Appendix A Derivations

Estimation with Many Decision-makers

To analyze the estimators in the main text, we follow the literature on many instruments (e.g., Bekker 1994; Chao, Swanson, Hausman, et al. 2012), and condition on the realizations of the instruments and covariates, $\{z_i, w_i\}_{i=1}^n$, throughout this section. To ease notation, we keep the conditioning implicit, writing, e.g., $E[y_i]$ rather than $E[y_i \mid \{z_i, w_i\}_{i=1}^n]$. To simplify the analysis, we also assume that the first-stage equation is correctly specified, so that the residual ν_i is mean zero, and not just uncorrelated with the instruments and covariates,

$$E[\nu_i] = 0. \tag{1}$$

Since we are assuming that the instrument is as-good-as-randomly assigned conditional on the covariates, it follows that $E[\varepsilon_i]$ depends only on w_i and not on z_i. To simplify derivations, we further assume that the dependence is linear:

$$E[\varepsilon_i] = w_i'\alpha \tag{2}$$

for some vector α. Similar to, for example, Chao, Swanson, Hausman, et al. (2012) we could allow for asymptotically negligible non-linearities in both eqs. (1) and (2) at the cost of complicating the derivations without affecting the results. Allowing for asymptotically non-negligible non-linearities requires treating z_i as random. Finally, we assume that the errors (ε_i, ν_i) are independent across i.

First consider the infeasible estimator $\hat{\beta}^*$. Since $\tilde{\ell}_i$ depends only on the instruments and covariates, it follows using eq. (1) that $E[\sum_i \tilde{\ell}_i x_i] = \sum_i \tilde{\ell}_i \ell_i = \sum_i \tilde{\ell}_i^2$. Similarly, using eq. (2), we have $E[\sum_i \tilde{\ell}_i \varepsilon_i] = \sum_i \tilde{\ell}_i w_i'\alpha = 0$, where the last equality uses the fact that $\tilde{\ell}_i$ and w_i are orthogonal by definition of a sample residual, $\sum_i \tilde{\ell}_i w_i = 0$. Approximating the expectation of a ratio by a ratio of expectations, we therefore obtain

$$E[\hat{\beta}^*] - \beta \approx \frac{0}{\sum_i \tilde{\ell}_i^2} = 0.$$

While using the ratio of expectations to approximate the expectation of a ratio may appear ad hoc, one can show that the approximation becomes exact asymptotically under the many instrument asymptotics proposed by Bekker (1994), where the dimension K of the instrument vector and the dimension L of the covariate vector grow with sample size—the results we give below mirror the asymptotic bias results given in Evdokimov and Kolesár (2018), for instance. Therefore, an alternative interpretation of the approximation $\approx$ is that it corresponds to the probability limit under the Bekker (1994) sequence.

To derive the bias expressions for the other estimators, it is convenient to use matrix notation. We denote the vectors of n treatment and outcome observations by x and y, respectively, and denote the matrix of covariates and instruments with rows w_i' and z_i' by W and Z, respectively. To state the results, it will be convenient to link the sum of squares of the relative leniency, $\sum_i \tilde{\ell}_i^2$, to the first-stage F statistic and the partial R^2. Recall that the first-stage F statistic for testing the null that $\pi = 0$ is, under homoskedastic errors, given by $F = \hat{\pi}'\tilde{Z}'\tilde{Z}\hat{\pi}/K\operatorname{var}(\nu_i)$, with expectation $E[F] = \sum_i \tilde{\ell}_i^2/K\operatorname{var}(\nu_i) + 1$. The population partial R^2 from adding the instruments to the first-stage regression is given by $R^2 = (\sum_i \tilde{\ell}_i^2)/(\sum_i \tilde{\ell}_i^2 + n\operatorname{var}(\nu_i))$ so

$$\frac{\sum_i \tilde{\ell}_i^2}{\operatorname{var}(\nu_i)} = K(E[F] - 1) = \frac{nR^2}{1 - R^2}. \tag{3}$$

This identity allows us to state the bias of a large class of estimators using the first-stage F statistic, as the following lemma shows.

Lemma 1. *Consider an estimator $\hat{\beta}_G = y'Gx/x'Gx$ based on a relative leniency measure Gx such that $Gx = \tilde{\ell} + G\nu$. Then, under homoskedastic errors, $E[\hat{\beta}_G] - \beta \approx \frac{\operatorname{tr}(G)\operatorname{cov}(\varepsilon_i,\nu_i)}{\sum_i \tilde{\ell}_i^2 + \operatorname{tr}(G)\operatorname{var}(\nu_i)} = \frac{\operatorname{tr}(G)}{K(E[F]-1)+\operatorname{tr}(G)} \frac{\operatorname{cov}(\varepsilon_i,\nu_i)}{\operatorname{var}(\nu_i)}$.*

The lemma follows by noting that $E[x'Gx] = \sum_i \tilde{\ell}_i^2 + E[\nu'G\nu] = \sum_i \tilde{\ell}_i^2 + \sum_i G_{ii}E[\nu_i^2]$, and $E[\varepsilon'Gx] = \sum_i G_{ii}\operatorname{cov}(\varepsilon_i, \nu_i)$. Under homoskedastic errors, $E[\nu_i^2]$ and $\operatorname{cov}(\varepsilon_i, \nu_i)$ do not depend on the index i, which yields the first expression. The second expression follows directly from eq. (3).

The ordinary least squares estimator $\hat{\beta}_{OLS} = \sum_i \tilde{x}_i y_i / \sum_i \tilde{x}_i^2$ uses the relative leniency measure Mx, where $M = I - W(W'W)^{-1}W'$ denotes the $n \times n$ annihilator matrix associated with the covariates. Since

the trace of M is $n-L$, where L is the covariate dimension, it follows from Lemma 1 that its bias is approximately

$$E[\hat{\beta}_{OLS}]-\beta \approx \frac{1-L/n}{K(E[F]-1)/n+1-L/n}\frac{\text{cov}(\varepsilon_i,\nu_i)}{\text{var}(\nu_i)} \approx (1-R^2)\frac{\text{cov}(\varepsilon_i,\nu_i)}{\text{var}(\nu_i)},$$

where the second approximation uses eq. (3) and the assumption that L/n is close to zero.

The 2SLS estimator uses the relative leniency measure $\tilde{z}_i\hat{\pi}$, which in matrix notation may be written Hx, with $H=\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'$ denoting the projection matrix associated with the instrument $\tilde{z}_i$. This matrix has trace equal to K, which yields

$$E[\hat{\beta}_{TSLS}]-\beta \approx \frac{1}{E[F]}\frac{\text{cov}(\varepsilon_i,\nu_i)}{\text{var}(\nu_i)}$$

Hence, the relative bias of 2SLS, $(E[\hat{\beta}_{TSLS}]-\beta)/(E[\hat{\beta}_{OLS}]-\beta)$, is approximately given by $\frac{1}{(1-R^2)E[F]}$, as claimed in the main text.

To show approximate unbiasedness of UJIVE, observe that we can write its leniency measure using matrix notation as $G_{UJIVE}x$ with $G_{UJIVE}=H-\text{diag}(H_{ii}/(M_{ii}-H_{ii}))(M-H)$. Note that this matrix formula also implies that UJIVE can be implemented in one step: one only needs to compute the diagonals of the projection matrices H and $M-H$, the fitted values Hx, and the residuals $(M-H)x$, all of which obtain easily from regressing x_i onto $\tilde{z}_i$ and onto z_i, w_i. The unbiasedness property then follows directly from Lemma 1 by noting that $\text{tr}(G_{UJIVE})=0$.

For the JIVE1 estimator studied in Angrist, Imbens, and Krueger (1999), the relative leniency measure can be written as $G_{JIVE}x$ with $G_{JIVE}=M(I-D_Q)^{-1}(H_Q-D_Q)$, where $H_Q=H+W(W'W)^{-1}W'$ is the projection matrix associated with the full vector of right-hand side variables (z_i,w_i), and D_Q is its diagonal. Since $\text{tr}(G)=\text{tr}((I-D_Q)^{-1}(H-D_QM))=-L$, we obtain

$$E[\hat{\beta}_{JIVE}-\beta] \approx -\frac{L}{K(E[F]-1)-L}\frac{\text{cov}(\varepsilon_i,\nu_i)}{\text{var}(\nu_i)},$$

which equals the 2SLS bias times $-\frac{E[F]}{K(E[F]-1)/L-1}$, as claimed in the main text.

For IJIVE, the G matrix takes the form $G_{IJIVE}=M(I-\text{diag}(H))^{-1}(H-\text{diag}(H))M$, with trace equal to $\text{tr}((I-\text{diag}(H))^{-1}(H-\text{diag}(H)M))=\sum_i\frac{H_{ii}(1-M_{ii})}{1-H_{ii}}$. While this quantity is always positive, under balanced design, i.e., when the diagonal elements of H_{ii} are all approximately equal, it simplifies to $LK/(n-K)$, which is much smaller than $\text{tr}(G_{JIVE})$.

To derive the bias-corrected 2SLS estimator, observe that the proof of Lemma 1 shows that under homoskedasticity, the numerator and denominator of the 2SLS formula have expectations $\frac{1}{n}\sum_i E[\hat{\ell}_i y_i]=\frac{1}{n}\sum_i\tilde{\ell}_i^2\beta+K\,\text{cov}(\nu_i\beta+\varepsilon_i,\nu_i)/n$, and $\frac{1}{n}\sum_i E[\hat{\ell}_i x_i]=\frac{1}{n}\sum_i\tilde{\ell}_i^2+K\,\text{var}(\nu_i)/n$, respectively. An unbiased estimator of $\text{var}(\nu_i)$ obtains as the degrees-of-freedom adjusted sample variance of the first-stage residuals, $\hat{\nu}=(M-H)x$, given by $\widehat{\text{var}}(\nu_i)=\sum_i\hat{\nu}_i^2/(n-K-L)=x'(M-H)x/(n-K-L)$. Since $\nu_i\beta+\varepsilon_i$ corresponds to the reduced form residual from regressing y onto Z and W, the degrees-of-freedom adjusted sample covariance between the first-stage residuals and reduced-form residuals, $(M-H)y$, given by $y'(M-H)x/(n-K-L)$, gives an unbiased estimator of $\text{cov}(\nu_i\beta+\varepsilon_i,\nu_i)$. Subtracting estimates of 2SLS numerator and denominator bias gives the bias-corrected 2SLS estimator,

$$\hat{\beta}_{B2SLS}=\frac{\sum_i\hat{\ell}_i y_i-Ky'(M-H)x/(n-K-L)}{\sum_i\hat{\ell}_i x_i-Kx'(M-H)x/(n-K-L)}=\frac{y'Hx-Ky'(M-H)x/(n-K-L)}{x'Hx-Kx'(M-H)x/(n-K-L)},$$

which uses the leniency measure $G_{B2SLS}x$ with $G_{B2SLS}=H-K(M-H)/(n-K-L)$. This has trace equal to 0, and $\hat{\beta}_{B2SLS}$ is therefore unbiased under homoskedasticity. This version of the bias-corrected 2SLS estimator corresponds to the version studied in Kolesár, Chetty, et al. (2015).

Finally, we show that the UJIVE estimator and the FEJIV estimator of Chao, Swanson, and Woutersen (2023) can be interpreted as minimizing the mean-squared error of the estimated leniency under homoskedasticity. It follows from the proof of Lemma 1 that a leniency measure Gx is unbiased under heteroskedasticity if the diagonal of G is zero, $G_{ii}=0$, and if $GW=0$ and $G\tilde{Z}=\tilde{Z}$ (the latter two conditions are equiva-

lent to the condition $Gx = \tilde{\ell} + G\nu$ in Lemma 1). Under these conditions, the mean squared error of the leniency measure is $E[(Gx-\tilde{\ell})'(Gx-\tilde{\ell})] = E[\nu'G'G\nu]$. When ν_i is homoskedastic, this expectation simplifies to $\text{var}(\nu_i)\,\text{tr}(G'G)$. Therefore, minimizing the mean squared error subject to an unbiasedness constraint is equivalent to minimizing $\text{tr}(G'G)$ subject to $\text{diag}(G) = 0$, $GW = 0$, and $G\tilde{Z} = \tilde{Z}$. The first-order condition for the Lagrangian associated with this problem is given by

$$G' = W\Pi' + \tilde{Z}A' + \text{diag}(\lambda),$$

where Π is the matrix of the Lagrange multipliers associated with the constraint $GW = 0$, A is the matrix of the Lagrange multipliers associated with the constraint $G\tilde{Z} = \tilde{Z}$, and λ is the vector of Lagrange multipliers associated with the constraint $\text{diag}(G) = 0$. Multiplying the first-order condition by W' and $\tilde{Z}'$, respectively, allows us to solve for Π and A: $\Pi' = -(W'W)^{-1}W'\,\text{diag}(\lambda)$ and $A' = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'(I - \text{diag}(\lambda))$. Plugging this back into the first-order condition then yields $G' = H + (M - H)\,\text{diag}(\lambda)$. Since the diagonal of G is zero, this implies $\lambda_i = -H_{ii}/(M_{ii} - H_{ii})$. Thus, $G_{UJIVE} = H - \text{diag}(H_{ii}/(M_{ii} - H_{ii}))(M - H)$ solves the minimization problem: the UJIVE leniency measure minimizes the mean squared error subject to an unbiasedness constraint.

Now consider the same minimization problem, but subject to the additional constraint that $W'G = 0$ and $\tilde{Z}'G = \tilde{Z}'$. The first additional constraint implies that $W'Gx = 0$, so that the leniency measure is orthogonal to the covariates. The second additional constraint ensures that the in-sample covariance of the leniency measure with $\tilde{Z}$ is the same as the in-sample covariance of x with $\tilde{Z}$: $\tilde{Z}'Gx = \tilde{Z}'x$. These additional constraints imply that the resulting estimator corresponds to the MINQUE estimator in Rao (1970); they also ensure that the optimal G matrix is symmetric. Under these additional constraints, the first-order condition becomes

$$G' = W\Pi' + \tilde{\Pi}'W + \tilde{Z}A' + \tilde{A}'\tilde{Z}' + \text{diag}(\lambda),$$

where $\tilde{\Pi}$ is the matrix of Lagrange multipliers associated with the constraint $W'G = 0$ and $\tilde{A}$ is the matrix of Lagrange multipliers associated with the constraint $\tilde{Z}'G = \tilde{Z}'$. Solving for the Lagrange multipliers as before, we obtain $G' = H - (M - H)\,\text{diag}(\lambda)(M - H)$. Since the diagonal of G is zero, this implies $\text{diag}(H) = \text{diag}((M - H)\,\text{diag}(\lambda)(M - H))$, which is equivalent to

$$\text{diag}(H) = ((M - H) \odot (M - H))\lambda, \tag{4}$$

where $A \odot B$ denotes the Hadamard (elementwise) product of two matrices, $(A \odot B)_{ij} = A_{ij}B_{ij}$. Provided that a solution λ to the linear system in eq. (4) exists (a sufficient, but not a necessary, condition is that $(M - H) \odot (M - H)$ is invertible) we therefore obtain the solution

$$G_{FEJIV} = H - (M - H)\,\text{diag}(\lambda)(M - H),$$

which corresponds precisely to the FEJIV estimator of Chao, Swanson, and Woutersen (2023). Because the minimization imposes additional constraints, it follows that when ν_i is homoskedastic, $E[(G_{UJIVE}x - \tilde{\ell})'(G_{UJIVE}x - \tilde{\ell})] \leq E[(G_{FEJIV}x - \tilde{\ell})'(G_{FEJIV}x - \tilde{\ell})]$. Since the linear system in eq. (4) has dimension n, solving it may be challenging in large datasets (if n is in the tens or hundreds of thousands, which is common in leniency applications). However, the additional constraint $W'G = 0$ ensures that the FEJIV estimator is invariant to adding linear functions of the covariates to the outcome equation: this desirable property is not shared by UJIVE.

Heterogeneous Effects

We now relax the assumption that treatment effects are constant and give a simple statement of the local average treatment effect theorem of Imbens and Angrist (1994). As before, assume that there are K examiners, with $k(i)$ denoting the examiner assigned to observation i, and z_i denoting a vector of indicators for examiner assignment. In a departure from the previous section, we no longer condition on the instruments and covariates, but treat the sample $\{y_i, x_i, z_i, w_i\}_{i=1}^n$ as *iid*. Let $\ddot{\ell}_i = \ddot{z}_i'\pi$ denote the population version of the relative leniency, where $\ddot{z}_i$ is the population residual from regressing z_i onto w_i (while the relative leniency measure $\tilde{\ell}_i = \tilde{z}_i'\pi$ uses the sample residual, $\ddot{\ell}_i$ uses the population residual). Since $\ddot{z}_i$ depends only on w_i and the examiner assignment, we may write $\ddot{\ell}_i = \ddot{\ell}(k(i), w_i)$. We assume that the conditional mean

of z_i given w_i is linear, so that $E[\ddot{z}_i \mid w_i] = 0$. However, we relax the assumption that the first-stage is linear, so that $\ddot{\ell}(k, w_i)$ is an approximation to, but does not necessarily equal, the true relative leniency $\ell^*(k, w_i) - E[\ell^*(k(i), w_i)]$, where $\ell^*(k, w) = E[x_i \mid k(i) = k, w_i = w]$ denotes the true absolute leniency.

Using $\ddot{\ell}_i$ as an instrument in a population IV regression of y_i onto x_i without any controls identifies

$$\beta^* = \frac{E[\ddot{\ell}_i y_i]}{E[\ddot{\ell}_i x_i]}.$$

The next result shows that β^* can be written as a weighted average of simple IV regressions restricted to the subpopulation with $w_i = w$ and $k(i) = k$ or $k(i) = j$. These simple IV regressions identify

$$\beta(k, j, w) = \frac{E[y_i \mid k(i) = k, w_i = w] - E[y_i \mid k(i) = j, w_i = w]}{E[x_i \mid k(i) = k, w_i = w] - E[x_i \mid k(i) = j, w_i = w]}.$$

Lemma 2. *Suppose that $E[\ddot{z}_i \mid w_i] = 0$. Then leniency IV estimand β^* may be written as a weighted average of pairwise IV regressions,*

$$\beta^* = \frac{\sum_{k>j} E[\omega(k, j, w_i)\beta(k, j, w_i)]}{\sum_{k>j} E[\omega(k, j, w_i)]}, \quad \omega(j, k, w) = p_j(w)p_k(w)(\ddot{\ell}(k, w) - \ddot{\ell}(j, w))(\ell^*(k, w) - \ell^*(j, w)). \quad (5)$$

Proof. Recall the identity $E[(A_i - A_{i'})(B_i - B_{i'})] = 2\operatorname{cov}(A_i, B_i)$, where (A_i, B_i) is a pair of random variables, and $(A_{i'}, B_{i'})$ is an independent copy. Let $\beta_N(k, j, w) = E[y_i \mid k(i) = k, w_i = w] - E[y_i \mid k(i) = j, w_i = w]$ denote the numerator of $\beta(j, k, w)$. By iterated expectations, $E[\ddot{\ell}_i y_i] = E[\operatorname{cov}(\ddot{\ell}_i, y_i) \mid w_i]$, so by the covariance identity,

$$\begin{aligned}
E[\ddot{\ell}_i y_i] &= \frac{1}{2} E[E[(\ddot{\ell}_i - \ddot{\ell}_{i'})(y_i - y_{i'}) \mid w_i = w_{i'}]] \\
&= \frac{1}{2} E[E[(\ddot{\ell}(k(i), w_i) - \ddot{\ell}(k(i'), w_i))\beta_N(k(i), k(i'), w_i) \mid w_i = w_{i'}]] \\
&= \frac{1}{2} \sum_{k,j} E[p_j(w_i)p_k(w_i)(\ddot{\ell}(k, w_i) - \ddot{\ell}(j, w_i))\beta_N(k, j, w_i)] \\
&= \sum_{k>j} E[p_j(w_i)p_k(w_i)(\ddot{\ell}(k, w_i) - \ddot{\ell}(j, w_i))\beta_N(k, j, w_i)].
\end{aligned}$$

An analogous argument applied to the denominator yields the result. □

The result shows that leniency IV identifies a weighted average of simple IV regressions, where we restrict the sample to observations assigned to one of two examiners in a given pair (j, k), and restrict the covariates to equal a particular value. If the first stage is correctly specified, $\ddot{\ell}(k, w) - \ddot{\ell}(j, w) = \ell^*(k, w) - \ell^*(j, w)$, in which case the weights are proportional to the product of the conditional assignment probabilities to j and k times the squared differences in examiner leniency.

Lemma 2 is an algebraic result, and holds without any assumptions on the model. To give a causal interpretation to β^*, we need a causal interpretation for $\beta(k, j, w)$. To this end, if the instruments are as-good-as-randomly assigned conditional on w_i, and an exclusion restriction holds (but we do not impose monotonicity), Proposition 3 in Angrist, Imbens, and Rubin (1996) shows that $\beta(k, j, w)$ is given by a weighted difference between treatment effects for compliers (those who are treated when assigned to k but not when assigned to j), and defiers (those who are treated when assigned to j but not when assigned to k). The weight on defiers is negative and equal to the proportion of defiers divided by the leniency difference, while the weight on compliers is positive and equals the proportion of compliers divided by the leniency difference:

$$\beta(k, j, w) = \frac{E[\beta_i C_i(k, j, w)]}{\ell^*(k, w) - \ell^*(j, w)}, \quad (6)$$

where β_i is the individual treatment effect, and $C_i(k, j, w)$ denotes a variable that equals 1 if an observation is a complier and -1 if they are a defier, so that $\ell^*(k, w) - \ell^*(j, w) = E[C_i(k, j, w)]$. Under monotonicity, there are no defiers, so that $\beta(k, j, w)$ corresponds to the average treatment effect for compliers. Combining

this interpretation with Lemma 2 then yields the version of the local average treatment effect theorem we referred to in the main text.

The next result generalizes the local average treatment effect theorem to allow for the presence of defiers. To state the result, let $x_i(k)$ denote the potential treatment indicator that equals 1 if examiner k would treat observation i, so that $C_i(k, j, w) = x_i(k) - x_i(j)$.

Lemma 3. *Suppose that* $E[\ddot{z}_i \mid w_i] = 0$ *and that eq. (6) holds. Then* $\beta^* = E[\lambda_i \beta_i]/E[\lambda_i]$*, where*

$$\lambda_i = \text{cov}(\ddot{\ell}(k(i), w_i), x_i(k(i)) \mid i) = (\bar{\ell}_i(1) - \bar{\ell}_i(0))(1 - \bar{x}_i)\bar{x}_i.$$

Here $\bar{x}_i = \sum_k p_k(w_i) x_i(k)$ *is the average treatment rate of unit* i*,* $\bar{\ell}_i(1) = \sum_k p_k(w_i)\ddot{\ell}(k, w_i)x_i(k)/\bar{x}_i$ *is the average relative leniency of examiners who would treat* i*, and* $\bar{\ell}_i(0) = \sum_k p_k(w_i)\ddot{\ell}(k, w_i)(1 - x_i(k))/(1 - \bar{x}_i)$ *is the average leniency of those who wouldn't.*

Proof. By Lemma 2, eq. (6) and iterated expectations,

$$E[\ddot{\ell}_i y_i] = \sum_{k>j} E[\omega(k, j, w_i)\beta(k, j, w_i)] = \sum_{k>j} E[p_j(w_i)p_k(w_i)(\ddot{\ell}(k, w_i) - \ddot{\ell}(j, w_i))(x_i(k) - x_i(j))\beta_i] = E[\lambda_i \beta_i],$$

where $\lambda_i = \sum_{k>j} p_j(w_i)p_k(w_i)(\ddot{\ell}(k, w_i) - \ddot{\ell}(j, w_i))(x_i(k) - x_i(j))$, while $E[\ddot{\ell}_i x_i] = E[\lambda_i]$. Using the definitions of $\bar{\ell}_i(1), \bar{\ell}_i(0)$ and $\bar{x}_i$, we have $\sum_j \ddot{\ell}(j, w_i)p_j(w_i) = \bar{x}_i\bar{\ell}_i(1) + (1 - \bar{x}_i)\bar{\ell}_i(0)$, so that

$$\begin{aligned}
\lambda_i &= \frac{1}{2}\sum_{k,j} p_j(w_i)p_k(w_i)(\ddot{\ell}(k, w_i) - \ddot{\ell}(j, w_i))(x_i(k) - x_i(j)) \\
&= \sum_{k,j} p_j(w_i)p_k(w_i)(\ddot{\ell}(k, w_i) - \ddot{\ell}(j, w_i))x_i(k) \\
&= \sum_k p_k(w_i)\ddot{\ell}(k, w_i)x_i(k) - \sum_j \ddot{\ell}(j, w_i)p_j(w_i)\sum_k p_k(w_i)x_i(k) \\
&= \bar{x}_i\bar{\ell}_i(1) - [\bar{x}_i\bar{\ell}_i(1) + (1 - \bar{x}_i)\bar{\ell}_i(0)]\bar{x}_i.
\end{aligned}$$

Note that by the covariance identity used in the proof of Lemma 2, $\lambda_i = \text{cov}(\ddot{\ell}(k(i), w_i), x_i(k(i)) \mid i)$, which yields the result. □

Lemma 3 shows that β^* identifies a weighted average of individual treatment effects, with weights that are zero for non-responders—individuals whose treatment status doesn't depend on examiner assignment (for whom $(1 - \bar{x}_i)\bar{x}_i = 0$). A necessary and sufficient condition for the weights on the remaining individuals to be positive is that the relative leniency measure of examiners who would treat i exceeds that of those who wouldn't, $\bar{\ell}_i(1) \geq \bar{\ell}_i(0)$. If the only covariate is an intercept, $w_i = 1$, and the first-stage is correctly specified, the condition $\bar{\ell}_i(1) \geq \bar{\ell}_i(0)$ is equivalent to the average monotonicity condition in Frandsen, Lefgren, and Leslie (2023), $\text{cov}(x_i(k(i)), \ell(k(i), 1) \mid i) \geq 0$. This equivalence was noted in Sigstad (2026). Lemma 3 generalizes the setup in Frandsen, Lefgren, and Leslie (2023) by allowing for covariates and for the first-stage equation to be misspecified. Under misspecification, it may be the case that the average monotonicity condition doesn't hold, even if it does hold for the true relative leniency, $\ell^*(k, w_i) - E[\ell^*(k(i), w_i)]$.

Inference

To derive a standard error for UJIVE that allows for treatment effect heterogeneity, let

$$y_i = z_i'\pi_Y + w_i'\delta_Y + \nu_{Yi}, \qquad E[\nu_{Yi}] = 0 \tag{7}$$

denote the *reduced form* regression of the outcome on instruments and covariates. To avoid technical complications, as in the estimation section of the Appendix, we condition on the instruments and covariates, and in analogy to eq. (1), we assume that the reduced form is correctly specified. The parameter of interest is

given by the estimand of the infeasible estimator $\hat{\beta}^*$,

$$\beta^* = \frac{\sum_i E[\tilde{\ell}_i y_i]}{\sum_i E[\tilde{\ell}_i x_i]} = \frac{\sum_i \tilde{\ell}_i z_i \pi_Y}{\sum_i \tilde{\ell}_i^2}.$$

If treatment effects are constant, so that eq. (2) holds, then it follows from the outcome equation and eq. (7) that $\pi_Y = \pi\beta$, and $\beta^* = \beta$. Recall that the UJIVE estimator is given by $\hat{\beta}_{UJIVE} = y' G_{UJIVE} x / x' G_{UJIVE} x$. Since $G_{UJIVE} x = \tilde{\ell} + G_{UJIVE}\nu$, it follows from the first stage equation and eq. (7) that we may decompose the numerator and denominator of the estimator as

$$\begin{pmatrix} y' G_{UJIVE} x \\ x' G_{UJIVE} x \end{pmatrix} = \sum_i \begin{pmatrix} \tilde{\ell}_i y_i \\ \tilde{\ell}_i x_i \end{pmatrix} + \sum_i \begin{pmatrix} r_{Y,i}\nu_i \\ r_i \nu_i \end{pmatrix} + \sum_{i,j} \begin{pmatrix} G_{ij}\nu_{Y,i}\nu_j \\ G_{ij}\nu_i\nu_j \end{pmatrix},$$

where G_{ij} is the (i,j) element of G_{UJIVE}, while $r_{Y,i} = \sum_j G_{ji}(z_j'\pi_Y + w_j'\delta_Y)$ and $r_i = \sum_j G_{ji}(z_j'\pi + w_j'\delta)$ denote the "signal" in the reduced form and the first stage. The second and third terms represent additional noise components in the estimator relative to the infeasible estimator $\hat{\beta}^*$. Since the third term is uncorrelated with the first two, it follows that the covariance matrix of the left-hand side may be written

$$\Sigma = \sum_i \operatorname{var}\begin{pmatrix} \tilde{\ell}_i y_i + r_{Y,i}\nu_i \\ \tilde{\ell}_i x_i + r_i \nu_i \end{pmatrix} + \sum_{i,j} \begin{pmatrix} G_{ij}^2 \sigma_{y,i}^2 \sigma_{x,j}^2 + G_{ij}G_{ji}\sigma_{xy,i}\sigma_{xy,j} & G_{ij}(G_{ij}+G_{ji})\sigma_{yx,i}\sigma_{x,j}^2 \\ G_{ij}(G_{ij}+G_{ji})\sigma_{yx,i}\sigma_{x,j}^2 & G_{ij}(G_{ij}+G_{ji})\sigma_{x,i}^2\sigma_{x,j}^2 \end{pmatrix}$$

where $\sigma_{y,i}^2 = \operatorname{var}(\nu_{Y,i})$, $\sigma_{x,i}^2 = \operatorname{var}(\nu_i)$, and $\sigma_{yx,i} = \operatorname{cov}(\nu_{Y,i}, \nu_i)$. Furthermore, the numerator and denominator can be shown to be asymptotically normal by a martingale central limit theorem (e.g., Evdokimov and Kolesár 2018, Lemma D.5) so that

$$\begin{pmatrix} y' G_{UJIVE} x \\ x' G_{UJIVE} x \end{pmatrix} \approx \mathcal{N}\left(\begin{pmatrix} \beta^* \sum_i \tilde{\ell}_i^2 \\ \sum_i \tilde{\ell}_i^2 \end{pmatrix}, \Sigma \right). \tag{8}$$

It then follows by an application of the delta method that in large samples, $\hat{\beta}_{UJIVE}$ has an approximately normal distribution,

$$\hat{\beta}_{UJIVE} \approx \mathcal{N}(\beta^*, \sigma_{UJIVE}^2),$$

where

$$\sigma_{UJIVE}^2 = \frac{\sum_i \operatorname{var}(\tilde{\ell}_i(y_i - x_i\beta^*) + (r_{Y,i} - r_i\beta^*)\nu_i) + \sum_{i,j}(G_{ij}^2 \sigma_{y-x\beta^*,i}^2 \sigma_{x,j}^2 + G_{ij}G_{ji}\sigma_{y-x\beta^*,x,i}\sigma_{y-x\beta^*,x,j})}{(\sum_i \tilde{\ell}_i^2)^2}$$

In contrast, by the same arguments, the standard error for the infeasible estimator is given by the square root of $\sigma_*^2 = \sum_i \operatorname{var}(\tilde{\ell}_i(y_i - x_i\beta^*))/(\sum_i \tilde{\ell}_i^2)^2$.

The $(r_{Y,i} - r_i\beta^*)\nu_i$ term in the numerator arises due to variation in complier treatment effects across complier groups. In particular, under regularity conditions, $\operatorname{var}((r_{Y,i} - r_i\beta^*)\nu_i) \approx \operatorname{var}(\tilde{z}_i'(\pi_Y - \pi\beta^*)\nu_i)$. Under constant treatment effects, the reduced form coefficients are proportional to the first stage, $\pi_Y = \pi\beta$, so that the first term in the above display may be replaced by $\sum_i \operatorname{var}(\tilde{\ell}_i(y_i - x_i\beta))$, or equivalently $\sum_i \operatorname{var}(\varepsilon_i\tilde{\ell}_i)$. The last term in the numerator is the Bekker (1994) many instrument term. It is negligible if the instruments are strong enough in the sense that $E[F]$ is large.

Note that Equation (8) corresponds exactly to Equation (3) in Angrist and Kolesár (2024) who consider an instrumental variables regression with a single instrument, with the variation in the first stage leniency, $\sum_i \tilde{\ell}_i^2$ playing the role of the first-stage regression coefficient on the single instrument, and Σ playing the role of the covariance between the reduced form and the first stage. It then follows from their analysis that delta-method based inference is not overly optimistic even if the instruments are weak, provided that the correlation between $(y - x\beta^*)' G_{UJIVE} x$ and $x' G_{UJIVE} x$, given by

$$\rho = \frac{\Sigma_{12} - \Sigma_{22}\beta^*}{\sqrt{\Sigma_{22}(\Sigma_{11} - 2\beta^*\Sigma_{12} + \beta^{*2}\Sigma_{22})}}$$

is not too large: as long as $|\rho| < 0.76$, rejection rates for nominal 5% tests stay below 10% regardless of

the instrument strength. In contrast to the single-instrument case, Σ now contains additional terms relative to the infeasible estimator, so that ρ no longer maps to the endogeneity parameter under homoskedasticity. Nonetheless, since Σ is consistently estimable, for any particular null hypothesis of interest, one can plug in the hypothesized value of β^* into the above expression along with estimates of Σ to verify that using the UJIVE standard errors doesn't lead to overrejection. Alternatively, since ρ is monotone in β^*, bounding the treatment effect parameter yields bounds in the plausible values of ρ: if these exclude large values of ρ, confidence intervals based on UJIVE standard errors will have good coverage rates.

Appendix B Data Appendix

The data in our application comes from Farre-Mensa, Hegde, and Ljungqvist (2020b), the replication file for Farre-Mensa, Hegde, and Ljungqvist (2020a). Our reanalysis sample contains 32,514 first-time patent applications filed after 2001 with final decisions by 2013.[18]

The variables used in the analyses (and listed in Table 1) are defined as follows. *# subsequent applications* is the number of applications with a filing date greater than the first-action date of a firm's first application and *any subsequent application* equals one if any applications were made. *# subsequent approved applications* is the number of approved applications with a filing date greater than the first-action date of a firm's first application and *any subsequent approved application* equals one if any applications were approved. *# citations to subsequent patents* is the number of citations received by all subsequent patent applications over the five years after each patent application's public disclosure date, and *any citation to subsequent patents* equals one if any such citation was received.

Patent class groups are defined by related subject matter, as classified by the US Patent Office (see https://www.uspto.gov/sites/default/files/patents/resources/classification/classescombined.pdf). *Patent class group I* includes chemical and related arts, *Patent class group II* includes communications, radiant energy, weapons, electrical, and computer arts, and *Patent class group III* includes material science, mechanical manufacturing and power, and related arts. *# independent claims in application* counts independent claims made in the patent application. *# VC rounds before application* is the number of venture capital funding rounds the startup had secured prior to its first patent application. *# startups in HQ state* counts the total number of startups headquartered in the same state as the applicant in the same year as the applicant. *Approval by European Patent Office* and *Approval by Japanese Patent Office* take value 1 if a patent was approved by a European and Japanese patent office, respectively, take value 0 if an application was made but not approved, and are missing if no application was made. *Years from application to first action* counts the years between the patent application and the first action on the patent application.

[18] Our analysis uses the full sample from the original paper as found in its replication package. The initial dataset has 34,435 observations. We drop 1 observation that is missing citation data. We then drop 1,851 observations with singleton covariates or instruments, and 69 observations with leverage of 1. This causes us to drop 378 collinear controls and 1,676 collinear IVs. See the code documentation at https://github.com/kolesarm/ManyIV for discussion on how this recursive algorithm is implemented to drop these observations and variables.

Table B.1: Treatment Effect Estimates with Precision Controls

	UJIVE	2SLS		OLS
	Examiner indicators (1)	FMHL approval rate (2)	Examiner indicators (3)	(4)
Any subsequent application	0.203	0.278	0.250	0.250
	(0.064)	(0.025)	(0.016)	(0.006)
Log(1 + # subsequent applications)	0.401	0.478	0.413	0.385
	(0.116)	(0.039)	(0.029)	(0.010)
Any subsequent approved application	0.275	0.254	0.248	0.231
	(0.059)	(0.022)	(0.015)	(0.005)
Log(1 + # subsequent approved applications)	0.403	0.370	0.343	0.304
	(0.095)	(0.031)	(0.023)	(0.008)
Any citation to subsequent patents	0.216	0.220	0.187	0.176
	(0.057)	(0.021)	(0.014)	(0.005)
Log(1 + # citations to subsequent patents)	0.514	0.499	0.410	0.363
	(0.146)	(0.047)	(0.035)	(0.012)

Notes: This table reports estimates of the effect of application approval on each outcome in Panel A of Table 1. This table adds additional precision controls: HQ state fixed effect; the number of independent claims in the application, along with a missing-value dummy, with the claims count set to 0 when missing and the dummy flags those observations; the number of VC rounds before the application, and the number of startups in the HQ state. Column 1 estimates effects with UJIVE, instrumenting with examiner indicators. Columns 2 and 3 instead estimate effects with 2SLS, instrumenting with examiners' approval rate as computed by Farre-Mensa, Hegde, and Ljungqvist (2020a) and examiner indicators respectively. Column 4 estimates effects by ordinary least squares. All estimates control for art unit-by-year fixed effects. The sample size is 32,514 in all specifications. Standard errors, reported in parentheses, are robust to heteroskedasticity and treatment effect heterogeneity.

Table B.2: Complier Characteristics with Precision Controls

	Sample Mean (1)	Complier Mean (2)	# Obs. (3)
Patent class group I	0.162	0.215	32,514
		(0.037)	
Patent class group II	0.410	0.416	32,514
		(0.044)	
Patent class group III	0.445	0.375	32,514
		(0.042)	
# independent claims in application	3.730	3.622	27,226
		(0.223)	
# VC rounds before application	0.124	0.178	32,514
		(0.047)	
# startups in HQ state	317.3	343.0	32,514
		(25.5)	
Approval by European Patent Office	0.372	0.142	3,287
		(0.111)	
Approval by Japanese Patent Office	0.398	0.601	1,197
		(0.479)	

Notes: This table reports means of each covariate in Panel C of Table 1 for the full sample and for compliers. This table adds additional precision controls: HQ state fixed effect; the number of independent claims in the application, along with a missing-value dummy, with claims count is set to 0 when missing and the dummy flags those observations; the number of VC rounds before the application, and the number of startups in the HQ state. Complier means are estimated with UJIVE as described in the text. Standard errors, reported in parentheses, are robust to heteroskedasticity and treatment effect heterogeneity.