# AI in Econometric Theory: An Example

Isaiah Andrews  Anna Mikusheva*

October 26, 2026

**Abstract**

This note contains a story and a paper. The paper is about the admissibility of two-stage least squares with weak instruments, a topic we have worked on together since 2019. The story is about how we (or, rather, our LLMs) wrote the paper, and some thoughts on what this may mean for our field.

Since 2019 we have worked on and off on the joint paper "GMM is Inadmissible Under Bounded Identification Strength" (Andrews and Mikusheva, 2024). This paper started from the observation that common GMM estimators are discontinuous at data realizations where the sample objective has a non-unique minimizer, and the intuition that in settings with weak identification one should thus be able to do better with alternative estimators. Motivated by this intuition, we attempted to prove that GMM estimators are dominated in the sense that there exist alternative estimators that are weakly better under all data-generating processes and strictly better under some. We were able to establish some results in this spirit, but subject to important caveats, namely that (a) the quantity of interest and the strength of identification were bounded in appropriate senses and (b) we could not analytically characterize a dominating estimator. We submitted the paper to journals for potential publication, but it was rejected multiple times due in part to these limitations.[1] Both before and after these rejections, we tried to establish stronger results for the special case of two-stage least squares under squared error loss, but repeatedly failed to either relax our conditions or analytically characterize a dominating estimator.

Motivated by the progress of AI in math (e.g., Alpöge, 2026; Ramos et al., 2026; OpenAI, 2026a,b), we recently experimented with using a commercially available frontier AI model (specifically, GPT-6 Astra) to try to solve these questions for two-stage least squares with

*We thank Jiafeng Chen, Paul Goldsmith-Pinkham, Ashesh Rambachan, Jonathan Roth, and Jesse Shapiro for helpful comments.

[1] We received a revise and resubmit in 2024, though we have been negligent in turning it around.

squared error loss. We found that the model could (a) prove that two-stage least squares is dominated without the boundedness conditions our proofs imposed and moreover (b) provide a variety of closed-form expressions for dominating estimators. Indeed, the model could go further, proving dominance in a more general setting that nests models with heterogeneous treatment effects, where the target parameter is defined precisely as the two-stage least squares estimand, a more demanding setting for dominance results.[2] We then used a combination of GPT-6 Astra and Claude Opus 5.5 to expand the results into the note below, which is entirely AI-written, though we asked for revisions in some places and take full responsibility for any errors.[3] Were we to write the note ourselves, we would make a number of further edits for clarity, emphasis, and style, but we provide it as-is in the spirit of illustrating current capabilities.

Given recent AI-driven results in math, that AI is now capable of proving what was, to us, a difficult result is entirely unsurprising. Nevertheless, it is a striking change from our own experience with AI models in the relatively recent past: for instance, despite sustained experimentation, neither of us found them particularly helpful for our own proof writing prior to early 2026 (though we found them helpful for proof checking starting in 2025). We think recent advances in capabilities, and the further advances which appear likely in the near future, call for a serious and ongoing re-evaluation of how our field assesses research outputs and allocates resources going forward. We hope that, by talking about our own experiences as established researchers in the field, we might both encourage others to do the same and help contribute to what will no doubt be a much wider discussion on these and related questions across fields.

Before moving to the AI-written note, we want to be clear that we do not view this advance in capabilities as a tragedy. A big part of why we each chose to work in econometrics is that we think the statistical methods used in economics research matter for the quality of the answers obtained, which are in turn consequential for a variety of decisions by both private and public actors. Given that belief, we also think that better tools for producing knowledge in econometrics (and in a wide variety of other domains) are a very good thing, albeit one which demands appropriate institutional and policy responses. Acknowledging the changes which have already taken place seems a useful early step towards formulating such responses.

[2]The current generation of models was not able to solve some follow-up questions we posed. In particular, the models did not succeed in analytically characterizing an estimator which dominates two-stage least squares and is itself undominated by anything else. We plan to return to that question after future model releases, and are particularly interested in whether such an alternative estimator might produce larger performance gains in empirically relevant settings than those found below.

[3]We have both carefully reviewed the note and proofs, in addition to using AI for verification. A Lean formalization is available here.

# Two-Stage Least Squares is Inadmissible with Weak Instruments

**Abstract**

Two-stage least squares (TSLS) is the standard estimator in linear instrumental variables regression, and with weak instruments its behavior is poorly described by conventional approximations. We study the Gaussian limit experiment with known covariance and unrestricted reduced-form and first-stage coefficients, which accommodates heterogeneous treatment effects, and take the TSLS estimand as the target. TSLS is known to have infinite mean squared error when there are fewer than three instruments, and we show that it is inadmissible under squared error loss whenever there are at least three, with no restriction on identification strength. The argument is constructive: moving the TSLS estimate toward a bounded statistic when the observed first stage is close to zero lowers risk at every parameter value, and we give two simple dominating estimators with explicit tuning. In simulations calibrated to published specifications, the variance reductions when the reduced form and first stage are zero are large with three instruments but negligible with 21 or more, and the improvements in mean squared error at the estimated coefficients are small and concentrated in the specifications with the smallest first-stage Wald statistics.

## 1 Introduction

Two-stage least squares (TSLS) is the standard estimator in linear instrumental variables (IV) regression. When instruments are weak, in the sense that the first stage is small relative to sampling uncertainty, conventional approximations to the behavior of TSLS are unreliable (Staiger and Stock, 1997; Stock and Wright, 2000); see Andrews et al. (2019) for a review. With normal errors, TSLS has finite moments only up to the degree of overidentification (Kinal, 1980), so its mean squared error is finite only when there are at least three instruments. With fewer instruments, any estimator with finite second moments dominates TSLS

under squared error loss.[1]

A decision-theoretic literature studies estimation with weak instruments. Chamberlain (2007) uses invariance and minimax arguments to derive estimators and confidence sets in a linear IV model with a single structural coefficient. Hirano and Porter (2015) show that no estimator can be mean-unbiased, quantile-unbiased, or translation equivariant (in their distributional sense) when the first-stage coefficients are unrestricted near zero, while Andrews and Armstrong (2017) construct an unbiased estimator when the sign of the first stage is known. Closer to our approach, other work improves on TSLS by shrinkage, under different criteria and asymptotic frameworks. Hansen (2017) shows that a weighted average of TSLS and ordinary least squares, with weight determined by a Hausman statistic, has lower asymptotic risk than TSLS, under a particular weighted quadratic loss and local exogeneity, when there are more than two endogenous regressors. Spiess (2017) shows that combining James–Stein shrinkage of the first stage with a control-function second stage yields an estimator that dominates TSLS in bias, with one endogenous regressor and at least four instruments. Andrews and Mikusheva (2022) derive the limit experiment for weakly identified generalized method of moments (GMM) models and study optimal decision rules there. Most directly related, Andrews and Mikusheva (2024) show that GMM estimators, including TSLS, are inadmissible in this limit experiment under any bounded, uniformly convex loss when identification strength is bounded. Their argument uses a complete class theorem (Brown, 1986) to show that admissible estimators must be uniformly continuous in the sample moments, which GMM estimators are not.

While this establishes that TSLS can be improved, it leaves open how. As Andrews and Mikusheva (2024) write, "Our proof for inadmissibility is non-constructive, in the sense that it does not deliver a dominating estimator." Their result also requires a bounded loss and a bound on identification strength. It therefore does not cover squared error loss on an unrestricted parameter space, under which TSLS has finite risk whenever there are at least three instruments.

A second gap concerns heterogeneous treatment effects. With multiple instruments and heterogeneous effects, TSLS estimates a weighted average of local average treatment effects under the assumptions of Imbens and Angrist (1994) (Imbens and Angrist, 1994; Angrist and Imbens, 1995). The constant-effect restriction that the reduced-form coefficients are proportional to the first-stage coefficients then generally fails. In the linear IV model, the decision-theoretic results discussed above impose this restriction (more generally, they as-

[1]The requirement of at least three instruments recalls the inadmissibility of the sample mean of a normal vector in three or more dimensions (Stein, 1956; James and Stein, 1961). Here $k \geq 3$ is the condition for TSLS to have finite risk.

sume correctly specified moment conditions). The weak-instrument-robust tests of Anderson and Rubin (1949), Kleibergen (2002), and Moreira (2003) are likewise derived for the constant-effect model. When heterogeneity makes the reduced-form coefficients nonproportional to the first-stage coefficients, the null hypotheses these tests are built around fail for every value of the coefficient, so the tests do not target the TSLS estimand. Andrews et al. (2019) note more broadly that existing weak-instrument-robust procedures assume constant effects. Recent work has begun to address this gap. Lee (2018) gives variance estimators for TSLS when instruments identify different local average treatment effects, under strong instruments, and Evdokimov and Kolesár (2019) study inference with many instruments and heterogeneous effects. Yap (2025) and Boot and Nibbering (2024) develop inference that is robust to many weak instruments under heterogeneity, and Vestre (2026) develops weak-instrument-robust tests for the TSLS estimand with a fixed number of instruments. Kleibergen and Zhan (2025) develop tests robust to both weak identification and misspecification for the pseudo-true value that minimizes the population continuously updated GMM objective, a target that with heterogeneous effects generally differs from the TSLS estimand. Our results concern estimation rather than inference, and apply to the TSLS estimand under heterogeneous effects with a fixed number of possibly weak instruments.

This note studies estimation of the TSLS estimand in the Gaussian limit experiment for linear IV. The covariance of the reduced-form and first-stage estimates is known, and we impose no restriction linking the reduced-form and first-stage coefficients, so the analysis allows for heterogeneous effects. We ask whether TSLS can be improved under squared error loss, at every parameter value, without any bound on identification strength. We show that it can, for every number of instruments $k \geq 3$, every covariance matrix, and every positive definite weight matrix. Since the constant-effect model is a special case of our parameter space, the same conclusion holds there, for the structural coefficient.

Our first result is a local stabilization argument. When the first-stage Wald statistic is small enough, replacing the TSLS estimate by a bounded statistic lowers conditional risk at every parameter value. Consequently, any estimator that moves TSLS toward this statistic on that event, and only there, by a fraction that depends only on the Wald statistic, strictly dominates TSLS. The stabilization region is small, so these estimators coincide with TSLS except on an event of probability at most $\Pr(\chi^2_k \leq 1/2)$, about 0.08 when $k = 3$. We further show that every estimator dominating TSLS must satisfy a calibration condition when the first stage is uninformative, and that some admissible estimator dominates TSLS, although our existence argument does not characterize it.

Our second result studies two simple dominating estimators, a truncated-linear and an exponential shrinkage of the TSLS estimate toward the same bounded statistic. Each has an

explicit tuning condition, and we obtain sharper tunings by verifying finite sets of sufficient conditions in exact arithmetic. When $W$ is proportional to $\Sigma_{\pi\pi}^{-1}$, as for TSLS under homoskedasticity, we also bound the variance reduction at a zero reduced form and first stage available from any estimator of this form that dominates TSLS. Only the truncated-linear shape can attain this bound, which shrinks rapidly with the number of instruments. Neither estimator dominates the other in general, and we make no claim that either is admissible.

Finally, we evaluate these estimators in designs calibrated to the 70 specifications with at least three instruments in the sample of Andrews et al. (2019). When both the reduced form and the first stage are zero, the certified estimators reduce variance by about three quarters on average in the 30 specifications with three instruments, but negligibly in the 40 with 21 or more. At the estimated coefficients, by contrast, the improvements in mean squared error are below 0.1% on average and at most 3.7%, and they are concentrated in the two specifications with the smallest first-stage Wald statistics.

The main limitation of our analysis is that our dominance results are exact statements about the Gaussian limit experiment with known covariance. We do not establish risk dominance for feasible estimators that use estimated covariance and weight matrices, and since squared error loss is unbounded, such results would not follow from convergence to the limit experiment alone.

Section 2 introduces the setting. Section 3 shows that TSLS is inadmissible and characterizes dominating estimators. Section 4 develops the two simple dominators, and Section 5 reports simulations calibrated to published specifications.

## 2 Setting

We study estimation in the Gaussian limit experiment for linear instrumental variables (IV) regression with weak instruments. Consider a scalar outcome, a scalar endogenous regressor, and $k$ instruments. Let $\hat{\rho}$ and $\hat{\pi}$ denote the coefficients from regressing the outcome and the endogenous regressor, respectively, on the instruments, after partialling out any exogenous controls. Under the weak-instrument asymptotics of Staiger and Stock (1997), the population coefficients shrink with the sample size at the root-$n$ rate, and suitably scaled estimates are asymptotically normal with a consistently estimable covariance matrix. Motivated by this approximation, we treat the scaled estimates as exactly normal with known covariance, and observe

$$\begin{pmatrix} \hat{\rho} \\ \hat{\pi} \end{pmatrix} \sim N\left( \begin{pmatrix} \rho \\ \pi \end{pmatrix}, \Sigma \right), \qquad \Sigma = \begin{pmatrix} \Sigma_{\rho\rho} & \Sigma_{\rho\pi} \\ \Sigma_{\pi\rho} & \Sigma_{\pi\pi} \end{pmatrix}, \tag{1}$$

where $\Sigma$ is positive definite. We place no restriction on the relationship between the reduced-form coefficients $\rho$ and the first-stage coefficients $\pi$, and take the parameter space to be $\rho \in \mathbb{R}^k$ and $\pi \in \mathbb{R}^k \setminus \{0\}$.

For a fixed positive definite weight matrix $W$, our target is

$$\theta(\rho, \pi) = \frac{\pi' W \rho}{\pi' W \pi}, \tag{2}$$

and the estimator we study is its plug-in analog, $\delta_0 = \hat{\pi}' W \hat{\rho} / \hat{\pi}' W \hat{\pi}$. When $W$ is the second moment matrix of the instruments, $\delta_0$ is the two-stage least squares (TSLS) estimator and $\theta$ is the TSLS estimand. If treatment effects are constant and the linear IV model is correctly specified, then $\rho = \beta\pi$ and $\theta = \beta$. With heterogeneous effects and multiple instruments, by contrast, $\rho$ need not be proportional to $\pi$. Under the monotonicity assumption of Imbens and Angrist (1994), TSLS then estimates a weighted average of local average treatment effects (Angrist and Imbens, 1995), although with covariates this interpretation requires additional conditions on how they are specified (Blandhol et al., 2026). With multiple instruments this assumption is restrictive, and under the weaker partial monotonicity condition of Mogstad et al. (2021) the weights can be negative. More generally, when the instrument-specific estimands are defined, $\theta$ is a linear combination of them with weights that sum to one (Rotemberg, 1983; Andrews, 2019), which Goldsmith-Pinkham et al. (2020) call Rotemberg weights. Leaving $\rho$ and $\pi$ unrestricted allows for all of these cases.

We evaluate an estimator $\delta = \delta(\hat{\rho}, \hat{\pi})$ by its mean squared error, $R_{\rho,\pi}(\delta) = \mathbb{E}_{\rho,\pi}[(\delta - \theta(\rho, \pi))^2]$, which is defined for $\pi \neq 0$. An estimator $\delta$ dominates $\delta'$ if $R_{\rho,\pi}(\delta) \leq R_{\rho,\pi}(\delta')$ for all $(\rho, \pi)$, with strict inequality for some $(\rho, \pi)$, and strictly dominates $\delta'$ if the inequality is strict for all $(\rho, \pi)$. An estimator is admissible if no other estimator dominates it.[2]

To understand the behavior of TSLS, it helps to purge the reduced-form estimate of its correlation with the first-stage estimate. Let $\Gamma = \Sigma_{\rho\pi}\Sigma_{\pi\pi}^{-1}$ and $\hat{a} = \hat{\rho} - \Gamma\hat{\pi}$. Then $\hat{a} \sim N(a, S)$ independently of $\hat{\pi} \sim N(\pi, \Sigma_{\pi\pi})$, where $a = \rho - \Gamma\pi$ and $S = \Sigma_{\rho\rho} - \Sigma_{\rho\pi}\Sigma_{\pi\pi}^{-1}\Sigma_{\pi\rho}$. Since $(\hat{a}, \hat{\pi})$ is a one-to-one transformation of $(\hat{\rho}, \hat{\pi})$, we can equivalently write estimators as functions of $(\hat{a}, \hat{\pi})$. Define

$$b(v) = \frac{v' W \Gamma v}{v' W v}, \qquad \ell(v) = \frac{W v}{v' W v}.$$

Substituting $\hat{\rho} = \hat{a} + \Gamma\hat{\pi}$ and $\rho = a + \Gamma\pi$ shows that

$$\delta_0 = b(\hat{\pi}) + \ell(\hat{\pi})'\hat{a}, \qquad \theta(\rho, \pi) = b(\pi) + \ell(\pi)'a. \tag{3}$$

[2]Under squared error loss, randomized estimators are weakly dominated by their conditional means given the data, so it is without loss to consider nonrandomized estimators.

Decomposition (3) splits TSLS into two parts. The first, $b(\hat{\pi})$, is a ratio of quadratic forms in $\hat{\pi}$, and so is bounded, taking values in an interval $[b_-, b_+]$ determined by $W$ and $\Gamma$. When $\Sigma$ has the Kronecker form implied by homoskedasticity, $\Gamma$ is proportional to the identity and $b(\hat{\pi})$ is constant, equal to the probability limit of ordinary least squares under weak-instrument asymptotics. The second part carries the information in the reduced form. Conditional on $\hat{\pi}$ it is normal with variance $\ell(\hat{\pi})' S \ell(\hat{\pi})$, which is of order $|\hat{\pi}|^{-2}$ and so explodes as the first-stage estimate approaches zero. Since $\mathbb{E}_{\rho,\pi}|\hat{\pi}|^{-2}$ is finite if and only if $k \geq 3$, TSLS has finite risk at every parameter value if and only if $k \geq 3$, consistent with the moment results of Kinal (1980). With $k \leq 2$ instruments any estimator with finite second moments dominates TSLS, so we assume $k \geq 3$ throughout.

Our results depend on $\Sigma$ and $W$ through three constants. Let $K = \Sigma_{\pi\pi}^{1/2} W \Sigma_{\pi\pi}^{1/2}$, where $\Sigma_{\pi\pi}^{1/2}$ is the symmetric square root, and define

$$\kappa = \frac{\lambda_{\max}(K)}{\lambda_{\min}(K)}, \qquad s_* = \lambda_{\min}\left(\Sigma_{\pi\pi}^{-1/2} S \Sigma_{\pi\pi}^{-1/2}\right), \qquad H = \frac{b_+ - b_-}{2}.$$

The condition number $\kappa \geq 1$ measures how far $W$ is from being proportional to $\Sigma_{\pi\pi}^{-1}$, with $\kappa = 1$ for TSLS under homoskedasticity. The constant $s_* > 0$ measures the reduced-form noise that remains after purging the first stage, relative to the first-stage noise. The half-width $H$ measures how much $b$ varies, and is zero under homoskedasticity.[3] Finally, let $Q = \hat{\pi}' \Sigma_{\pi\pi}^{-1} \hat{\pi}$ denote the Wald statistic for testing $\pi = 0$, which equals $k$ times the first-stage $F$ statistic when the covariance is known. The distribution of $Q$ is noncentral $\chi^2_k$ with noncentrality $\pi' \Sigma_{\pi\pi}^{-1} \pi$.

# 3 TSLS is Inadmissible

Decomposition (3) shows that TSLS adds to the bounded statistic $b(\hat{\pi})$ a term whose conditional variance is of order $1/Q$. When the observed first stage is close to zero, this term contributes mostly noise. Our first result shows that, once $Q$ is small enough, replacing TSLS by $b(\hat{\pi})$ lowers conditional risk at every parameter value.

**Lemma 1.** *Let* $\bar{q} = s_* / \{2(H + \sqrt{H^2 + \kappa s_*})^2\}$. *For all* $\rho \in \mathbb{R}^k$, *all* $\pi \neq 0$, *and all* $q \in (0, \bar{q}]$,[4]

$$\mathbb{E}_{\rho,\pi}\left[(b(\hat{\pi}) - \theta)^2 - (\delta_0 - \theta)^2 \mid Q = q\right] \leq -\frac{s_*}{2q}.$$

[3] All of our results remain valid if $\kappa$ and $H$ are replaced by upper bounds and $s_*$ by a positive lower bound. Such bounds can be certified exactly by rational matrix inequalities when $\Sigma$ and $W$ are rational.

[4] Here and below, conditional expectations given $Q = q$ are computed under the conditional distribution in which the direction of $\Sigma_{\pi\pi}^{-1/2}\hat{\pi}$ has density proportional to $\exp(\sqrt{q}\,\pi' \Sigma_{\pi\pi}^{-1/2} n)$ with respect to the uniform measure on the unit sphere. This is a version of the conditional distribution given $Q$ for every $q > 0$.

Lemma 1 conditions on the observed Wald statistic, not on the parameters. The inequality holds for every $(\rho, \pi)$, including parameter values with a strong first stage and a large target, on the event that $Q$ falls below $\bar{q}$. Replacing TSLS by $b(\hat{\pi})$ removes the term $\ell(\hat{\pi})'\hat{a}$, whose conditional variance is at least $s_*/Q$. It also changes the conditional bias. For targets $\theta$ near the range of $b$, the change in squared bias is bounded, and is outweighed by the variance reduction when $Q \leq \bar{q}$. For targets far from that range, the squared bias of $b(\hat{\pi})$ can be large. Averaging over the direction of $\hat{\pi}$, however, removing $\ell(\hat{\pi})'\hat{a}$ reduces the squared conditional bias as well once $Q$ is small. Appendix A gives the proof.

Lemma 1 immediately yields a class of dominating estimators. For a measurable function $w : (0, \infty) \to [0, 1]$, consider the estimator

$$\delta_w = \delta_0 - w(Q)\{\delta_0 - b(\hat{\pi})\} = b(\hat{\pi}) + \{1 - w(Q)\}\ell(\hat{\pi})'\hat{a}, \tag{4}$$

which moves TSLS toward $b(\hat{\pi})$ by a fraction $w(Q)$ that depends on the observed first-stage strength.

**Theorem 1.** *Suppose $w(q) = 0$ for $q > \bar{q}$, and $w$ is positive on a set of positive Lebesgue measure. Then for all $\rho \in \mathbb{R}^k$ and $\pi \neq 0$,*

$$R_{\rho,\pi}(\delta_0) - R_{\rho,\pi}(\delta_w) \geq \frac{s_*}{2}\mathbb{E}_\pi\left[\frac{w(Q)}{Q}\right] > 0.$$

Since squared error is convex and $\delta_w$ is a convex combination of $\delta_0$ and $b(\hat{\pi})$ with weights that depend only on $Q$, the conditional risk of $\delta_w$ given $Q$ is at most the corresponding combination of the conditional risks of $\delta_0$ and $b(\hat{\pi})$. Theorem 1 then follows from Lemma 1. The estimator $\delta_w$ coincides with TSLS whenever $Q > \bar{q}$, so it modifies TSLS only when the observed first stage is very close to zero. Since $\bar{q} \leq 1/(2\kappa) \leq 1/2$ and the distribution of $Q$ is stochastically increasing in the noncentrality, $\delta_w$ differs from TSLS with probability at most $\Pr(\chi^2_k \leq 1/2)$, which is about 0.08 for $k = 3$. The lower bound on the improvement does not depend on the reduced-form coefficients $\rho$, and depends on $\pi$ only through the noncentrality $\pi'\Sigma_{\pi\pi}^{-1}\pi$. It is positive at every parameter value but decays rapidly as the first stage strengthens, since the event $Q \leq \bar{q}$ then becomes rare.

Theorem 1 applies for any covariance matrix and weight matrix, and taking $w(q) = (1 - q/\bar{q})_+$ satisfies its conditions. It therefore implies the following.

**Corollary 1.** *For every $k \geq 3$, every positive definite $\Sigma$, and every positive definite $W$, TSLS is strictly dominated under squared error loss, and so is inadmissible. In particular, in the constant-effects model $\rho = \beta\pi$, with $\beta \in \mathbb{R}$ and $\pi \in \mathbb{R}^k \setminus \{0\}$, the estimator $\delta_w$ strictly dominates TSLS as an estimator of $\beta$.*

The second part of Corollary 1 follows because the constant-effects model is a subset of our parameter space, on which $\theta = \beta$. Andrews and Mikusheva (2024) show that GMM estimators, including TSLS, are inadmissible in this model under any bounded, uniformly convex loss when identification strength is bounded. Corollary 1 differs from their result for TSLS in three respects. First, it uses squared error loss, which is unbounded. Second, it places no bound on identification strength, or on $\beta$. Third, it is constructive, delivering an explicit dominating estimator. Their result, by contrast, covers general GMM estimators and a broad class of loss functions. The first part of Corollary 1 shows that TSLS remains inadmissible without constant effects, as an estimator of its own estimand.

Theorem 1 exhibits one class of dominating estimators. Our next result shows that every estimator dominating TSLS shares a feature with it, which concerns its behavior when both the reduced form and the first stage are uninformative. Let $m = \mathbb{E}_{V\sim N(0,\Sigma_{\pi\pi})}[b(V)]$ denote the mean of $b(\hat{\pi})$ when $\pi = 0$.

**Theorem 2.** *Let $\delta$ be an estimator, written as a function of $(\hat{a}, \hat{\pi})$, with $R_{\rho,\pi}(\delta) \leq R_{\rho,\pi}(\delta_0)$ for all $\rho \in \mathbb{R}^k$ and $\pi \neq 0$. Then*

1. *$\mathbb{E}_{V\sim N(0,\Sigma_{\pi\pi})}[\delta(u, V)] = m$ for Lebesgue-almost every $u \in \mathbb{R}^k$; and*
2. *$\mathbb{E}_{\rho,0}[\delta^2] \leq \mathbb{E}_{\rho,0}[\delta_0^2]$ for every $\rho \in \mathbb{R}^k$.*

The first part of Theorem 2 requires that when the first stage is completely uninformative, a dominating estimator average to the same constant $m$ over the first-stage noise, whatever the value of the purged reduced form. TSLS satisfies this condition because $b$ is even and $\ell$ is odd, and so does any estimator of the form $b(\hat{\pi}) + f(\hat{\pi})\ell(\hat{\pi})'\hat{a}$ with $f$ bounded and even, including $\delta_w$. The proof considers sequences of parameter values along which $\pi$ tends to zero while the target $\theta$ is held fixed at an arbitrary value. The risk comparison along these sequences restricts the mean of $\delta$ at $\pi = 0$, and completeness of the normal location family then delivers the conclusion for almost every $u$. Although the target is undefined at $\pi = 0$, the second part shows that a dominating estimator must also have second moments no larger than those of TSLS there.

Theorem 1 shows that TSLS is inadmissible, but does not deliver an admissible estimator. Our final result in this section shows that an admissible estimator dominating TSLS exists. Let $\mathbb{E}_{0,0}$ denote expectations when $\rho = 0$ and $\pi = 0$.

**Theorem 3.** *Let $\tilde{\delta}$ be an estimator that dominates $\delta_0$, and let $\mathcal{D}_{\tilde{\delta}}$ be the set of estimators $\delta$ with $R_{\rho,\pi}(\delta) \leq R_{\rho,\pi}(\tilde{\delta})$ for all $\rho \in \mathbb{R}^k$ and $\pi \neq 0$. There exists an estimator $\delta^* \in \mathcal{D}_{\tilde{\delta}}$, unique up to almost-sure equality, that minimizes $\mathbb{E}_{0,0}[(\delta - m)^2]$ over $\mathcal{D}_{\tilde{\delta}}$, and $\delta^*$ is admissible.*

*In particular, taking $\tilde{\delta} = \delta_w$ from Theorem 1 yields an admissible estimator that strictly dominates TSLS at every parameter value.*

The proof of Theorem 3 is a projection argument. By Theorem 2, every element of $\mathcal{D}_{\tilde{\delta}}$ has a finite second moment when $\rho = 0$ and $\pi = 0$. Moreover, $\mathcal{D}_{\tilde{\delta}}$ is closed and convex in the corresponding $L^2$ space, so the minimizer exists and is unique. Any estimator that dominates $\delta^*$ also lies in $\mathcal{D}_{\tilde{\delta}}$. Taking limits along parameter sequences that approach $\rho = 0$ and $\pi = 0$ with the target held at $m$ shows that such an estimator has no larger value of the objective, so by uniqueness it coincides with $\delta^*$. It shows that admissible improvements on TSLS exist, but does not characterize them, and in particular does not imply that $\delta_w$, or the estimators in Section 4, are themselves admissible.

# 4 Two Simple Dominators

Theorem 1 shows that moving TSLS toward $b(\hat{\pi})$ on the event $Q \leq \bar{q}$, by a fraction that depends only on $Q$, improves on it. We next study two simple estimators of this type, each with a single tuning parameter. Both belong to the family

$$\delta_f = b(\hat{\pi}) + f(Q)\ell(\hat{\pi})'\hat{a}, \tag{5}$$

where $f : [0, \infty) \to [0, 1]$. The estimator $\delta_f$ equals TSLS when $f \equiv 1$ and $b(\hat{\pi})$ when $f \equiv 0$. Since $f(Q)$ is an even function of $\hat{\pi}$, every such estimator satisfies the necessary condition in Theorem 2. Unlike TSLS, both estimators below have finite moments of all orders for every $k$, since $f(Q)\ell(\hat{\pi})$ is bounded whenever $f(q)$ is bounded by a multiple of $q$ near zero.

The first estimator shrinks linearly toward $b(\hat{\pi})$ when $Q$ falls below a cutoff $Q_*$:

$$f_c(q) = \min\{1, q/Q_*\}.$$

This is the estimator $\delta_w$ with $w(q) = (1 - q/Q_*)_+$. By Theorem 1, it strictly dominates TSLS whenever $Q_* \leq \bar{q}$, with $R_{\rho,\pi}(\delta_0) - R_{\rho,\pi}(\delta_{f_c}) \geq (s_*/2)\mathbb{E}_\pi[(Q^{-1} - Q_*^{-1})_+]$. We refer to it as the truncated-linear estimator.

The second estimator shrinks toward $b(\hat{\pi})$ by a factor that decays exponentially in $Q$:

$$f_e(q) = 1 - \exp(-\tau q), \qquad \tau > 0.$$

Unlike the truncated-linear estimator it modifies TSLS for every value of $Q$, so Theorem 1 does not apply. A separate argument shows that it also dominates TSLS when $\tau$ is large enough.

**Proposition 1.** *Let* $c = 4\kappa(k-1)H^2/s_*$. *If* $1+2\tau > 2\kappa(k-2)+c$, *then for all* $\rho \in \mathbb{R}^k$ *and* $\pi \neq 0$,

$$R_{\rho,\pi}(\delta_0) - R_{\rho,\pi}(\delta_{f_e}) \geq s_* \left\{1 - \frac{c}{1+2\tau-2\kappa(k-2)}\right\} \mathbb{E}_\pi\left[\frac{e^{-\tau Q}}{Q}\right] > 0.$$

The improvement from the exponential estimator is again positive at every parameter value, and its lower bound does not depend on $\rho$ and depends on $\pi$ only through the noncentrality $\pi'\Sigma_{\pi\pi}^{-1}\pi$. The threshold for $\tau$ grows linearly with the number of instruments. Some linear growth is necessary when $W$ is proportional to $\Sigma_{\pi\pi}^{-1}$: Step 1 of the proof of Proposition 2 shows that dominance then requires $1+2\tau > k-2$. For TSLS under homoskedasticity, where $\kappa = 1$ and $H = 0$, the sufficient condition in Proposition 1 reduces to $\tau > k - 5/2$, while the necessary condition is $\tau > (k-3)/2$. The proof, in Appendix A, writes the improvement as a quadratic function of the purged reduced-form mean $a$ and shows that this quadratic is positive for every $a$. The coefficients of the quadratic are controlled by expanding the distribution of the direction of $\hat{\pi}$ in a power series in the first-stage mean, and the condition on $\tau$ ensures that the quadratic term dominates the linear term.

The conditions in Theorem 1 and Proposition 1 are sufficient, and conservative. Appendix B gives sharper sufficient conditions for both estimators. These take the form of a finite set of inequalities on one-dimensional moments of the taper, together with a constant computed from $\Sigma_{\pi\pi}$ and $W$, and they imply dominance at every parameter value. We verify them in exact rational arithmetic, so the resulting tuning parameters are certified to deliver dominance rather than checked on a grid of parameter values. In the three-instrument designs of Section 5, for example, the certified cutoffs $Q_*$ are roughly 6 to 13 times larger than $\bar{q}$.

The improvements of both estimators are easiest to describe when $\rho = 0$ and $\pi = 0$, where the target is undefined but all of the estimators in (5) have mean $m$. Since the length and direction of the standard normal vector $\Sigma_{\pi\pi}^{-1/2}\hat{\pi}$ are independent when $\pi = 0$,

$$\mathbb{E}_{0,0}[(\delta_f - m)^2] = \mathrm{Var}_{0,0}(b(\hat{\pi})) + C_S\, \mathbb{E}\left[f(X)^2\right], \qquad X \sim \chi^2_{k-2}, \tag{6}$$

where $C_S = \mathbb{E}_{0,0}[\ell(\hat{\pi})' S \ell(\hat{\pi})]$. The estimators in (5) therefore reduce only the second component of the TSLS variance, and the fraction of $C_S$ they remove depends only on $k$ and the taper $f$. For the truncated-linear estimator, the fraction of $C_S$ removed is $F_{k-2}(Q_*) - k(k-2)F_{k+2}(Q_*)/Q_*^2$, where $F_j$ is the $\chi^2_j$ distribution function. For the exponential estimator it is $2(1+2\tau)^{-(k-2)/2} - (1+4\tau)^{-(k-2)/2}$.

How large can these improvements be? Our next result bounds the variance reduction available from any estimator of the form (5) that dominates TSLS. For $c > 0$ let $f_c(q) =$

$\min\{1, q/c\}$, let $X \sim \chi^2_{k-2}$, and let $c_k$ denote the unique $c > 1$ that solves

$$\mathbb{E}\left[(1 - X/c)(1 + X/c - 2X)\, 1\{X < c\}\right] = 0.$$

**Proposition 2.** *Suppose $W$ is proportional to $\Sigma_{\pi\pi}^{-1}$. If $f : [0, \infty) \to [0, 1]$ is measurable and $R_{\rho,\pi}(\delta_f) \leq R_{\rho,\pi}(\delta_0)$ for all $\rho \in \mathbb{R}^k$ and $\pi \neq 0$, then*

$$\mathbb{E}_{0,0}[(\delta_0 - m)^2] - \mathbb{E}_{0,0}[(\delta_f - m)^2] \leq \psi_k\, \mathbb{E}_{0,0}[(\delta_0 - m)^2], \qquad \psi_k = \mathbb{E}\left[1 - f_{c_k}(X)^2\right].$$

*Equality holds if and only if $f = f_{c_k}$ Lebesgue-almost everywhere on $(0, \infty)$ and $b$ is constant on $\mathbb{R}^k \setminus \{0\}$.*

The condition that $W$ be proportional to $\Sigma_{\pi\pi}^{-1}$ holds for TSLS under homoskedasticity. The proof first shows that any dominating taper must satisfy $\mathbb{E}[1 - f(X)^2] \geq 2\mathbb{E}[\{1 - f(X)\}X]$, by considering parameter values with $\pi$ approaching zero and $a$ large. It then maximizes the variance reduction subject to this constraint, which is a concave program whose solution is the truncated-linear taper with cutoff $c_k$. The bound $\psi_k$ equals 0.89 for $k = 3$, 0.52 for $k = 4$, 0.25 for $k = 5$, 0.014 for $k = 8$, and 0.0015 for $k = 10$, and is below $10^{-9}$ for $k \geq 21$, since $\psi_{21} \approx 3.6 \times 10^{-10}$ and $\psi_k$ is nonincreasing in $k$. Proposition 2 thus shows that, when $W$ is proportional to $\Sigma_{\pi\pi}^{-1}$, estimators of the form (5) that dominate TSLS can deliver large variance reductions at $\rho = \pi = 0$ only when the number of instruments is small. It also shows that among such estimators the truncated-linear shape is the only candidate to attain the bound. In the designs of Section 5 with $k \geq 21$, which nearly satisfy this condition ($\kappa \leq 1.016$), the certified truncated-linear cutoffs are within 1% of $c_k$.

Neither estimator dominates the other in general. Appendix C gives a homoskedastic design with three instruments, and tuning parameters satisfying the conditions of Theorem 1 and Proposition 1, for which each estimator has strictly lower risk than the other at some parameter values. The truncated-linear estimator leaves TSLS unchanged unless $Q$ is small, and is the only shape that can attain the bound in Proposition 2. The exponential estimator has less conservative explicit tuning: in Table 1 its explicit version removes 26% of the variance at $\rho = \pi = 0$ on average, against 12% for the explicit truncated-linear estimator. In the two specifications in Section 5 where improvements away from zero are non-negligible, it also improves more than the truncated-linear estimator with the same type of tuning. We make no claim that either estimator is admissible or optimal.

# 5 Simulations Calibrated to Published Specifications

The results so far establish that TSLS can be improved at every parameter value by estimators that modify it only, or mainly, when the observed first stage is close to zero. To gauge their size in practice, we evaluate the four estimators from Section 4 in designs calibrated to published IV specifications. These are the truncated-linear and exponential estimators, each with explicit and with certified tuning.

We use the specifications from the American Economic Review between 2014 and 2018 collected by Andrews et al. (2019) that have at least three instruments, so that TSLS has finite risk. This yields 70 specifications. Thirty come from Stephens and Yang (2014) and have three instruments. The remaining forty come from Young (2014) and have 21, 28, or 59 instruments. Two papers is a narrow base, and we view the results as illustrative. For each specification we set $\Sigma$ equal to the estimated covariance matrix of the reduced-form and first-stage coefficients, and set $W$ to the TSLS weight. The four estimators are tuned once for each specification and then held fixed. For the explicit tunings we set $Q_* = \bar{q}$ and set $\tau$ to $1 + 10^{-4}$ times the threshold $\kappa(k-2) + c/2 - 1/2$ in Proposition 1, using the certified bounds on $\kappa$, $s_*$, and $H$ described in Appendix B.

We report three comparisons, each as the percentage reduction relative to TSLS. The first is the variance at $\rho = 0$ and $\pi = 0$, computed from (6). The second is the mean squared error at the estimated coefficients, $(\rho, \pi) = (\hat{\rho}, \hat{\pi})$, with the target recomputed at those values. The third averages the mean squared error over the posterior for $(\rho, \pi)$ under a flat prior, $N((\hat{\rho}', \hat{\pi}')', \Sigma)$. We report the ratio of posterior-average risks, rather than the average of ratios, with each draw's target computed at that draw. Appendix D describes the numerical integration, including the importance sampling we use for the posterior averages.

Table 1 reports the results. Each column averages over two very different groups of specifications. When $\rho = 0$ and $\pi = 0$, the certified estimators reduce variance by 75% and 73% on average in the 30 three-instrument specifications, and by up to 87%. In the 40 specifications with 21 or more instruments the reductions are below $10^{-7}$ percent, in line with $\psi_k < 10^{-9}$ for $k \geq 21$. At the estimated coefficients the improvements are below 0.1% on average and at most 3.7%. Averaging over the posterior yields somewhat larger improvements, below 0.5% on average and up to 22%. These are driven almost entirely by two specifications, which reach 10% and 22% for the certified exponential estimator. All other specifications show improvements below 0.1%.

Two features of these designs explain the pattern away from $\rho = \pi = 0$. First, the estimators modify TSLS appreciably only when the Wald statistic $Q$ is small in absolute terms: the truncated-linear estimator only when $Q \leq Q_*$, and the exponential estimator by

| | Variance at $\rho = \pi = 0$ | MSE at estimates | Posterior-average MSE |
|---|---|---|---|
| Truncated-linear, explicit | 12.36 (14.74) | 0.00 (0.02) | 0.10 (0.64) |
| | [36.34] | [0.17] | [5.08] |
| Truncated-linear, certified | 32.19 (38.16) | 0.03 (0.20) | 0.37 (2.35) |
| | [86.52] | [1.64] | [18.29] |
| Exponential, explicit | 26.07 (31.19) | 0.03 (0.20) | 0.32 (2.09) |
| | [77.82] | [1.65] | [16.45] |
| Exponential, certified | 31.17 (36.89) | 0.08 (0.48) | 0.46 (2.87) |
| | [85.80] | [3.71] | [21.91] |

Table 1: Percentage reduction relative to TSLS across the 70 specifications with at least three instruments in the Andrews et al. (2019) sample. Entries are the mean, with the standard deviation across specifications in parentheses and the maximum in brackets. The 40 specifications with 21 or more instruments contribute essentially nothing to any column.

a factor $e^{-\tau Q}$. For large $k$ this is much more restrictive than a small first-stage $F$ statistic $Q/k$. The specifications from Young (2014) have weak first stages by the rule of thumb of Staiger and Stock (1997), with first-stage $F$ statistics between 1.4 and 6.8 in the data. Since $k$ is large, however, the observed Wald statistics range from about 50 to 400, and these are the noncentralities of $Q$ at the estimated coefficients. Values of $Q$ small enough for the estimators to act then essentially never arise. Second, the observed Wald statistics in the three-instrument specifications from Stephens and Yang (2014) range from 19 to 282. The two specifications with non-negligible improvements have the two smallest, 19 and 25, at which $Q \leq Q_*$ has probability about $2 \times 10^{-3}$ and $3 \times 10^{-4}$ under the certified truncated-linear tuning. For the other 28, whose Wald statistics exceed 58, this probability is below $10^{-9}$. The ranking of the estimators also differs across columns. The certified truncated-linear estimator has the lowest variance at $\rho = \pi = 0$ in most three-instrument specifications, while the certified exponential estimator has the largest improvements in the two specifications where the improvements away from zero are non-negligible.

# 6 Conclusion

We show that TSLS is inadmissible under squared error loss whenever there are at least three instruments, for any covariance and weight matrix, and whether or not treatment effects are constant. Simple estimators that stabilize TSLS when the observed first stage is close to zero improve on it at every parameter value. In designs calibrated to published specifications, however, the resulting improvements in mean squared error are small except in the two specifications with the smallest first-stage Wald statistics.

# A Proofs

**Notation.** Throughout, $k \geq 3$. Write $Z = \Sigma_{\pi\pi}^{-1/2}\hat{\pi} \sim N(\eta, I_k)$ with $\eta = \Sigma_{\pi\pi}^{-1/2}\pi$, and $Y = \Sigma_{\pi\pi}^{-1/2}\hat{a} \sim N(\xi, \Omega)$ with $\xi = \Sigma_{\pi\pi}^{-1/2}a$ and $\Omega = \Sigma_{\pi\pi}^{-1/2}S\Sigma_{\pi\pi}^{-1/2}$, so that $Z$ and $Y$ are independent. Let $K = \Sigma_{\pi\pi}^{1/2}W\Sigma_{\pi\pi}^{1/2}$, $L(z) = Kz/z'Kz$, and $\tilde{b}(z) = b(\Sigma_{\pi\pi}^{1/2}z)$. Then $Q = |Z|^2$, $\ell(\hat{\pi})'\hat{a} = L(Z)'Y$, $\delta_0 = \tilde{b}(Z) + L(Z)'Y$, and $\theta = \tilde{b}(\eta) + L(\eta)'\xi$. The function $\tilde{b}$ is even and takes values in $[b_-, b_+]$, while $L$ is odd and satisfies $L(rz) = L(z)/r$ for $r > 0$, $n'L(n) = 1$, and $|n|^{-1} \leq |L(n)| \leq \kappa|n|^{-1}$. Since $\mathbb{E}|Z|^{-2} < \infty$ for $k \geq 3$, every expectation below is finite. For $\eta = re$ with $|e| = 1$, write $\theta - \tilde{b}(\eta) = L(\eta)'\xi = \eta'K\xi/\eta'K\eta$, so that

$$e'K\xi = \{\theta - \tilde{b}(\eta)\}\,|\eta|\,e'Ke. \tag{7}$$

**Proof of Lemma 1.** Write $Z = rn$ with $r = \sqrt{Q}$ and $|n| = 1$, and let $f(n) = L(n)'\xi$ and $s(n) = L(n)'\Omega L(n)$. Since $|L(n)| \geq 1$, $s(n) \geq s_*$. Conditional on $Z = rn$, $\delta_0 - \theta = \tilde{b}(n) - \theta + L(n)'Y/r$, so

$$\mathbb{E}\left[(\tilde{b}(Z) - \theta)^2 - (\delta_0 - \theta)^2 \mid Z = rn\right] = -\frac{2}{r}\{\tilde{b}(n) - \theta\}f(n) - \frac{f(n)^2}{r^2} - \frac{s(n)}{r^2}. \tag{8}$$

Let $b_c = (b_+ + b_-)/2$, $t = \theta - b_c$, and $T = \sqrt{H^2 + \kappa s_*}$, and note that $\bar{q} = s_*/\{2(H+T)^2\} = (T-H)/\{2\kappa(T+H)\}$ since $T^2 - H^2 = \kappa s_*$. Fix $r^2 \leq \bar{q}$.

Suppose first that $|t| \leq T$. Then $|\tilde{b}(n) - \theta| \leq H + T$, and the right side of (8) equals $(\tilde{b}(n) - \theta)^2 - (\tilde{b}(n) - \theta + f(n)/r)^2 - s(n)/r^2 \leq (H+T)^2 - s_*/r^2 \leq -s_*/(2r^2)$ for every $n$.

Suppose instead that $|t| > T$. Conditional on $|Z| = r$, the direction $n$ has density proportional to $\exp(r\eta'n)$ with respect to the uniform measure $\sigma$ on the unit sphere. Since $\tilde{b}$ is even and $f$ is odd, pairing $n$ with $-n$ gives

$$\mathbb{E}\left[(8) \mid Q = r^2\right] = \frac{1}{Z_r}\int \Big[ -\frac{2}{r}\{\tilde{b}(n) - \theta\}f(n)\sinh(r\eta'n) - \frac{f(n)^2 + s(n)}{r^2}\cosh(r\eta'n)\Big]\, d\sigma(n),$$

where $Z_r = \int \cosh(r\eta'n)\, d\sigma(n)$. Let

$$A_r = \int f(n)^2 \cosh(r\eta'n)\, d\sigma(n), \qquad J_r = \int (\eta'n)^2 \cosh(r\eta'n)\, d\sigma(n).$$

Since $|\sinh x| \leq |x|\cosh x$ and $|\tilde{b}(n) - \theta| \leq |t| + H$, the Cauchy–Schwarz inequality gives

$$\left|\frac{2}{r}\int \{\tilde{b}(n) - \theta\}f(n)\sinh(r\eta'n)\, d\sigma(n)\right| \leq 2(|t| + H)\sqrt{A_r J_r}.$$

To bound $J_r$ by $A_r$, let $e = \eta/|\eta|$ and $N_r = \int nn' \cosh(r\eta' n)\, d\sigma(n)$. By rotational symmetry about $e$, $N_r = \lambda_\parallel ee' + \lambda_\perp(I - ee')$ with $\lambda_\perp \geq 0$, so $J_r = |\eta|^2 \lambda_\parallel$. Since $n'Kn \leq \lambda_{\max}(K)$, (7) gives

$$A_r \geq \frac{\xi' K N_r K \xi}{\lambda_{\max}(K)^2} \geq \frac{\lambda_\parallel (e'K\xi)^2}{\lambda_{\max}(K)^2} = \frac{J_r\{\theta - \tilde{b}(\eta)\}^2 (e'Ke)^2}{\lambda_{\max}(K)^2} \geq \frac{(|t| - H)^2}{\kappa^2} J_r,$$

where the last step uses $|\theta - \tilde{b}(\eta)| \geq |t| - H > 0$ and $e'Ke \geq \lambda_{\min}(K)$. Hence $\sqrt{A_r J_r} \leq \kappa A_r/(|t| - H)$, and

$$\mathbb{E}\left[(8) \mid Q = r^2\right] \leq -\frac{s_*}{r^2} + \frac{A_r}{Z_r}\left\{-\frac{1}{r^2} + \frac{2\kappa(|t| + H)}{|t| - H}\right\}.$$

Since $(|t| - H)/(|t| + H)$ is nondecreasing in $|t|$, it is at least $(T - H)/(T + H) = 2\kappa\bar{q}$, so the term in braces is nonpositive when $r^2 \leq \bar{q}$. In both cases the conditional expectation is at most $-s_*/(2r^2)$. □

**Proof of Theorem 1.** Write $w = w(Q)$. Since $\delta_w - \theta = (1 - w)(\delta_0 - \theta) + w\{\tilde{b}(Z) - \theta\}$ and squared error is convex, $(\delta_0 - \theta)^2 - (\delta_w - \theta)^2 \geq w[(\delta_0 - \theta)^2 - \{\tilde{b}(Z) - \theta\}^2]$. Taking expectations conditional on $Q$ and applying Lemma 1 on the event $Q \leq \bar{q}$, outside of which $w = 0$, gives $\mathbb{E}[(\delta_0 - \theta)^2 - (\delta_w - \theta)^2 \mid Q] \geq s_* w(Q)/(2Q)$. Taking expectations proves the inequality. Since $Q$ has a positive density on $(0, \infty)$ and $w$ is positive on a set of positive measure, $\mathbb{E}_\pi[w(Q)/Q] > 0$. □

**Proof of Corollary 1.** The function $w(q) = (1 - q/\bar{q})_+$ satisfies the conditions of Theorem 1. The constant-effects model is the subset of the parameter space with $\rho = \beta\pi$, on which $\theta = \beta$. □

The proofs of Theorems 2 and 3 use the following continuity property. We work with $(\hat{a}, \hat{\pi}) \sim N(\mu, \Lambda)$, where $\mu = (a', \pi')'$ and $\Lambda = \mathrm{diag}(S, \Sigma_{\pi\pi})$, and note that $\rho = a$ when $\pi = 0$.

**Lemma 2.** *Suppose $\mathbb{E}_{a,\pi}[\delta^2] < \infty$ for all $a$ and all $\pi \neq 0$. Then $\mathbb{E}_{a,\pi}[\delta^2] < \infty$ for all $(a, \pi) \in \mathbb{R}^{2k}$, and $\mathbb{E}_{a,\pi}[\delta]$ and $\mathbb{E}_{a,\pi}[\delta^2]$ are continuous on $\mathbb{R}^{2k}$.*

*Proof.* The density of $N(\mu, \Lambda)$ is $\phi_0(x)\exp(\tau'x - \tau'\Lambda\tau/2)$, where $\tau = \Lambda^{-1}\mu$. Fix a compact set of values of $\tau$, and enclose it in a box whose vertices $\tau_1, \ldots, \tau_J$ all have nonzero last $k$ coordinates. Since $\tau \mapsto \exp(\tau'x)$ is convex, its maximum over the box is attained at a vertex, so $\exp(\tau'x) \leq \sum_j \exp(\tau_j'x)$. Consequently, uniformly over the compact set, the density at $\mu$ is bounded by a constant times $\sum_j \phi_{\mu_j}(x)$, where $\mu_j = \Lambda\tau_j$ has a nonzero first-stage component. The right side integrates $1 + |\delta| + \delta^2$, and the conclusion follows by dominated convergence. □

**Proof of Theorem 2.** Since $R_{\rho,\pi}(\delta) \le R_{\rho,\pi}(\delta_0) < \infty$, $\delta$ has a finite second moment at every parameter value with $\pi \neq 0$, so Lemma 2 applies to both $\delta$ and $\delta_0$. Fix $a \in \mathbb{R}^k$ and $t \in \mathbb{R}$, choose $v \neq 0$ with $v'Wa = 0$, and for $\varepsilon > 0$ let $\pi_\varepsilon = \varepsilon v$ and $a_\varepsilon = a + \varepsilon(tI - \Gamma)v$. By (3),

$$\theta = b(v) + \frac{v'Wa + \varepsilon v'W(tI - \Gamma)v}{\varepsilon v'Wv} = t$$

along this sequence. Letting $\varepsilon \downarrow 0$ in $R(\delta) \le R(\delta_0)$ and applying Lemma 2 gives $\mathbb{E}_{a,0}[(\delta - t)^2] \le \mathbb{E}_{a,0}[(\delta_0 - t)^2]$ for every $t \in \mathbb{R}$. The difference between the two sides is affine in $t$, so $\mathbb{E}_{a,0}[\delta] = \mathbb{E}_{a,0}[\delta_0]$ and $\mathbb{E}_{a,0}[\delta^2] \le \mathbb{E}_{a,0}[\delta_0^2]$, which proves the second part. Since $b$ is even, $\ell$ is odd, and $\mathbb{E}|\ell(\hat{\pi})| < \infty$, $\mathbb{E}_{a,0}[\delta_0] = m$. Let $r(u) = \mathbb{E}_{V \sim N(0,\Sigma_{\pi\pi})}[\delta(u, V)] - m$. By Jensen's inequality $r$ is square-integrable under $N(a, S)$ for every $a$, and we have shown that $\mathbb{E}_{\hat{a} \sim N(a,S)}[r(\hat{a})] = 0$ for every $a \in \mathbb{R}^k$. Completeness of the normal location family (e.g. Lehmann and Romano, 2005, Theorem 4.3.1) then implies $r(u) = 0$ for almost every $u$. □

**Proof of Theorem 3.** Every $\delta \in \mathcal{D}_{\tilde{\delta}}$ satisfies $R(\delta) \le R(\tilde{\delta}) \le R(\delta_0)$, so by Theorem 2 it is square-integrable under $P_{0,0}$, the distribution with $\rho = 0$ and $\pi = 0$. We view $\mathcal{D}_{\tilde{\delta}}$ as a subset of $L^2(P_{0,0})$, identifying estimators that agree $P_{0,0}$-almost surely. All of the normal distributions in (1) are mutually absolutely continuous, so such estimators have the same risk at every parameter value. The set $\mathcal{D}_{\tilde{\delta}}$ is nonempty since it contains $\tilde{\delta}$, and convex since squared error is convex. It is closed: if $\delta_n \to \delta$ in $L^2(P_{0,0})$, a subsequence converges $P_{0,0}$-almost surely, hence almost surely under every distribution in the model, and Fatou's lemma gives $R(\delta) \le \liminf R(\delta_n) \le R(\tilde{\delta})$. The objective $\mathbb{E}_{0,0}[(\delta - m)^2]$ is the squared distance from $\delta$ to the constant $m$, so it has a unique minimizer $\delta^*$ over the closed convex set $\mathcal{D}_{\tilde{\delta}}$.

Suppose $\delta'$ dominates $\delta^*$. Then $R(\delta') \le R(\delta^*) \le R(\tilde{\delta})$, so $\delta' \in \mathcal{D}_{\tilde{\delta}}$. For any $v \neq 0$, along $\pi_\varepsilon = \varepsilon v$ and $a_\varepsilon = \varepsilon(mI - \Gamma)v$ the target equals $m$ by (3), and both $\rho_\varepsilon$ and $\pi_\varepsilon$ tend to zero. Letting $\varepsilon \downarrow 0$ in $R(\delta') \le R(\delta^*)$ and applying Lemma 2 gives $\mathbb{E}_{0,0}[(\delta' - m)^2] \le \mathbb{E}_{0,0}[(\delta^* - m)^2]$. By uniqueness of the minimizer $\delta' = \delta^*$ almost surely, so the two have the same risk everywhere, contradicting the assumption that $\delta'$ dominates $\delta^*$. Hence $\delta^*$ is admissible. If $\tilde{\delta} = \delta_w$, then $R(\delta^*) \le R(\delta_w) < R(\delta_0)$ at every parameter value. □

**Gain identity.** The proofs of Propositions 1 and 2 use an expression for the improvement of $\delta_f$ over TSLS. Let $g = 1 - f$ and $h = 1 - f^2 = 2g - g^2$, both evaluated at $Q$. Conditional on $Z$, $(\delta_0 - \theta)^2 - (\delta_f - \theta)^2 = h\,(L'Y)^2 + 2g\{\tilde{b}(Z) - \theta\}L'Y$, where $L = L(Z)$. Taking expectations over $Y$, substituting $\theta = \tilde{b}(\eta) + L(\eta)'\xi$, and then taking expectations over $Z$ gives

$$R(\delta_0) - R(\delta_f) = \xi' A \xi + 2\xi' z + c_0, \tag{9}$$

where $A = T_h - m_g L(\eta)' - L(\eta)m_g'$, with $T_h = \mathbb{E}[hLL']$, $m_g = \mathbb{E}[gL]$, $z = \mathbb{E}[g\{\tilde{b}(Z) - \tilde{b}(\eta)\}L]$, and $c_0 = \mathbb{E}[hL'\Omega L]$. Here expectations are over $Z \sim N(\eta, I)$. Two facts are used repeatedly. First, the density of $\chi^2_k$ divided by its argument is the density of $\chi^2_{k-2}$ divided by $k-2$, so $\mathbb{E}_{\chi^2_k}[u(X)/X] = \mathbb{E}_{\chi^2_{k-2}}[u(X)]/(k-2)$. Second, under $N(0, I_k)$ the length $|Z|$ and the direction $Z/|Z|$ are independent.

**Proof of Proposition 1.** Let $g(q) = e^{-\tau q}$, $\alpha = 1+2\tau$, $d = k-2$, and $J = \mathbb{E}[g/Q]$. Let $n$ be uniform on the unit sphere. Write $\eta = re$ with $|e| = 1$, and define $M_j = \mathbb{E}_n[L(n)L(n)'(e'n)^{2j}]$ and $v_j = \mathbb{E}_n[L(n)(e'n)^{2j+1}]$ for $j \geq 0$.

We first bound the quadratic term. Matrix Cauchy–Schwarz gives $v_j' M_j^{-1} v_j \leq \omega_j = \mathbb{E}_n[(e'n)^{2j+2}]$. Since $n'Kn \leq \lambda_{\max}(K)$, $M_j \succeq \lambda_{\max}(K)^{-2} K N_j K$ with $N_j = \mathbb{E}_n[nn'(e'n)^{2j}]$, and $N_j e = \omega_j e$ by symmetry, so $L(e)' M_j^{-1} L(e) \leq \kappa^2/\omega_j$. Since the largest eigenvalue of $xy' + yx'$ is at most $2|x||y|$, these bounds give

$$v_j L(e)' + L(e)v_j' \preceq 2\kappa M_j. \tag{10}$$

Expanding $\exp(re'Z)$ in the density of $N(re, I)$ relative to $N(0, I)$, and keeping the even and odd terms respectively, gives the absolutely convergent series

$$\begin{aligned} T_h &= e^{-r^2/2} \sum_{j \geq 0} \frac{r^{2j}}{(2j)!} \mathbb{E}[h(X)X^{j-1}] M_j, \\ m_g L(\eta)' + L(\eta)m_g' &= e^{-r^2/2} \sum_{j \geq 0} \frac{r^{2j}}{(2j)!} \frac{\mathbb{E}[g(X)X^j]}{2j+1} \\ &\qquad \times \{v_j L(e)' + L(e)v_j'\}, \end{aligned}$$

where $X \sim \chi^2_k$. Gamma integration gives $\mathbb{E}[g(X)X^j] = \{(k+2j-2)/\alpha\}\mathbb{E}[g(X)X^{j-1}]$, and $\mathbb{E}[h(X)X^{j-1}] \geq \mathbb{E}[g(X)X^{j-1}]$ since $h \geq g$. Since $(k+2j-2)/(2j+1) \leq d$ for $k \geq 3$, (10) yields $A \succeq (1 - 2\kappa d/\alpha)T_h$, which is positive definite under the condition of the proposition.

We next bound the linear term. Let $\Delta = \tilde{b}(Z) - \tilde{b}(\eta)$. In the coordinates $x = K^{1/2}z$, $\tilde{b}$ is a Rayleigh quotient of a symmetric matrix with eigenvalues in $[b_-, b_+]$, so the difference between its values at two unit vectors is at most $2H$ times the sine of the angle between them. Since $\sin\angle(K^{1/2}z, K^{1/2}\eta) \leq \kappa^{1/2} \sin\angle(z, \eta)$, we obtain $\Delta^2 \leq 4\kappa H^2 \sin^2\angle(Z, \eta)$. Let $Z_i$ be a coordinate of $Z$ orthogonal to $\eta$. Gaussian integration by parts applied to $g(Q)Z_i/Q$ gives $\mathbb{E}[(g - 2g')Z_i^2/Q] + 2\mathbb{E}[gZ_i^2/Q^2] = \mathbb{E}[g/Q]$. The boundary term near the origin vanishes since $k \geq 3$. Since $g - 2g' = \alpha g$, summing over the $k-1$ orthogonal coordinates gives

$\mathbb{E}[g \sin^2 \angle(Z, \eta)] \le (k-1)J/\alpha$. By matrix Cauchy–Schwarz and $g^2/h = g/(2-g) \le g$,

$$zz' \preceq T_h \mathbb{E}\left[\frac{g^2}{h}\Delta^2\right] \preceq \frac{4\kappa(k-1)H^2}{\alpha} J\, T_h = \frac{cs_*}{\alpha} J\, T_h.$$

Combining with the bound on $A$, $z'A^{-1}z \le cs_*J/(\alpha - 2\kappa d)$. Finally, $c_0 \ge s_* \mathbb{E}[h/Q] \ge s_* J$ since $L'\Omega L \ge s_*/Q$ and $h \ge g$. Completing the square in (9),

$$R(\delta_0) - R(\delta_{f_e}) \ge c_0 - z'A^{-1}z \ge s_* J \left\{1 - \frac{c}{\alpha - 2\kappa d}\right\} > 0. \qquad \square$$

**Proof of Proposition 2.** Since $W$ is proportional to $\Sigma_{\pi\pi}^{-1}$, $K$ is proportional to the identity and $L(z) = z/|z|^2$. Let $X \sim \chi^2_{k-2}$, $g = 1 - f$, and $h = 1 - f^2$.

Step 1 (a necessary condition). Since $\xi$ is unrestricted, dominance and (9) imply that $A$ is positive semidefinite at every $\eta \neq 0$: otherwise $R(\delta_0) - R(\delta_f)$ would be negative for $\xi$ a large multiple of a direction with $\xi' A \xi < 0$. Take $\eta = re_1$, so that $L(\eta)_1 = 1/r$ and $e_1' A e_1 = \mathbb{E}_{re_1}[hZ_1^2/Q^2] - (2/r)\mathbb{E}_{re_1}[gZ_1/Q]$. Write $\mathbb{E}_{re_1}[u(Z)] = e^{-r^2/2}\mathbb{E}_0[u(Z)e^{rZ_1}]$. In the first term the integrand is bounded by $e^{|Z|}/|Z|^2$, which is integrable. In the second, $gZ_1/Q$ is odd, so $(2/r)\mathbb{E}_0[gZ_1 e^{rZ_1}/Q] = 2\mathbb{E}_0[gZ_1 \sinh(rZ_1)/(rQ)]$. Since $|\sinh(rx)/r| \le |x|e^{|x|}$ for $0 < r \le 1$, the integrand of the last expectation is bounded by $e^{|Z|}$. By dominated convergence and independence of the length and direction of $Z$ under $N(0, I)$,

$$\lim_{r \downarrow 0} e_1' A e_1 = \frac{1}{k}\left\{\mathbb{E}_{\chi^2_k}\left[\frac{h(Q)}{Q}\right] - 2\mathbb{E}_{\chi^2_k}[g(Q)]\right\} = \frac{1}{k(k-2)}\left\{\mathbb{E}[h(X)] - 2\mathbb{E}[g(X)X]\right\} \ge 0.$$

Step 2 (the variance reduction). Under $\rho = 0$ and $\pi = 0$, $Y \sim N(0, \Omega)$ is independent of $Z \sim N(0, I)$, so by (6), $\mathbb{E}_{0,0}[(\delta_0 - m)^2] = \mathrm{Var}_{0,0}(b(\hat{\pi})) + C_S$ and $\mathbb{E}_{0,0}[(\delta_0 - m)^2] - \mathbb{E}_{0,0}[(\delta_f - m)^2] = C_S \mathbb{E}[h(X)]$. Here $C_S > 0$, since $S$ is positive definite and $\ell(\hat{\pi}) \neq 0$ almost surely. It therefore suffices to show that $\mathbb{E}[h(X)] \le \psi_k$ for every $g$ satisfying Step 1, with equality if and only if $f = f_{c_k}$ almost everywhere. The variance reduction is then $C_S \mathbb{E}[h(X)] \le C_S \psi_k \le \psi_k \mathbb{E}_{0,0}[(\delta_0 - m)^2]$. Since $\psi_k > 0$, both inequalities hold with equality if and only if $f = f_{c_k}$ almost everywhere and $\mathrm{Var}_{0,0}(b(\hat{\pi})) = 0$. Since $b$ is continuous away from zero and $\hat{\pi}$ has a positive density, the variance is zero if and only if $b$ is constant on $\mathbb{R}^k \setminus \{0\}$.

Step 3 (duality). Let $g$ satisfy Step 1 and let $\mu > 0$. Then $\mathbb{E}[h(X)] \le \mathbb{E}[(1+\mu)h(X) - 2\mu g(X)X]$. For each $x \ge 0$ the function $\gamma \mapsto (1+\mu)(2\gamma - \gamma^2) - 2\mu\gamma x$ is strictly concave, with unique maximizer over $[0, 1]$ equal to $(1 - x/c)_+$, where $c = (1+\mu)/\mu > 1$. Let $g_c = (1 - x/c)_+$ and $h_c = 2g_c - g_c^2$, and let $\Phi(c) = \mathbb{E}[h_c(X) - 2g_c(X)X]$. Then $\mathbb{E}[h(X)] \le \mathbb{E}[h_c(X)] + \mu\Phi(c)$. If $\Phi(c) = 0$, then $g_c$ itself satisfies Step 1 and the bound equals $\mathbb{E}[h_c(X)]$. Equality for a feasible $g$ requires $g$ to attain the pointwise maximum almost everywhere, so $g = g_c$ almost

everywhere.

Step 4 (the root). The function $\Phi$ is continuous by dominated convergence. For $c \le 1$ and $x < c$, $1 + x/c - 2x \ge 1 - x > 0$, so $\Phi > 0$ on $(0, 1]$. As $c \to \infty$, $\Phi(c) \to \mathbb{E}[1 - 2X] = 1 - 2(k-2) < 0$. Hence $\Phi$ has a root $c_k > 1$. If $c$ and $c'$ were two roots, Step 3 would give $g_{c'} = g_c$ almost everywhere, and since $X$ has a positive density on $(0, \infty)$, $c' = c$. Writing $1 - f_{c_k}^2 = h_{c_k}$, Steps 2 to 4 prove the proposition. Finally, $(1+\mu)h_c(x) - 2\mu g_c(x)x = (1+\mu)g_c(x)^2$ for every $x \ge 0$, and $1 + \mu = c/(c-1)$. Step 3 applied to $g = g_{c_k}$ therefore gives $\psi_k \le \{c/(c-1)\}\mathbb{E}[g_c(X)^2]$ for every $c > 1$, with equality at $c = c_k$ because $\Phi(c_k) = 0$. Since $g_c(x)^2$ is nonincreasing in $x$ and $\chi^2_{k-2}$ is stochastically increasing in $k$, $\psi_k$ is nonincreasing in $k$. □

# B Sharper Certificates

This appendix gives the sufficient conditions behind the certified tuning parameters in Section 5. We use the notation of Appendix A. Let $X \sim \chi^2_{k-2}$, and for a taper $f$ let $g = 1 - f$, $h = 1 - f^2$, $U_m = \mathbb{E}[g(X)X^m]$, and $H_m = \mathbb{E}[h(X)X^m]$ for $m \ge 0$. With $n$ uniform on the unit sphere, let $M_0 = \mathbb{E}_n[L(n)L(n)']$ and $N = \mathbb{E}_n[L(n)n']$, and let $C_0$ be any constant such that

$$N e\, L(e)' + L(e)\, e' N' \preceq C_0 M_0 \qquad \text{for all unit vectors } e. \tag{11}$$

By (10) with $j = 0$, $C_0 = 2\kappa$ is always valid, but smaller values can often be certified, as discussed below. Recall $c = 4\kappa(k-1)H^2/s_*$.

**Proposition 3.** *Let $g : [0, \infty) \to [0, 1]$ be nonincreasing, absolutely continuous, not almost everywhere zero, and zero on $(Q_*, \infty)$. Suppose there exist $\zeta > 0$, $\bar{r} > 0$, and $\nu \in (0, 1)$ such that*

1. *$-2g'(q) \ge (\zeta - 1)g(q)$ for almost every $q$;*
2. *$H_0 \ge \bar{r} U_0$;*
3. *$H_0 - \nu U_0 \ge C_0 U_1$;*
4. *$(2m+1)(H_m - \nu U_m) \ge 2\kappa U_{m+1}$ for $1 \le m < \bar{M}$, where $\bar{M} = \max\{1, \lceil \kappa Q_*/(1-\nu) - 1/2 \rceil\}$; and*
5. *$\bar{r}\nu > c/\zeta$.*

*Then $\delta_f$ strictly dominates TSLS.*

*Proof.* We follow the proof of Proposition 1, with sharper bounds at each step. Converting the series expansions there to moments of $X$, the coefficient of $M_j$ in $T_h$ is proportional to $H_j$, and that of $v_j L(e)' + L(e)v_j'$ in $m_g L(\eta)' + L(\eta)m_g'$ is proportional to $U_{j+1}/(2j+1)$, with the same positive proportionality factor. Using (11) for $j = 0$ and (10) for $j \geq 1$, conditions 3 and 4 imply that each coefficient of $A$ dominates $\nu$ times the corresponding coefficient of $T_g = \mathbb{E}[gLL']$. For $m \geq \bar{M}$ this holds automatically, since $H_m \geq U_m$ and $U_{m+1} \leq Q_* U_m$. Hence $A \succeq \nu T_g$. Since $2-g$ is nondecreasing, tilting the distribution proportional to $g(X)$ by $X^m$ increases the mean of $2-g$, so $H_m/U_m \geq H_0/U_0 \geq \bar{r}$ for all $m$. Applying this in the series gives $T_h \succeq \bar{r} T_g$ and $c_0 \geq s_* \bar{r} J$, where $J = \mathbb{E}[g/Q]$. Under condition 1, $g - 2g' \geq \zeta g$, so the integration by parts argument gives $\mathbb{E}[g \sin^2 \angle(Z, \eta)] \leq (k-1)J/\zeta$. Matrix Cauchy–Schwarz with weight $g$ gives $zz' \preceq (cs_* J/\zeta) T_g$, so $z' A^{-1} z \leq cs_* J/(\zeta\nu)$. Completing the square in (9) gives $R(\delta_0) - R(\delta_f) \geq s_* J(\bar{r} - c/(\zeta\nu)) > 0$. Since $g$ vanishes at $Q_*$ and is absolutely continuous, the integration by parts has no boundary term at the cutoff. □

**The truncated-linear estimator.** For $g(q) = (1 - q/Q_*)_+$ condition 1 holds with $\zeta = 1 + 2/Q_*$. With $x = q/Q_*$ and $I_\beta(Q_*) = \int_0^1 x^{\beta-1}(1-x)e^{-Q_* x/2} dx$, we have $H_m/U_m = 1 + e_m$ and $U_{m+1}/U_m = Q_* e_m$, where $e_m = I_{(k-2)/2+m+1}(Q_*)/I_{(k-2)/2+m}(Q_*)$. Conditions 2 to 5 therefore reduce to finitely many inequalities between one-dimensional integrals, which we bound above and below by rational numbers using alternating Taylor expansions of $e^{-y}$. We locate numerically the supremum of the cutoffs passing these conditions, reduce it by a relative reserve of $10^{-5}$, and verify all conditions at the reduced value in exact arithmetic.

**The exponential estimator.** For $g(q) = e^{-\tau q}$, condition 1 holds with $\zeta = 1 + 2\tau$. Gamma integration gives $H_m/U_m = 2 - v^{(k-2)/2+m}$ with $v = (1 + 2\tau)/(1 + 4\tau)$, and $U_{m+1}/U_m = (k-2+2m)/(1+2\tau)$. Since $H_m/U_m$ increases in $m$ and $(k-2+2m)/(2m+1)$ is nonincreasing for $k \geq 3$, condition 4 for every $m \geq 1$ follows from the case $m = 1$. The argument of Proposition 3 then applies without the support restriction. With $d = k - 2$ and $\alpha = 1 + 2\tau$, so that $v = \alpha/(2\alpha - 1)$, and with $\bar{r} = 2 - v^{d/2}$, it therefore suffices that some $\nu \in (0, 1)$ satisfy

$$\bar{r} - \nu \geq \frac{C_0 d}{\alpha}, \qquad 3\left(2 - v^{d/2+1} - \nu\right) \geq \frac{2\kappa(d+2)}{\alpha}, \qquad \bar{r}\nu > \frac{c}{\alpha}.$$

We take the smallest $\tau$, up to a small reserve, that passes these conditions.

**Certified constants.** The constants $\kappa$, $s_*$, and $H$ enter only through upper bounds on $\kappa$ and $H$ and a lower bound on $s_*$. We obtain these from rational matrix inequalities, $m_K \Sigma_{\pi\pi}^{-1} \preceq W \preceq M_K \Sigma_{\pi\pi}^{-1}$, $S \succeq \underline{s}\, \Sigma_{\pi\pi}$, and $b_- W \preceq (W\Gamma + \Gamma' W)/2 \preceq b_+ W$, which give $\kappa \leq M_K/m_K$, $s_* \geq \underline{s}$, and $H \leq (b_+ - b_-)/2$. We check these inequalities by exact $LDL'$

factorization. For $C_0$, let $K = O\Lambda O'$ with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$. For $V \sim N(0, I_k)$ and $s = V'\Lambda V$, define $D_{1j} = \mathbb{E}[V_j^2/s]$ and $D_{2j} = \mathbb{E}[V_j^2/s^2]$. Then (11) holds whenever, for some $t > 0$,

$$C_0 \geq \max_j \frac{(tD_{1j} + 1)^2}{2t(k-2)D_{2j}\lambda_j}.$$

This follows from the inequality $xy' + yx' \preceq (tx + y)(tx + y)'/(2t)$, after a congruence that diagonalizes the problem. The quantities $D_{1j}$ and $D_{2j}$ are one-dimensional integrals, which we enclose by interval Taylor integration. The eigenvalues are enclosed using a validated eigenbasis, and the final inequalities are checked in exact arithmetic. Across our designs the certified values of $C_0$ range from 2.0 to 2.7, while $2\kappa$ reaches about 9. As an independent check, we recomputed $\kappa$, $s_*$, $H$, and $C_0$ from the exact covariance and weight matrices for all 70 designs, and confirmed that all four tuned estimators pass their certificates in every design.

# C Neither Estimator Dominates the Other

Let $k = 3$, $\Sigma = I_6$, and $W = I_3$. Then $\Gamma = 0$, $S = I_3$, $b \equiv 0$, $\kappa = 1$, $s_* = 1$, $H = 0$, and $\bar{q} = 1/2$. Consider the truncated-linear estimator with $Q_* = 1/2$, which dominates TSLS by Theorem 1. Consider also the exponential estimator with $\tau = 3/5$, which dominates TSLS by Proposition 1, since $1 + 2\tau = 11/5 > 2$.

Let $a = \beta e_1$ and $\pi = re_1$. Since $b \equiv 0$, $z = 0$ in (9), so the improvement over TSLS is $\beta^2 e_1' A e_1 + c_0$. The limit computed in the proof of Proposition 2 applies here, and the same dominated convergence argument gives $c_0 \to \mathbb{E}_{\chi^2_3}[h(Q)/Q]$ as $r \downarrow 0$. Hence, as $r \downarrow 0$, the improvement of $\delta_f$ converges to

$$G_f(\beta) = \mathcal{H}_f + \frac{\beta^2}{3}(\mathcal{H}_f - 2\mathcal{M}_f), \qquad \mathcal{H}_f = \mathbb{E}_{\chi^2_3}\left[\frac{1 - f(Q)^2}{Q}\right], \quad \mathcal{M}_f = \mathbb{E}_{\chi^2_3}[1 - f(Q)].$$

For $\beta \neq 0$ both risks diverge as $r \downarrow 0$, since the target $\beta/r$ diverges, but their difference converges.

For the exponential estimator, with $t = 1 + 2\tau = 11/5$ and $u = 1 + 4\tau = 17/5$, gamma integration gives $\mathcal{H}_{f_e} = 2t^{-1/2} - u^{-1/2}$ and $\mathcal{M}_{f_e} = t^{-3/2}$. For the truncated-linear estimator, the substitution $Q = s^2/2$ gives

$$\mathcal{H}_{f_c} = \pi^{-1/2} \int_0^1 (1 - s^4) e^{-s^2/4} ds, \qquad \mathcal{H}_{f_c} - 2\mathcal{M}_{f_c} = \pi^{-1/2} \int_0^1 (1 - s^2) e^{-s^2/4} ds,$$

where, here and in the bounds below, $\pi \approx 3.14159$ denotes the mathematical constant rather

than the first-stage coefficient. The following bounds use only rational arithmetic and the facts $3 < \pi < 3.15$, $1.4832 < \sqrt{11/5} < 1.4833$, and $1.8439 < \sqrt{17/5} < 1.8440$. At $\beta = 0$, $G_{f_e}(0) = 2t^{-1/2} - u^{-1/2} > 2/1.4833 - 1/1.8439 > 0.806$, while $G_{f_c}(0) \le \pi^{-1/2} \int_0^1 (1 - s^4) ds < 0.8/\sqrt{3} < 0.47$.

At $\beta = 3$, $G_{f_e}(3) = 8t^{-1/2} - 4u^{-1/2} - 6t^{-3/2} < 8/1.4832 - 4/1.8440 - 6/(2.2 \cdot 1.4833) < 1.386$. Since $4 - 3s^2 - s^4 \ge 0$ on $[0, 1]$ and $e^{-y} \ge 1 - y$, we have $G_{f_c}(3) = \pi^{-1/2} \int_0^1 (4 - 3s^2 - s^4) e^{-s^2/4} ds \ge \pi^{-1/2} \int_0^1 (4 - 3s^2 - s^4)(1 - s^2/4) ds$. The right side equals $(4 - 4/3 - 1/20 + 1/28)/\sqrt{\pi} > 2.652/\sqrt{3.15} > 1.49$. Hence for all sufficiently small $r > 0$, the exponential estimator has strictly lower risk than the truncated-linear estimator at $(a, \pi) = (0, re_1)$, and strictly higher risk at $(a, \pi) = (3e_1, re_1)$. Numerically, the ordering already holds at $r = 0.2$. There the improvements over TSLS are 0.793 for the exponential and 0.418 for the truncated-linear estimator when $\beta = 0$, and 1.367 and 1.473 when $\beta = 3$.

# D Numerical Methods

**Risk of TSLS and the exponential estimator.** Both risks reduce to one-dimensional integrals. Given $\hat{\pi}$, the reduced form enters quadratically and can be integrated exactly. The remaining expectations involve ratios of quadratic forms in $\hat{\pi}$, multiplied for the exponential estimator by $\exp(-\tau Q)$. Writing $1/s = \int_0^\infty e^{-xs} dx$ and $1/s^2 = \int_0^\infty x e^{-xs} dx$ for $s = \hat{\pi}' W \hat{\pi}$, each such expectation is a one-dimensional integral of a Gaussian moment, which we evaluate by adaptive quadrature. We compute the improvement of the exponential estimator directly, rather than as a difference of risks, which preserves accuracy when the improvement is small.

**Risk of the truncated-linear estimator.** After integrating out the reduced form, the improvement over TSLS is an integral over $\hat{\pi}$ restricted to the ball $Q \le Q_*$. In polar coordinates the radial integral is over $[0, \sqrt{Q_*}]$, and pairing antipodal directions replaces the exponential tilt by hyperbolic sines and cosines. With three instruments we use a product Gauss–Legendre rule, in the radius and in the cosine of the angle to $\Sigma_{\pi\pi}^{-1/2} \pi$, combined with a uniform azimuth grid. We compare a coarse grid with a fine one. With more instruments we combine the same radial rule with Monte Carlo over directions. The resulting MSE improvements, at the estimates and at the posterior draws, are below $10^{-21}$ as a fraction of TSLS risk. We also bound them analytically using $|L(z)| \le \kappa/|z|$ and the stochastic ordering of noncentral $\chi^2$ distributions.

**Posterior averages.** With three instruments, the risk of TSLS generically grows like $1/|\pi|^2$ as $\pi \to 0$, as does the squared target. Its posterior average is therefore finite, but

its square is not integrable, so a plain Monte Carlo average over posterior draws has infinite variance. We instead draw the first-stage coefficients from a mixture of the posterior and a distribution whose density near the origin is proportional to $|\Sigma_{\pi\pi}^{-1/2}\pi|^{-2}$, and draw $a$ from its exact posterior. Reweighting by the ratio of densities yields an unbiased estimator with finite variance, and the delta method gives standard errors for the ratio of averages. Using 20,000 draws for each design, the standard error of each design's posterior-average percentage reduction is below 0.36 percentage points. A separate plain Monte Carlo run with 40,000 posterior draws agrees with the importance-sampling estimates to within 1.6 importance-sampling standard errors in the two specifications where the posterior improvements are non-negligible. For designs with more instruments, the squared risk is integrable and we use 400 posterior draws.

**Validation.** Our TSLS risks and certified exponential improvements reproduce, to a relative error below $10^{-10}$, values saved from an earlier and separate run of the same computation. Our quadrature for the truncated-linear estimator matches an independent implementation to about $10^{-14}$. It also matches the limit derived in Appendix C. A brute-force simulation from the raw formulas, with $10^7$ draws in the two specifications with non-negligible improvements, matches the computed risk of each of the four estimators to within two simulation standard errors. Unlike the loss of TSLS with three instruments, their losses have finite variance.

# References


Levent Alpöge. Post on X announcing a counterexample to the Jacobian conjecture, July 20, 2026. URL https://x.com/__alpoge__/status/2079028340955197566.

T. W. Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20(1):46–63, 1949. doi: 10.1214/aoms/1177730090.

Isaiah Andrews. On the structure of IV estimands. *Journal of Econometrics*, 211(1):294–307, 2019. doi: 10.1016/j.jeconom.2018.12.017.

Isaiah Andrews and Timothy B. Armstrong. Unbiased instrumental variables estimation under known first-stage sign. *Quantitative Economics*, 8(2):479–503, 2017. doi: 10.3982/QE700.

Isaiah Andrews and Anna Mikusheva. Optimal decision rules for weak GMM. *Econometrica*, 90(2):715–748, 2022. doi: 10.3982/ECTA18678.

Isaiah Andrews and Anna Mikusheva. GMM is inadmissible under bounded identification strength. Working paper. An earlier version is arXiv:2204.12462, "GMM is Inadmissible Under Weak Identification", 2024.

Isaiah Andrews, James H. Stock, and Liyang Sun. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11:727–753, 2019. doi: 10.1146/annurev-economics-080218-025643.

Joshua D. Angrist and Guido W. Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995. doi: 10.1080/01621459.1995.10476535.

Christine Blandhol, John Bonney, Magne Mogstad, and Alexander Torgovitsky. When is TSLS actually LATE? *Review of Economic Studies*, 2026. doi: 10.1093/restud/rdag029. Advance access, April 2026.

Tom Boot and Didier Nibbering. Inference on LATEs with covariates. Working paper, arXiv:2402.12607, 2024.

Lawrence D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, volume 9 of *IMS Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1986. doi: 10.1214/lnms/1215466757.

Gary Chamberlain. Decision theory applied to an instrumental variables model. *Econometrica*, 75(3):609–652, 2007. doi: 10.1111/j.1468-0262.2007.00764.x.

Kirill S. Evdokimov and Michal Kolesár. Inference in instrumental variables analysis with heterogeneous treatment effects. Working paper, Princeton University Department of Economics 2018-16; version of January 2019, 2019.

Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624, 2020. doi: 10.1257/aer.20181047.

Bruce E. Hansen. Stein-like 2SLS estimator. *Econometric Reviews*, 36(6–9):840–852, 2017. doi: 10.1080/07474938.2017.1307579.

Keisuke Hirano and Jack R. Porter. Location properties of point estimators in linear instrumental variables and related models. *Econometric Reviews*, 34(6–10):720–733, 2015. doi: 10.1080/07474938.2014.956573.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994. doi: 10.2307/2951620.

W. James and Charles Stein. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 361–379. University of California Press, 1961.

Terrence W. Kinal. The existence of moments of $k$-class estimators. *Econometrica*, 48(1): 241–249, 1980. doi: 10.2307/1912027.

Frank Kleibergen. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803, 2002. doi: 10.1111/1468-0262.00353.

Frank Kleibergen and Zhaoguo Zhan. Double robust inference for continuous updating GMM. *Quantitative Economics*, 16(1):295–327, 2025. doi: 10.3982/QE2347.

Seojeong Lee. A consistent variance estimator for 2SLS when instruments identify different LATEs. *Journal of Business & Economic Statistics*, 36(3):400–410, 2018. doi: 10.1080/07350015.2016.1186555.

E. L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer, New York, third edition, 2005.

Magne Mogstad, Alexander Torgovitsky, and Christopher R. Walters. The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review*, 111(11):3663–3698, 2021. doi: 10.1257/aer.20190221.

Marcelo J. Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048, 2003. doi: 10.1111/1468-0262.00438.

OpenAI. On the Navier–Stokes Millennium Prize Problem. OpenAI research post, September 8, 2026a. URL https://openai.com/index/navier-stokes-solution/.

OpenAI. Sharing AI progress in mathematics. OpenAI research post, October 6, 2026b. URL https://openai.com/index/sharing-ai-progress-in-mathematics/. The manuscripts are at https://github.com/openai/math.

Arthur Freitas Ramos, David Barros Hulak, and Ruy Jose Guerra Barretto de Queiroz. Formal verification of an explicit counterexample to the Jacobian conjecture. *Archive of Formal Proofs*, July 2026. URL https://isa-afp.org/entries/Jacobian_Counterexample.html.

Julio J. Rotemberg. Instrument variable estimation of misspecified models. Working Paper 1508-83, Sloan School of Management, Massachusetts Institute of Technology, 1983.

Jann Spiess. Bias reduction in instrumental variable estimation through first-stage shrinkage. Working paper, arXiv:1708.06443, 2017.

Douglas Staiger and James H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, 1997. doi: 10.2307/2171753.

Charles Stein. 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. University of California Press, 1956.

Melvin Stephens, Jr. and Dou-Yan Yang. Compulsory education and the benefits of schooling. *American Economic Review*, 104(6):1777–1792, 2014. doi: 10.1257/aer.104.6.1777.

James H. Stock and Jonathan H. Wright. GMM with weak identification. *Econometrica*, 68 (5):1055–1096, 2000. doi: 10.1111/1468-0262.00151.

Arnstein Vestre. Inference on the TSLS estimand with weak instruments and treatment effect heterogeneity. Working paper, arXiv:2606.07871, 2026.

Luther Yap. Inference with many weak instruments and heterogeneity. Working paper, arXiv:2408.11193, 2025.

Alwyn Young. Structural transformation, the mismeasurement of productivity growth, and the cost disease of services. *American Economic Review*, 104(11):3635–3667, 2014. doi: 10.1257/aer.104.11.3635.