JMP Revisions Following Econometrica Comments

Francis J. DiTraglia

November 7, 2013

Abstract

This document collects all of the new material that I plan to add to my JMP following the referee reports from Econometrica. Once I've put this material together, I'll merge it with the existing version and send it to ReStud.

Referee Comments and My Responses

The two main objections to my paper were as follows:

1. The FMSC provides an asymptotically unbiased estimate of MSE, but the FMSC itself is random, even in the limit. It is not true that the FMSC converges in probability to actual AMSE because of this additional randomness.

2. If the valid model is identified, why not just use it? If it's not identified, FMSC doesn't work. It's not clear that the gains from this procedure would be large in practice since the FMSC has a variance. See point # 1.

3. This approach is not robust to weak identification.

4. The procedure for correcting confidence intervals is too computationally intensive and aims to get the size correct without even looking at the width.

5. Finite sample MSE doesn't always exist for these kinds of models. Maybe this isn't the best measure to look at. What about OLS versus IV?

6. There are some missing references.

7. Some objections to the simulation experiment. In particular, no widths are reported for confidence intervals.

The most important objections are 1 and 2, possibly followed by 3. The other points can be fairly easily dealt with and I will do so below.

Objection # 1 This comment boils down to "I don't like efficient model selection: you should use consistent model selection." It is indeed true that the FMSC is random even in the limit and that it does not converge to actual AMSE: it is merely an asymptotically unbiased estimator of AMSE. However, this is the whole point: I explicitly want to *avoid* doing consistent selection because the associated risk properties are so poor. I need to make this clearer by citing, for example, the Leeb and Pötscher paper from 2008 that Larry Wasserman discussed on his blog.

In a certain sense, this comment amounts to a critique of my use of local asymptotics. I should make very clear that this device is widely used in econometrics to study, among other things, local power, local-to-unit roots, etc. I should provide references for this, including Schorfheide and Moon. Furthermore, a great many model selection procedures are random in the limit: Mallows C_p and AIC are two well-known examples, but many people have worked with selection in this framework. I should cite Bruce Hansen as well as Frank Schorfheide's VAR paper and Schorfheide and Moon.

I should also relate my framework to the idea of uniform asymptotic validity as discussed in Andrews and Guggenberger (2010) and Schorfheide and Moon.

Objection # 2 This comment suggests that: (a) my simulations weren't convincing enough, and (b) I wasn't clear enough about the sort of situation for which the FMSC is designed. The solution is to do more and better simulations and to be clearer! The point of the FMSC is that we often find ourselves in a setting where there are various "plausible" assumptions we could use in estimation, some of which are weaker and some of which are stronger. Typically, we worry that the weaker assumptions might not provide sufficient information to study the question we're interested in, which we worry that the stronger assumptions might not quite be true. You can think of this as a kind of "prior knowledge" that violations of the stronger assumptions are "small" which is pretty much exactly what the local mis-specification idea encodes. I can relate this to the idea of "plausibly exogenous" as well as Schorfheide and Moon.

I should be clear about the fact, and I need to find the citation for this, that model selection cannot uniformly beat the "full" model. (In this case, the full model is the set of moment conditions based on the weakest assumptions.) However, selection *can* beat the "full" model over large regions of the parameter space, so it's really a questino of where you think you might be a priori. I should argue that the whole idea is to use

my method when you consider it likely that you might be in the relevant region of the parameter space. I should also show in simulations that the cost you pay when you are *not* in this region isn't too high.

Objection # 3 This point is less important but also harder to handle. I'm pretty sure that there's no straightforward way to include weak identification in my framework directly although it's a very compelling idea: when you have weak identification it might make a lot of sense to use a slightly endogenous instrument. Even though I don't think it's possible to incorporate weak instrument *asymptotics*, however, I can still evaluate how my proposed *procedure* deals with weak instruments. There are at least two ways to do this. The first is by carrying out a simulation study. The second is by looking trying to relate the FMSC to some other well-known tests or procedures. In the IV versus OLS case, for example, I know that the result is a Hausman test with a non-standard critical value. I seem to recall that Dufour has a paper in which he argues that this is a good idea when you might have weak instruments. Basically the idea here is to look at how the FMSC implicitly trades off instrument strength and validity when we take the procedure *outside* of its asymptotic framework.

Overall Thoughts and To Do List Besides responding to the Objections 1–3, here are some other things I should do, roughly in order of importance:

1. Should defend my assumption that we have a minimal set of correct MCs by referencing the literature, including Chen etc.

2. Change the notation to allow us to use arbitrary subsets of the moment conditions. This will accomodate the OLS/IV example.

3. Allow the weighting matrix W to be indexed by S.

4. Fold in the IV/OLS example. This will allow me to discuss a number of interesting points, including optimal estimator averaging, weak instruments, relationship to testing, etc. Also allows a brief consideration of what happens when we have no valid moment conditions.

5. Do a better job with the choosing instruments example.

6. New and better simulation experiments. Look at median absolute deviation as well as trimmed MSE, etc. Try to cover more of the parameter space, etc. Pictures rather than tables. Everything needs to be replicable for ReStud!

7. Re-do the empirical example with improved code for the confidence interval. Everything needs to be replicable for ReStud!

8. Possibly add a second empirical example for OLS versus IV.

9. Look at the proposed references and think about including them.

10. Might want to be slightly more careful about regularity conditions. See for example Schorfheide and Moon.

1 New Notation for Moment Selection Vector

This is straightforward: just need to redefine S and Ξ_S. See the GFIC paper.

2 Two Running Examples in the Paper

Two simple but empirically relevant examples we'll consider throughout the paper. Helpful because they make the intuition clear and also interesting in their own right. To be clear, FMSC applies to GMM in general, not just to linear models like these.

Example #1: Choosing Instrumental Variables Consider the linear model

$$y_i = \mathbf{x}_i'\beta + \epsilon_i \tag{1}$$

where some or all of the regressors are endogenous. Suppose we have a vector of valid instruments $\mathbf{z}_i^{(1)}$ and another vector of "suspect" instruments $\mathbf{w}_i$ that are likely to be highly relevant but may well be slightly endogenous. These could be "plausibly exogenous." Should we include $\mathbf{z}_i$ in the instrument set for use in estimation? Arises in various settings. One concerns exogeneity assumptions in a panel data setting: strict exogeneity versus predeterminedness.

Example #2: Least Squares versus Instrumental Variables This example is similar to the first one, but illustrates that we don't have to structure the problem in terms of choosing over-identifying restrictions: we can select over fundamentally different estimators. Suppose we want to estimate the effect of an endogenous regressor x in a linear model of the form

$$y_i = \mathbf{w}_i'\theta + \beta x_i + \epsilon_i \tag{2}$$

where $\mathbf{w}_i$ is a vector of endogenous control regressors. In this case the target parameter is β. Suppose we have a vector of valid instruments $\mathbf{z}_i$. Should we use OLS or IV? Dufour: "IV is like Amputation; it should be a last resort to save the patient." Easily generalized to more than one endogenous regressor, but this example is very common in many settings, particularly treatment effects with microdata. Cite the Nevo and Rosen and Plausibly Exogenous papers. Another issue is weak instruments and whether it is possible to *combine* OLS and IV.

3 Local Mis-specification for the Examples

Example #1: Choosing Instrumental Variables

$$E_n \left[\begin{array}{c} \mathbf{w}_i(y_i - \mathbf{x}_i\beta) \\ \mathbf{z}_i(y_i - \mathbf{x}_i\beta) \end{array} \right] = \left[\begin{array}{c} \mathbf{0} \\ \boldsymbol{\tau}/\sqrt{n} \end{array} \right] \tag{3}$$

Example #2: Least Squares versus Instrumental Variables

$$E_n \left[\begin{array}{c} \mathbf{w}_i(y_i - \mathbf{w}_i'\theta - \beta x_i) \\ \mathbf{z}_i(y_i - \mathbf{w}_i'\theta - \beta x_i) \\ x_i(y_i - \mathbf{w}_i'\theta - \beta x_i) \end{array} \right] = \left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \\ \tau/\sqrt{n} \end{array} \right] \tag{4}$$

In our consideration of this example in the remainder of the paper we will assume, without loss of generality, that there are no exogenous regressors $\mathbf{w}_i$. If there are any, they can always be "projected out" of y, x and $\mathbf{z}$.

4 Lemma: A CLT for Local Mis-specification

Covers both of the examples from the paper in an iid setting (microdata, panel, etc.) Can be extended to handle dependence by using something other than Lindeberg-Feller.

Lemma 4.1 (CLT Under Local Mis-specification)**.** *Let* $\{\mathbf{w}_i, \mathbf{z}_i, \epsilon_i \colon 1 \leq i \leq n, n = 1, 2, \ldots\}$ *be a triangular array of random variables such that*

(a) $(\mathbf{w}_i, \mathbf{z}_i, \epsilon_i) \sim$ *iid within each row of the array (i.e. for fixed* n*)*

(b) $E_n[\epsilon_i \mathbf{w}_i] = \mathbf{0}$

(c) $E_n[\epsilon_i \mathbf{z}_i] = \boldsymbol{\tau}/\sqrt{n}$

(d) $E_n\left[\left|\epsilon_i \mathbf{w}_i\right|^{2+\eta}\right] < C$ *and* $E_n\left[\left|\epsilon_i \mathbf{z}_i\right|^{2+\eta}\right] < C$ *for all* n *and some* $\eta > 0$ *where* $C < \infty$.

(e) $\Omega = \lim_{n\to\infty} E_n \begin{bmatrix} \epsilon_i^2 \mathbf{w}_i \mathbf{w}_i' & \epsilon_i^2 \mathbf{w}_i \mathbf{z}_i' \\ \epsilon_i^2 \mathbf{z}_i \mathbf{w}_i' & \epsilon_i^2 \mathbf{z}_i \mathbf{z}_i' \end{bmatrix}$ *exists and is both finite and positive definite.*

Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} \epsilon_i \mathbf{w}_i \\ \epsilon_i \mathbf{z}_i \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} \mathbf{0} \\ \boldsymbol{\tau} \end{bmatrix}, \Omega \right)$$

Proof. We proceed by verifying the conditions of the Lindeberg-Feller Central Limit Theorem (van der Vaart 2.27) for the sequence of random vectors $Y_{ni} = n^{-1/2}\left(\epsilon_i \mathbf{w}_i', \epsilon_i \mathbf{z}_i'\right)'$. Since each row of the triangular array contains n iid observations,

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} Var_n(Y_{ni}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1^n} Var_n(Y_{ni}) = \lim_{n\to\infty} Var_n(Y_{ni})$$

and under the local mis-specification assumption,

$$Var_n(Y_{ni}) = E_n \begin{bmatrix} \epsilon_i^2 \mathbf{w}_i \mathbf{w}_i' & \epsilon_i^2 \mathbf{w}_i \mathbf{z}_i' \\ \epsilon_i^2 \mathbf{z}_i \mathbf{w}_i' & \epsilon_i^2 \mathbf{z}_i \mathbf{z}_i' \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\tau}\boldsymbol{\tau}'/n \end{bmatrix} \to \Omega.$$

Similarly, the sum from the Lindeberg Condition simplifies to

$$\sum_{i=1}^{k_n} E\left[\left|Y_{ni}\right|^2 \mathbf{1}\left\{\left|Y_{ni}\right| > \xi\right\}\right] = E_n\left[\epsilon_i^2 \left(\left|\mathbf{w}_i\right|^2 + \left|\mathbf{z}_i\right|^2\right) \mathbf{1}\left\{A_n\right\}\right] \to 0$$

as $n \to \infty$ for any $\xi > 0$, where

$$A_n = \left\{ \left|\epsilon_i\right| \left(\left|\mathbf{w}_i\right|^2 + \left|\mathbf{z}_i\right|^2\right)^{1/2} > n^{1/2}\xi \right\}$$

Now, by Hölder's Inequality followed by Minkowski's Inequality (White 3.4, 3.11)

$$\begin{aligned} E_n\left[\epsilon_i^2 \left(\left|\mathbf{w}_i\right|^2 + \left|\mathbf{z}_i\right|^2\right) \mathbf{1}\left\{A_n\right\}\right] &= E_n\left[\left(\left|\epsilon_i \mathbf{w}_i\right|^2 + \left|\epsilon_i \mathbf{z}_i\right|^2\right) \mathbf{1}\left\{A_n\right\}\right] \\ &\leq E_n\left[\left|\left|\epsilon_i \mathbf{w}_i\right|^2 + \left|\epsilon_i \mathbf{z}_i\right|^2\right|^p\right]^{1/p} E_n\left[\left|\mathbf{1}\{A_n\}\right|^q\right]^{1/q} \\ &= E_n\left[\left|\left|\epsilon_i \mathbf{w}_i\right|^2 + \left|\epsilon_i \mathbf{z}_i\right|^2\right|^p\right]^{1/p} P(A_n)^{1/q} \\ &\leq \left(E_n\left[\left|\epsilon_i \mathbf{w}_i\right|^{2p}\right]^{1/p} + E_n\left[\left|\epsilon_i \mathbf{z}_i\right|^{2p}\right]^{1/p}\right) P(A_n)^{1/q} \end{aligned}$$

provided that $p > 1$, $1/p + 1/q = 1$ and all the relevant moments exist. By the Generalized Chebyshev Inequality (White 2.41)

$$\begin{aligned} P(A_n) &\leq \left(\frac{1}{n\xi^2}\right) E_n\left[\epsilon_i^2 \left(\left|\mathbf{w}_i\right|^2 + \left|\mathbf{z}_i\right|^2\right)\right] \\ &= \left(\frac{1}{n\xi^2}\right) E_n\left[\left|\epsilon_i \mathbf{w}_i\right|^2 + \left|\epsilon_i \mathbf{z}_i\right|^2\right] \leq \frac{2C}{n\xi^2} \end{aligned}$$

Combining these and taking $p = 1 + \eta$, $q = (1 + \eta)/\eta$, we have

$$E_n\left[\epsilon_i^2\left(|\mathbf{w}_i|^2 + |\mathbf{z}_i|^2\right)\mathbf{1}\left\{A_n\right\}\right] \;\leq\; 2C^{\frac{1}{1+\eta}}\left(\frac{2C}{n\xi^2}\right)^{\frac{\eta}{\eta+1}} \to 0.$$

Therefore, by the Lindeberg-Feller Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\begin{array}{c}\epsilon_i\mathbf{w}_i\\ \epsilon_i\mathbf{z}_i - \boldsymbol{\tau}/\sqrt{n}\end{array}\right] = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\begin{array}{c}\epsilon_i\mathbf{w}_i\\ \epsilon_i\mathbf{z}_i\end{array}\right] - \left[\begin{array}{c}\mathbf{0}\\ \boldsymbol{\tau}\end{array}\right]\right) \xrightarrow{d} N(\mathbf{0}, \Omega).$$

□

5 Derivations for OLS vs. IV Example

Without loss of generality, we may assume that there are no exogenous regressors or, equivalently, that they have been "projected out." Thus, the DGP is:

$$y_i = \beta x_i + \epsilon_i \tag{5}$$
$$x_i = \mathbf{z}_i'\boldsymbol{\pi} + v_i \tag{6}$$

and the local mis-specification assumption becomes

$$E_n\left[\begin{array}{c}\mathbf{z}_i\epsilon_i\\ x_i\epsilon_i\end{array}\right] = \left[\begin{array}{c}\mathbf{0}\\ \tau/\sqrt{n}\end{array}\right]. \tag{7}$$

Stacking observations,

$$\mathbf{y} = \mathbf{x}\beta + \boldsymbol{\epsilon} \tag{8}$$
$$\mathbf{x} = Z\boldsymbol{\pi} + \mathbf{v} \tag{9}$$

where $Z' = (\mathbf{z}_1, \ldots, \mathbf{z}_n), \mathbf{x}' = (x_1, \ldots, x_n)$ and so on. We consider two estimators of the scalar β: the ordinary least squares (OLS) estimator $\widehat{\beta}$ and the Generalized Instrumental Variable (GIV) estimator $\widetilde{\beta}$

$$\widehat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \tag{10}$$
$$\widetilde{\beta} = (\mathbf{x}'ZW_nZ'\mathbf{x})^{-1}\mathbf{x}'ZW_nZ'\mathbf{y} \tag{11}$$

where W_n is a positive definite weighting matrix.

Theorem 5.1. *Suppose that*

(a) The DGP with triangular array.

(b) Local Mis-specification.

(c) Fourth and a bit moments of everything.

(d) $W_n \xrightarrow{p} W$*, a positive definite matrix.*

(e) Limit of variance for v_i *exists and is positive.*

(f) The limit $\lim_{n\to\infty} E_n[\mathbf{z}_i\mathbf{z}_i'] = Q_z$ *exists and is positive definite*

(g) The instruments $\mathbf{z}_i$ *are relevant:* $|\boldsymbol{\pi}| > 0$

Then,

$$\begin{bmatrix} \sqrt{n}\left(\widehat{\beta}-\beta\right) \\ \sqrt{n}\left(\widetilde{\beta}-\beta\right) \\ \widehat{\tau} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \widehat{K} & \mathbf{0}' \\ 0 & \widetilde{K} \\ 1 & -\widehat{K}^{-1}\widetilde{K} \end{bmatrix} \left(\begin{bmatrix} \tau \\ \mathbf{0} \end{bmatrix} + M \right)$$

where $M \sim \mathcal{N}(0,\Omega)$ *and*

$$\widehat{K} = \left(\boldsymbol{\pi}' Q_z \boldsymbol{\pi} + \sigma_v^2\right)^{-1} \tag{12}$$

$$\widetilde{K} = \left(\boldsymbol{\pi}' Q_z W Q_z \boldsymbol{\pi}\right)^{-1} \boldsymbol{\pi}' Q_z W \tag{13}$$

$$\Omega = \lim_{n\to\infty} E_n \begin{bmatrix} \epsilon_i^2 x_i^2 & \epsilon_i^2 x_i \mathbf{z}_i' \\ \epsilon_i^2 x_i \mathbf{z}_i & \epsilon_i^2 \mathbf{z}_i \mathbf{z}_i' \end{bmatrix} \tag{14}$$

Proof. Substituting the DGP and rearranging, $\sqrt{n}\left(\widehat{\beta}-\beta\right) = \widehat{K}_n \boldsymbol{m}_n$ and similarly $\sqrt{n}\left(\widetilde{\beta}-\beta\right) = \widetilde{K}_n \boldsymbol{m}_n$, where

$$\begin{aligned} \widehat{K}_n &= \left(\mathbf{x}'\mathbf{x}/n\right)^{-1} \\ \widetilde{K}_n &= \left[\left(\mathbf{x}'Z/n\right) W_n \left(Z'\mathbf{x}/n\right)\right]^{-1} \left(\mathbf{x}'Z/n\right) W_n \\ \boldsymbol{m}_n &= \begin{bmatrix} \mathbf{x}'\boldsymbol{\epsilon}/\sqrt{n} \\ Z'\boldsymbol{\epsilon}/\sqrt{n} \end{bmatrix} \end{aligned}$$

Moreover,

$$\begin{aligned} \widehat{\tau} &= \sqrt{n}\left[\mathbf{x}'(\mathbf{y}-\mathbf{x}\widetilde{\beta})/n\right] = \mathbf{x}'\boldsymbol{\epsilon}/\sqrt{n} - (\mathbf{x}'\mathbf{x}/n)\sqrt{n}(\widetilde{\beta}-\beta) \\ &= \begin{bmatrix} 1 & \widehat{K}_n^{-1}\widetilde{K}_n \end{bmatrix} \boldsymbol{m}_n \end{aligned}$$

Since

$$\begin{aligned}\frac{\mathbf{x}'\mathbf{x}}{n} &= \boldsymbol{\pi}'\left(\frac{Z'Z}{n}\right)\boldsymbol{\pi} + \left(\frac{\mathbf{v}'Z}{n}\right)\boldsymbol{\pi} + \boldsymbol{\pi}'\left(\frac{Z'\mathbf{v}}{n}\right) + \frac{\mathbf{v}'\mathbf{v}}{n} \\ \frac{\mathbf{x}'Z}{n} &= \boldsymbol{\pi}'\left(\frac{Z'Z}{n}\right) + \frac{\mathbf{v}'Z}{n}\end{aligned}$$

it suffices to examine $Z'Z/n$, $Z'\mathbf{v}/n$ and $\mathbf{v}'\mathbf{v}/n$. Because the triangular array is iid in each row, uniformly bounded fourth moments are sufficient for L_2 convergence (see e.g. Davidson 19.1) which implies convergence in probability. Thus,

$$\begin{aligned}Z'Z/n &\xrightarrow{p} \lim_{n\to\infty} E_n[\mathbf{z}_i\mathbf{z}_i'] = Q_z \\ Z'\mathbf{v}/z &\xrightarrow{p} \lim_{n\to\infty} E_n[\mathbf{z}_i v_i] = 0 \\ \mathbf{v}'\mathbf{v}/n &\xrightarrow{p} \lim_{n\to\infty} E_n[v_i^2] = \sigma_v^2\end{aligned}$$

It follows that $\widehat{K}_n \xrightarrow{p} \widehat{K}$ and $\widetilde{K}_n \xrightarrow{p} \widetilde{K}$. Finally, we apply Lemma 4.1 to $\boldsymbol{m}_n$. The required bounds $E_n\left[|\epsilon_i\mathbf{z}_i|^{2+\eta}\right] < C$ and $E_n\left[|\epsilon_i x_i|^{2+\eta}\right] < C$ follow from our assumptions on the moments of v_i, ϵ_i and $\mathbf{z}_i$ by Minkowski's Inequality and the Cauchy-Schwarz Inequality. □

5.1 A Simplification - 2SLS

Preceding result covers any iid setting. To get more intuition and compare to well-known procedures consider a simplification so that 2SLS is the efficient GIV estimator and OLS is fully efficient.

Corollary 5.1. *Suppose that*

(a) $W_n = (Z'Z/n)^{-1}$

(b) $\sigma_\epsilon^2 = \lim_{n\to\infty} E_n[\epsilon_i^2]$ *exists and is positive*

(c) $\Omega = \sigma_\epsilon^2\left(\lim_{n\to\infty} E_n\begin{bmatrix} x_i^2 & x_i\mathbf{z}_i' \\ x_i\mathbf{z}_i & \mathbf{z}_i\mathbf{z}_i' \end{bmatrix}\right)$

Then, under the conditions of Theorem 5.1

$$\begin{bmatrix} \sqrt{n}\left(\widehat{\beta}-\beta\right) \\ \sqrt{n}\left(\widetilde{\beta}-\beta\right) \\ \widehat{\tau} \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} \tau/\sigma_x^2 \\ 0 \\ \tau \end{bmatrix}, \sigma_\epsilon^2\begin{bmatrix} 1/\sigma_x^2 & 1/\sigma_x^2 & 0 \\ 1/\sigma_x^2 & 1/\gamma^2 & 1-1/(\gamma^2\sigma_x^2) \\ 0 & 1-1/(\gamma^2\sigma_x^2) & \sigma_x^2(\sigma_x^2/\gamma^2-1) \end{bmatrix}\right)$$

where $\sigma_x^2 = \gamma^2 + \sigma_v^2$ *and* $\gamma^2 = \boldsymbol{\pi}'Q_z\boldsymbol{\pi}$.

Proof. First,

$$\Omega = \sigma_\epsilon^2 \begin{bmatrix} \boldsymbol{\pi}' Q_z \boldsymbol{\pi} + \sigma_v^2 & \boldsymbol{\pi}' Q_z \\ Q_z \boldsymbol{\pi} & Q_z \end{bmatrix} = \sigma_\epsilon^2 \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \sigma_\epsilon^2 V$$

Since $W_n = (Z'Z/n)^{-1} \xrightarrow{p} Q_z^{-1}$, we have $\widetilde{K} = (\boldsymbol{\pi}' Q_z \boldsymbol{\pi})^{-1} \boldsymbol{\pi}'$. Defining

$$\Sigma = \begin{bmatrix} \widehat{K} & \mathbf{0}' \\ \mathbf{0} & \widetilde{K} \\ 1 & -\widehat{K}^{-1}\widetilde{K} \end{bmatrix}$$

we have,

$$\begin{aligned} \sigma_\epsilon^2 \Sigma V \Sigma' &= \sigma_\epsilon^2 \begin{bmatrix} \widehat{K} & \mathbf{0}' \\ \mathbf{0} & \widetilde{K} \\ 1 & -\widehat{K}^{-1}\widetilde{K} \end{bmatrix} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \begin{bmatrix} \widehat{K}' & \mathbf{0}' & 1 \\ \mathbf{0} & \widetilde{K}' & -\widetilde{K}\widehat{K}^{-1} \end{bmatrix} \\ &= \sigma_\epsilon^2 \begin{bmatrix} 1/\sigma_x^2 & 1/\sigma_x^2 & 0 \\ 1/\sigma_x^2 & 1/\gamma^2 & 1 - 1/(\gamma^2\sigma_x^2) \\ 0 & 1 - 1/(\gamma^2\sigma_x^2) & \sigma_x^2(\sigma_x^2/\gamma^2 - 1) \end{bmatrix}. \end{aligned}$$

□

5.2 Infeasible FMSC: OLS versus 2SLS

We see that the variance of the OLS estimator is always strictly lower than that of the 2SLS estimator since $\sigma_\epsilon^2/\sigma_x^2 = \sigma_\epsilon^2/(\gamma^2 + \sigma_v^2)$. The AMSE of the OLS and 2SLS estimators takes a particularly simple form:

$$\text{AMSE(OLS)} = \frac{\tau^2}{\sigma_x^4} + \frac{\sigma_\epsilon^2}{\sigma_x^2} \tag{15}$$

$$\text{AMSE(2SLS)} = \frac{\sigma_\epsilon^2}{\gamma^2} \tag{16}$$

where $\sigma_x^2 = \gamma^2 + \sigma_v^2$ and $\gamma^2 = \boldsymbol{\pi}' Q_z \boldsymbol{\pi}$. The AMSE of the OLS estimator is lower than that of the OLS estimator when

$$\begin{aligned}
\frac{\tau^2}{\sigma_x^4} + \frac{\sigma_\epsilon^2}{\sigma_x^2} &< \frac{\sigma_\epsilon^2}{\gamma^2} \\
\frac{\tau^2}{\sigma_\epsilon^2 \sigma_x^2} + 1 &< \frac{\sigma_x^2}{\gamma^2} \\
\tau^2 &< \sigma_\epsilon^2 \sigma_x^2 \left(\frac{\sigma_x^2}{\gamma^2} - 1 \right) \\
\tau^2 &< \sigma_x^2 \sigma_\epsilon^2 \left(\frac{\sigma_x^2 - \gamma^2}{\gamma^2} \right) \\
\tau^2 &< \sigma_x^2 \sigma_\epsilon^2 \left(\frac{\sigma_v^2}{\gamma^2} \right) \\
\tau^2 &< \sigma_v^2 \sigma_\epsilon^2 \left(\frac{\sigma_x^2}{\gamma^2} \right) \\
|\tau| &< \sigma_v \sigma_\epsilon \sqrt{\frac{\pi' Q_z \pi + \sigma_v^2}{\pi' Q_z \pi}} \\
|\tau| &< \sigma_v \sigma_\epsilon \sqrt{1 + \frac{1}{\kappa^2}}
\end{aligned}$$

where $\kappa^2 = (\boldsymbol{\pi}' Q_z \boldsymbol{\pi})/\sigma_v^2$, is the signal-to-noise ratio in the first stage. We see that the FMSC trades the endogeneity of x, as measured by τ, against the strength of the instruments, as measured by the concentration parameter κ^2.

To get a better sense of the form of this trade-off, we can "re-interpret" this cutoff *as though* it were finite-sample rule. Recall that we defined $E_n[x_i \epsilon_i] = \tau/\sqrt{n}$. Hence,

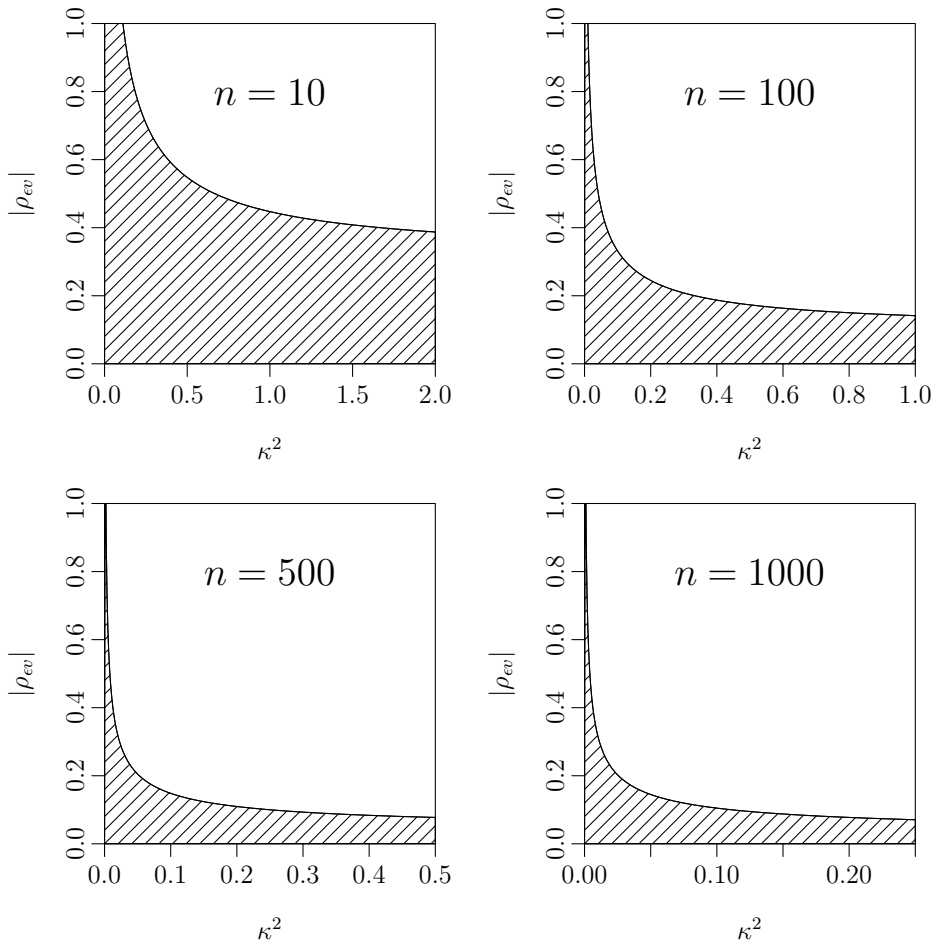
$$\tau = \sqrt{n} E_n[x_i \epsilon_i] = \sqrt{n} E_n\left[\left(\mathbf{z}_i' \pi + v_i \right) \epsilon_i \right] = \sqrt{n} E_n[v_i \epsilon_i] = \sqrt{n} Cor_n(v_i, \epsilon_i) \sigma_v \sigma_\epsilon$$

provided that we do not envision σ_v and σ_ϵ changing with sample size. Thus, for a *fixed* sample size n, we can think of τ as the correlation between ϵ_i and v_i scaled by $\sqrt{n}$ and their respective standard deviations. Dropping the n subscript on the correlation and substituting into the preceding inequality, we have

$$|\rho| < \sqrt{\frac{1}{n} \left(1 + \frac{1}{\kappa^2} \right)}$$

where ρ is the finite sample correlation between v_i and ϵ_i. Thus, interpreted as a finite sample rule, the FMSC tells us to weigh the endogeneity of x_i as measured by ρ against the sample size and the concentration parameter κ^2. We should only use the

2SLS estimator when the concentration parameter is sufficiently large, the sample size is sufficiently large, or x is sufficiently endogenous. The following figure depicts the region in which OLS is favored for different sample sizes.



As the sample size grows, the boundary moves towards the origin, meaning that we are more likely to use 2SLS. However, each of regions asymptotes at the origin so that when instruments are weak, we choose OLS. This makes intuitive sense. Notice that we change the horizontal axis limits to make it easier to see the threshold for weak instruments depending on sample size.

5.3 AMSE-Optimal Averaging

We showed above that the infeasible version of the FMSC selects OLS when

$$|\tau| < \sigma_v \sigma_\epsilon \sqrt{\frac{\pi' Q_z \pi + \sigma_v^2}{\pi' Q_z \pi}}$$

But selection is a blunt instrument. A better idea is to *combine* the OLS and 2SLS estimator by taking a weighted average and choose the weights to minimize AMSE. The problem is

$$\omega^* = \underset{\omega\in[0,1]}{\text{argmin}}\ \text{AMSE}\left(\omega\widehat{\beta} + (1-\omega)\widetilde{\beta}\right)$$

Let $\widehat{\beta}(\omega) = \omega\widehat{\beta} + (1-\omega)\widetilde{\beta}$. Then,

$$\text{Bias}\left(\widehat{\beta}(\omega)\right) \;=\; \text{Bias}(\widehat{\beta}) + \text{Bias}(\widetilde{\beta}) = \omega\left(\frac{\tau}{\sigma_x^2}\right)$$

and

$$\begin{aligned}
Var\left(\widehat{\beta}(\omega)\right) &= \left[\begin{array}{cc} \omega & 1-\omega \end{array}\right]\sigma_\epsilon^2\left[\begin{array}{cc} 1/\sigma_x^2 & 1/\sigma_x^2 \\ 1/\sigma_x^2 & 1/\gamma^2 \end{array}\right]\left[\begin{array}{c} \omega \\ 1-\omega \end{array}\right] \\
&= \sigma_\epsilon^2/\sigma_x^2\left[\begin{array}{cc} \omega & 1-\omega \end{array}\right]\left[\begin{array}{cc} 1 & 1 \\ 1 & \sigma_x^2/\gamma^2 \end{array}\right]\left[\begin{array}{c} \omega \\ 1-\omega \end{array}\right] \\
&= \frac{\sigma_\epsilon^2}{\sigma_x^2}\left\{\omega + (1-\omega)\left[\omega + (1-\omega)\sigma_x^2/\gamma^2\right]\right\} \\
&= \frac{\sigma_\epsilon^2}{\sigma_x^2}\left[\omega + (1-\omega)\omega + (1-\omega)^2\sigma_x^2/\gamma^2\right] \\
&= \frac{\sigma_\epsilon^2}{\sigma_x^2}\left[2\omega - \omega^2 + \frac{\sigma_x^2}{\gamma^2}\left(1 - 2\omega + \omega^2\right)\right] \\
&= \frac{\sigma_\epsilon^2}{\sigma_x^2}\left[(2\omega - \omega^2)\left(1 - \frac{\sigma_x^2}{\gamma^2}\right) + \frac{\sigma_x^2}{\gamma^2}\right] \\
&= \frac{\sigma_\epsilon^2}{\sigma_x^2}\left[(2\omega - \omega^2)\left(1 - \frac{\sigma_x^2}{\gamma^2}\right) + \frac{\sigma_x^2}{\gamma^2}\right]
\end{aligned}$$

and accordingly

$$\text{AMSE}\left(\widehat{\beta}(\omega)\right) \;=$$

Assumption 5.1 (DGP)**.** *Let* $\left\{(\mathbf{z}'_{ni}, \epsilon_{ni}, v_{ni})' : 1 \leq i \leq n, n = 1, 2, \ldots\right\}$ *be a triangular array of random variables such that* $(\mathbf{z}'_{ni}, \epsilon_{ni}, v_{ni})' \sim iid\,(0, \mathcal{V}_n)$ *for fixed* n *where*

$$\mathcal{V}_n = \begin{bmatrix} Q_Z & 0 & 0 \\ 0 & \sigma_\epsilon^2 & \tau/\sqrt{n} \\ 0 & \tau/\sqrt{n} & \sigma_v^2 \end{bmatrix} > 0 \tag{17}$$

$$y_{ni} = \beta x_{ni} + \epsilon_{ni} \tag{18}$$

$$x_{ni} = \mathbf{z}'_{ni}\pi + v_{ni} \tag{19}$$

and τ, π, β *are constants.*

Assumption 5.2 (Regularity Conditions)**.** *For some* $\eta > 0$*,*

(a) $\sup_n E\left[|\epsilon_i v_i|^{2+\eta}\right] < \infty$ *and*

(b) $\sup_n E\left[|\epsilon_{ni} z_{ni\ell}|^{2+\eta}\right] < \infty$ *for each component* $z_{ni\ell}$ *of* $\mathbf{z}_{ni}$*.*

For convenience state a Lemma that we'll use for all the limit distributions in the paper. Provides sufficient conditions that specialize the Lindeberg-Feller CLT to the triangular array defined in Assumption 5.1. These are basically standard.

6 Derivations for Choosing Instruments Example

$$y_i = \mathbf{x}'_i\boldsymbol{\beta} + \epsilon_i \tag{20}$$

$$\mathbf{x}_i = \Pi_w \mathbf{w}_i + \Pi_z \mathbf{z}_i + \mathbf{v}_i \tag{21}$$

$$E_n[\mathbf{z}_i(y_i - \mathbf{x}'_i\boldsymbol{\beta})] = E_n[\mathbf{z}_i\epsilon_i] = \boldsymbol{\tau}/\sqrt{n}$$

$$E_n[\mathbf{w}_i(y_i - \mathbf{x}'_i\boldsymbol{\beta})] = E_n[\mathbf{w}_i\epsilon_i] = \mathbf{0}$$

$$E_n[\mathbf{x}_i\epsilon_i] = \Pi_z\boldsymbol{\tau}/\sqrt{n} + E_n[\epsilon_i\mathbf{v}_i]$$

Need to think about how to control the endogeneity of $\mathbf{x}$ as I vary other parameters. Perhaps set

$$E_n[\epsilon_i\mathbf{v}_i] = \boldsymbol{\gamma}^{\mathbf{2}} - \Pi_z\boldsymbol{\tau}/\sqrt{n}$$

so that the endogeneity of $\mathbf{x}$ is held constant at $\boldsymbol{\gamma}^{\mathbf{2}}$? Does this have any effect on the FMSC or does it only matter in the simulation?

Should be able to write this down in terms of a tradeoff