Queen's Economics Department Working Paper No. 1031

The Case Against JIVE

Russell Davidson
McGill University

James MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

9-2004

The Case Against JIVE

Russell Davidson and **James G. MacKinnon**

Abstract

We perform an extensive series of Monte Carlo experiments to compare the performance of two variants of the "Jackknife Instrumental Variables Estimator," or JIVE, with that of the more familiar 2SLS and LIML estimators. We find no evidence to suggest that JIVE should ever be used. It is always more dispersed than 2SLS, often very much so, and it is almost always inferior to LIML in all respects. Interestingly, JIVE seems to perform particularly badly when the instruments are weak.

Contact Information (Davidson):

Department of Economics
McGill University
Montreal, Quebec, Canada
H3A 2T7

GREQAM
Centre de la Vieille Charité
2 rue de la Charité
13002 Marseille, France

email: **russell.davidson@mcgill.ca**

Contact Information (MacKinnon):

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

email: **jgm@econ.queensu.ca**

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. We are grateful to three referees and participants at the Montreal Econometric Workshop for comments on earlier versions.

Revised, September 2004

1. Introduction

The finite-sample properties of instrumental variables estimators can be very poor, especially when the sample size is small and/or the instruments are weak. See Nelson and Startz (1990a, 1990b) and Staiger and Stock (1997) among many others. It was suggested, first by Phillips and Hale (1977), and later by Angrist, Imbens, and Krueger (1999) and Blomquist and Dahlberg (1999), henceforth AIK and BD, respectively, that finite-sample properties might be improved by replacing the usual fitted values from the reduced form regression(s) by "omit-one" fitted values which omit observation t when estimating the t^{th} fitted value. The idea is that this should eliminate the correlation between the fitted values and the structural-equation errors.

AIK and BD proposed more than one "Jackknife Instrumental Variables Estimator", or JIVE, estimator based on these omit-one fitted values, although these estimators do not really make use of the jackknife. A few simulations designed to study the finite-sample properties of estimators of this type were performed by AIK, and greater numbers by BD. Flores-Lagunes (2002) and Hahn, Hausman, and Kuersteiner (2004) (henceforth HHK) were the first to provide simulation evidence that JIVE can perform poorly. HHK also proposed a genuine jackknife estimator called JN2SLS which should not be confused with JIVE.

In this paper, we perform an extensive series of Monte Carlo experiments that compare the finite-sample properties of two forms of JIVE with those of 2SLS and LIML. In order to keep the experimental design manageable, we limit ourselves to a model with one structural and one reduced-form equation. Within this limitation, our experiments are much more comprehensive than any comparable ones that we are aware of. We vary the sample size, the number of instruments, the weakness of the instruments, and the correlation between the errors of the reduced-form and structural equations. All of these factors affect the performance of the estimators we study, often dramatically so.

HHK provide simulation evidence that, like LIML, JIVE has no moments, a fact proved in Davidson and MacKinnon (2004). Here, we show that only in very limited parts of the parameter space is JIVE systematically better behaved than LIML, the contrary being true elsewhere. Since we also find that LIML does not clearly dominate 2SLS, there seems little reason to prefer JIVE to either of these estimators.

2. IV, JIVE, and LIML Estimators

We are interested in one structural equation from a linear simultaneous equations model. The structural equation can be written as

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta}_1 + \boldsymbol{Y}\boldsymbol{\beta}_2 + \boldsymbol{u} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{X} \equiv [\boldsymbol{Z}\ \boldsymbol{Y}]$, with $\boldsymbol{Z}$ an $n \times k_1$ matrix of observations on exogenous variables, $\boldsymbol{Y}$ an $n \times k_2$ matrix of observations on endogenous variables, and $\boldsymbol{\beta} \equiv [\boldsymbol{\beta}_1^\top\ \boldsymbol{\beta}_2^\top]^\top$ a

vector of unknown parameters. The rest of the system is treated as an unrestricted reduced form:

$$\boldsymbol{Y} = \boldsymbol{W\Pi} + \boldsymbol{V}, \tag{2}$$

where there are l instruments in the matrix $\boldsymbol{W} \equiv [\boldsymbol{Z}\ \boldsymbol{W}_2]$. The errors in (1) and (2) are assumed IID, with expectation zero conditional on $\boldsymbol{W}$, and nonzero correlation between the errors for the same observation.

The 2SLS, or generalized IV, estimator can be written compactly as

$$\hat{\boldsymbol{\beta}}^{\mathrm{IV}} = (\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{y},$$

where $\boldsymbol{P}_{\boldsymbol{W}} \equiv \boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top$ is the orthogonal projection on to the space spanned by the columns of $\boldsymbol{W}$. The matrix of fitted values implicitly used as instruments for the endogenous explanatory variables $\boldsymbol{Y}$ is $\hat{\boldsymbol{Y}} \equiv \boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{Y} = \boldsymbol{W}\hat{\boldsymbol{\Pi}}$, where $\hat{\boldsymbol{\Pi}}$ is the matrix of OLS estimates from the first-stage regressions (2). In contrast, the matrix of omit-one fitted values used by JIVE is $\tilde{\boldsymbol{Y}}$, of which the t^{th} row is $\boldsymbol{W}_t \hat{\boldsymbol{\Pi}}^{(t)}$, where $\hat{\boldsymbol{\Pi}}^{(t)}$ is the matrix of OLS estimates computed without observation t. It can be shown that $\tilde{\boldsymbol{Y}}_t = \big((\boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{Y})_t - h_t \boldsymbol{Y}_t\big)/(1 - h_t)$, where h_t is the t^{th} diagonal element of $\boldsymbol{P}_{\boldsymbol{W}}$.

We consider two JIVE estimators. The principal one is

$$\hat{\boldsymbol{\beta}}^{\mathrm{JIVE}} = (\tilde{\boldsymbol{X}}^\top \boldsymbol{X})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{y}, \text{ where } \tilde{\boldsymbol{X}} \equiv [\boldsymbol{Z}\ \tilde{\boldsymbol{Y}}]. \tag{3}$$

This estimator was called JIVE1 by AIK and UJIVE by BD. We will adopt the former notation. AIK also propose a very similar estimator, which they call JIVE2. It uses the matrix with typical row $\ddot{\boldsymbol{Y}}_t \equiv \big((\boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{Y})_t - h_t \boldsymbol{Y}_t\big)/(1 - 1/n)$ instead of $\tilde{\boldsymbol{Y}}$ in (3). In most of our simulations, the results for JIVE2 were extremely similar to those for JIVE1, although JIVE2 generally seemed to perform a little bit less well. We therefore present results for JIVE1 only, except for one case, with a different experimental design, in which JIVE1 and JIVE2 differed substantially. Note, however, that the two JIVE estimators may differ noticeably in any particular sample.

BD also consider the estimator $\hat{\boldsymbol{\beta}}^{\mathrm{JLS}} \equiv (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{y}$, which is simply the OLS estimator from a regression of $\boldsymbol{y}$ on $\tilde{\boldsymbol{X}}$. It has finite-sample properties extremely different from those of the two JIVE estimators. Since $\tilde{\boldsymbol{X}}$ can be thought of as $\boldsymbol{X}$ measured with error, JLS suffers from a sort of errors-in-variables problem. Unlike the 2SLS estimator, which is always biased towards the inconsistent OLS estimator, JLS seems to be biased in the direction of zero, just like the OLS estimator when the explanatory variable is measured with error. In our experiments, JLS was almost always both more biased and less precise than 2SLS, even when there was no correlation between the reduced-form and structural error terms. Limited results are presented in Figures 12 and 13, which are available from the JAE Data Archive website.

A classic alternative to the 2SLS estimator is the LIML estimator

$$\hat{\boldsymbol{\beta}}^{\text{LIML}} = \left(\boldsymbol{X}^\top(\mathbf{I} - \hat{\kappa}\boldsymbol{M}_{\boldsymbol{W}})\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top(\mathbf{I} - \hat{\kappa}\boldsymbol{M}_{\boldsymbol{W}})\boldsymbol{y},$$

where $\hat{\kappa}$ is the value of the ratio

$$\frac{(\boldsymbol{y} - \boldsymbol{Y}\boldsymbol{\beta}_2)^\top \boldsymbol{M}_{\boldsymbol{Z}}(\boldsymbol{y} - \boldsymbol{Y}\boldsymbol{\beta}_2)}{(\boldsymbol{y} - \boldsymbol{Y}\boldsymbol{\beta}_2)^\top \boldsymbol{M}_{\boldsymbol{W}}(\boldsymbol{y} - \boldsymbol{Y}\boldsymbol{\beta}_2)},$$

minimized with respect to $\boldsymbol{\beta}_2$. Here $\boldsymbol{M}_{\boldsymbol{Z}}$ and $\boldsymbol{M}_{\boldsymbol{W}}$ are orthogonal projections that annihilate the columns of $\boldsymbol{Z}$ and $\boldsymbol{W}$ respectively. We also study this estimator.

3. The Simulation Experiments

The design of our experiments is just about as simple as possible. There is one structural equation, with dependent variable $\boldsymbol{y}$ generated by the equation

$$\boldsymbol{y} = \beta_1\boldsymbol{\iota} + \beta_2\boldsymbol{x} + \boldsymbol{u}.$$

Here $\boldsymbol{\iota}$ is a vector of ones, and the endogenous explanatory variable $\boldsymbol{x}$ is generated as a function of the instruments in the $n \times l$ matrix $\boldsymbol{W}$ by the reduced-form equation

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{\pi} + \boldsymbol{v}. \tag{4}$$

It is shown in Davidson and MacKinnon (2004) that the distributions of all the estimators covered in our experiments are the same for this model, with n observations, as for a model in which the structural equation has k_1 exogenous explanatory variables, one endogenous explanatory variable, and a sample size of $n - k_1 + 1$.

The first column of $\boldsymbol{W}$ is $\boldsymbol{\iota}$, and the remaining columns, except when otherwise noted, are IID standard normal random variables, redrawn for each replication. Only the orthogonal projection $\boldsymbol{P}_{\boldsymbol{W}}$ affects the estimators we consider, and so it is of no interest to change $\boldsymbol{W}$ in ways that do not affect $\boldsymbol{P}_{\boldsymbol{W}}$. The number of overidentifying restrictions is $r = l - 2$. The elements of $\boldsymbol{u}$ and $\boldsymbol{v}$ have variances σ_u^2 and σ_v^2, respectively, and correlation ρ.

The t statistic for β_2 to be equal to its true value does not depend on the values of β_1, β_2, and σ_u. The bias of the IV estimator, when it exists, does not depend on β_1 or β_2, and is proportional to σ_u. Thus, without loss of generality, we set $\beta_1 = 1$, $\beta_2 = 1$, and $\sigma_u^2 = 1$ in all experiments. It can also be seen that multiplying σ_v and the elements of the vector $\boldsymbol{\pi}$ by the same constant affects only the scale of the variable $\boldsymbol{x}$, and so has only a scale effect on the estimators of β_2 and no effect at all on the t statistics.

What does have an important effect is the ratio of $\|\boldsymbol{\pi}\|^2$ to σ_v^2. This ratio can be interpreted as the signal-to-noise ratio in the reduced-form equation (4). In order to capture the effect of the ratio on the distributions we study, we normalize the quantity $\|\boldsymbol{\pi}\|^2 + \sigma_v^2$ to unity. In addition, without loss of generality, the values of the π_j, for $j \geq 2$, are constrained to be equal, with $\pi_1 = 0$. The parameter that does vary is

denoted by R^2_∞: It is the limiting R^2 of the reduced-form regression as $n \to \infty$. Since $n^{-1}\boldsymbol{W}^\top\boldsymbol{W}$ tends to an identity matrix as $n \to \infty$, we see that $R^2_\infty = 1/(1+\sigma_v^2/\|\boldsymbol{\pi}\|^2)$. It is conveniently restricted to the $[0, 1]$ interval and is a monotonically increasing function of the "concentration parameter" that is often used in the weak instruments literature; see Stock, Wright, and Yogo (2002). A small value of nR^2_∞ implies that the instruments are weak. In our experiments, therefore, we vary the sample size and three parameters: the number r of overidentifying restrictions, the correlation ρ, and R^2_∞.

All experiments used 500,000 replications. When either ρ or R^2_∞ was varied, we performed experiments for every value from 0.00 to 0.99 at intervals of 0.01; only the absolute value of ρ affects anything other than the sign of any bias. When $R^2_\infty = 0$, β_2 is not asymptotically identified. This case is included to show what happens in the limit as the instruments become infinitely weak. When r was varied, we considered all values from 0 to 16. Every experiment was performed six times, for samples of size 25, 50, 100, 200, 400, and 800.

To keep our experiments within reasonable bounds, when we varied one parameter, we held the other two fixed at certain base values, chosen so as to make estimation challenging. The base value of ρ was 0.9, because all three estimators generally performed worse as ρ was increased. The base value of R^2_∞ was 0.1, which implies that the instruments are very weak when n is small. The base value of r was 5, a compromise chosen since computational costs rise sharply with r, and since results for very small r sometimes differ markedly from those for larger values.

Because the LIML and JIVE estimators have no moments, we report as a measure of the central tendency of each estimator its *median bias*, that is, the 0.5 quantile of the $\hat{\beta}_2$ estimates minus the true value β_{20}. As a measure of dispersion, we report the *nine decile range*, that is, the 0.95 quantile minus the 0.05 quantile. The nine decile range is the width of an interval in which the estimates lie 90% of the time.

4. Results of the Experiments

All results are reported graphically. To save space, only one figure is printed here. The rest may be found on the JAE Data Archive website at `www.econ.queensu.ca/jae`. The results of the principal experiments are reported in Figures 1 through 9. Each figure contains six panels, one for each of the sample sizes 25, 50, 100, 200, 400, and 800. Readers should be careful to check the vertical scales of the various panels, which are not always the same for different sample sizes. Figure A, which is printed here, contains the panels for $n = 50$ from Figures 1 through 6.

The first three figures concern median bias. Figure 1 suggests that the median bias of the 2SLS estimator is proportional to ρ and inversely proportional to n. These results are consistent with the theoretical results in Phillips (1984). In contrast, the median biases of LIML and JIVE1 are evidently nonlinear functions of ρ. That of LIML is indiscernible for $n \geq 200$, but that of JIVE1, as with 2SLS, appears to be $O(n^{-1})$, positive for small samples and negative for large ones. Figure 2 shows that

the median biases of all three estimators generally decline as R^2_∞ increases. For 2SLS and LIML, the decline is monotonic. For JIVE1, however, the sign always changes from positive to negative as R^2_∞ increases, before finally approaching zero. When nR^2_∞ is very small, JIVE1 is actually slightly more biased than 2SLS. Figure 3 shows the dependence of median bias on the number of overidentifying restrictions. For 2SLS and LIML, it increases monotonically with r, although, for $n \geq 100$, the median bias of LIML appears to be essentially zero. That of JIVE1 can be of either sign and can be larger in absolute value than that of 2SLS. It is always negative for the larger sample sizes. Notice that 2SLS and LIML are identical when $r = 0$, but JIVE1 is quite different, and always more biased.

We conclude that JIVE1 is usually worse than LIML in terms of median bias, and for larger sample sizes almost always very much worse. However, because the sign of the bias changes, it is always possible to find parameter values for which JIVE1 is not biased. Although JIVE1 usually outperforms 2SLS according to this criterion, there are numerous cases in which it fails to do so. Unlike LIML, JIVE1 may be biased in either the same direction as 2SLS or in the opposite direction.

Figures 4 through 6 deal with the dispersion of the three estimators, as measured by the nine decile range. In Figure 4, where ρ varies on the horizontal axes, JIVE1 is always more dispersed then LIML, which in turn is always more dispersed than 2SLS. The differences are very large when n is small, although they are quite modest when $n = 800$. Figure 5 shows that the same pattern holds for most values of R^2_∞. However, LIML is more dispersed than JIVE1 for a set of extremely small values, this set becoming smaller as n increases. The phenomenon evidently occurs only when the instruments are extremely weak, which is also the case in which the median bias of JIVE1 is very large. Figure 6 shows that JIVE1 is always more dispersed than LIML except for large values of r when $n = 25$. 2SLS, the only estimator with moments, is always less dispersed than LIML, which is often very much less dispersed than JIVE1. It is clear from all three figures that JIVE1 and LIML are much more dispersed than 2SLS when the instruments are weak. Thus it is evident that, in line with a conclusion reached in HHK, using an estimator with no moments is not a way to solve the weak instruments problem.

The next three figures concern rejection frequencies for two-tailed asymptotic tests based on pseudo-t statistics for β_2 to equal its true value. On average, in Figures 7 through 9, LIML seems to yield the most reliable inferences, although there are a number of cases in which JIVE1 is more reliable. Because the base case value of ρ is 0.9, overrejection is more common in the figures than underrejection. However, it can be seen from Figure 7 that all three estimators underreject for smaller values of ρ. This is particularly severe for JIVE1. For values of ρ less than about 0.35, even 2SLS yields more reliable inferences than JIVE1. Figures 7 through 9 thus provide little support for using JIVE1 rather than LIML for testing purposes.

For all the simulation results reported so far, the instruments were IID normal. This implies that the design of the $\boldsymbol{W}$ matrix is, on average, balanced. Since the h_t play a key role in distinguishing between JIVE1 and JIVE2, it is of interest to see what happens when the design of $\boldsymbol{W}$ is not balanced. To this end, we performed a second

series of experiments in which the instruments were IID lognormal. Limited results from these experiments, for the case $n = 50$, are shown in the right-hand columns of Figures 10 and 11. For comparison, and to support our claim that JIVE1 and JIVE2 are extremely similar when the instruments are normally distributed, the left-hand columns contain corresponding results from the original experiments. As can be seen from Figures 10 and 11, the distribution of the 2SLS estimator is only slightly affected by the design of the $\boldsymbol{W}$ matrix. The same is true of the LIML estimator, results for which are not shown. However, the results for JIVE1 and, especially, JIVE2 are very sensitive to the design of $\boldsymbol{W}$, and the two can yield quite different results. However, there is nothing in the figures, or in any of the results that we do not report, to suggest that changing the design of $\boldsymbol{W}$ will cause either version of JIVE to perform better relative to IV and LIML.

5. Conclusion

We have presented the results of a rather extensive set of Monte Carlo experiments on the finite-sample performance of four single-equation estimators for linear simultaneous equations models, two variants of JIVE, and the familiar 2SLS and LIML estimators. The results show clearly that, in most regions of the parameter space that we have studied, JIVE is inferior to LIML with regard to median bias, dispersion, and reliability of inference. Thus, if an investigator wishes to use an estimator that has no moments, it is usually better to use LIML than JIVE. Our results, however, do not provide unambiguous support for the use of LIML. It is always more dispersed than 2SLS, often dramatically so when the instruments are weak. Other estimators that have moments, such as Fuller's (1977) modification of LIML and the JN2SLS estimator proposed by HHK, outperform LIML in many circumstances, as the latter paper shows. It would be of interest to investigate additional estimators using the experimental design of this paper. On account of space limitations, we have not done so here.

References

Angrist JD, Imbens GW, Krueger AB. 1999. Jackknife instrumental variables estimation, *Journal of Applied Econometrics* **14**, 57–67.

Blomquist S, Dahlberg M. 1999. Small sample properties of LIML and jackknife IV estimators: Experiments with weak instruments, *Journal of Applied Econometrics* **14**, 69–88.

Davidson R, MacKinnon JG. 2004. Moments of IV and JIVE estimators, mimeo, Queen's University.

Flores-Lagunes, A. 2002. Estimation methods robust to weak instruments, University of Arizona working paper.

Fuller WA. 1977. Some properties of a modification of the limited information estimator, *Econometrica* **45**, 939–953.

Hahn J, Hausman JA, Kuersteiner G. 2004. Estimation with weak instruments: Accuracy of higher order bias and MSE approximations, *Econometrics Journal* **7**, 272–306.

Nelson CR, Startz R. 1990a. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one, *Journal of Business* **63**, S125–140.

Nelson CR, Startz R. 1990b. Some further results on the exact small sample properties of the instrumental variables estimator, *Econometrica* **58**, 967–976.

Phillips GDA, Hale C. 1977. The bias of instrumental variable estimators of simultaneous equation systems, *International Economic Review* **18**, 219–228.

Phillips, PCB. 1984. Exact small sample theory in the simultaneous equations model. In *Handbook of Econometrics*, Volume 1, Griliches Z, Intriligator M (eds.) Amsterdam: North-Holland.

Staiger D, Stock JH. 1997. Instrumental variables regressions with weak instruments, *Econometrica* **65**: 557–586.

Stock JH, Wright JH, Yogo M. 2002. A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business and Economic Statistics* **20**, 518–529.

Reply

Russell Davidson and James G. MacKinnon

We welcome the comments on our paper by Ackerberg and Devereux (AD) and Blomquist and Dahlberg (BD), and we are happy to take this opportunity to respond briefly to them.

We agree wholeheartedly with both AD and BD that is it unprofitable to try to obtain meaningful empirical results when all available instruments are very weak. At least in our experiments, however, it is only in this case that JIVE1 ever outperforms LIML in terms of dispersion; see Figure 5 and the top left panel of Figure 6. Thus we disagree with BD that our "categorical rejection of JIVE" is not in accord with our simulation results. BD are, however, absolutely correct to point out that our experiments deal with a very simple case and do not tell the whole story.

AD discuss two quite new estimators, IJIVE and UIJIVE, which apparently improve upon the JIVE1 estimator. Their comment suggests that our conclusion that JIVE1 is inferior to LIML does not necessarily apply to these estimators. In order to investigate the properties of these two estimators, we performed a new set of simulation experiments, with the same design as those we describe in the paper. The results of the new experiments are presented graphically in Figures 1a–6a, which, like Figures 1–13 from the paper itself, are available from the JAE Data Archive website at `www.econ.queensu.ca/jae/`. The new figures deal with the same cases as Figures 1–6 of the paper. Results for IJIVE and UIJIVE are added, and, for readability, results for 2SLS are removed.

The superiority of IJIVE and, especially, UIJIVE relative to JIVE1 emerges quite clearly from the new figures. What is particularly interesting is that UIJIVE tends to be substantially less dispersed than the other JIVE estimators. When the instruments are weak, it is often much less dispersed than LIML. Our tentative conclusion is that UIJIVE is the best JIVE estimator to date and may well be worth using in practice.

The simulation results suggest that IJIVE and UIJIVE, like JIVE1 and LIML, have no moments, and we have confirmed this analytically. This leads us to question the interpretation of the results in Phillips and Hale (1977) and in Ackerberg and Devereux (2003) that purport to yield approximate biases for these estimators. These results are based on stochastic expansions which can be represented schematically as

$$n^{1/2}(\hat{\beta}_2 - \beta_{20}) = t_0 + n^{-1/2}t_1 + o_p(n^{-1/2}), \tag{5}$$

where $\hat{\beta}_2$, as in the paper, is an estimator of the coefficient of the endogenous regressor in the second-stage regression, β_{20} is the true value of the parameter, t_0 is a random variable that has a normal distribution with zero expectation, and t_1 is a random variable with a nonzero expectation, which provides the approximate bias of the estimator.

The catch is that the $o_p(n^{-1/2})$ remainder in (5) has no moments. However, the fact that it does tend to zero in probability implies that the distribution of the first two

terms of the truncated expansion converges to that of the estimator itself for large sample sizes n. We have expressed the stochastic expansion as an expansion in powers of $n^{-1/2}$, but it is just as possible to express it in the form of small-sigma asymptotics, as is done by Phillips and Hale. The two expansions appear to be equivalent, at least to order $n^{-1/2}$.

The techniques proposed in Appendix 1 of Phillips and Hale for simplifying the calculation of JIVE-type estimators can equally well be applied to IJIVE and UIJIVE. In all cases, the instruments $\tilde{\boldsymbol{Y}}_t$ can be expressed in terms of the fitted values $\boldsymbol{P_W Y}$ of the first-stage regressions of 2SLS and the diagonal elements h_t of $\boldsymbol{P_W}$, as we show for JIVE1 in the paper. Since IJIVE is just JIVE1 using projected variables, the same formulas can be used for it. For UIJIVE, it is easy to see that the instruments are given by a very similar formula, namely,

$$\tilde{\boldsymbol{Y}}_t = \frac{1}{1 - h_t + \omega}\big((\boldsymbol{P_W Y})_t - (h_t - \omega)\boldsymbol{Y}_t\big), \tag{6}$$

in which h_t is replaced by $h_t - \omega$, with $\omega = 2/n$ in the case with just one included endogenous variable, or $\omega = (g+1)/n$ more generally when there are g included endogenous variables. Of course, here $\boldsymbol{Y}$ and $\boldsymbol{W}$ should be interpreted as matrices of endogenous variables and instruments that have been projected off the included exogenous variables.

The formula (6) suggests that, whereas JIVE1 and IJIVE involve "omit-1" fitted values, UIJIVE involves "omit-less-than-1" fitted values. Using this formula together with equation (3) of our paper to calculate the UIJIVE estimator is orders of magnitude faster, for large n, than using the elegant formula in Ackerberg and Devereux (2003), which involves manipulating $n \times n$ matrices.

A point that emerges clearly from our simulations and those presented by AD in their comment is that, although UIJIVE is constructed so as to reduce mean bias, it does not do nearly so well as regards median bias. This implies that the estimator is significantly skewed, a point that would be worth subsequent investigation. It would also be interesting to look more closely at the modified LIML estimator proposed by Fuller (1977), for which moments exist, to see how well it performs relative to UIJIVE.

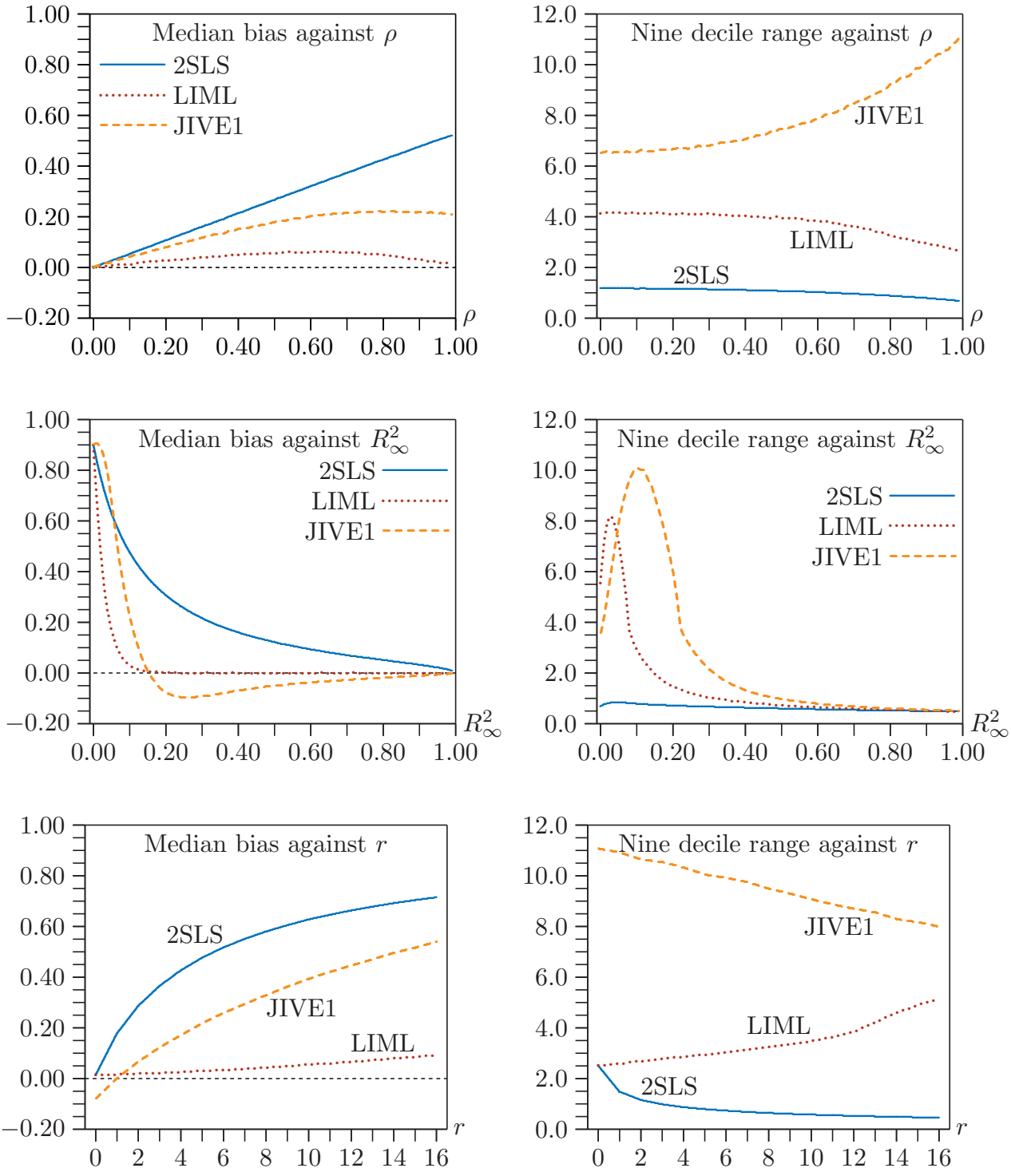

Figure A. Median bias and nine decile range when $n = 50$

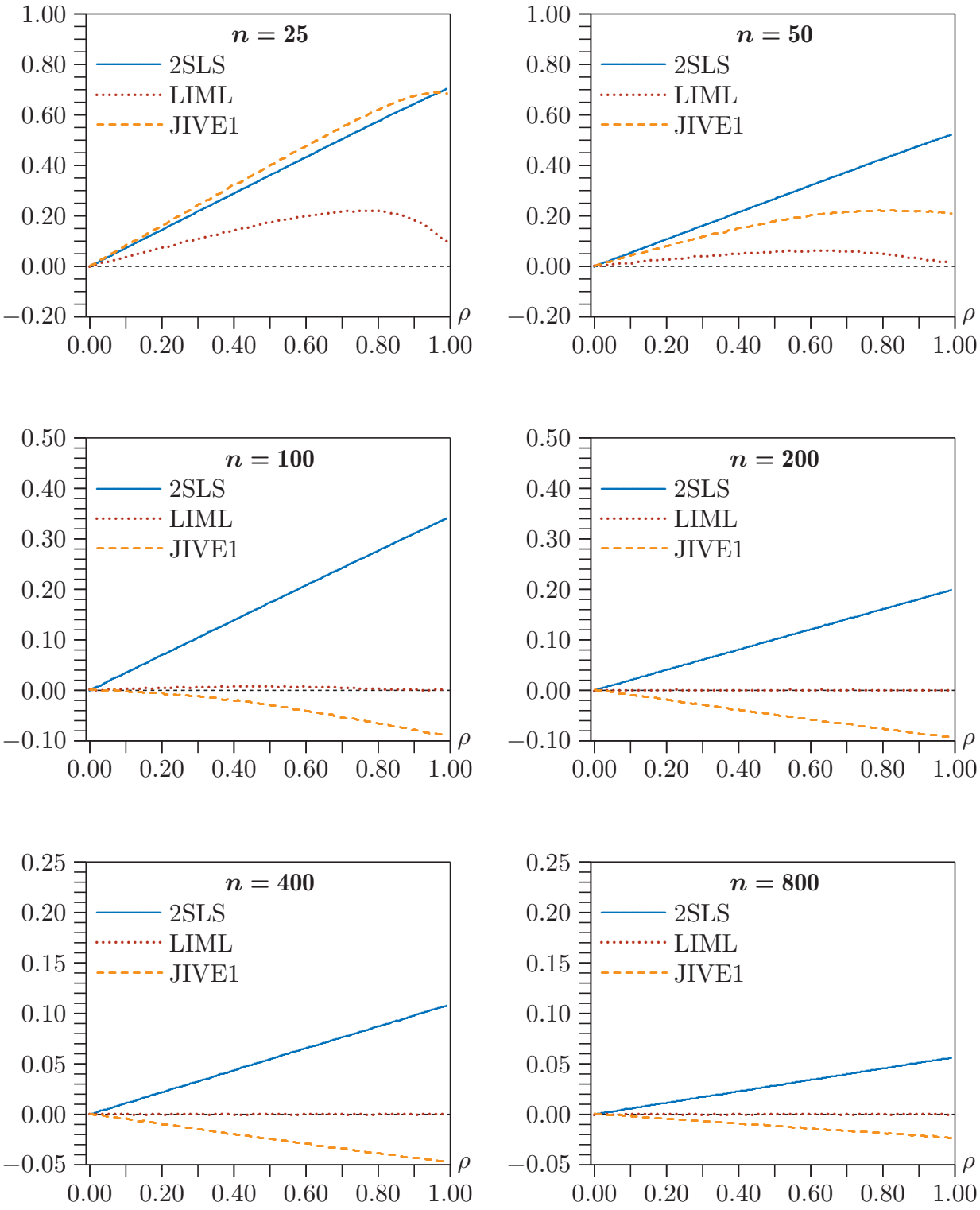

Figure 1. Median bias of three estimators, $r = 5$, $R^2_\infty = 0.1$

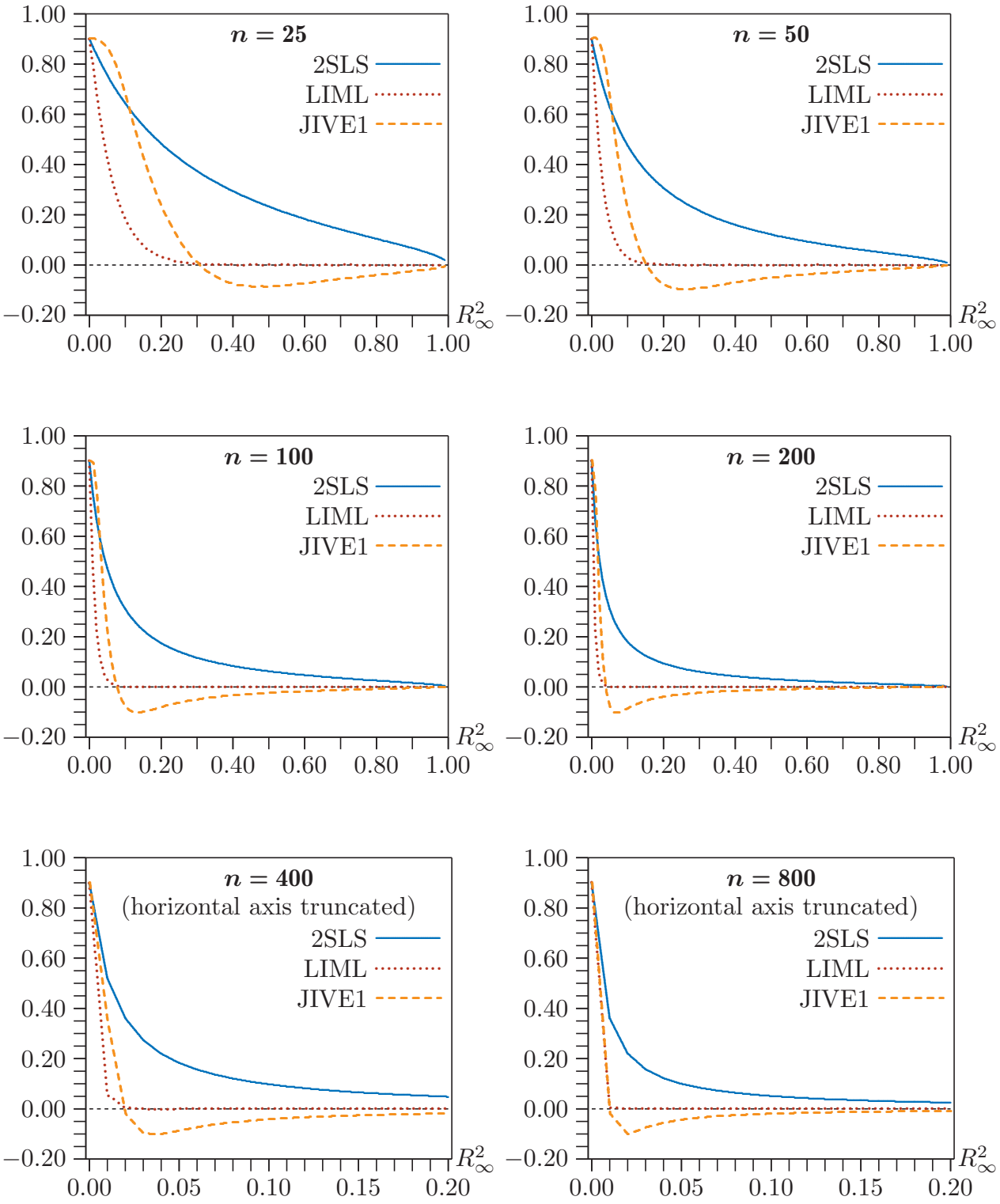

Figure 2. Median bias of three estimators, $r = 5$, $\rho = 0.9$

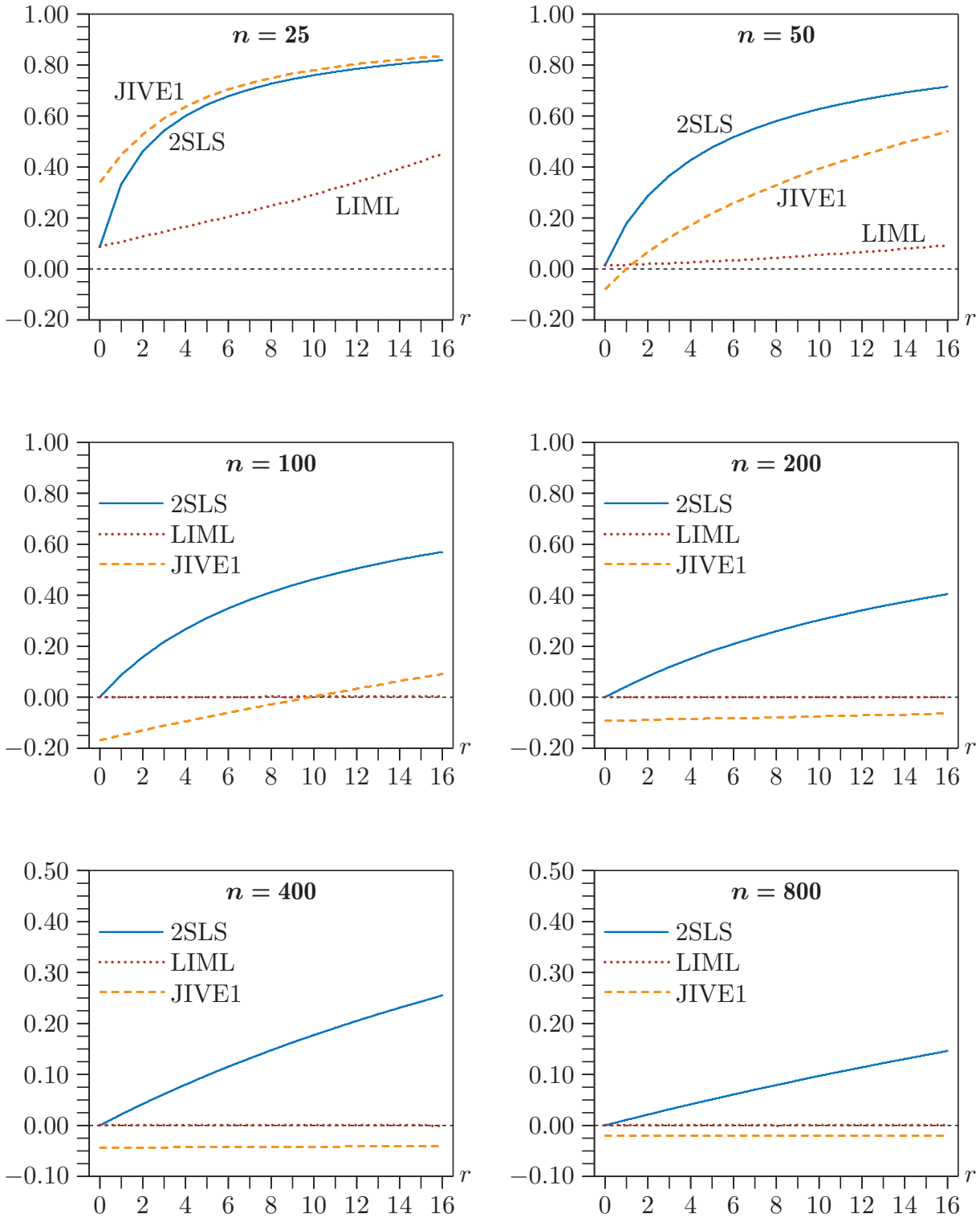

Figure 3. Median bias of three estimators, $R^2_\infty = 0.1$, $\rho = 0.9$

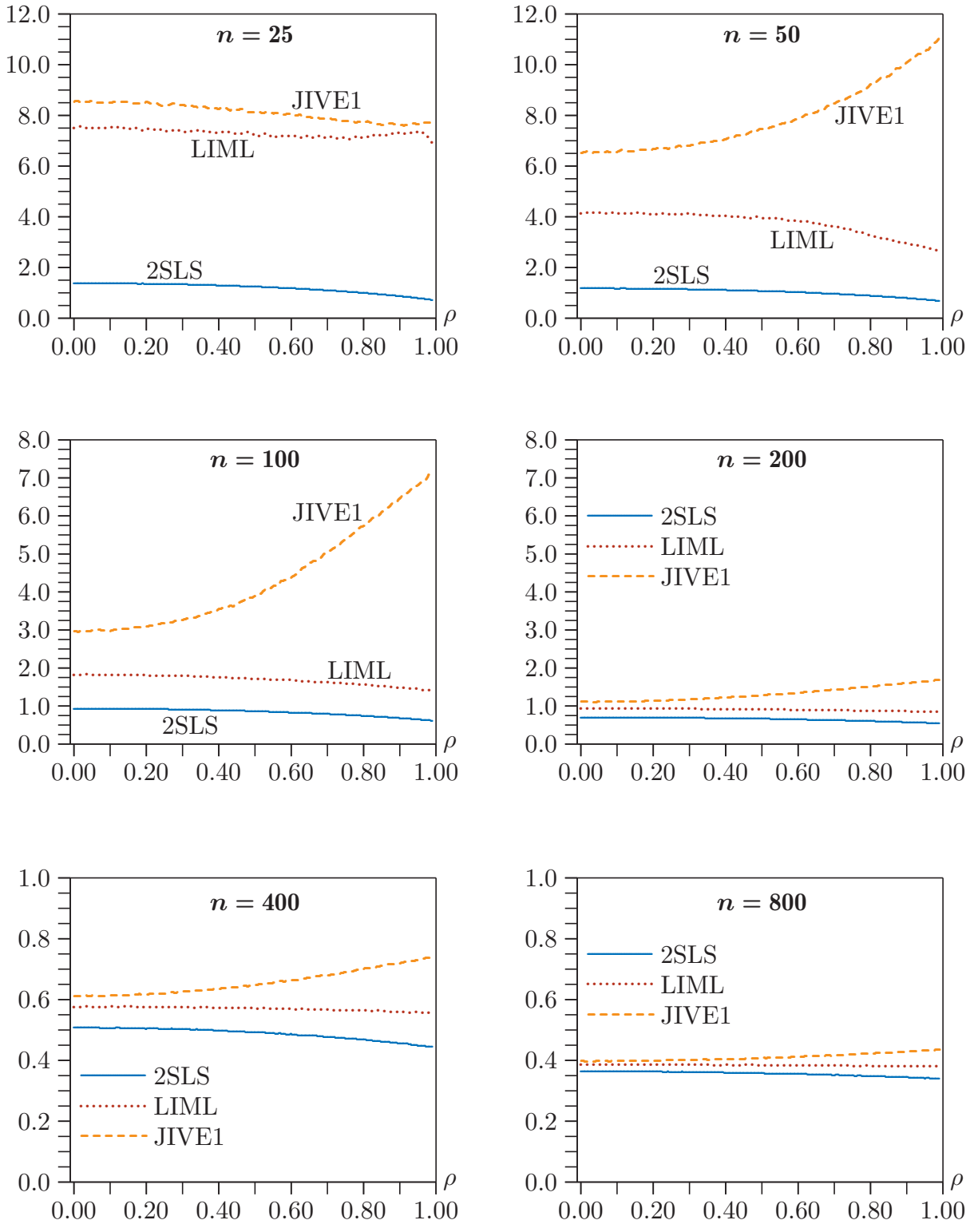

Figure 4. Nine decile range of three estimators, $r = 5$, $R^2_\infty = 0.1$

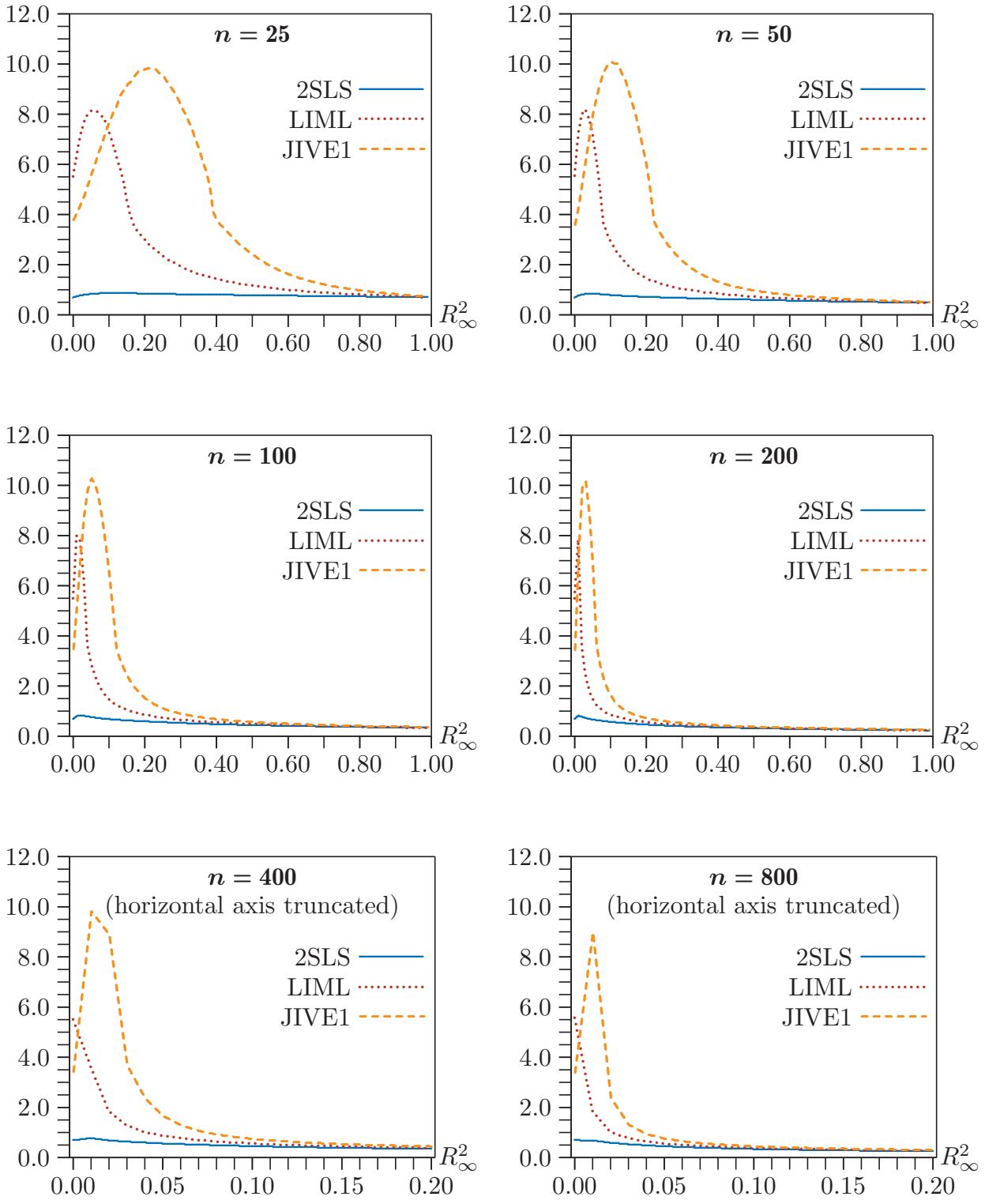

Figure 5. Nine decile range of three estimators, $r = 5$, $\rho = 0.9$

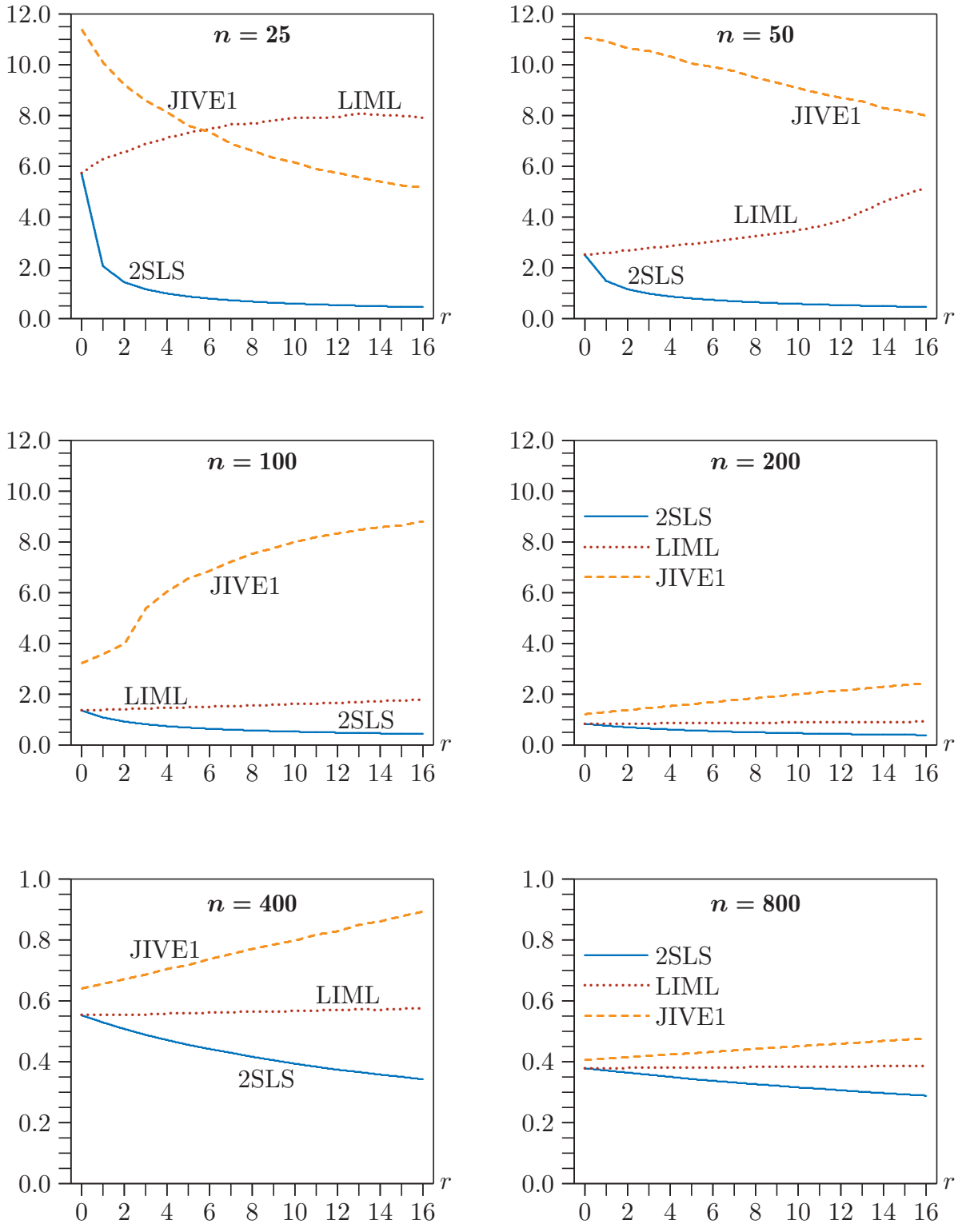

Figure 6. Nine decile range of three estimators, $R^2_\infty = 0.1$, $r = 0.9$

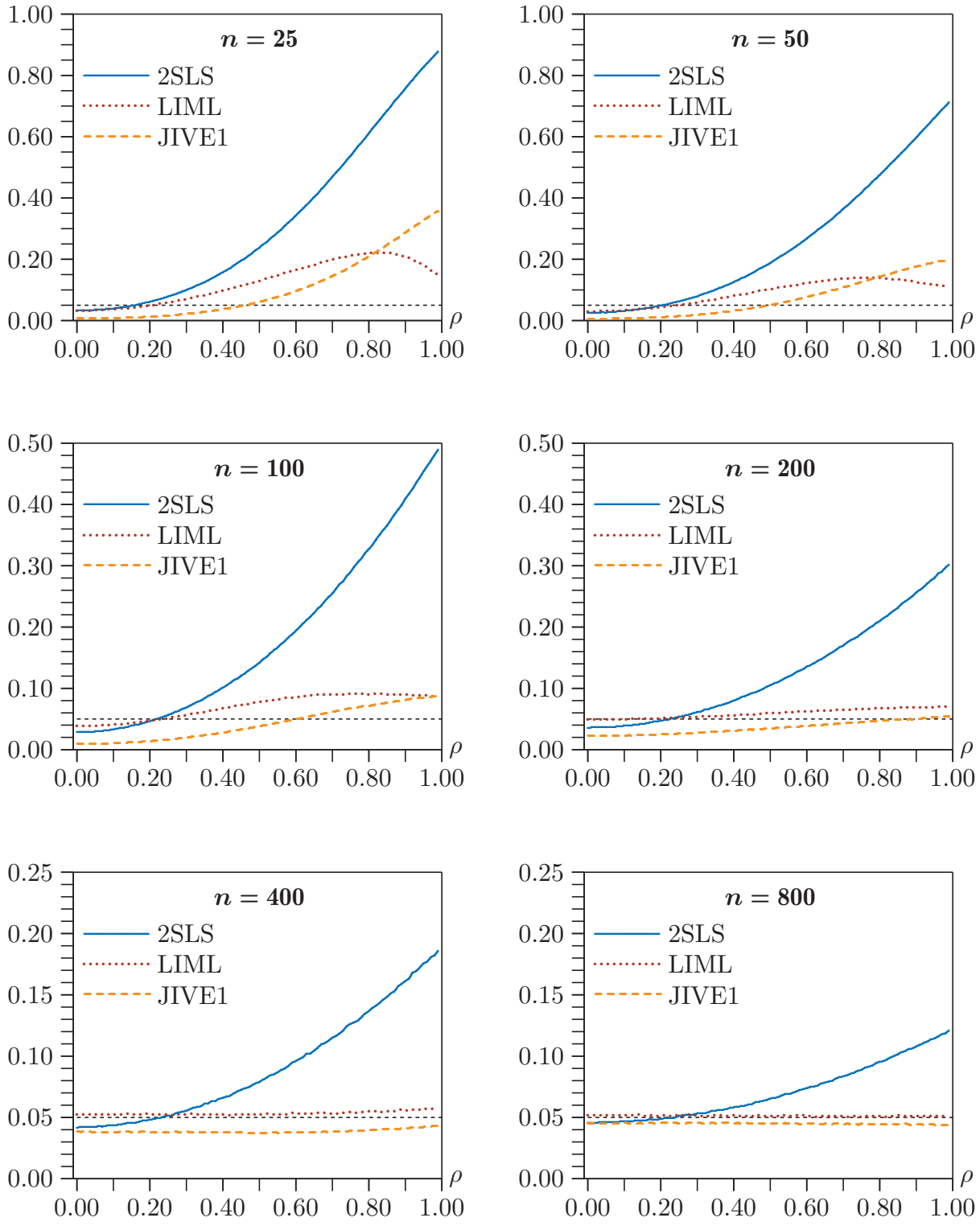

Figure 7. Rejection frequencies for asymptotic t tests at .05 level, $r = 5$, $R^2_\infty = 0.1$

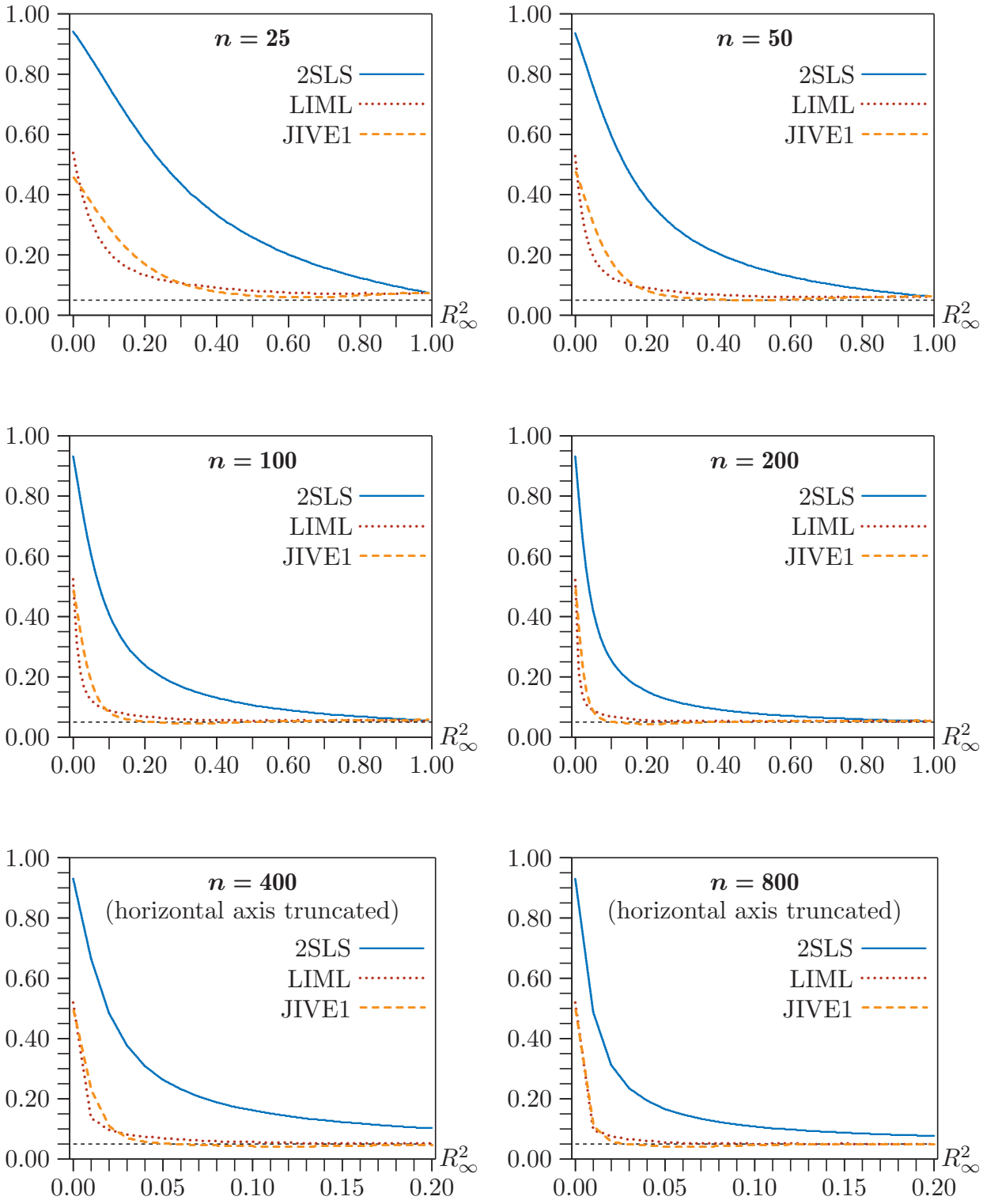

Figure 8. Rejection frequencies for asymptotic t tests at .05 level, $r = 5$, $\rho = 0.9$

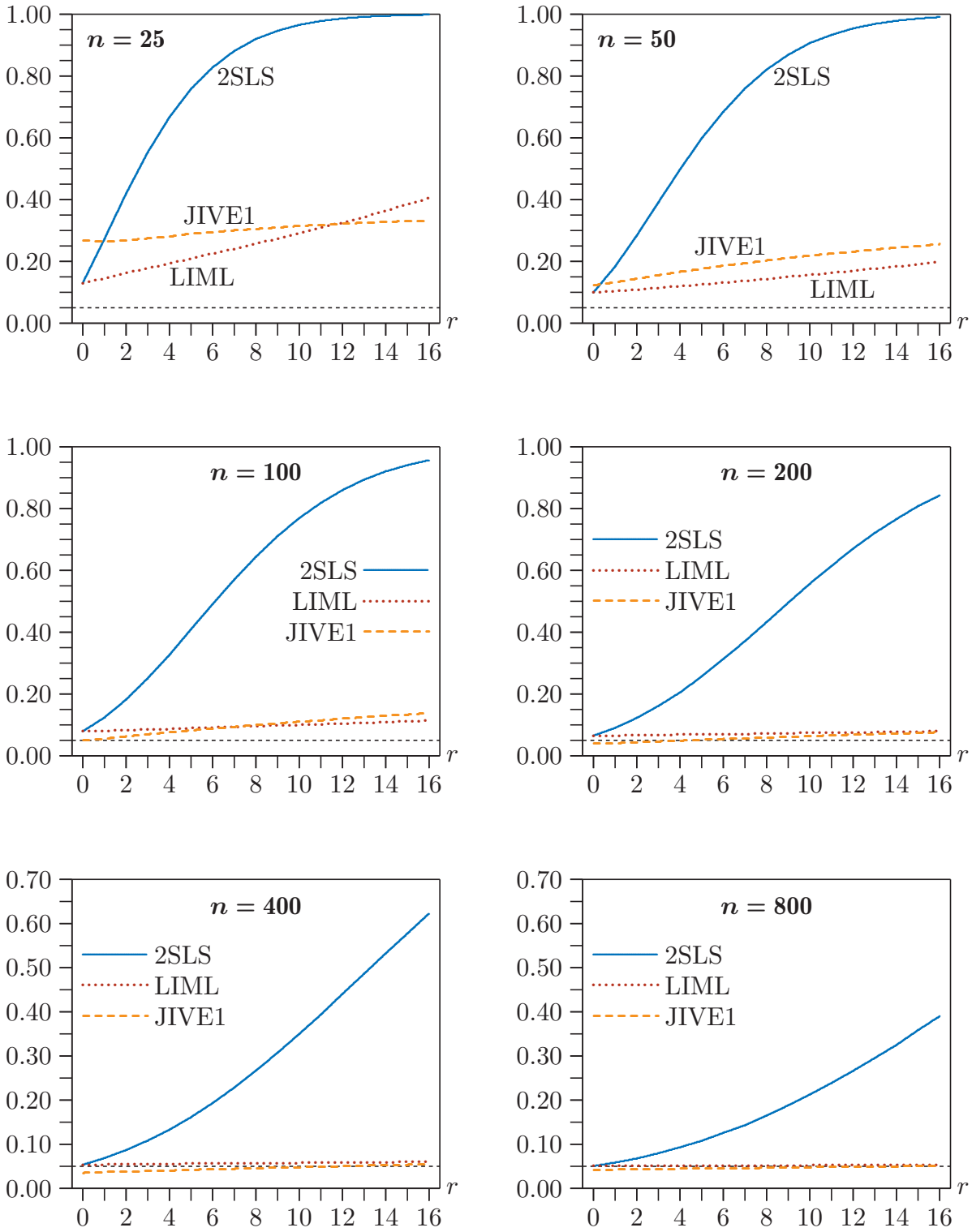

Figure 9. Rejection frequencies for asymptotic t tests at .05 level, $R^2_\infty = 0.1$, $\rho = 0.9$

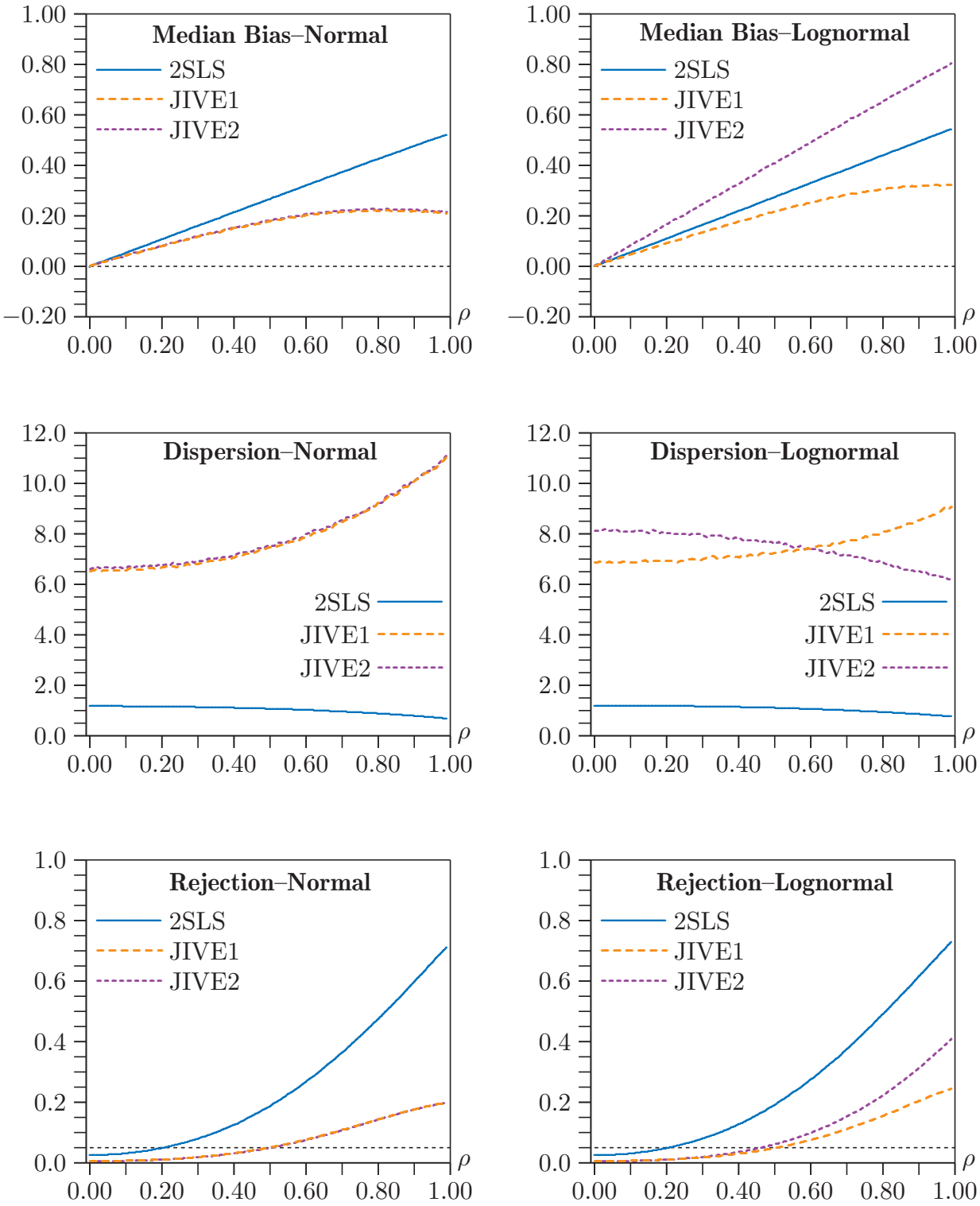

Figure 10. Effects of the Distribution of the Instruments, $n = 50$, $r = 5$, $R^2_\infty = 0.1$

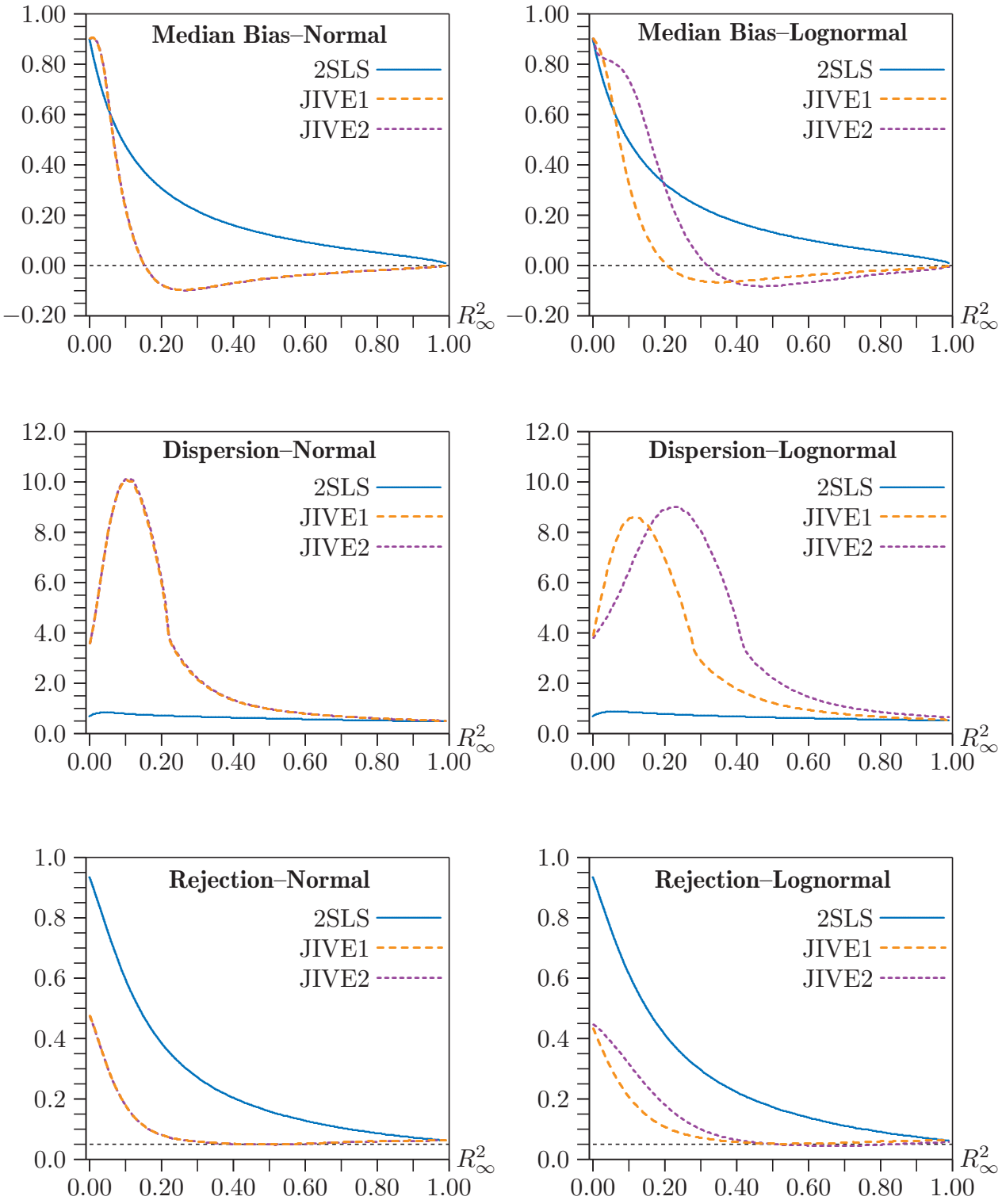

Figure 11. Effects of the Distribution of the Instruments, $n = 50$, $r = 5$, $\rho = 0.9$

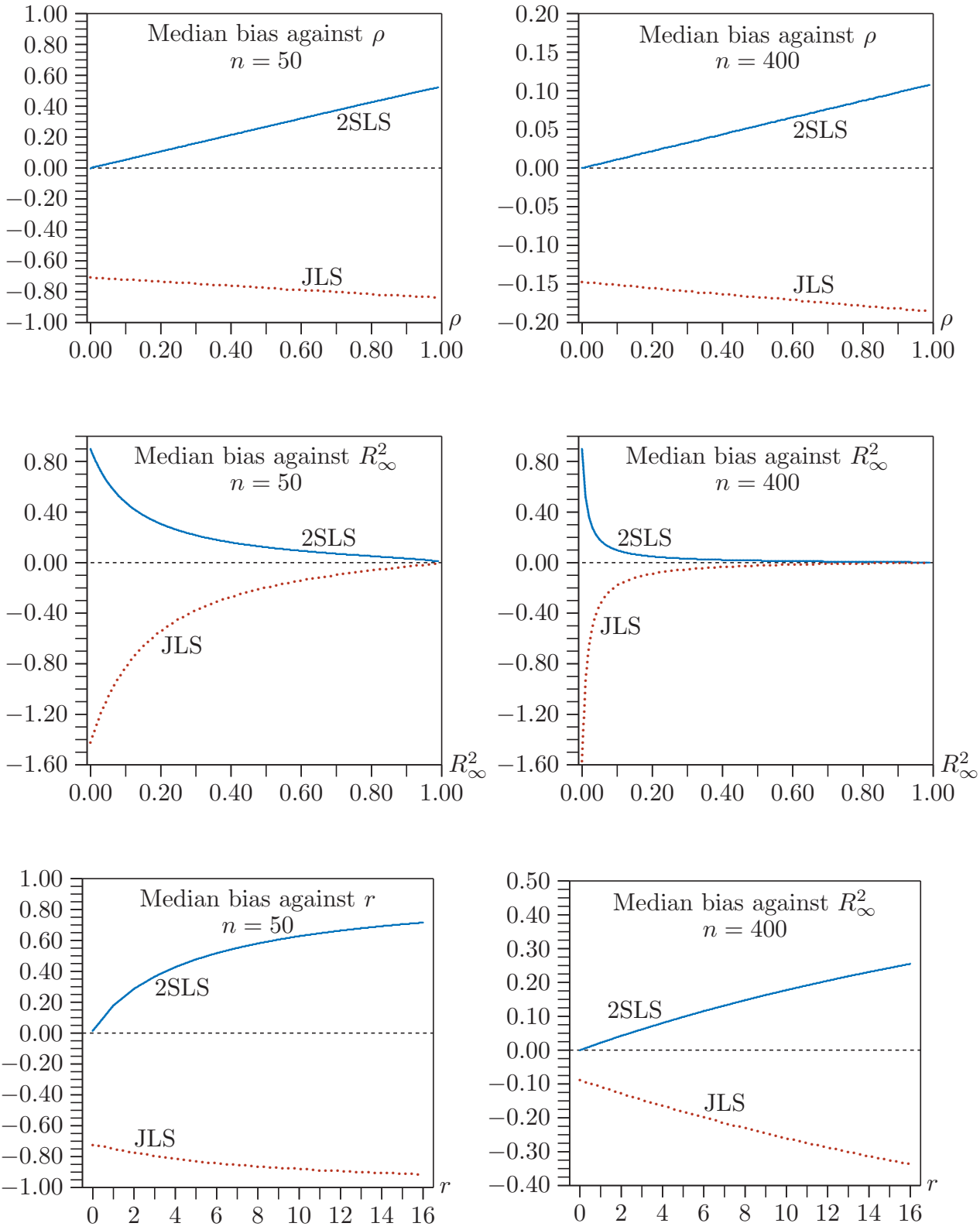

Figure 12. Median bias of 2SLS and JLS

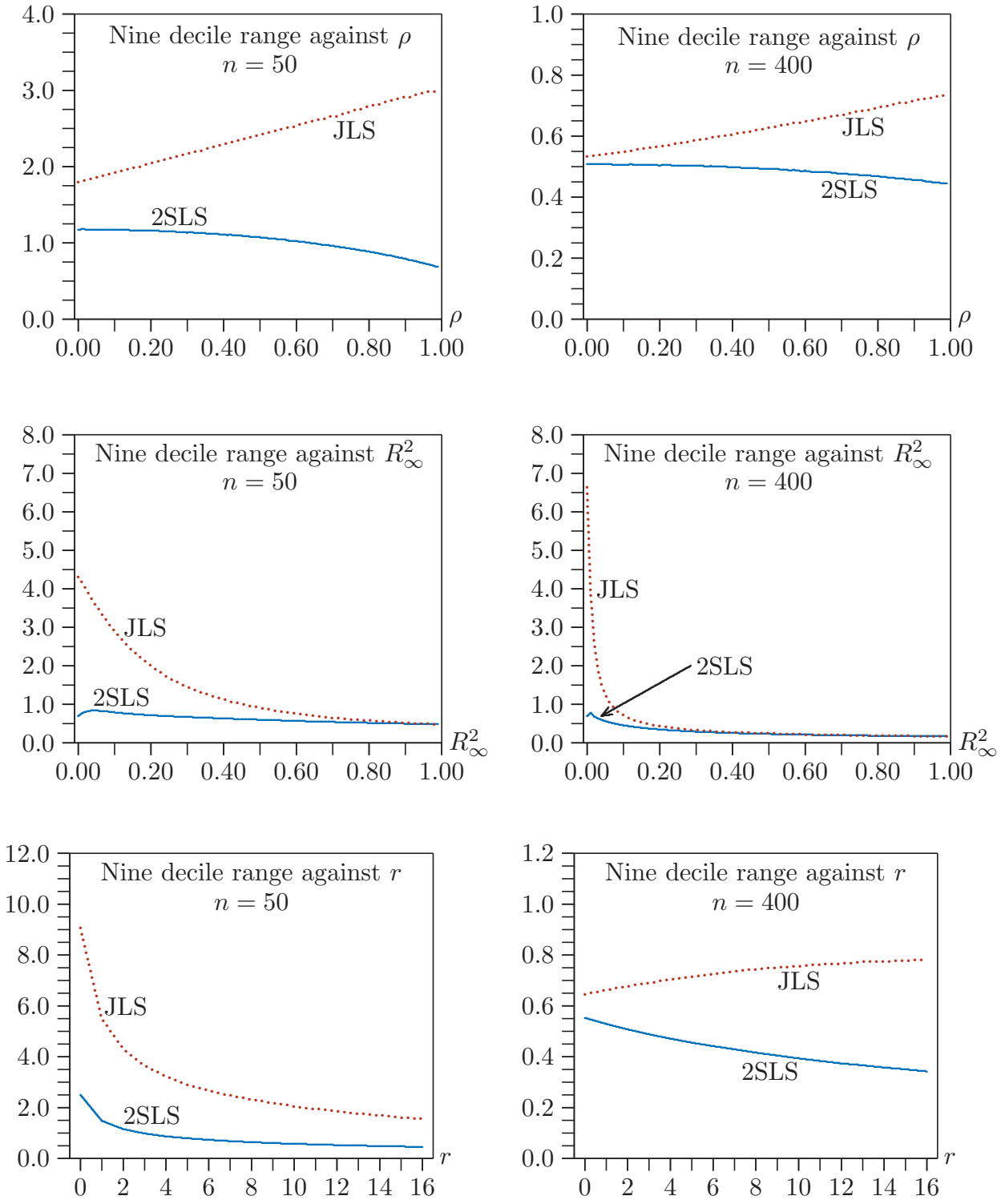

Figure 13. Nine decile range of 2SLS and JLS

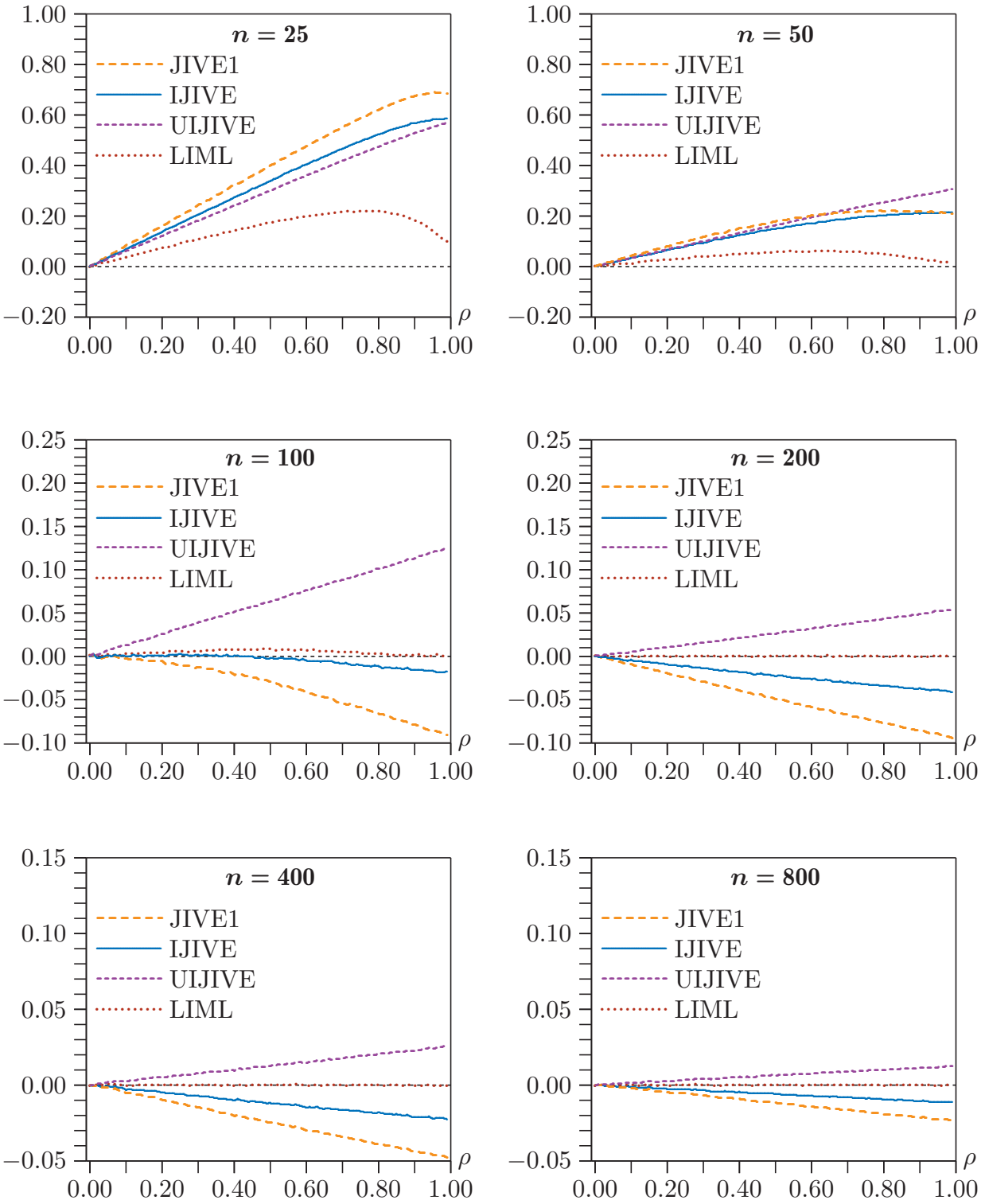

Figure 1a. Median bias of four estimators, $r = 5$, $R^2_\infty = 0.1$

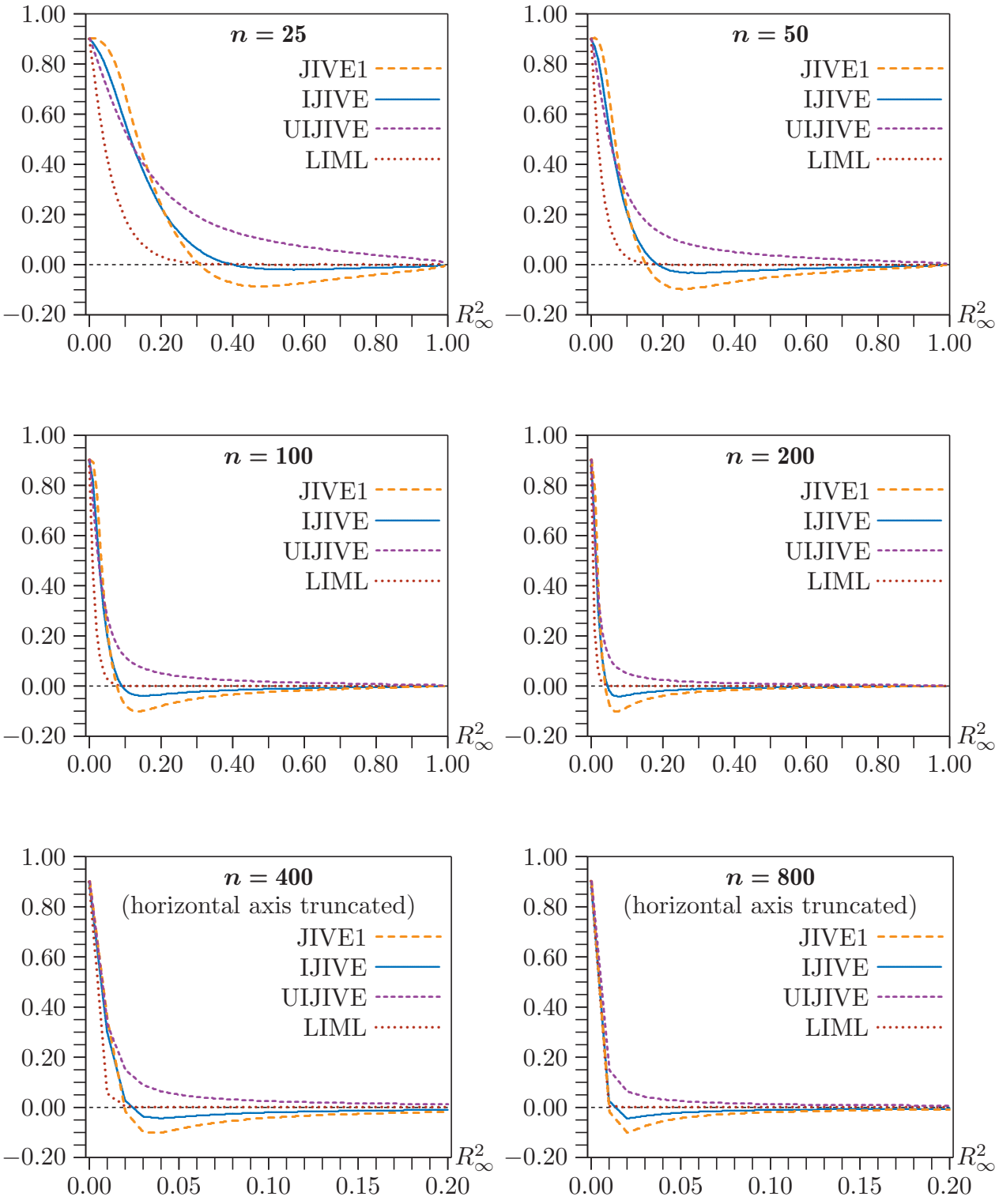

Figure 2a. Median bias of four estimators, $r = 5$, $\rho = 0.9$

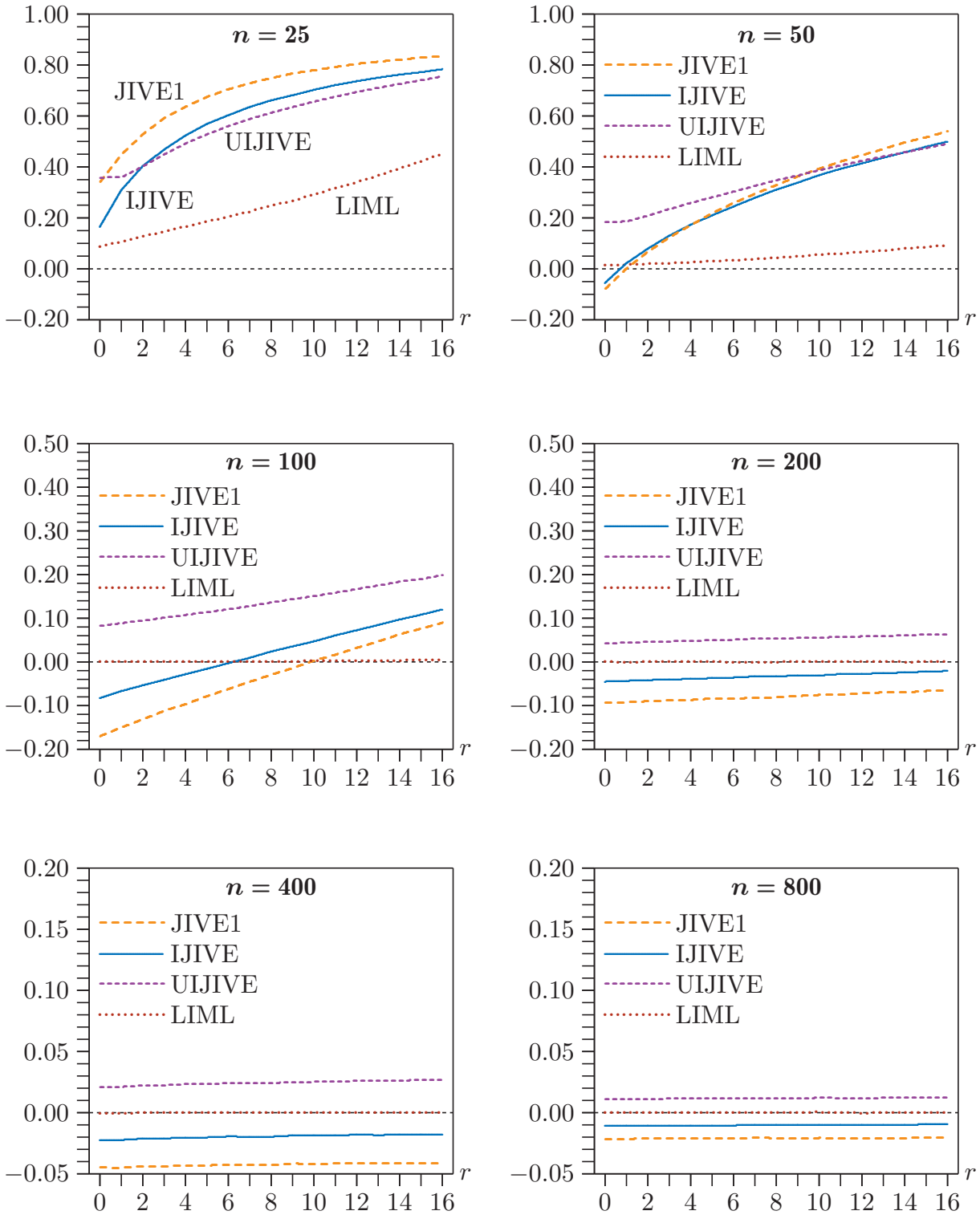

Figure 3a. Median bias of four estimators, $R^2_\infty = 0.1$, $\rho = 0.9$

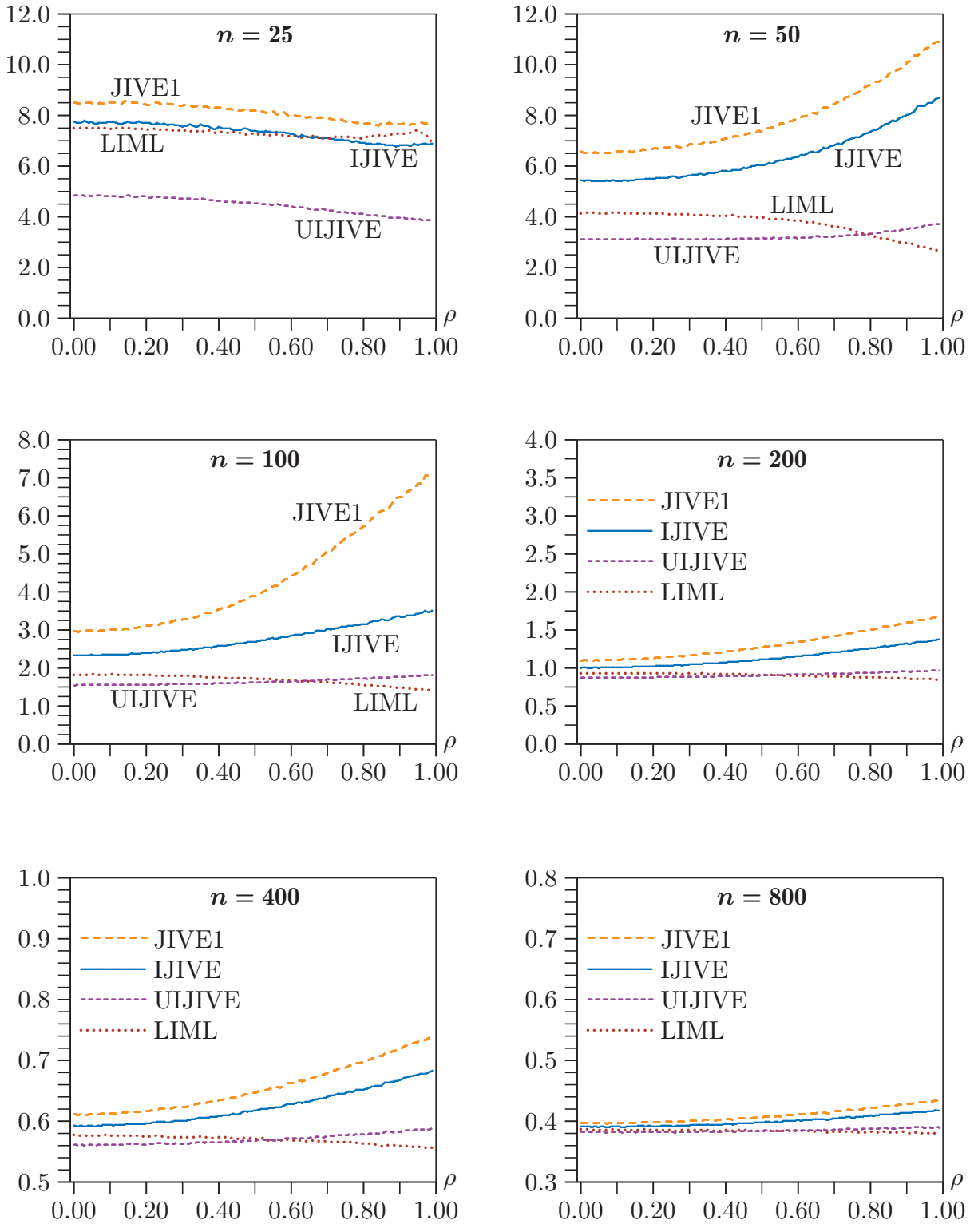

Figure 4a. Nine decile range of four estimators, $r = 5$, $R^2_\infty = 0.1$

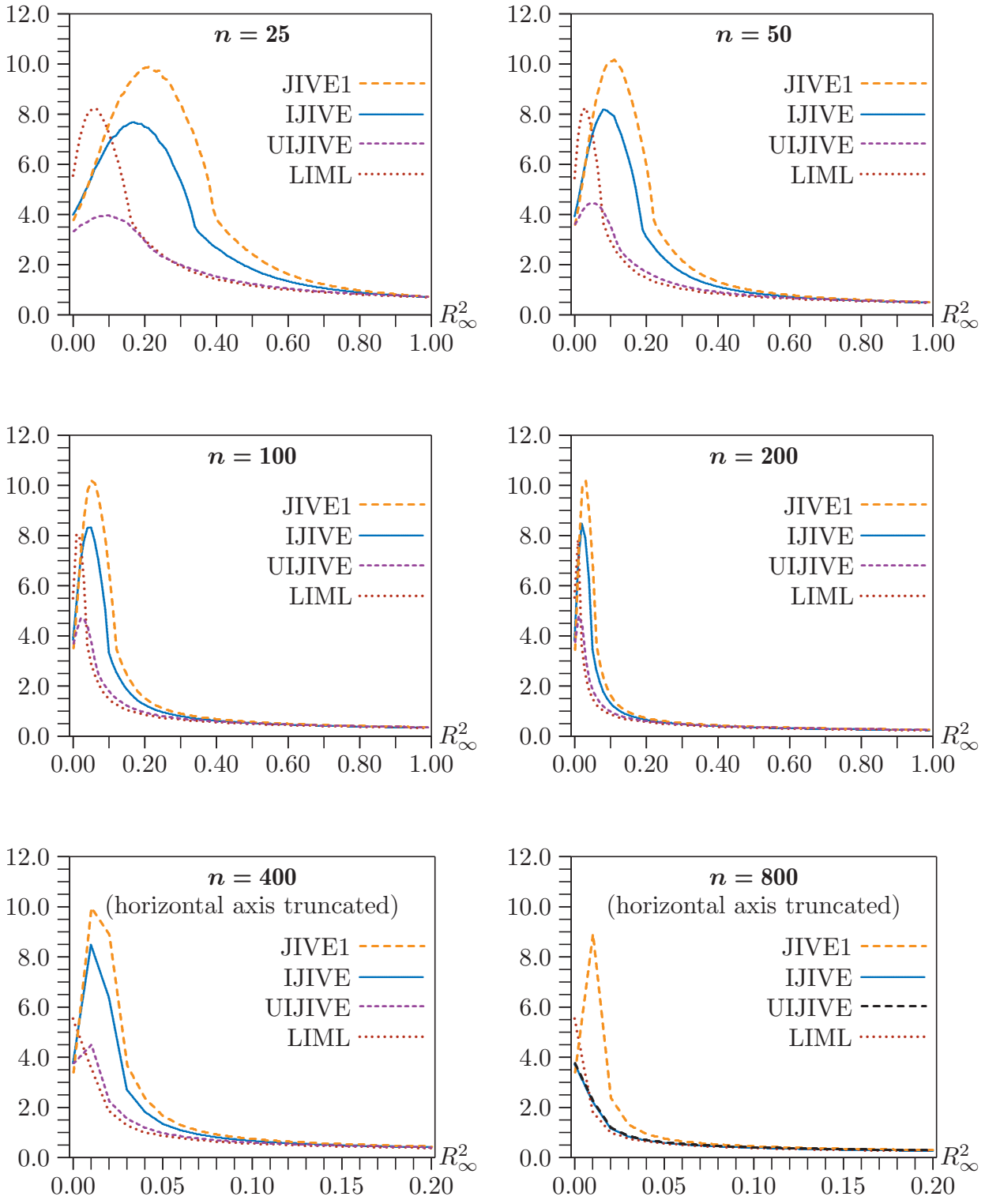

Figure 5a. Nine decile range of four estimators, $r = 5$, $\rho = 0.9$

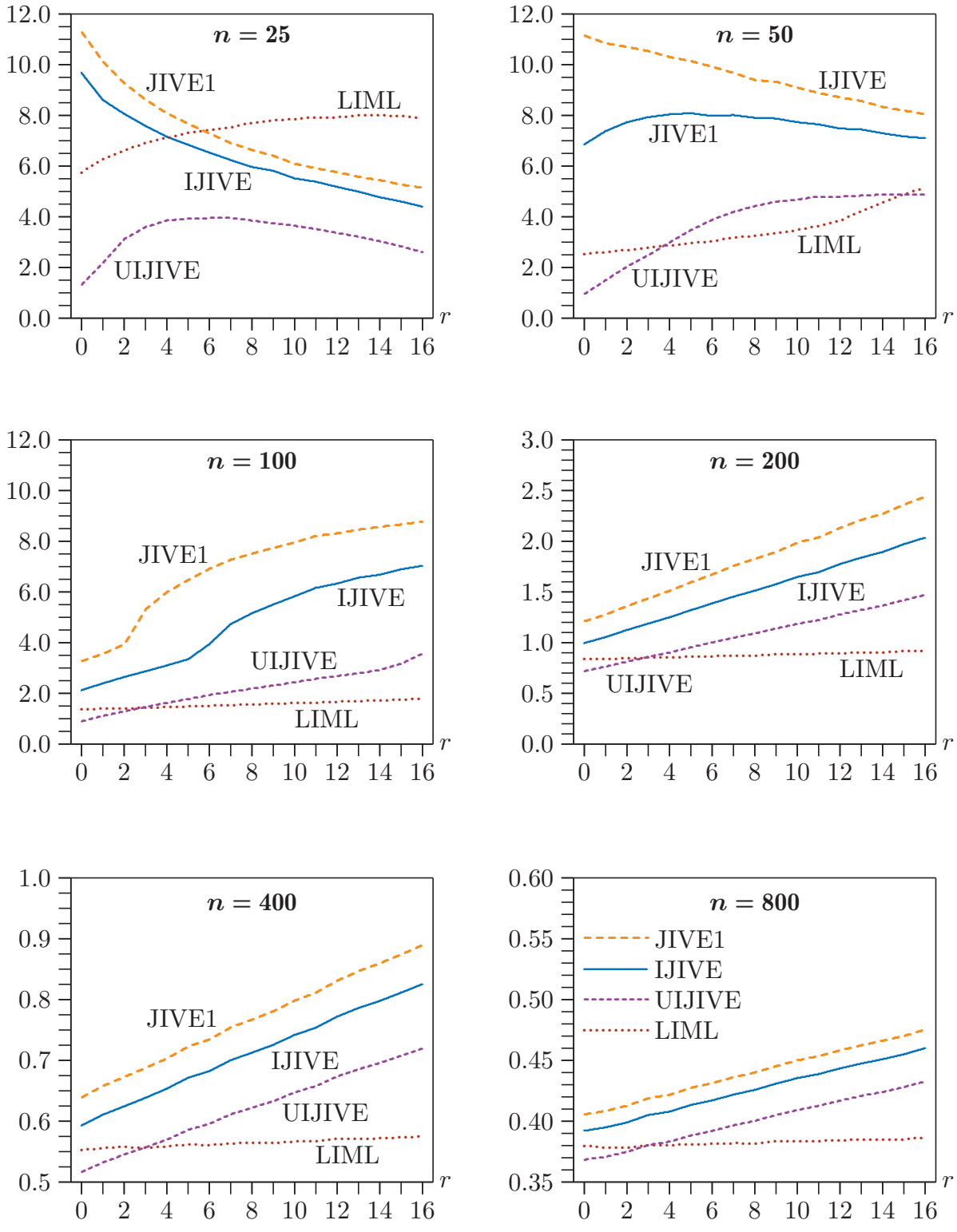

Figure 6a. Nine decile range of four estimators, $R^2_\infty = 0.1$, $r = 0.9$