_i}\right\}\right\}.$$

Rewrite the averaging estimator as

$$\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_m - \mathrm{E}\left[\hat{\boldsymbol{\beta}}^m\right] + \sum_{i=1}^{m}\left[(1-\hat{\omega}^i)\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i\right]. \tag{A.42}$$

Then, note that $p = \mathrm{tr}\left(\boldsymbol{\Sigma}_u^{-1}\left[\boldsymbol{\Sigma}_k + \sum_{i=1}^{m}\boldsymbol{\Sigma}_{\delta_i}\right]\right)$, and

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\left[(1-\hat{\omega}_i)\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i\right]'\boldsymbol{\Sigma}_u^{-1}\left[(1-\hat{\omega}_i)\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i\right]$$

$$= \sum_{i,j}(\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i)'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\delta}}_j - \boldsymbol{\delta}_j) + \sum_{i,j}\hat{\omega}_i\hat{\omega}_j\hat{\boldsymbol{\delta}}_i'\boldsymbol{\Sigma}_u^{-1}\hat{\boldsymbol{\delta}}_j - 2\sum_{i,j}\hat{\omega}_i\hat{\boldsymbol{\delta}}_i\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\delta}}_j - \boldsymbol{\delta}_j)$$

$$= \sum_{i=1}^{m}(\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i)'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i) - 2\sum_{i=1}\hat{\omega}_i\hat{\boldsymbol{\delta}}_i\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i). \tag{A.43}$$

This gives for the difference $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$

$$
\begin{aligned}
D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = p^{-1/2} \Big\{ & (\hat{\boldsymbol{\beta}}_m - \mathrm{E}[\hat{\boldsymbol{\beta}}_m])' \boldsymbol{\Sigma}_u^{-1} (\hat{\boldsymbol{\beta}}_m - \mathrm{E}[\hat{\boldsymbol{\beta}}_m]) - \mathrm{tr}(\boldsymbol{\Sigma}_k \boldsymbol{\Sigma}_u^{-1}) \\
& - 2 \sum_{i=1}^{m} (\hat{\boldsymbol{\beta}}_m - \mathrm{E}[\hat{\boldsymbol{\beta}}_m])' \boldsymbol{\Sigma}_u^{-1} [(1 - w_i)\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i] \\
& + \sum_{i=1}^{m} \Big\{ [\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i] \boldsymbol{\Sigma}_u^{-1} [\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i] - \mathrm{tr}(\boldsymbol{\Sigma}_u^{-1} \boldsymbol{\Sigma}_{\delta_i}) \Big\} \\
& -2 \sum_{i=1}^{m} \tau_i \left\{ \frac{\hat{\boldsymbol{\delta}}_i \boldsymbol{\Sigma}_u^{-1} (\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i)}{\hat{q}_i} - \frac{s_i}{\hat{q}_i} + 2 \frac{\hat{\boldsymbol{\delta}}_i \boldsymbol{\Sigma}_u^{-1} \boldsymbol{\Sigma}_{\delta_i} \boldsymbol{\Sigma}_u^{-1} \hat{\boldsymbol{\delta}}_i}{\hat{q}_i^2} \right\} \Big\}
\end{aligned} \tag{A.44}
$$

Equation (A.38) shows that the variance of $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$ is the sum of the variances of the individual components. These are provided by the proof of Theorem 1 in Appendix A.3. Asymptotic normality follows as before using the consistency of the averaging weights. The proof in Appendix A.4 can be applied term by term to obtain Theorem 4. ■

A.7 Proof of Lemma 5

The proof of Lemma 5 consists of five steps: (1) derive the expression for $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$, (2) show consistency of the estimator of the error variance and averaging weights, (3) derive the limiting distribution of the key quantities appearing in $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$, (4) show that the quantities in step (3) are asymptotically uncorrelated, (5) conclude. Throughout, we will use the notation $(d, n \to \infty)$ to indicate joint limits. Since $d = r$, this is equivalent to $(r, n \to \infty)$.

Throughout, the scalar M is a finite constant that can differ between lines.

Step 1: $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$

The unrestricted estimator and the restricted estimator are given by

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{\varepsilon} \\
\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} &= \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R} (\boldsymbol{R}' (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R})^{-1} \boldsymbol{R}' (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{\varepsilon} \\
&\quad - n^{-1/2} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R} (\boldsymbol{R}' (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R})^{-1} \boldsymbol{R}' \boldsymbol{h}.
\end{aligned} \tag{A.45}
$$

The difference $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}$ then satisfies,

$$
\hat{\boldsymbol{\delta}} - \boldsymbol{\delta} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R} (\boldsymbol{R}' (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R})^{-1} \boldsymbol{R}' (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{\varepsilon}. \tag{A.46}
$$

Define the following three quantities

$$
\begin{aligned}
D_1 &= \hat{\sigma}^{-2}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]) - (p-r), \\
D_2 &= (1-2\hat{\omega})\left[\hat{\sigma}^{-2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - r\right], \\
D_3 &= -2\hat{\sigma}^{-2}\hat{\omega}\boldsymbol{\delta}'\boldsymbol{X}'\boldsymbol{\varepsilon} - 4\hat{\omega}.
\end{aligned} \tag{A.47}
$$

We now show that the difference between the mean squared error and the unbiased risk estimate is the sum of D_1, D_2, and D_3. Write the averaging estimator as

$$
\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta} = \tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}] + (1-\hat{\omega})(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - \hat{\omega}\boldsymbol{\delta}. \tag{A.48}
$$

Then,

$$
\begin{aligned}
p^{1/2}D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) &= \hat{\sigma}^{-2}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta}) - p + (r-2)\hat{\omega} \\
&= D_1 + \hat{\sigma}^{-2}(1-\hat{\omega})^2(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + \hat{\sigma}^{-2}\hat{\omega}^2\boldsymbol{\delta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\delta} \\
&\quad - 2\hat{\sigma}^{-2}\hat{\omega}\boldsymbol{\delta}'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]) \\
&\quad - 2\hat{\sigma}^{-2}\hat{\omega}(1-\hat{\omega})(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\delta} - r + (r-2)\hat{\omega} \\
&= D_1 + D_2 + \hat{\sigma}^{-2}\hat{\omega}^2\left[\boldsymbol{\delta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\delta} + (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + 2(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\delta}\right] \\
&\quad - 2\hat{\sigma}^{-2}\hat{\omega}\boldsymbol{\delta}'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}] + \hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - r + (r-2)\hat{\omega} + r(1-2\hat{\omega}) \\
&= D_1 + D_2 + \hat{\sigma}^{-2}\hat{\omega}^2\hat{\boldsymbol{\delta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\delta}} - (r-2)\hat{\omega} + D_3 \\
&= D_1 + D_2 + D_3,
\end{aligned} \tag{A.49}
$$

where to obtain the second line we use that $(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}]) = 0$, and to obtain the final line we use that $\hat{\omega} = (r-2)\hat{\sigma}^2/\hat{\boldsymbol{\delta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\delta}}$.

Using the expressions for the restricted estimator (A.45) and the difference $\hat{\boldsymbol{\delta}}$ in (A.46), we have

$$
D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta}) = \left(\frac{p-r}{p}\right)^{1/2} G_1 + \left(\frac{r}{p}\right)^{1/2}\left[(1-2\hat{\omega})G_2 - 2\hat{\omega}N_1\right] - 4p^{-1/2}\hat{\omega}, \tag{A.50}
$$

where, defining $\boldsymbol{P}_{LR} = \boldsymbol{L}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{L}'$ with $\boldsymbol{L}$ such that $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \boldsymbol{L}'\boldsymbol{L}$, and $\boldsymbol{M}_{LR} = \boldsymbol{I} - \boldsymbol{P}_{LR}$,

$$
\begin{aligned}
G_1 &= (p-r)^{-1/2}\left[\hat{\sigma}^{-2}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{M}_{LR}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'\boldsymbol{\varepsilon} - (p-r)\right], \\
G_2 &= r^{-1/2}\left[\hat{\sigma}^{-2}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}_{LR}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'\boldsymbol{\varepsilon} - r\right], \\
N_1 &= (nr)^{-1/2}\hat{\sigma}^{-2}\boldsymbol{h}'\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}.
\end{aligned} \tag{A.51}
$$

Below we show that the weights converge to a finite constant, and hence, the last term of (A.50) is $O_p(p^{-1/2})$.

Step 2: Consistency of error variance estimator and averaging weights

Lemma A.3 *Under Assumption LR1–LR3, as* $(d, n \to \infty)$,
(a) $\hat{\sigma}^2 \to_p \sigma^2$,
(b) $\hat{\omega} \to_p \omega = (c+1)^{-1}$ *with* c *defined in Assumption LR2.*

Proof: Part (a). The error variance estimator is $\hat{\sigma}^2 = \frac{1}{n-p}\boldsymbol{\varepsilon}'\boldsymbol{M}_X\boldsymbol{\varepsilon}$. We have

$$\begin{aligned} \mathrm{E}[\hat{\sigma}^2] &= \sigma^2, && \text{(A.52)} \\ \mathrm{var}(\hat{\sigma}^2) &= \frac{1}{(n-p)^2}\left[(\mathrm{E}[(\boldsymbol{\varepsilon}'\boldsymbol{M}_X\boldsymbol{\varepsilon})^2] - \mathrm{E}[\boldsymbol{\varepsilon}'\boldsymbol{M}_X\boldsymbol{\varepsilon}]^2\right] && \text{(A.53)} \\ &\leq \frac{1}{(n-p)^2}\left[\mathrm{E}[(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})^2] - \sigma^4(n-p)^2\right] \\ &\leq \frac{1}{n^2(1-p/n)^2}\left[(\mathrm{E}[\varepsilon_i^4] - 3\mathrm{E}[\varepsilon_i^2]^2)n + \sigma^4 n^2 + 2\sigma^4 n - \sigma^4 n^2(1-p/n)^2\right] \\ &= O(n^{-1}), \end{aligned}$$

where the last line uses that $p/n \to 0$ and the fact that ε_i has bounded fourth moment. Part (a) now follows from Chebyshev's inequality.

Part(b). Consider

$$\begin{aligned} r^{-1}\hat{\boldsymbol{\delta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\delta}} &= r^{-1}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \\ &\quad + (rn)^{-1}\boldsymbol{h}'\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{h} && \text{(A.54)} \\ &\quad + 2r^{-1}n^{-1/2}\boldsymbol{h}'(\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}. \end{aligned}$$

The limit of the second term is $\sigma^2 c$ by Assumption LR2. For the first term, define $\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and note that $\mathrm{tr}(\boldsymbol{A}) = r$. Then,

$$\begin{aligned} r^{-1}\mathrm{E}[\boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon}] &= r^{-1}\mathrm{E}[\mathrm{E}[\boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon}|\boldsymbol{X}]] \\ &= r^{-1}\sigma^2\mathrm{E}[\mathrm{tr}(\boldsymbol{A})] && \text{(A.55)} \\ &= \sigma^2. \end{aligned}$$

For the variance, by Ullah (2004) Appendix A5, and using that conditional on $\boldsymbol{X}$,

ε_i has bounded fourth moment,

$$
\begin{aligned}
\operatorname{var}(r^{-1}\boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon}) &= r^{-2}\mathrm{E}\left[\mathrm{E}[(\boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon})^2|\boldsymbol{X}]\right] - \sigma^4 \\
&\leq r^{-2}M\mathrm{E}[\operatorname{tr}(\boldsymbol{A}(\boldsymbol{I}_n \odot \boldsymbol{A}))] + 2\sigma^4 r^{-1} \\
&= r^{-2}M\mathrm{E}[\operatorname{tr}((\boldsymbol{I}_n \odot \boldsymbol{A})^{1/2}\boldsymbol{A}(\boldsymbol{I}_n \odot \boldsymbol{A})^{1/2})] + 2\sigma^4 r^{-1} \\
&\leq r^{-2}M\mathrm{E}[\operatorname{tr}(\boldsymbol{A})] + 2\sigma^4 r^{-1} \\
&= O(r^{-1}),
\end{aligned} \tag{A.56}
$$

where the inequality on the fourth line uses that

$$
\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}_{(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{R}}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}' \preceq \boldsymbol{P}_X \preceq \boldsymbol{I}_n. \tag{A.57}
$$

By Chebyshev's inequality, as $(d, n \to \infty)$, $r^{-1}\boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon} \to_p \sigma^2$.

The final term of (A.54) has expected value equal to zero, and

$$
\begin{aligned}
&\operatorname{var}(r^{-1}n^{-1/2}\boldsymbol{h}'(\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}) \\
&= \sigma^2 \cdot r^{-1} \cdot \mathrm{E}\left[r^{-1}\boldsymbol{h}'\boldsymbol{R}(\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{h}\right] \\
&= O(r^{-1}).
\end{aligned} \tag{A.58}
$$

Then, as $(d, n \to \infty)$, $r^{-1}n^{-1/2}\boldsymbol{h}'(\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \to_p 0$.

We have now established that, as $(d, n \to \infty)$, $r^{-1}\hat{\boldsymbol{\delta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\delta}} \to_p c+1$ and hence

$$
\hat{\omega} = \frac{r-2}{r}\frac{\hat{\sigma}^2 r}{\hat{\boldsymbol{\delta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\delta}}} \to_p \frac{1}{c+1}. \tag{A.59}
$$

This concludes the proof of Lemma A.3. ■

Step 2: Limiting distribution of G_1, G_2, and N_1 in (A.51)

Define $\boldsymbol{P}_X = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, and $\boldsymbol{P}_{G2} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}_{LR}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'$, and $\boldsymbol{P}_{G1} = \boldsymbol{P}_X - \boldsymbol{P}_{G2}$. We have $\boldsymbol{P}_{G1}, \boldsymbol{P}_{G2} \preceq \boldsymbol{P}_X$. Also, $\boldsymbol{P}_{G1}^2 = \boldsymbol{P}_{G1}$ and $\boldsymbol{P}_{G2}^2 = \boldsymbol{P}_{G2}$. Finally, $\operatorname{rk}(\boldsymbol{P}_X) = \operatorname{tr}(\boldsymbol{P}_X) = p$, $\operatorname{rk}(\boldsymbol{P}_{G2}) = \operatorname{tr}(\boldsymbol{P}_{G2}) = \operatorname{tr}(\boldsymbol{P}_{LR}) = \operatorname{rk}(\boldsymbol{P}_{LR}) = r$, and $\operatorname{rk}(\boldsymbol{P}_{G1}) = \operatorname{tr}(\boldsymbol{P}_{G1}) = p - r$.

G_1 only contributes to $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$ when $r/p \to c$ with $c \in (0,1)$. In deriving the distribution of G_1, we therefore make this assumption. This implies that $p - r \to \infty$. If $r = p$, then G_1 does not appear, but the arguments below still apply to G_2 and N_1.

Both in G_1 and G_2 there is a quadratic form that can be split schematically as

$$\frac{\sigma^{-2}\boldsymbol{\varepsilon}'\boldsymbol{P}\boldsymbol{\varepsilon} - \text{rk}(\boldsymbol{P})}{\sqrt{\text{rk}(\boldsymbol{P})}} = \frac{\sum_{i=1}^{n} P_{ii}\left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right) + \sum_{i\neq j}\varepsilon_i\varepsilon_j P_{ij}}{\sqrt{\text{rk}(\boldsymbol{P})}} = A_1 + A_2. \tag{A.60}$$

where $\boldsymbol{P}$ can be $\boldsymbol{P}_{G1}$ or $\boldsymbol{P}_{G2}$ and P_{ij} denotes the i,j-th element of $\boldsymbol{P}$. We will first show that $A_1 = o_p(1)$, and then apply Lemma A.2 to A_2. Notice that $\text{E}[A_1] = 0$. Also, $\boldsymbol{P}_{G1}, \boldsymbol{P}_{G2} \preceq \boldsymbol{P}_X$, and by Assumption LR1,

$$[\boldsymbol{P}_X]_{ii} \leq \frac{p}{nb}\frac{1}{p}\boldsymbol{x}_i'\boldsymbol{x}_i, \quad a.s.n. \tag{A.61}$$

We can then bound the variance of A_1 as

$$\begin{aligned}
\text{var}(A_1) &= \text{rk}(\boldsymbol{P})^{-1}\text{var}\left(\sum_{i=1}^{n} P_{ii}\left(\varepsilon_i^2/\sigma^2 - 1\right)\right) \\
&= \text{rk}(\boldsymbol{P})^{-1}\sum_{i=1}^{n}\text{E}[P_{ii}^2]\text{E}\left[\text{E}\left[\left(\varepsilon_i^2/\sigma^2 - 1\right)^2 \middle| \boldsymbol{X}\right]\right] \\
&\leq M\text{rk}(\boldsymbol{P})^{-1}\sum_{i=1}^{n}\text{E}[P_{ii}^2] \\
&\leq M\text{rk}(\boldsymbol{P})^{-1}\sum_{i=1}^{n}\text{E}[[\boldsymbol{P}_X]_{ii}^2] \\
&\leq M\frac{p^2}{\text{rk}(\boldsymbol{P})n}\text{E}[(p^{-1}\boldsymbol{x}_i'\boldsymbol{x}_i)^2] \\
&\leq M\frac{p^2}{\text{rk}(\boldsymbol{P})n}.
\end{aligned} \tag{A.62}$$

where to obtain the fourth line, we use that $P_{ii} = \boldsymbol{e}_i'\boldsymbol{P}\boldsymbol{e}_i \leq \boldsymbol{e}_i'\boldsymbol{P}_X\boldsymbol{e}_i$. For G_1, $\text{rk}(\boldsymbol{P}_{G1}) = p(1 - r/p)$, so that when $p/n \to 0$, $A_1 = o_p(1)$. For G_2, $\text{rk}(\boldsymbol{P}_{G2}) = r$. By Assumption LR3, $r/p \to c$ with $c \in (0, 1]$, so that again $A_1 = o_p(1)$ when $p/n \to 0$.

We now turn to the distribution of A_2. Notice that for $\boldsymbol{P}_{G1}$ and $\boldsymbol{P}_{G2}$,

$$\sum_{i\neq j} P_{ij}^2 = \text{tr}(\boldsymbol{P}^2) - \sum_{i=1}^{n} P_{ii}^2 = \text{rk}(\boldsymbol{P}) - \sum_{i=1}^{n} P_{ii}^2. \tag{A.63}$$

Using Markov's inequality, we have that both for G_1 and G_2,

$$
\begin{aligned}
P\left(\operatorname{rk}(\boldsymbol{P})^{-1}\sum_{i=1}^{n}P_{ii}^2 \geq \epsilon\right) &\leq \frac{\operatorname{rk}(\boldsymbol{P})^{-1}\sum_{i=1}^{n}\mathrm{E}[P_{ii}^2]}{\epsilon} \\
&\leq \frac{p^2}{n\cdot\operatorname{rk}(\boldsymbol{P})}\frac{1}{\epsilon\cdot b}\mathrm{E}[(p^{-1}\boldsymbol{x}_i'\boldsymbol{x}_i)^2] \\
&\leq \frac{M}{\epsilon}\frac{p^2}{n\cdot\operatorname{rk}(\boldsymbol{P})}.
\end{aligned}
\tag{A.64}
$$

Since for both G_1 and G_2, $p/\operatorname{rk}(\boldsymbol{P}) \to c$, by Assumption LR3, the r.h.s. goes to zero as $p/n \to 0$ for any $\epsilon > 0$. This then implies that

$$
\frac{1}{\operatorname{rk}(\boldsymbol{P})}\sum_{i\neq j}P_{ij}^2 = 1 + o_p(1),
\tag{A.65}
$$

which simplifies the variance in Lemma A.2 to $V_n = 2\sigma^4$. We now check conditions (i)–(iv) of Lemma A.2 for $\boldsymbol{P}_{G1}$ and $\boldsymbol{P}_{G2}$. Conditions (ii)–(iv) hold by Assumption LR2 and Assumption LR3. From condition (i), we only need to verify the last part.

$$
\begin{aligned}
[\boldsymbol{P}]_{ii} &\leq \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i \\
&= \boldsymbol{x}_i'(\boldsymbol{x}_i\boldsymbol{x}_i' + \boldsymbol{X}_{-i}'\boldsymbol{X}_{-i})^{-1}\boldsymbol{x}_i \\
&= \boldsymbol{x}_i'\left[(\boldsymbol{X}_{-i}'\boldsymbol{X}_{-i})^{-1} - \frac{(\boldsymbol{X}_{-i}'\boldsymbol{X}_{-i})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'(\boldsymbol{X}_{-i}'\boldsymbol{X}_{-i})^{-1}}{1+\boldsymbol{x}_i'(\boldsymbol{X}_{-i}'\boldsymbol{X}_{-i})^{-1}\boldsymbol{x}_i}\right]\boldsymbol{x}_i \\
&= \frac{\boldsymbol{x}_i'(\boldsymbol{X}_{-i}'\boldsymbol{X}_{-i})^{-1}\boldsymbol{x}_i}{1+\boldsymbol{x}_i'(\boldsymbol{X}_{-i}'\boldsymbol{X}_{-i})^{-1}\boldsymbol{x}_i} \\
&< 1 \quad a.s.n.
\end{aligned}
\tag{A.66}
$$

By Lemma A.2 and the consistency of $\hat{\sigma}^2$, we now have

$$
G_1 \Rightarrow N(0,2), \quad G_2 \Rightarrow N(0,2).
\tag{A.67}
$$

It remains to derive the distribution of N_1 for which we use Theorem 2 of Phillips and Moon (1999). Define

$$
\begin{aligned}
Y_{i,p} &= \frac{1}{\sqrt{n}}w_{in}\varepsilon_i \\
w_{in} &= \frac{1}{\sqrt{r}}\boldsymbol{h}'\boldsymbol{R}(\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i,
\end{aligned}
\tag{A.68}
$$

with variance

$$\Omega_{i,p} = \mathrm{E}\left[Y_{i,p}^2\right] = \sigma^2 n^{-1} r^{-1} \mathrm{E}\left[\left(\boldsymbol{h}'\boldsymbol{R}(\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i\right)^2\right]. \tag{A.69}$$

We only have to consider the case where $\boldsymbol{h} \neq \boldsymbol{0}$, since otherwise N_1 does not appear in the expression for $D(\hat{\boldsymbol{\beta}}^a, \boldsymbol{\beta})$. Let

$$s_{n,p}^2 = \sum_{i=1}^{n} \Omega_{i,p} = \sigma^2 r^{-1} \mathrm{E}[\boldsymbol{h}'(\boldsymbol{R}(\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{h})] = \sigma^2 \mathrm{E}[c_{d,n}], \tag{A.70}$$

and define

$$\xi_{i,n,p} = \frac{Y_{i,p}}{s_{n,p}}. \tag{A.71}$$

Case 1: $\boldsymbol{R} = \boldsymbol{I}$. In this case,

$$w_{in} = w_i = \frac{1}{\sqrt{r}} \boldsymbol{h}'\boldsymbol{x}_i, \tag{A.72}$$

and $w_i \varepsilon_i$ is an i.i.d. sequence. Via standard arguments, we can directly verify the Lindeberg condition in Theorem 2 of Phillips and Moon (1999), which in our case is

$$\frac{1}{n\mathrm{E}[c_{d,n}]} \sum_{i=1}^{n} \mathrm{E}\left[w_i^2 \frac{\varepsilon_i^2}{\sigma^2} I\left[\frac{1}{\mathrm{E}[c_{d,n}]} w_i^2 \frac{\varepsilon_i^2}{\sigma^2} > \epsilon^2 n\right]\right] \to 0. \tag{A.73}$$

Since $c_{d,n}$ is bounded, it is sufficient to show

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[w_i^2 \frac{\varepsilon_i^2}{\sigma^2} I\left[w_i^2 \frac{\varepsilon_i^2}{\sigma^2} > \epsilon^2 n\right]\right] \to 0. \tag{A.74}$$

As $w_i \varepsilon_i$ is i.i.d., this holds if

$$\mathrm{E}\left[w_1^2 \frac{\varepsilon_1^2}{\sigma^2} I\left[w_1^2 \frac{\varepsilon_1^2}{\sigma^2} > \epsilon^2 n\right]\right] \to 0. \tag{A.75}$$

Using Markov's inequality,

$$\begin{aligned} \mathrm{P}\left[w_1^2 \frac{\varepsilon_1^2}{\sigma^2} > \epsilon^2 n\right] &\leq \frac{\mathrm{E}[w_1^2 \varepsilon_1^2]}{\epsilon^2 \sigma^2 n} \\ &\leq \frac{\frac{1}{r}\boldsymbol{h}'\boldsymbol{Q}_x\boldsymbol{h}}{\epsilon^2 n} \\ &\leq \frac{M}{n}. \end{aligned} \tag{A.76}$$

By Assumption LR2, $\mathrm{E}[w_{1n}^2\varepsilon_1^2/\sigma^2] = \mathrm{E}[w_1^2] < \infty$, so that by the dominated convergence theorem the Lindeberg condition (A.75) holds.

Case 2: $\boldsymbol{R} \neq \boldsymbol{I}$. In this case, we use part (c) of Assumption LR1 and condition on $\boldsymbol{X}$. We then need to show that

$$\frac{1}{nc_{d,n}}\sum_{i=1}^{n}\mathrm{E}\left[w_{in}^2\frac{\varepsilon_i^2}{\sigma^2}I\left[\frac{1}{c_{d,n}}w_{in}^2\frac{\varepsilon_i^2}{\sigma^2} > \epsilon^2 n\right]\bigg|\boldsymbol{X}\right] \to 0. \tag{A.77}$$

Write $m_n = \max_{i=1,\ldots,n}\frac{w_{in}^2}{nc_{d,n}}$, then it is sufficient to show that

$$\frac{1}{nc_{d,n}}\sum_{i=1}^{n}w_{in}^2\mathrm{E}\left[\frac{\varepsilon_i^2}{\sigma^2}I\left[\frac{\varepsilon_i^2}{\sigma^2} > \frac{\epsilon^2}{m_n}\right]\bigg|\boldsymbol{X}\right] = \mathrm{E}\left[\frac{\varepsilon_i^2}{\sigma^2}I\left[\frac{\varepsilon_i^2}{\sigma^2} > \frac{\epsilon^2}{m_n}\right]\bigg|\boldsymbol{X}\right] \to 0. \tag{A.78}$$

Since $\mathrm{E}[\varepsilon_i^2/\sigma^2] = 1 < \infty$, this holds if $m_n \to 0$. Using the Cauchy-Schwarz inequality, it is sufficient to show that,

$$\max_{i=1,\ldots,n}\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i \to 0. \tag{A.79}$$

This condition was already noted by Huber et al. (1973) to ensure normality of linear combinations of the least squares estimator in a linear regression model. The following shows that this condition indeed holds *a.s.n.* under Assumption LR1.

Since by Assumption LR1, $b\boldsymbol{I}_p \preceq n^{-1}\boldsymbol{X}'\boldsymbol{X}$ *a.s.n.* for some constant $b > 0$, it is sufficient to show that

$$\max_{i=1,\ldots,n}\boldsymbol{x}_i'\boldsymbol{x}_i/n \to 0, \quad a.s. \tag{A.80}$$

Denote $M_n = \max_{i=1,\ldots,n}|\boldsymbol{x}_i'\boldsymbol{x}_i/n - \mathrm{tr}(\boldsymbol{Q}_X)/n|$. We have

$$\begin{aligned}
\mathrm{P}(M_n > \epsilon) &= 1 - \mathrm{P}(M_n \leq \epsilon)\\
&= 1 - \mathrm{P}\left(|\boldsymbol{x}_i'\boldsymbol{x}_i/n - \mathrm{tr}(\boldsymbol{Q}_X)/n| \leq \epsilon\right)^n\\
&= 1 - \left[1 - \mathrm{P}\left(|\boldsymbol{x}_i'\boldsymbol{x}_i/n - \mathrm{tr}(\boldsymbol{Q}_X)/n| > \epsilon\right)\right]^n\\
&\leq 1 - \left[1 - (M\cdot p)/(n^2\cdot\epsilon^2)\right]^n\\
&= 1 - \exp(n\log(1 - (M/\epsilon^2)p/n^2))\\
&= 1 - \exp(o(1))\\
&= o(1),
\end{aligned} \tag{A.81}$$

where the inequality uses Chebyshev's inequality and the fact that $p^{-1}\mathrm{var}(\boldsymbol{x}_i'\boldsymbol{x}_i) \leq M$ by Assumption LR1. We have established that as $(p, n \to \infty)$, $M_n \to_p 0$.

Since the sequence M_n is monotone increasing, this implies $M_n \to 0$ $a.s.$ and the Lindeberg condition (A.78) holds $a.s.n.$

The above establishes that $\mathrm{P}(\sum_{i=1}^n \xi_{i,n,p} \leq y|\boldsymbol{X}) \to \Phi(y)$ $a.s.$, with $\Phi(y)$ the standard normal CDF evaluated at y. We now follow the argument at the top of p. 81 of Chao et al. (2012). The unconditional probability $\mathrm{P}(\sum_{i=1}^n \xi_{i,n,p}) = \mathrm{E}[\mathrm{P}(\sum_{i=1}^n \xi_{i,n,p} \leq y|\boldsymbol{X})]$. Since for some $\epsilon > 0$, $\sup_n \mathrm{E}[|\mathrm{P}(\sum_{i=1}^n \xi_{i,n,p} \leq y)|^{1+\epsilon}] < \infty$, the convergence also holds unconditionally, i.e. $\mathrm{P}(\sum_{i=1}^n \xi_{i,n,p}) \to \Phi(y)$. Finally, since $\mathrm{plim}_{(d,n\to\infty)} c_{d,n} = c$, we have that as $(d, n \to \infty)$, $N_1 \Rightarrow N(0, c)$ $a.s.$

Step 3: Covariance between G_1, G_2, and N_1

What remains is to bound the covariance between G_1, G_2, and N_1. Note that this covariance is identically zero when the errors are normal. Under non-normality, using Ullah (2004) Appendix A5, we have that the covariance between N_1 and G_2 is

$$
\begin{aligned}
\frac{r}{p}\mathrm{cov}(N_1, G_2) = {} & \mathrm{E}[\varepsilon_i^3]\frac{1}{p\sqrt{n}}\sum_{i=1}^n \mathrm{E}\left[n^{-1}\boldsymbol{x}_i'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}_{LR}(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{x}_i \times \right. \\
& \left. \boldsymbol{x}_i'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}(\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{h}\right] \\
\leq {} & \mathrm{E}[\varepsilon_i^3]\frac{1}{pn^{3/2}}\mathrm{E}\left\{\left[\sum_{i=1}^n \left(\boldsymbol{x}_i'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}_{LR}(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{x}_i\right)^2\right]^{1/2}\right. \\
& \left.\left[\sum_{i=1}^n \left(\boldsymbol{x}_i'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}(\boldsymbol{R}'(n^{-1}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{h}\right)^2\right]^{1/2}\right\} \\
\leq {} & M\mathrm{E}[\varepsilon_i^3]\frac{1}{n}\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n (p^{-1}\boldsymbol{x}_i'\boldsymbol{x}_i)^2\right)^{1/2}(nrc_{d,n})^{1/2}\right] \\
\leq {} & M\mathrm{E}[\varepsilon_i^3]\left(\frac{r}{n}\right)^{1/2}\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n (p^{-1}\boldsymbol{x}_i'\boldsymbol{x}_i)^2\right)^{1/2}\right] \\
\leq {} & M\mathrm{E}[\varepsilon_i^3]\left(\frac{r}{n}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[(p^{-1}\boldsymbol{x}_i'\boldsymbol{x}_i)^2\right]\right)^{1/2} \\
\leq {} & M\mathrm{E}[\varepsilon_i^3]\left(\frac{r}{n}\right)^{1/2},
\end{aligned}
$$

where the first inequality uses Cauchy-Schwarz inequality, the second inequality the definition of $c_{d,n}$, the fact that $\boldsymbol{P}_{LR}$ is a projection matrix, and the fact that the eigenvalues of $n^{-1}\boldsymbol{X}'\boldsymbol{X}$ are bounded, the third inequality used that $c_{d,n}$ is finite by assumption, the fourth inequality uses Jensen's inequality, and finally the expected value is finite by Assumption LR1.

The same argument holds for the covariance between G_1 and N_1. What remains is the covariance between G_1 and G_2. Again using Ullah (2004) Appendix A5 gives

$$\begin{aligned}\left(\frac{p-r}{p}\frac{r}{p}\right)^{1/2}\operatorname{cov}(G_1,G_2) &= \left(\mathrm{E}[\varepsilon_i^4]-3\sigma^4\right)\frac{1}{p}\sum_{i=1}^{n}\mathrm{E}\big[\big[\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}_{LR}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{x}_i\times \\ &\qquad \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{M}_{LR}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{x}_i\big] \\ &\le M\left(\mathrm{E}[\varepsilon_i^4]-3\sigma^4\right)\frac{1}{p}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i)^2] \\ &\le M\left(\mathrm{E}[\varepsilon_i^4]-3\sigma^4\right)\frac{p}{n^2}\sum_{i=1}^{n}\mathrm{E}[(p^{-1}\boldsymbol{x}_i'\boldsymbol{x}_i)^2] \\ &\le M\left(\mathrm{E}[\varepsilon_i^4]-3\sigma^4\right)\frac{p}{n}.\end{aligned}$$

Conclusion

Using the results above, we have that as $(d, n\to\infty)$,

$$D(\hat{\boldsymbol{\beta}}^a,\boldsymbol{\beta}) \Rightarrow N(0,\sigma^2(c)) \quad a.s. \tag{A.82}$$

The variance is the sum of the variances of G_1, G_2 and N_1, as scaled in (A.50),

$$\begin{aligned}\sigma^2(c) &= 2(1-a)+a\left[2\left(1-\frac{2}{c+1}\right)^2+4\frac{c}{(c+1)^2}\right] \\ &= 2-4a\left[\frac{1}{c+1}-\frac{1}{(c+1)^2}\right],\end{aligned} \tag{A.83}$$

which corresponds to Theorem 1. ■

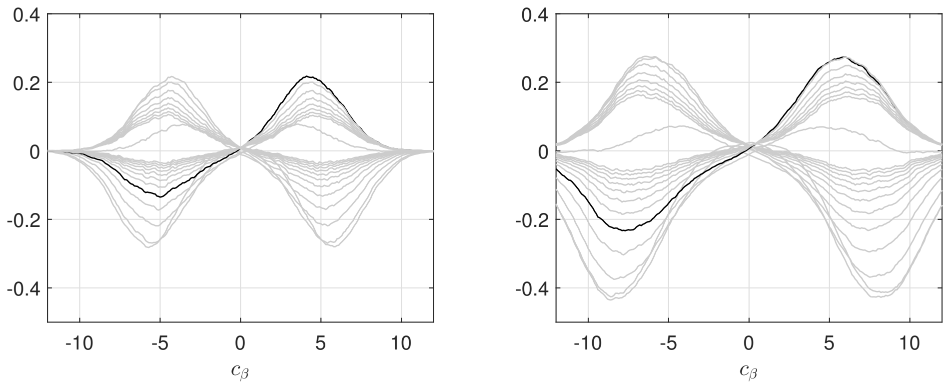
Appendix B Simulations: additional results

B.1 Benchmark confidence regions

The confidence regions by Casella and Hwang (1983) are given by Definition 1, with the averaging estimator as in (1)-(4) with the weights restricted to be less or equal then 1. Defining $R(x) = 1-\frac{p-2}{x^2}$, the critical values are

$$\hat{b}_{CH}^2 = \begin{cases} R(b_\chi)\left[b_\chi^2 - p\log(R(b_\chi))\right] & \text{if } n\hat{q}_n \le b_\chi^2, \\ R(n\hat{q}_n)[b_\chi^2 - p\log(R(n\hat{q}_n))] & \text{if } n\hat{q}_n > b_\chi^2. \end{cases} \tag{B.1}$$

Figure 4: Restricted estimator magnitude



Samworth (2005) Taylor expands $(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\beta}}^a - \boldsymbol{\beta})$ around $\boldsymbol{\beta} = \mathbf{0}$ to get

$$\hat{b}_S^2 = \min\left\{ w_\alpha(0) + \frac{1}{2} w_\alpha''(0) n \hat{q}_n, b_\chi^2 \right\}, \quad w_\alpha(0) = \left(b_\chi - \frac{p-2}{b_\chi} \right)^2,$$
$$w_\alpha(0)'' = \frac{2}{k}\left(1 - \frac{(p-2)}{b_\chi^2}\right)\left[\frac{(p-2)(p-1)}{b_\chi^2 + p - 2} - \frac{2(p-2)b_\chi^2}{(b_\chi^2 + p - 2)^2} + \frac{(p-2)^2}{b_\chi^2 + p - 2}\right] + \frac{2(p-2)(p-1)}{b_\chi^2 \cdot p}. \tag{B.2}$$

B.2 Restricted estimator magnitude

Figure 4 shows the power difference with the standard F test for the simulation experiment in Section 4.2. Gray lines correspond to different choices of the magnitude m in (43). The black line corresponds to $m = 3$. On the left, $p = 6$, on the right $p = 12$. The magnitude of the nuisance parameters and the correlation with the variables of interest are irrelevant as the fixed restricted vector is averaged with the unrestricted estimator that is orthogonalized with respect to the variables not of interest.

B.3 Instrumental variables model

Consider the following instrumental variables model also used in Hansen (2017).

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{X} &= \boldsymbol{Z}\boldsymbol{\Pi} + \boldsymbol{U}, \quad \boldsymbol{U} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k]', \quad [Z]_{ij} \sim N.i.d.(0,1) \\ \begin{pmatrix} \varepsilon_i \\ \boldsymbol{u}_i \end{pmatrix} &\sim N\left[\boldsymbol{0}, \begin{pmatrix} 1 & \sigma_u\rho/\sqrt{k} & \dots & \sigma_u\rho/\sqrt{k} \\ \sigma_u\rho/\sqrt{k} & \sigma_u^2 & 0 & 0 \\ \vdots & 0 & \sigma_u^2 & 0 \\ \sigma_u\rho/\sqrt{k} & 0 & 0 & \sigma_u^2 \end{pmatrix}\right], \end{aligned} \tag{B.3}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n \times (p+1)$ matrices. The first column of both $\boldsymbol{X}$ and $\boldsymbol{Z}$ is a vector of ones. The remaining elements of $\boldsymbol{Z}$ are independent standard normals. The sample size $n = 150$, the number of parameters of interest $p = 6$, and $\boldsymbol{\Pi} = \boldsymbol{I}_p$. The variance of the first stage error is varied as $\sigma_u^2 = \frac{n}{k}F^{-1}$, where $F = \{5, 10, 20\}$. The correlation between the error terms is varied over $\rho = \{0.3, 0.9\}$. Our goal is to perform efficient inference on $\boldsymbol{\beta}$, which is again given by (45).

For the unrestricted estimator, we consider the 2SLS estimator. Since the model is exactly identified, this is an appropriate choice. We consider two choices for the restricted estimator. First, we consider indirect restrictions by averaging with the OLS estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$. Second, we consider direct restrictions given by (43). Following the previous experiment, we again set $m = \{-3, 0, 3\}$. The error variance σ_ε^2 is estimated using the 2SLS estimator.

In Table 4, we report the coverage rate under a first stage signal that ranges over $F = \{5, 10, 20\}$. First consider the case where $c_\beta = 0$ and $\rho = 0.9$. The endogeneity creates an assymetry between the coverage when choosing the restricted estimator as in (43) with $m = -3$ and $m = 3$. The former results in overcoverage, while the latter in undercoverage. This effect only slowly decreases when F increases from 5 to 20. In general, coverage away from $c_\beta = 0$ is slightly too large when the sign of the restricted estimator corresponds to the true sign of c_β. Using the OLS estimator that ignores the endogeneity results in overcoverage when $\rho = 0.3$ and undercoverage when $\rho = 0.9$.

Figure 5 shows the power of the hypothesis test corresponding to the confidence regions. The left panels have weak endogeneity ($\rho = 0.3$). The top panel has a first stage signal strength $F = 5$, while the lower panel has $F = 20$. As for the linear regression model, the power difference when averaging the 2SLS estimator with the OLS estimator depends on the sign of c_β. Using a fixed restricted estimator

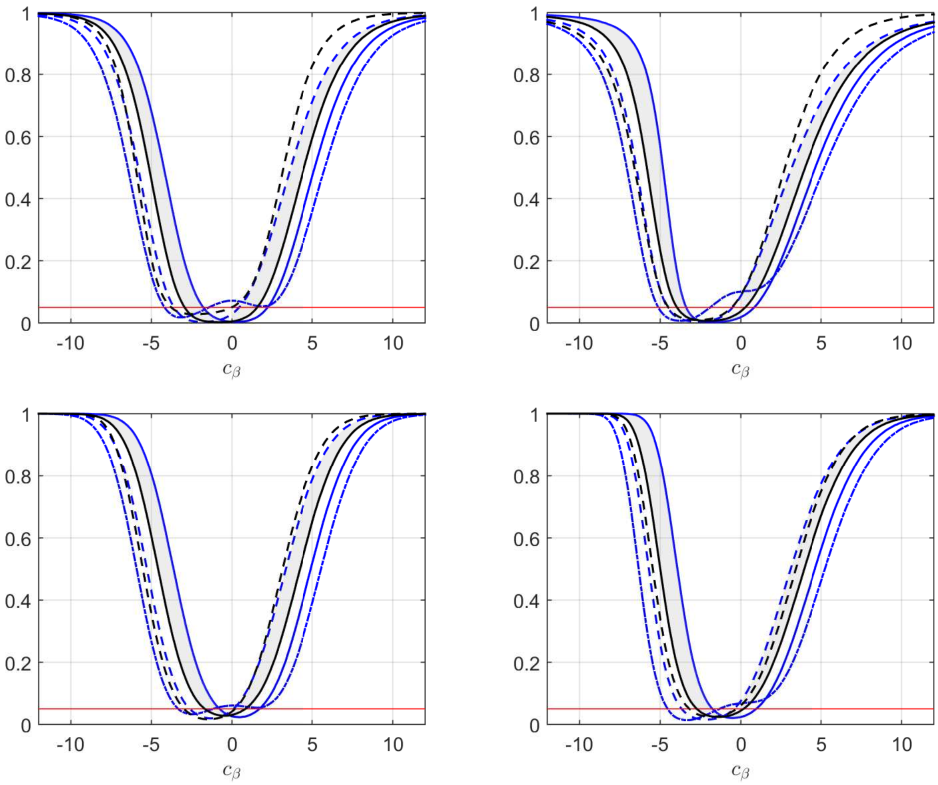
Table 4: Instrumental variables model: coverage rate.

n		$F=5$		$F=10$		$F=20$	
		$\rho=0.3$	$\rho=0.9$	$\rho=0.3$	$\rho=0.9$	$\rho=0.3$	$\rho=0.9$
150	$m=-3$	0.996	0.979	0.985	0.978	0.970	0.974
	$m=0$	0.924	0.890	0.937	0.924	0.939	0.931
	$m=3$	0.971	0.911	0.960	0.915	0.951	0.920
	OLS	0.957	0.915	0.964	0.930	0.960	0.937
	F	0.994	0.962	0.984	0.959	0.970	0.956
500	$m=-3$	0.997	0.981	0.987	0.980	0.974	0.976
	$m=0$	0.928	0.899	0.939	0.927	0.939	0.932
	$m=3$	0.971	0.913	0.961	0.918	0.955	0.924
	OLS	0.950	0.911	0.962	0.929	0.958	0.939
	F	0.995	0.962	0.984	0.959	0.969	0.956

Note: coverage rate at $\boldsymbol{\beta} = \mathbf{0}$ under (B.3) with sample size $n = \{150, 500\}$, number of parameters of interest $p = 6$, and first stage signal strength $F = \{5, 10, 20\}$. We report using direct restrictions (43) with $m = \{-3, 0, 3\}$, and indirect restrictions by averaging with the OLS estimator. The standard F-test is provided for reference. Nominal coverage equals 0.95.

with the correct sign can yield substantial power improvements as indicated by the gray area. The right panels increase the endogeneity to $\rho = 0.9$. This creates a strong asymmetry between $c_\beta > 0$ and $c_\beta < 0$. It appears that the estimates for $\boldsymbol{\beta}$ are upward biased, causing overestimation of the error variance σ_ε when c_β is slightly below zero. Here, most methods yield overcoverage with the exception of the restricted estimator with $m = 0$. This effect disappears when F increases from $F = 5$ in the top panel to $F = 20$ in the bottom panel. The power increase over the standard F-test based on the 2SLS estimator can be substantial. For example, in the lower right panel, when $c_\beta = -5$, and we use the fixed restricted estimator with $m = -3$, the power increases from 0.45 to 0.8.

Figure 5: Instrumental variables model: power.



Note: The figure shows power against $H_0 : \boldsymbol{\beta} = \mathbf{0}$ at a sample size of $n = 150$ and $p = 6$ parameters of interest. The black solid line shows the power from the usual F-test centered at the 2SLS estimator. The black dashed line when averaging with the restricted least squares estimator. The blue lines correspond to the restricted estimator (43) with $m = -3$ (solid), $m = 0$ (dash-dotted), $m = 3$ (dashed). The gray area highlights the power difference between using the fixed restricted estimator with the correct sign and the standard F-test. In the left panels we have weak endogeneity ($\rho = 0.2$), in the right panels $\rho = 0.9$. The first stage signal strength is $F = 5$ in the upper panel and $F = 20$ in the lower panel.

List of research reports

14001-OPERA: Germs, R. and N.D. van Foreest, Optimal control of production-inventory systems with constant and compound poisson demand

14002-EEF: Bao, T. and J. Duffy, Adaptive vs. eductive learning: Theory and evidence

14003-OPERA: Syntetos, A.A. and R.H. Teunter, On the calculation of safety stocks

14004-EEF: Bouwmeester, M.C., J. Oosterhaven and J.M. Rueda-Cantuche, Measuring the EU value added embodied in EU foreign exports by consolidating 27 national supply and use tables for 2000-2007

14005-OPERA: Prak, D.R.J., R.H. Teunter and J. Riezebos, Periodic review and continuous ordering

14006-EEF: Reijnders, L.S.M., The college gender gap reversal: Insights from a life-cycle perspective

14007-EEF: Reijnders, L.S.M., Child care subsidies with endogenous education and fertility

14008-EEF: Otter, P.W., J.P.A.M. Jacobs and A.H.J. den Reijer, A criterion for the number of factors in a data-rich environment

14009-EEF: Mierau, J.O. and E. Suari Andreu, Fiscal rules and government size in the European Union

14010-EEF: Dijkstra, P.T., M.A. Haan and M. Mulder, Industry structure and collusion with uniform yardstick competition: theory and experiments

14011-EEF: Huizingh, E. and M. Mulder, Effectiveness of regulatory interventions on firm behavior: a randomized field experiment with e-commerce firms

14012-GEM: Bressand, A., Proving the old spell wrong: New African hydrocarbon producers and the 'resource curse'

14013-EEF: Dijkstra P.T., Price leadership and unequal market sharing: Collusion in experimental markets

14014-EEF: Angelini, V., M. Bertoni, and L. Corazzini, Unpacking the determinants of life satisfaction: A survey experiment

14015-EEF: Heijdra, B.J., J.O. Mierau, and T. Trimborn, Stimulating annuity markets

14016-GEM: Bezemer, D., M. Grydaki, and L. Zhang, Is financial development bad for growth?

14017-EEF: De Cao, E. and C. Lutz, Sensitive survey questions: measuring attitudes regarding female circumcision through a list experiment

14018-EEF: De Cao, E., The height production function from birth to maturity

14019-EEF: Allers, M.A. and J.B. Geertsema, The effects of local government amalgamation on public spending and service levels. Evidence from 15 years of municipal boundary reform

14020-EEF: Kuper, G.H. and J.H. Veurink, Central bank independence and political pressure in the Greenspan era

14021-GEM: Samarina, A. and D. Bezemer, Do Capital Flows Change Domestic Credit Allocation?

14022-EEF: Soetevent, A.R. and L. Zhou, Loss Modification Incentives for Insurers Under ExpectedUtility and Loss Aversion
14023-EEF: Allers, M.A. and W. Vermeulen, Fiscal Equalization, Capitalization and the Flypaper Effect.

14024-GEM: Hoorn, A.A.J. van, Trust, Workplace Organization, and Comparative Economic Development.

14025-GEM: Bezemer, D., and L. Zhang, From Boom to Bust in de Credit Cycle: The Role of Mortgage Credit.

14026-GEM: Zhang, L., and D. Bezemer, How the Credit Cycle Affects Growth: The Role of Bank Balance Sheets.

14027-EEF: Bružikas, T., and A.R. Soetevent, Detailed Data and Changes in Market Structure: The Move to Unmanned Gasoline Service Stations.

14028-EEF: Bouwmeester, M.C., and B. Scholtens, Cross-border Spillovers from European Gas Infrastructure Investments.

14029-EEF: Lestano, and G.H. Kuper, Correlation Dynamics in East Asian Financial Markets.

14030-GEM: Bezemer, D.J., and M. Grydaki, Nonfinancial Sectors Debt and the U.S. Great Moderation.

14031-EEF: Hermes, N., and R. Lensink, Financial Liberalization and Capital Flight: Evidence from the African Continent.

14032-OPERA: Blok, C. de, A. Seepma, I. Roukema, D.P. van Donk, B. Keulen, and R. Otte, Digitalisering in Strafrechtketens: Ervaringen in Denemarken, Engeland, Oostenrijk en Estland vanuit een Supply Chain Perspectief.

14033-OPERA: Olde Keizer, M.C.A., and R.H. Teunter, Opportunistic condition-based maintenance and aperiodic inspections for a two-unit series system.

14034-EEF: Kuper, G.H., G. Sierksma, and F.C.R. Spieksma, Using Tennis Rankings to Predict Performance in Upcoming Tournaments

15001-EEF: Bao, T., X. Tian, X. Yu, Dictator Game with Indivisibility of Money

15002-GEM: Chen, Q., E. Dietzenbacher, and B. Los, The Effects of Ageing and Urbanization on China's Future Population and Labor Force

15003-EEF: Allers, M., B. van Ommeren, and B. Geertsema, Does intermunicipal cooperation create inefficiency? A comparison of interest rates paid by intermunicipal organizations, amalgamated municipalities and not recently amalgamated municipalities

15004-EEF: Dijkstra, P.T., M.A. Haan, and M. Mulder, Design of Yardstick Competition and Consumer Prices: Experimental Evidence

15005-EEF: Dijkstra, P.T., Price Leadership and Unequal Market Sharing: Collusion in Experimental Markets

15006-EEF: Anufriev, M., T. Bao, A. Sutin, and J. Tuinstra, Fee Structure, Return Chasing and Mutual Fund Choice: An Experiment

15007-EEF: Lamers, M., Depositor Discipline and Bank Failures in Local Markets During the Financial Crisis

15008-EEF: Oosterhaven, J., On de Doubtful Usability of the Inoperability IO Model

15009-GEM: Zhang, L. and D. Bezemer, A Global House of Debt Effect? Mortgages and Post-Crisis Recessions in Fifty Economies

15010-I&O: Hooghiemstra, R., N. Hermes, L. Oxelheim, and T. Randøy, The Impact of Board Internationalization on Earnings Management

15011-EEF: Haan, M.A., and W.H. Siekman, Winning Back the Unfaithful while Exploiting the Loyal: Retention Offers and Heterogeneous Switching Costs

15012-EEF: Haan, M.A., J.L. Moraga-González, and V. Petrikaite, Price and Match-Value Advertising with Directed Consumer Search

15013-EEF: Wiese, R., and S. Eriksen, Do Healthcare Financing Privatisations Curb Total Healthcare Expenditures? Evidence from OECD Countries

15014-EEF: Siekman, W.H., Directed Consumer Search

15015-GEM: Hoorn, A.A.J. van, Organizational Culture in the Financial Sector: Evidence from a Cross-Industry Analysis of Employee Personal Values and Career Success

15016-EEF: Te Bao, and C. Hommes, When Speculators Meet Constructors: Positive and Negative Feedback in Experimental Housing Markets

15017-EEF: Te Bao, and Xiaohua Yu, Memory and Discounting: Theory and Evidence

15018-EEF: Suari-Andreu, E., The Effect of House Price Changes on Household Saving Behaviour: A Theoretical and Empirical Study of the Dutch Case

15019-EEF: Bijlsma, M., J. Boone, and G. Zwart, Community Rating in Health Insurance: Trade-off between Coverage and Selection

15020-EEF: Mulder, M., and B. Scholtens, A Plant-level Analysis of the Spill-over Effects of the German *Energiewende*

15021-GEM: Samarina, A., L. Zhang, and D. Bezemer, Mortgages and Credit Cycle Divergence in Eurozone Economies

16001-GEM: Hoorn, A. van, How Are Migrant Employees Manages? An Integrated Analysis

16002-EEF: Soetevent, A.R., Te Bao, A.L. Schippers, A Commercial Gift for Charity

16003-GEM: Bouwmeerster, M.C., and J. Oosterhaven, Economic Impacts of Natural Gas Flow Disruptions

16004-MARK: Holtrop, N., J.E. Wieringa, M.J. Gijsenberg, and P. Stern, Competitive Reactions to Personal Selling: The Difference between Strategic and Tactical Actions

16005-EEF: Plantinga, A. and B. Scholtens, The Financial Impact of Divestment from Fossil Fuels

16006-GEM: Hoorn, A. van, Trust and Signals in Workplace Organization: Evidence from Job Autonomy Differentials between Immigrant Groups

16007-EEF: Willems, B. and G. Zwart, Regulatory Holidays and Optimal Network Expansion

16008-GEF: Hoorn, A. van, Reliability and Validity of the Happiness Approach to Measuring Preferences

16009-EEF: Hinloopen, J., and A.R. Soetevent, (Non-)Insurance Markets, Loss Size Manipulation and Competition: Experimental Evidence

16010-EEF: Bekker, P.A., A Generalized Dynamic Arbitrage Free Yield Model

16011-EEF: Mierau, J.A., and M. Mink, A Descriptive Model of Banking and Aggregate Demand

16012-EEF: Mulder, M. and B. Willems, Competition in Retail Electricity Markets: An Assessment of Ten Year Dutch Experience

16013-GEM: Rozite, K., D.J. Bezemer, and J.P.A.M. Jacobs, Towards a Financial Cycle for the US, 1873-2014

16014-EEF: Neuteleers, S., M. Mulder, and F. Hindriks, Assessing Fairness of Dynamic Grid Tariffs

16015-EEF: Soetevent, A.R., and T. Bružikas, Risk and Loss Aversion, Price Uncertainty and the Implications for Consumer Search

16016-HRM&OB: Meer, P.H. van der, and R. Wielers, Happiness, Unemployment and Self-esteem

16017-EEF: Mulder, M., and M. Pangan, Influence of Environmental Policy and Market Forces on Coal-fired Power Plants: Evidence on the Dutch Market over 2006-2014

16018-EEF: Zeng,Y., and M. Mulder, Exploring Interaction Effects of Climate Policies: A Model Analysis of the Power Market

16019-EEF: Ma, Yiqun, Demand Response Potential of Electricity End-users Facing Real Time Pricing

16020-GEM: Bezemer, D., and A. Samarina, Debt Shift, Financial Development and Income Inequality in Europe

16021-EEF: Elkhuizen, L, N. Hermes, and J. Jacobs, Financial Development, Financial Liberalization and Social Capital

16022-GEM: Gerritse, M., Does Trade Cause Institutional Change? Evidence from Countries South of the Suez Canal

16023-EEF: Rook, M., and M. Mulder, Implicit Premiums in Renewable-Energy Support Schemes

17001-EEF: Trinks, A., B. Scholtens, M. Mulder, and L. Dam, Divesting Fossil Fuels: The Implications for Investment Portfolios

17002-EEF: Angelini, V., and J.O. Mierau, Late-life Health Effects of Teenage Motherhood

17003-EEF: Jong-A-Pin, R., M. Laméris, and H. Garretsen, Political Preferences of (Un)happy Voters: Evidence Based on New Ideological Measures

17004-EEF: Jiang, X., N. Hermes, and A. Meesters, Financial Liberalization, the Institutional Environment and Bank Efficiency

17005-EEF: Kwaak, C. van der, Financial Fragility and Unconventional Central Bank Lending Operations

17006-EEF: Postelnicu, L. and N. Hermes, The Economic Value of Social Capital

17007-EEF: Ommeren, B.J.F. van, M.A. Allers, and M.H. Vellekoop, Choosing the Optimal Moment to Arrange a Loan

17008-EEF: Bekker, P.A., and K.E. Bouwman, A Unified Approach to Dynamic Mean-Variance Analysis in Discrete and Continuous Time

17009-EEF: Bekker, P.A., Interpretable Parsimonious Arbitrage-free Modeling of the Yield Curve

17010-GEM: Schasfoort, J., A. Godin, D. Bezemer, A. Caiani, and S. Kinsella, Monetary Policy Transmission in a Macroeconomic Agent-Based Model

17011-I&O: Bogt, H. ter, Accountability, Transparency and Control of Outsourced Public Sector Activities

17012-GEM: Bezemer, D., A. Samarina, and L. Zhang, The Shift in Bank Credit Allocation: New Data and New Findings

17013-EEF: Boer, W.I.J. de, R.H. Koning, and J.O. Mierau, Ex-ante and Ex-post Willingness-to-pay for Hosting a Major Cycling Event

17014-OPERA: Laan, N. van der, W. Romeijnders, and M.H. van der Vlerk, Higher-order Total Variation Bounds for Expectations of Periodic Functions and Simple Integer Recourse Approximations

17015-GEM: Oosterhaven, J., Key Sector Analysis: A Note on the Other Side of the Coin

17016-EEF: Romensen, G.J., A.R. Soetevent: Tailored Feedback and Worker Green Behavior: Field Evidence from Bus Drivers

17017-EEF: Trinks, A., G. Ibikunle, M. Mulder, and B. Scholtens, Greenhouse Gas Emissions Intensity and the Cost of Capital

17018-GEM: Qian, X. and A. Steiner, The Reinforcement Effect of International Reserves for Financial Stability

17019-GEM/EEF: Klasing, M.J. and P. Milionis, The International Epidemiological Transition and the Education Gender Gap

2018001-EEF: Keller, J.T., G.H. Kuper, and M. Mulder, Mergers of Gas Markets Areas and Competition amongst Transmission System Operators: Evidence on Booking Behaviour in the German Markets

2018002-EEF: Soetevent, A.R. and S. Adikyan, The Impact of Short-Term Goals on Long-Term Objectives: Evidence from Running Data

2018003-MARK: Gijsenberg, M.J. and P.C. Verhoef, Moving Forward: The Role of Marketing in Fostering Public Transport Usage

2018004-MARK: Gijsenberg, M.J. and V.R. Nijs, Advertising Timing: In-Phase or Out-of-Phase with Competitors?

2018005-EEF: Hulshof, D., C. Jepma, and M. Mulder, Performance of Markets for European Renewable Energy Certificates

2018006-EEF: Fosgaard, T.R., and A.R. Soetevent, Promises Undone: How Committed Pledges Impact Donations to Charity

2018007-EEF: Durán, N. and J.P. Elhorst, A Spatio-temporal-similarity and Common Factor Approach of Individual Housing Prices: The Impact of Many Small Earthquakes in the North of Netherlands

2018008-EEF: Hermes, N., and M. Hudon, Determinants of the Performance of Microfinance Institutions: A Systematic Review

2018009-EEF: Katz, M., and C. van der Kwaak, The Macroeconomic Effectiveness of Bank Bail-ins

2018010-OPERA: Prak, D., R.H. Teunter, M.Z. Babai, A.A. Syntetos, and J.E. Boylan, Forecasting and Inventory Control with Compound Poisson Demand Using Periodic Demand Data

2018011-EEF: Brock, B. de, Converting a Non-trivial Use Case into an SSD: An Exercise

2018012-EEF: Harvey, L.A., J.O. Mierau, and J. Rockey, Inequality in an Equal Society

2018013-OPERA: Romeijnders, W., and N. van der Laan, Inexact cutting planes for two-stage mixed-integer stochastic programs

2018014-EEF: Green, C.P., and S. Homroy, Bringing Connections Onboard: The Value of Political Influence

2018015-OPERA: Laan, N. van der, and W. Romeijnders, Generalized aplha-approximations for two-stage mixed-integer recourse models

2018016-GEM: Rozite, K., Financial and Real Integration between Mexico and the United States

2019001-EEF: Lugalla, I.M., J. Jacobs, and W. Westerman, Drivers of Women Entrepreneurs in Tourism in Tanzania: Capital, Goal Setting and Business Growth

2019002-EEF: Brock, E.O. de, On Incremental and Agile Development of (Information) Systems

2019003-OPERA: Laan, N. van der, R.H. Teunter, W. Romeijnders, and O.A. Kilic, The Data-driven Newsvendor Problem: Achieving On-target Service Levels.

2019004-EEF: Dijk, H., and J. Mierau, Mental Health over the Life Course: Evidence for a U-Shape?

2019005-EEF: Freriks, R.D., and J.O. Mierau, Heterogeneous Effects of Conditional School Resources on Health and Test Scores: Evidence from the Netherlands.

2019006-OPERA: Broek, M.A.J. uit het, R.H. Teunter, B. de Jonge, J. Veldman, Joint Condition-based Maintenance and Condition-based Production Optimization.

2019007-OPERA: Broek, M.A.J. uit het, R.H. Teunter, B. de Jonge, J. Veldman, Joint Condition-based Maintenance and Load-sharing Optimization for Multi-unit Systems with Economic Dependency

2019008-EEF: Keller, J.T. G.H. Kuper, and M. Mulder, Competition under Regulation: Do Regulated Gas Transmission System Operators in Merged Markets Compete on Network Tariffs?

2019009-EEF: Hulshof, D. and M. Mulder, Renewable Energy Use as Environmental CSR Behavior and the Impact on Firm Profit

2019010-EEF: Boot, T., Confidence Regions for Averaging Estimators www.rug.nl/feb