

Rheinische Friedrich-Wilhelms-Universität Bonn

Master Programme in Economics

**Term Paper for Research Module in  
Econometrics & Statistics**

**Weak Instrumental Variables:  
Limitations of Traditional 2SLS  
and  
Exploring Alternative Instrumental Variable Estimators**

Submitted by:

Aiwei Huang

Madhurima Chandra

Laura Malkhasyan

Supervised by:

JProf. Dr. Dominik Liebl

February 05, 2020

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Two-Stage Least Squares and its Limitations</b>	<b>4</b>
2.1	Basic Model . . . . .	4
2.2	Concentration Parameter . . . . .	5
2.3	Two-Stage Least Squares Estimator . . . . .	5
2.4	Limitations of 2SLS: Bias . . . . .	6
2.4.1	Bias in Single Instrument case . . . . .	6
2.4.2	Approximate Expression of Bias in Multiple-Instruments case . . . . .	7
2.4.3	Inconsistency of the 2SLS estimator . . . . .	8
<b>3</b>	<b>Testing for Weak Instruments</b>	<b>9</b>
3.1	Defining the ‘Weakness’ precisely . . . . .	9
3.2	First Stage F-statistic . . . . .	9
3.3	Anderson-Rubin Test . . . . .	10
<b>4</b>	<b>Alternative Estimators</b>	<b>11</b>
4.1	Jackknife Instrumental Variables Estimator . . . . .	11
4.2	Limited Information Maximum-Likelihood Estimator . . . . .	13
<b>5</b>	<b>Simulation Study</b>	<b>15</b>
5.1	Finite Sample Properties . . . . .	15
5.2	Performance of the Estimators . . . . .	16
<b>6</b>	<b>Application to Returns to Schooling</b>	<b>18</b>
<b>7</b>	<b>Conclusion</b>	<b>21</b>
<b>A</b>	<b>Formulae</b>	<b>i</b>
A.1	Deriving the approximate expression of 2SLS Bias . . . . .	i
A.2	Deriving the LIML Estimator . . . . .	iii
<b>B</b>	<b>Tables &amp; Figures</b>	<b>iv</b>

# Chapter 1

## Introduction

Instrumental variables estimation has gained considerable traction in recent decades as a tool for causal inference, particularly amongst empirical researchers. However, this has also highlighted the importance of taking a deeper look at the theoretical properties underlying one's estimator of choice, whether it is instrumental variable estimator or any other estimator. A case-in-point is the well-known 1991 quarter-of-birth study on returns to schooling estimates by Angrist and Krueger [1991] and the equally famous rebuttal of their results published by Bound et al. [1995].

The Bound et al. [1995] critique of the quarter-of-birth study in essence formed the starting point of the weak instruments literature. They provided simulation evidence that the instruments used by Angrist and Krueger [1991] were very weak and this in turn led to misleading results. Interestingly, their study was not the first time that the poor finite-sample behavior of the instrumental variables estimator (two-stage least squares estimator in particular) was discussed. Previously, numerous authors had presented results on the magnitude of the finite-sample bias of the 2SLS estimator towards OLS, such as Nagar [1959], Basmann [1960], Richardson [1968] and Sawa [1969]. However, their critique was the first to suggest some practices for the detection of possible presence of weak instruments (reporting first-stage F-statistic and the  $R^2$  of the first stage regression), which have since become standard practice for using instrumental variables in empirical studies.

The literature on weak instruments have since developed considerably and continues to do so. Staiger and Stock [1997] showed that conventional asymptotic approaches fail to provide good approximations for weak instruments, and formalized the problem of weak instruments by introducing 'weak-instrument asymptotics' which mimics the situation of weak instruments better. Stock et al. [2002] provide a comprehensive survey of the weak instruments literature, wherein they emphasize that the problem of bias of the two-stage least-squares estimator is not solely a small sample problem. It affects estimates carried out on large sample sizes as well, as is the case with the quarter-of-birth study which has a dataset of around 300,000 observations.

Further, given the poor finite-sample properties of the two-stage least squares estimator, numerous alternate estimators have also been proposed. These include the USSIV estimator suggested by Angrist and Krueger [1995], Fuller-k estimator by Fuller [1977], bias-adjusted 2SLS estimator by Donald and Newey [2001], JIVE estimators by Angrist et al. [1999] and Blomquist and Dahlberg [1999]. Another estimator, the LIML was formalized by Anderson et al. [1949]. In our paper, we take a deeper look at

---

two of these estimators - JIVE and LIML estimators.

We introduce the two conditions for instrumental variables estimation, the first of which will be a recurring theme in our paper, since it connects to the problem of weak instruments. Consider the standard population regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

where  $\varepsilon_i$  is the error term representing omitted factors that determine  $Y_i$ . Variables correlated with the error term are termed endogenous variables and those uncorrelated with the error term are exogenous. A valid instrument  $Z_i$  must satisfy two conditions:

- Instrument Relevance:  $\text{corr}(Z_i, X_i) \neq 0$ .

It is the implications of this condition, specifically the strength (or weakness) of this correlation between the instrument and the endogenous regressor, that we are interested in. Instruments that explain little of the variation in  $X$  are called weak instruments. A precise definition is introduced in section 3.1.

- Instrument Exogeneity:  $\text{corr}(Z_i, \varepsilon_i) = 0$

Note that this condition cannot be statistically tested, since it involves the covariance between  $Z_i$  and the unobserved  $\varepsilon_i$ .

If number of instruments ( $K$ ) equals the number of endogenous regressors ( $L$ ) we have the just or exactly identified case. If the number of instruments exceeds the number of endogenous regressors,  $K \geq L$ , then we have the over-identified case.

This paper makes three contributions. First, we provide a detailed theoretical discussion on the properties of the standard two-stage least squares estimator in the presence of weak instruments, and introduce and derive two alternative estimators. Second, we conduct Monte-Carlo simulations to compare the finite-sample behavior of the different estimators, particularly in the weak-instruments case. Third, we apply the estimators to a real-world context; we employ the different estimators to calculate returns to schooling.

The rest of the paper is sectioned as follows. Chapter 2 presents a derivation of the 2SLS estimator and its limitations in the weak instruments case. Chapter 3 discusses methods to test for weak instruments. Chapter 4 presents the alternative estimators (JIVE and LIML). Chapter 5 presents the simulation study. Chapter 6 presents the application to returns to schooling. Chapter 7 concludes.

## Chapter 2

# Two-Stage Least Squares and its Limitations

### 2.1 Basic Model

Throughout the paper, unless stated otherwise, we consider the following basic model with two equations, following the notation of Angrist et al. [1999]:

$$Y_i = X_i\beta + \varepsilon_i$$

$$X_i = Z_i\pi + \eta_i$$

$Y_i$  is scalar.  $X_i$  is  $1 \times L$  row vector of potentially endogenous regressors.  $Z_i$  is  $1 \times K$  row vector, with  $K \geq L$ .  $K - L$  is the number of over-identifying restrictions. In matrix notation:

$$\underset{(N \times 1)}{\mathbf{Y}} = \underset{(N \times L)}{\mathbf{X}}\beta + \underset{(N \times 1)}{\varepsilon} \quad (2.1)$$

$$\underset{(N \times L)}{\mathbf{X}} = \underset{(N \times K)}{\mathbf{Z}}\pi + \underset{(N \times L)}{\eta} \quad (2.2)$$

where (2.1) denotes the structural equation, (2.2) is the first stage.  $\mathbf{Y}$  and  $\varepsilon$  are  $N \times 1$  vectors.  $\mathbf{X}$  and  $\eta$  are  $N \times L$  matrices, and  $\mathbf{Z}$  is  $N \times K$  matrix. If the vectors of regressors and that of instruments have  $M$  common elements, then  $M$  columns of the  $N \times L$  matrix  $\eta$  have their elements as zero.

The following assumptions hold for our model:

1. Conditional on instruments  $Z_i$ , the error term  $\varepsilon_i$  has expectation zero ( $E[\varepsilon_i|Z_i] = 0$ ) and variance  $\sigma^2$ .
2.  $E[\eta|\mathbf{Z}] = 0$  and  $E[\eta_i'\eta_i|\mathbf{Z}] = \Sigma_\eta$ , with rank  $L - M$ .
3.  $E[\varepsilon_i\eta_i'|\mathbf{Z}] = \sigma_{\varepsilon\eta}$ , where  $\sigma_{\varepsilon\eta}$  is an  $L$ -dimensional column vector.
4. All observations of  $(Y_i, X_i, Z_i)$  are independent and identically distributed.

## 2.2 Concentration Parameter

The Concentration Parameter  $\mu^2$ , given by:

$$\mu^2 = \frac{\pi' Z' Z \pi}{\sigma_\eta^2} \quad (2.3)$$

measures the strength of the instruments. It is unitless. An important question to be addressed at this point is that since  $\pi$ , and thus  $\mu^2$ , is unknown, how is the researcher supposed to know whether  $\mu^2$  value is low enough for the instrument under study to be weak? This is addressed by the fact that the concentration parameter can be interpreted in terms of the first stage F statistic, discussed in Section 3. If the sample size is large,  $E(F) \cong \mu^2/K + 1$ , and thus  $F - 1$  can be treated as an estimator of  $\mu^2/K$ , which gives us a convenient way to test for weak instruments.

The concept of  $\mu^2$  is essentially the starting point for understanding the weak instruments literature. Rothenberg [1984] showed that the concentration parameter plays the role we commonly associate with the sample size or number of observations: as  $\mu^2$  becomes large, the normal distribution becomes a good approximation for the 2SLS distribution. Likewise a small concentration parameter leads to a non-normal distribution of 2SLS, and the estimator becomes biased. Thus  $\mu^2$  can be interpreted as the effective sample size.

## 2.3 Two-Stage Least Squares Estimator

It is helpful to characterize the 2SLS and (later on the jackknife estimator as well), as a feasible version of the ideal but infeasible instrumental variables estimator. Adhering to the notation defined in section 2.1, we derive the formula for the 2SLS estimator using this characterization. The derivation is adopted, with modifications to the notation, from Hansen [2019] and Angrist et al. [1999].

First, we estimate the ideal estimator  $\hat{\beta}_{OPT}$  using the optimal instrument  $Z\pi$ , by ordinary least-squares:

$$\hat{\beta}_{OPT} = ((Z\pi)'X)^{-1}((Z\pi)'Y) \quad (2.4)$$

$\pi$  is unknown however, and so the optimal estimator  $\hat{\beta}_{OPT}$  is infeasible.

We use an estimate of  $\pi$  instead, and denote it by  $\hat{\pi}$ .  $\hat{\pi}$  is estimated from the reduced form regression which yields  $\hat{\pi} = (Z'Z)^{-1}(Z'X)$ . Substituting  $\hat{\pi}$  into (2.4),

$$\hat{\beta}_{2SLS} = ((Z\hat{\pi})'(Z\hat{\pi}))^{-1}((Z\hat{\pi})'Y) \quad (2.5)$$

$$\hat{\beta}_{2SLS} = ((Z(Z'Z)^{-1}(Z'X))'(Z(Z'Z)^{-1}(Z'X)))^{-1}((Z(Z'Z)^{-1}(Z'X))'Y) \quad (2.6)$$

$$\hat{\beta}_{2SLS} = ((X'Z)(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \quad (2.7)$$

---

This is the two-stage least squares estimator.  $\hat{\beta}_{2SLS}$  can be thought of as an estimator with the constructed instrument  $(\mathbf{Z}\hat{\pi})$  where  $\hat{\pi} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})$ .

Representing (2.7) in terms of the projection matrix  $P_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ :

$$\hat{\beta}_{2SLS} = ((\mathbf{X}'P_Z\mathbf{X})^{-1}\mathbf{X}'P_Z\mathbf{Y}) \quad (2.8)$$

The two-stage least squares is by far the most widely used estimator for instrumental variables estimation, however, as we briefly discussed in chapter 1, it has certain concerning limitations, which we expand upon in the next section.

## 2.4 Limitations of 2SLS: Bias

In the weak instruments case, the 2SLS estimator fails to provide unbiased, reliable estimates. As shown in Section 2.3, we arrived at the 2SLS estimator using an estimate of  $\pi$  since the true value of the first-stage coefficients is unknown, which implies a certain amount of over-fitting of the first-stage equation, leading to bias in the direction of the expectation of the OLS estimator of  $\beta$ .

There is considerable theoretical research noting the poor finite-sample properties of instrumental variables estimator. Results on the magnitude of this bias have been provided by Nagar [1959], Richardson [1968], Sawa [1969] and Buse [1992]. Further, as shown by noteworthy results of Bound et al. [1995], with weak instruments particularly, the finite sample behavior worsens and the estimator can become biased as well as inconsistent.

### 2.4.1 Bias in Single Instrument case

To visualize the problem intuitively, first let us consider the simple case of a single (weak) instrument and endogenous regressor:

$$Y_i = X_i\beta + \varepsilon_i$$

$$X_i = Z_i\pi_1 + \eta_i$$

In this case, the 2SLS estimator is simply given by the ratio of the two covariances:

$$\hat{\beta}_{2SLS} = \frac{\sigma_{Y_i Z_i}}{\sigma_{X_i Z_i}} \quad (2.9)$$

However, given the fact that our instrument is very weak,  $\sigma_{X_i Z_i} = 0$ . Hence,  $\hat{\beta}_{2SLS}$  does not exist.

In the simplest case, the 2SLS estimator is just the ratio of two covariances, and with weak instruments, the 2SLS or general instrumental variables estimator does not exist.

### 2.4.2 Approximate Expression of Bias in Multiple-Instruments case

In this section we provide two different expressions approximating the bias of the two-stage least squares estimator in the general case and show how it centers around ordinary least squares.

Defining the ‘relative bias’ of 2SLS to be its bias relative to the inconsistency of OLS, Buse [1992] derived an expression for the approximate bias of  $\hat{\beta}_{2SLS}$  using power series approximations. This result holds even when the errors are not normally distributed:

$$\frac{\sigma_{\varepsilon\eta}(K-2)}{\pi'\mathbf{Z}'\mathbf{Z}\pi} \quad (2.10)$$

where  $N$  is the sample size and  $K$  is the number of excluded instruments. This expression is approximately inversely proportional to  $\mu^2/(K-2)$ , as shown:

$$\frac{\sigma_{\varepsilon\eta}\sigma_{\eta}^2(K-2)}{\sigma_{\eta}^2\pi'\mathbf{Z}'\mathbf{Z}\pi} = \frac{\sigma_{\varepsilon\eta}(K-2)}{\sigma_{\eta}^2\mu^2} \quad (2.11)$$

where  $\mu^2$  denotes the concentration parameter (measure of the strength of the instrument as discussed in Section 1.). Note that in the above equation, the term  $\frac{\sigma_{\varepsilon\eta}}{\sigma_{\eta}^2}$  approximately equals the asymptotic bias of the OLS estimator when the instrument explains little of the variation of  $\mathbf{X}$ . Hence, from equation (2.10) it is clearly seen that for  $K > 2$ , the bias of the 2SLS estimator relative to OLS is inversely proportional to the concentration parameter, hence, weaker the instrument(s), lower is the concentration parameter, and higher is the bias of 2SLS towards the OLS estimate.

It can be seen from the above formulation that increasing the number of instruments with the explanatory power remaining constant, causes the relative bias of the 2SLS to only increase.

Before moving on to the second expression, it is important to briefly introduce alternate asymptotic representations that are generally used for the weak instruments case. As discussed by Stock et al. [2002], for weak instruments, conventional asymptotic approximations to finite-sample distributions are quite poor. Two alternate asymptotic methods commonly employed are **weak-instrument asymptotics** (involving a sequence of models chosen to keep  $\mu^2/K$  constant as sample size  $N \rightarrow \infty$ ) pioneered by Staiger and Stock [1997] and **many-instrument or group asymptotics** (involving sequence of models with fixed instruments and normal errors, where  $K$  is proportional to  $N$  and  $\mu^2/K$  converges to a constant finite limit), first proposed by Bekker [1994]. In the 1995 working paper version of Angrist et al. [1999], it was referred to as group asymptotics.

We present an expression for the approximate bias of the 2SLS estimator using group asymptotics, wherein we let the number of instruments grow proportional to rate of the sample size. This keeps the instruments weak. The detailed derivation of this expression is provided in the appendix A.1.

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\varepsilon\eta}}{\sigma_{\eta}^2} \frac{1}{F+1} \quad (2.12)$$



---

where  $F$  is the population analog of the F-statistic for the joint significance of the instruments in the first-stage regression.  $\frac{\sigma_{\varepsilon\eta}}{\sigma_{\eta}^2}$  approximately equals the asymptotic bias of the OLS estimator. Given a weak first-stage (weak instruments case),  $F \rightarrow 0$ , and we can see that the bias approaches  $\frac{\sigma_{\varepsilon\eta}}{\sigma_{\eta}^2}$ . With a strong first-stage,  $F \rightarrow \infty$  and then the 2SLS bias goes to 0.

### 2.4.3 Inconsistency of the 2SLS estimator

In this section we relate the two conditions for instrumental variables estimation to the consistency property of the 2SLS estimator. If the weak correlation (between the instrument(s) and the endogenous variable) is coupled together with even a small violation of the second condition of instrumental variables estimation (which is the instrument exogeneity condition) then we have an inconsistent estimator. This insight was first discussed by Bound et al. [1995].

To represent the problem discussed above, let us consider the probability limit of the 2SLS estimator:

$$plim \hat{\beta}_{2SLS} = \beta + \frac{\sigma_{\hat{\mathbf{X}},\varepsilon}}{\sigma_{\hat{\mathbf{X}}}^2} \quad (2.13)$$

where  $\hat{\mathbf{X}}$  is the projection of  $\mathbf{X}$  onto  $\mathbf{Z}$ , and  $\sigma_{\hat{\mathbf{X}},\varepsilon}$  is the covariance between  $\hat{\mathbf{X}}$  and  $\varepsilon$ .

From equation (2.13) we can intuitively understand that if  $\sigma_{\hat{\mathbf{X}}}^2$  is small, which implies that we have a weak instrument, then as long as  $\sigma_{\hat{\mathbf{X}},\varepsilon}$  is zero the estimator will be consistent. However, suppose we have a weak instrument and also,  $\sigma_{\hat{\mathbf{X}},\varepsilon}$  is small but non-zero, then even that small correlation between  $\mathbf{Z}$  (and thus,  $\hat{\mathbf{X}}$ ) and the structural error term  $\varepsilon$  can lead to a large inconsistency.

Hence, with weak instruments, even moderate correlation between instrument and structural error term can magnify the inconsistency of IV estimator.

## Chapter 3

# Testing for Weak Instruments

Testing for presence of weak instruments is, at the time of writing, an active field of research. For a detailed overview, see Stock et al. [2002]. For the purpose of our study, we limit our attention to two tests - the widely-used first-stage F-statistic and the Anderson-Rubin Test, which has gained resurgence in recent years in light of new developments in instrumental variables research.

### 3.1 Defining the ‘Weakness’ precisely

Stock and Yogo [2002] posit that the definition of weak instruments depends on the inferential task to be carried out, and cannot be resolved in the abstract. One approach is to define a set of instruments to be weak if  $\mu^2/K$  is small enough that inferences based on conventional normal approximating distributions are misleading. For instance, if a researcher wants their 2SLS estimate bias to be small, one measure of whether an instrument(s) is strong is whether  $\mu^2/K$  is large enough such that the 2SLS relative bias (relative to the bias of ordinary least squares) is below a certain threshold, for example the relative bias is below 10%. Hence, to be deemed a ‘weak’ instrument, the 2SLS estimate using that instrument should have relative bias above 10%. The definition we discussed (and use for our simulation) is based on relative bias, another definition (for instance on size of test) may result in a different cut-off value.

### 3.2 First Stage F-statistic

The first-stage F-statistic is the F-statistic testing the hypothesis that the coefficients on the instruments equal zero ( $\pi = 0$ ) in the first stage of two stage least squares. Stock and Yogo [2002] show that the definition of weak instruments discussed above implies a threshold value for  $\mu^2/K$ , under weak asymptotics. A weak instrument will have a  $\mu^2/K$  value (and hence, an F-statistic value, since  $F-1$  can be treated as an estimator of  $\mu^2/K$  as discussed in Section 2.2) lower than the threshold. For the case of a single endogenous regressor, Staiger and Stock [1997] provide a rule-of thumb threshold of 10: a value less than 10 indicates that the instruments are weak, in which case the 2SLS estimator is biased and 2SLS t-statistics and confidence intervals are unreliable.

---

Stock et al. [2002] provide a table listing critical values of the first-stage F-statistic such that the relative bias of 2SLS estimates is greater than 10%, for different numbers of instruments. The authors arrived at those critical values based on weak-instrument asymptotic approximations. We include a subset of this table (which is relevant for our simulations) as Table B.1 in appendix B for reference.

### 3.3 Anderson-Rubin Test

The AR test is a hypothesis test that has the property of being valid whether instruments are strong, weak or even irrelevant ( $\pi = 0$ ). It tests the null hypothesis  $\beta = \beta_0$  using the statistic. It was proposed by Anderson et al. [1949].

$$AR(\beta) = \frac{(\mathbf{Y} - \mathbf{X}\beta)' P_z (\mathbf{Y} - \mathbf{X}\beta) / K}{(\mathbf{Y} - \mathbf{X}\beta)' \mathbf{M}_Z (\mathbf{Y} - \mathbf{X}\beta) / (N - K)} \quad (3.1)$$

One definition of the LIML estimator is that it minimizes  $AR(\beta)$ . With fixed instruments and normal errors, the quadratic forms in the numerator and denominator of (3.1) are independent chi-squared random variables under the null hypothesis, and  $AR(\beta_0)$  has an exact  $F_{K, T-K}$  null distribution. Under the more general conditions of weak-instrument asymptotics,  $AR(\beta_0) \xrightarrow{d} \chi_k^2 / K$  under the null hypothesis, regardless of the value of  $\mu^2 / K$ . Thus the AR statistic provides a fully robust test of the hypothesis  $\beta = \beta_0$ .

The set of values of  $\beta$  that are not rejected by a 5% Anderson–Rubin test will constitute a 95% confidence set for  $\beta$ . The logic behind the Anderson–Rubin statistic is that it never assumes instrument relevance, and the AR confidence set will have a coverage probability of 95% in large samples, regardless of the strength or weakness of instruments. In light of the importance given to the problem of weak instruments in recent years, this test has gained traction among econometricians, who increasingly advocate for its use for robust inference with weak instruments (See Staiger and Stock [1997]). Particularly, recent research has shown the AR confidence set to be optimal in the single-endogenous-regressor just-identified setting.

# Chapter 4

## Alternative Estimators

To tackle the problem of finite-sample bias of IV, which as we have shown, is particularly problematic in the presence of weak instruments. In this section we present a detailed overview of two alternative estimators which in theory exhibit better finite sample properties.

Both of these estimators, the Jackknife IV estimator and the LIML estimator, fall under the broader class of k-class estimators. These estimators are partially robust ie less sensitive to weak instruments; they are more reliable in comparison to 2SLS estimates.

### 4.1 Jackknife Instrumental Variables Estimator

For an intuitive sense of the functioning of the jackknife instrumental variables estimator, consider the term ‘jackknife’ as presented in statistical literature. The jackknife estimator of a parameter is found by leaving out each observation from a dataset, calculating the estimate and computing the average of these calculations. Given a sample of size  $N$ , the jackknife estimate is found by aggregating the estimates of each sub-sample of size  $(N - 1)$ . In a similar vein, the estimator we describe in this section replaces the usual fitted values from the reduced form regression of the two-stage least squares by ‘omit-one’ fitted values.

The key feature of the JIVE estimator is that it eliminates the correlation between the fitted values and the structural equation errors, as explain in detail below. The fitted value of the standard 2SLS estimator is only asymptotically independent of the structural error ( $\varepsilon_i$ ), but the JIVE is independent even in finite samples.

The derivation we present in this section is adopted from Angrist et al. [1999]. To derive the expression formally, consider again the expressions for  $\hat{\beta}_{OPT}$  and  $\hat{\beta}_{2SLS}$  from section 2.2, equations (2.4) and (2.5).  $\hat{\beta}_{OPT}$  employed the optimal instrument  $\mathbf{Z}\pi$  and  $\hat{\beta}_{2SLS}$  used an estimate of  $\mathbf{Z}\pi$ , denoted by  $\mathbf{Z}\hat{\pi}$ . Now we introduce the new estimator for  $\beta$  by using a different estimate of the optimal instrument,  $Z_i\hat{\pi}$ .

From the formulation of approximate bias of the 2SLS estimator we presented in section 2.4.2 (equation 2.11), it is clear that increasing the number of instruments while keeping the explanatory/predictive power of the instruments constant, leads to an increase in the bias of  $\hat{\beta}_{2SLS}$ . However, we can see that for the optimal estimator, increasing the number of instruments while keeping  $Z_i\pi$

fixed will have no effect on the properties of  $\hat{\beta}_{OPT}$ . Thus, for finite samples in the presence of many instruments,  $\hat{\beta}_{2SLS}$  fares much worse than  $\hat{\beta}_{OPT}$ .

Recall the representation of the 2SLS estimator using the projection matrix  $P_Z$  from Section 2 (equation 2.8). We can see that the first-stage fitted values  $\mathbf{Z}\hat{\pi}$  can be written as:

$$\mathbf{Z}\hat{\pi} = P_Z \mathbf{X} = \mathbf{Z}\pi + P_Z \eta \quad (4.1)$$

The term  $P_Z \eta$  in the above equation is correlated with the error term of the first-stage  $\eta$  (see equation 1.2) and hence with the structural error term  $\varepsilon$ . Put differently, since  $\hat{\pi}$  is estimated on the full sample which includes the  $i$ th observation, it is correlated with  $\eta_i$ , which is correlated with  $\varepsilon_i$ . This correlation is given by:

$$E[\varepsilon_i Z_i \hat{\pi}] = E[E[\varepsilon_i Z_i \hat{\pi} | \mathbf{Z}]] = E[Z_i (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'_i \cdot E[\varepsilon_i \eta_i | \mathbf{Z}]] = E[Z_i (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'_i \cdot \sigma'_{\varepsilon\eta}] = (K/N) \cdot \sigma'_{\varepsilon\eta} \quad (4.2)$$

Due to this correlation between  $\hat{\pi}$  and  $\varepsilon_i$ ,  $\hat{\beta}_{2SLS}$  is biased for  $\beta$ . While the correlation disappears asymptotically, it holds implications in finite samples.

In devising the new instrument, we attempt to keep the correlation in (4.2) equal to zero. The problem stems from  $\hat{\pi}$  being estimated on the full sample which includes the  $i$ th observation. For the new estimator which has the constructed instrument  $Z_i \tilde{\pi}$ ,  $\tilde{\pi}$  is estimated not on the full sample but on the sample with the  $i$ th observation removed. Therefore, the new estimated instrument will be independent of  $\varepsilon_i$  even in finite samples.

The  $i$ th row of the estimated instrument for 2SLS,  $\mathbf{Z}\hat{\pi}$ , where  $\hat{\pi} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})$ , is given by

$$Z_i \hat{\pi} = Z_i (\mathbf{Z}'\mathbf{Z})^{-1} (\mathbf{Z}'\mathbf{X}) \quad (4.3)$$

Remove the  $i$ th row from the matrices of regressors and instruments and denote them by  $\mathbf{X}(i)$  and  $\mathbf{Z}(i)$ . Thus JIVE removes the dependence between  $Z_i \hat{\pi}$  and the regressor  $X_i$ .

The corresponding estimate of  $\pi$  and the constructed instrument will be:

$$\tilde{\pi}(i) = (\mathbf{Z}(i)' \mathbf{Z}(i))^{-1} (\mathbf{Z}(i)' \mathbf{X}(i)) \quad (4.4)$$

$$Z_i \tilde{\pi}(i) = Z_i (\mathbf{Z}(i)' \mathbf{Z}(i))^{-1} (\mathbf{Z}(i)' \mathbf{X}(i) Y) \quad (4.5)$$

$\varepsilon_i$  and  $X_j$  are independent when  $i \neq j$ , so it follows that

$$E[\varepsilon_i Z_i \tilde{\pi}(i)] = E[Z_i (\mathbf{Z}(i)' \mathbf{Z}(i))^{-1} (\mathbf{Z}(i)') E[\mathbf{X}(i) \varepsilon_i | \mathbf{Z}]] = 0 \quad (4.6)$$

Hence, now we have  $E[\varepsilon_i Z_i \tilde{\pi}(i)] = 0$ , so we have removed the correlation between the fitted values and the structural error presented in equation (4.2).

The JIVE estimator is equal to:

$$\hat{\beta}_{JIVE} = (\hat{\mathbf{X}}'_{JIVE} \mathbf{X})^{-1} (\hat{\mathbf{X}}'_{JIVE} \mathbf{Y}) \quad (4.7)$$

where  $\hat{\mathbf{X}}_{JIVE}$  is  $N \times L$  dimensional matrix with the  $i$  th row  $Z_i \tilde{\pi}(i)$ .

Blomquist and Dahlberg [1999] summarize the construction of the JIVE estimator by providing the following algorithm:

---

**Algorithm 1** Jackknife Instrumental Variables Estimator

---

- 1: Use all observations but the  $i$ th to estimate parameters of the first-stage equation.
  - 2: Combine the estimated first stage parameters with the instruments for the  $i$ th observation,  $Z_i$ , to construct a fitted value for the  $i$ th observation  $X_i$ .
  - 3: Repeat steps 1 and 2 for all  $N$  observations.
  - 4: Regress  $Y$  on the fitted values and the exogenous regressors.
- 

The estimator for  $Z_i \pi$ ,  $Z_i \tilde{\pi}(i)$  is consistent.  $\hat{\beta}_{JIVE}$  has the same probability limit and first-order asymptotic distribution as  $\hat{\beta}_{OPT}$  and  $\hat{\beta}_{2SLS}$ , under conventional fixed model asymptotics (Stock et al. [2002]).

## 4.2 Limited Information Maximum-Likelihood Estimator

Limited Information Maximum Likelihood (LIML) is an alternative method to estimate the parameters of the structural equation. It was formalized by Anderson et al. [1949]. The derivation of the LIML estimator is shown in the appendix, from equations (A.13) to (A.19).

We follow the same model as described in Section 2, and extend it as shown below:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon = \mathbf{X}_0\beta_0 + \mathbf{X}_1\beta_1 + \varepsilon \quad (4.8)$$

$$\mathbf{X}_1 = \mathbf{Z}\pi + \eta = \mathbf{Z}_0\pi_0 + \mathbf{Z}_1\pi_1 + \eta \quad (4.9)$$

where  $\mathbf{Z}_0 = \mathbf{X}_0$ .

Here,  $\mathbf{X}_0$  is the matrix of exogenous variables and  $\mathbf{X}_1$  of endogenous variables.  $\mathbf{Z}_0 = \mathbf{X}_0$  is the matrix of included exogenous variables,  $\mathbf{Z}_1$  is the matrix of excluded exogenous variables. Because the LIML estimator is based on the structural equation for  $\mathbf{Y}$  combined with the first-stage equation for  $\mathbf{X}_1$ , it is called ‘limited information’. LIML estimator is given by:

$$\hat{\beta}_{LIML} = (\mathbf{X}'(\mathbf{I} - \hat{\kappa}\mathbf{M}_Z)\mathbf{X})^{-1} \mathbf{X}'(\mathbf{I} - \hat{\kappa}\mathbf{M}_Z)\mathbf{Y} \quad (4.10)$$

---

The LIML estimator has some excellent properties when the number of excluded instruments (number of columns in matrix  $\mathbf{X}_1$ ) in the first-stage equation and the sample size are large. Although the LIML estimator and the 2SLS estimator are asymptotically equivalent in the standard large sample theory, they are quite different in case of many instruments or many weak instruments. It has no finite moments, which implies that its density tends to have very thick tails. Anderson et al. [2010] show that the LIML estimator shows asymptotic optimality with many weak instruments. The LIML estimator is asymptotically efficient in higher order, while the 2SLS estimator is inconsistent, shown by Kunitomo [1987].

# Chapter 5

## Simulation Study

In this chapter, we firstly explore the finite sample behavior of different IV estimators as well as OLS estimator by Monte Carlo simulation. Then, we discuss the performance of all these estimators in 4 models, namely Just-identification with strong IV, Just-identification with weak IV, Over-identification with strong IVs, Over-identification with weak IVs. The experiment presented here follows from the simulations presented in Angrist et al. [1999], Davidson and MacKinnon [2006].

### 5.1 Finite Sample Properties

To maintain simplicity in the design of our experiment, we let the error terms of both reduced-form and structural equation to be homoscedastic, all the regressions to be linear.

The structural equation is defined as follow,

$$Y = \beta_0 \iota + \beta_1 X + \varepsilon \quad (5.1)$$

where  $\iota$  is a vector of ones and  $X$  is the endogenous variable, which is a one-dimensional column vector generated by the reduced-form equation,

$$X = Z\pi + \eta = \pi_0 \iota + \sum_{i=1}^{K-1} z_i \pi_i + \eta \quad (5.2)$$

here, the first column of matrix  $Z$  is also  $\iota$ , the remaining columns  $z_i$  are IID variables with mean zero and unit variance.

Without loss of generality, we set  $\beta_0 = \beta_1 = 1$ ,  $\sigma_\varepsilon^2 = 1$ ,  $\pi_0 = 0$ ,  $\pi_1, \dots, \pi_{K-1}$  to be equal. Under the exogeneity condition,  $Z$  is uncorrelated with  $\varepsilon$ , therefore we know that the correlation between  $X$  and  $\varepsilon$  is only through  $\eta$ . The correlation coefficient between  $\varepsilon$  and  $\eta$  is denoted by  $\rho$  (the covariance between the two is denoted by  $\sigma_{\varepsilon\eta}$ ).

We carry out 3 experiments - we explore how variation in  $\rho$ , the strength of instruments and also the number of instruments impacts the finite sample behavior of the estimators under study. The limiting  $R^2$  of the reduced-form equation, denoted by  $R_\infty^2$ , is given by  $R_\infty^2 = 1/(1 + \sigma_\eta^2 / \|\pi\|^2)$ .

We normalized  $\sigma_\eta^2 + \|\pi\|^2$  to 1 so as to restrict  $R_\infty^2 \in [0, 1]$ .  $R_\infty^2$  is a monotonically increasing function of the ‘Concentration Parameter’, discussed in Section 2. Matrix  $Z$  has  $K$  columns and  $K - 2$



over-identifying restrictions. We vary  $\rho$  and  $R_\infty^2$  at interval of 0.01.

Every experiment is performed with sample size 25, 50, 100, 200, 400, 800. All the experiments are replicated 1000 times. Since LIML and JIVE do not have first and second moments, we employ Median Bias, 0.5 quantile of the estimation minus the true value, to evaluate the central tendency of different estimators.

In the first experiment, we vary  $\rho$ , keep  $K = 7$ ,  $R_\infty^2 = 0.1$  (implies  $\|\pi\|^2 = 0.1, \sigma_\eta^2 = 0.9$ ) constant. Result is shown in Figure B.1. The median bias of OLS, 2SLS and JIVE proportionally increases to  $\rho$  for all sample sizes. For LIML, as the sample size reaches 200 and above, its median bias is negligible.

In the second experiment, we vary  $R_\infty^2$  and keep  $K = 7$ ,  $\rho = 0.9$  fixed. Figure B.2 clearly shows that the growth of  $R_\infty^2$  leads to a decline in the median bias for all four estimators. It is noteworthy that at sample sizes of 100 and above, LIML shows much lower median bias (compared to other estimators) even at small values of  $R_\infty^2$ . JIVE shows an interesting trend, it fluctuates around OLS, while maintaining the same overall trend as OLS - decrease in median bias as  $R_\infty^2$  decreases.

We change the value of  $K$  in the third experiment and keep  $R_\infty^2 = 0.1$ ,  $\rho = 0.9$  fixed. In Figure B.3, except for OLS and LIML, increase in number of instruments  $K$  leads to growth of median bias. At small sample size (50 or lower), the median bias of LIML increases with  $K$ , however, at sizes 100 and above, the median bias appears to be stable around zero.

## 5.2 Performance of the Estimators

We use 4 sets of parameters corresponding to 4 models. The parameter-setting is determined through two hypothesis tests. We employ the first-stage F statistic to detect the strength of instruments and the Anderson-Rubin statistic as a fully robust test for  $\beta_1$  in the two cases with weak instruments. We apply the same model in the previous section but fix sample size to be equal to 200 and replicate 5000 times. To compare estimators, median bias and coverage probability for 95% confidence interval (estimated value plus or minus 1.96 times the asymptotic standard error) are employed. Coverage probability is the proportion of the time that the interval contains the true value, in our case  $\beta_1 = 1$ . Unlike OLS, 2SLS and LIML, the asymptotic standard error of JIVE is calculated based on  $\hat{X}_{JIVE}$ .

- *Model 1: Just-identification with strong instrument*

Let  $\beta_0 = \beta_1 = 1$ ,  $\pi_0 = 0$ ,  $\pi_1 = 0.3$ . Here,  $K = 2$  and

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix} \right)$$

- *Model 2: Just-identification with weak instrument*

---

Let  $\beta_0 = \beta_1 = 1$ ,  $\pi_0 = 0$ ,  $\pi_1 = 0.2$ . Here,  $K = 2$  and

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{pmatrix} \right)$$

- *Model 3: Over-identification with strong instruments*

Let  $\beta_0 = \beta_1 = 1$ ,  $\pi_0 = 0$ ,  $\pi_1 = \pi_2 = \dots = \pi_{15} = 0.3$ . Here,  $K = 16$  and

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & 0.10 \\ 0.10 & 0.25 \end{pmatrix} \right)$$

- *Model 4: Over-identification with weak instruments*

Let  $\beta_0 = \beta_1 = 1$ ,  $\pi_0 = 0$ ,  $\pi_1 = \pi_2 = \dots = \pi_{15} = 0.1$ . Here,  $K = 16$  and

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix} \right)$$

Figures B.4 to B.7 present the distribution of  $\beta_0$  and  $\beta_1$  in Models 1-4 respectively. In Model 1, the OLS estimator of  $\beta_1$  is notably biased. In Model 2, the OLS estimator of  $\beta_1$  is even more biased. Besides, the 2SLS, LIML and JIVE display a very wide range of estimates for both  $\beta_0$  and  $\beta_1$ . Since we are more interested in the slope coefficient ( $\beta_1$ ), we generate Figures B.8 to B.11 to look at  $\beta_1$  in more detail. After we exclude some outliers and focus on the interval  $[-3, 5]$ , in Figure B.9 we observe notable negative skew in 2SLS, LIML and JIVE estimators.

In figure B.6, corresponding to Model 3, we see that  $\beta_1$  of the OLS estimator shows a remarkable reduction of bias compared to Model 2. In addition, for all estimators, the distribution of  $\beta_1$  is concentrated on a small range which contains the true value. However, in Model 4, the bias of OLS estimate of increases again and the distribution of LIML and JIVE estimates disperses slightly. We include a series of quantiles around  $\beta_1$  in Table B.2 to display the dispersion. The JIVE estimator returns a surprisingly large number of outliers on both tails in the just-identified with weak IV model. If we had a smaller number of replications, then in spite of the high coverage probability, we might have seen some extreme JIVE estimates, quite far from the true value.

According to the results of our MC simulations, we find that the LIML estimator performs well (in terms of median bias) in most of models, despite the fact that JIVE has the highest coverage probability in all the models. We observe that JIVE does not dominate LIML in any case/model. Between 2SLS and JIVE, we do not observe JIVE performing uniformly better than 2SLS in the models we considered. While we do observe JIVE do well in overidentified case as discussed in Angrist et al. [1999], however, our findings are more in line with results of Davidson and MacKinnon [2006].

## Chapter 6

# Application to Returns to Schooling

In this section, we revisit the famous paper and its ('provocative' as termed by Bound et al. [1995]) results that led to the beginning of the weak instruments literature: Angrist and Krueger [1991], who use quarter of birth as an instrument for estimating the impact of educational attainment on earnings. On the same data, we confirm the weakness of the instruments used in the paper using the first-stage F-statistic and proceed to apply estimators more robust to weak instruments.

Endogeneity of education is a well-known problem that economists face while estimating the effect of education on earnings. The reason for the endogeneity is omitted variables, such as the 'ability', which can be correlated with both educational attainment and earnings of an individual. As a result, OLS will give biased estimates for the return to education. To correct this bias quarter of birth is included in the regression as an instrumental variable. Association between the quarter of birth and schooling is explained by compulsory schooling requirements in the United States. According to school start age policy, children are required to enter school in the fall of the calendar year in which they turn 6. While compulsory schooling laws allow students to leave school after they turned 16. As a result, students who were born earlier in the calendar year tend to attend school for a shorter period of time than those born at the end of the year. So interaction of the two requirements of schooling laws generates variation in educational attainment for the students who graduate right after their 16th birthday.

Quarter of birth must satisfy the instrument relevance and exogeneity conditions to be a valid instrument for educational attainment. The relevance condition implies that the instrument must be correlated with the endogenous variable, in this case with the years of schooling. Higher the correlation between these two variables, stronger is the first stage in the two-stage least squares estimation. To satisfy the second condition quarter of birth must effect earnings only via its effect on schooling years. In general, it is not possible to test this condition statistically. Angrist and Krueger argue that student's birthday can not be correlated with other personal features which may affect earnings, thus the variation in education due to the individual's birthday is exogenous (Angrist and Krueger [1991]).

For our estimations, we used the dataset from Angrist and Krueger [1991] which is taken from 1980 US Census. The sample consists of 329,509 men, who were born in 1930-1939. The dataset includes information on the quarter of birth, year of birth, state of birth, years of schooling and earn-

ings for this sample. All figures and tables are presented in Appendix B.

Figure B.12 shows the relation between quarter of birth and schooling (first-stage). The graph indicates that there is an upward trend in average years of schooling for men born in 1930-1939. There is also a persistent seasonal pattern in education. Men born at the beginning of the year tend to have less schooling on average than those who were born later in the year.

Figure B.13 illustrates the reduced form, which is the relationship between the quarter of birth and wages. Here we also notice a pattern where the 3rd and 4th quarter of births correspond to a higher log weekly wages.

Our general model is given by:

$$E_i = \sum_c Y_{ic} \delta_c + \sum_c \sum_j Y_{ic} Q_{ij} \Theta_{jc} + \epsilon_i \quad (6.1)$$

$$\ln W_i = \sum_c Y_{ic} \xi_c + \rho E_i + \mu_i \quad (6.2)$$

Here  $E_i$  is the schooling of  $i$ th individual,  $Y_{ic}$  is a dummy variable, denoting if the  $i$ th individual was born in  $c$ th year,  $Q_{ij}$  is a dummy variable showing quarter of birth of the  $i$ th individual and  $W_i$  is the weakly wage. In Table B.3 we present results of the first stage estimations. The first and third columns of the table indicate that individuals born in the last quarter of the year had about 0.10 year more schooling compared to the men born in the first to third quarters. The second, fourth and fifth columns show the estimates of each quarter of birth in comparison with the first quarter. Naturally, the largest difference is between the last and the first quarter, which is around 0.15 year, independent of including year of birth and state of birth as control variables in the regression.

Table B.4 attempts to replicate the main results of Table 2 from Angrist et al. [1999]. We compare 2SLS, LIML and JIVE estimators for three different specifications of the model based on Angrist and Krueger [1991] data. Additionally, we report the first-stage F-statistic and  $R^2$  values.

Column 1 shows the results of our estimation wherein we regress wages on schooling, including three quarter of birth dummies as instruments and nine year of birth dummies as control variables. Here, our first stage F-statistic is greater than the critical value (for the case of  $K = 3$  where  $K$  is the number of instruments) of 9.08, according to Table B.1 provided in Appendix B. Hence in the first model the excluded instruments are strong.

Moving to column 2, to control for the age-related trends we include interactions of the year of birth dummies with quarter of birth dummies as instruments in the second model. This leads to a slightly lower 2SLS estimate as well as a slight decrease in standard errors. However, the value of the first-stage F-statistic reduces dramatically to 4.91. The small value of the F-statistic indicates that the excluded instruments and educational attainment are only weakly correlated, which as we discussed in section 2.4.2 can lead to finite-sample bias in the 2SLS estimate. Looking at the  $R^2$  in columns

---

(1) and (2), we notice that compared to the case of the first column, the explanatory power of the instruments does not increase much in the second specification.

Finally, in the last specification (column 3), we increase the number of instruments to 180, taking year of birth  $\times$  quarter of birth, and state of birth  $\times$  quarter of birth interactions as instruments. The reasoning behind increasing the number of instruments by including interaction terms is to ensure that seasonal differences do not vary by state and birth year. While this results in about 40 percent reduction in standard errors compared to the column (2), however, the value of the first-stage F-statistic again decreases, compared to the preceding model. This indicates that while attempting to increase the precision of the 2SLS estimates, the weakness of the instruments, however, leads to the 2SLS estimates becoming biased.

In Table B.4 we show the estimates of the alternative estimators discussed previously (LIML and JIVE) which are considered to have better finite sample properties when the instruments are weak. As we can see, JIVE estimates vary considerably from the other two.

Given the encouraging performance of the LIML estimator in our simulation study, and considering the closeness of the LIML and 2SLS estimates in our application to returns to schooling, we suspect that in this specific case, LIML and 2SLS probably give more reliable results than JIVE. However, we do not at all recommend using 2SLS estimator in case of weak instruments in general, considering the theoretical discussion and simulation results presented previously.

## Chapter 7

### Conclusion

In our study, we hope to have provided the reader a concise but comprehensive overview of selected aspects of the weak instruments literature.

When we face the issue of weak instruments, the best solution is of course to find better, stronger instruments. However, in empirical practice this is easier said than done. Hence, we think research into alternative estimators which can give more reliable estimates than the two-stage least squares estimator in the weak instruments case, is very relevant. From our small study, we find the LIML estimator to perform the best when the correlation between the instrument and the endogenous explanatory variable is low. We posit the LIML be a possible solution in the (fairly common) case that a researcher has a weak instrument(s) and is not in a position to find other (stronger) instruments. Development of methods for robust inference of weak instruments such as the Anderson-Rubin statistic is a very promising area of further research.

# Appendix A

## Formulae

### A.1 Deriving the approximate expression of 2SLS Bias

This derivation has been adopted from Pischke [2018]. Start with the representation of the 2SLS estimator as shown in equation (2.8).

$$\hat{\beta}_{2SLS} = (\mathbf{X}'P_Z\mathbf{X})^{-1}\mathbf{X}'P_Z\mathbf{Y} = \beta + (\mathbf{X}'P_Z\mathbf{X})^{-1}\mathbf{X}'P_Z\varepsilon \quad (\text{A.1})$$

where  $P_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$  is the projection matrix. So the bias of  $\hat{\beta}_{2SLS}$  will be

$$\hat{\beta}_{2SLS} - \beta = (\mathbf{X}'P_Z\mathbf{X})^{-1}(\pi'\mathbf{Z}' + \eta')P_Z\varepsilon = (\mathbf{X}'P_Z\mathbf{X})^{-1}\pi'\mathbf{Z}'\varepsilon + (\mathbf{X}'P_Z\mathbf{X})^{-1}\eta'P_Z\varepsilon \quad (\text{A.2})$$

Using group asymptotics the expectation of this expression can be presented as:

$$E[\hat{\beta}_{2SLS} - \beta] \approx (E[\mathbf{X}'P_Z\mathbf{X}])^{-1}E[\pi'\mathbf{Z}'\varepsilon] + (E[\mathbf{X}'P_Z\mathbf{X}])^{-1}E[\eta'P_Z\varepsilon] \quad (\text{A.3})$$

$Z_i$  instruments are uncorrelated with  $\varepsilon_i$  and  $\eta_i$ , so  $E[\pi'\mathbf{Z}'\varepsilon] = 0$  and we will have

$$E[\hat{\beta}_{2SLS} - \beta] \approx (E[\mathbf{X}'P_Z\mathbf{X}])^{-1}E[\pi'\mathbf{Z}'\varepsilon] + (E[\mathbf{X}'P_Z\mathbf{X}])^{-1}E[\eta'P_Z\varepsilon] = (E[\mathbf{X}'P_Z\mathbf{X}])^{-1}E[\eta'P_Z\varepsilon] \quad (\text{A.4})$$

Substituting the first stage equation  $\mathbf{X} = \mathbf{Z}\pi + \eta$  we have

$$E[\hat{\beta}_{2SLS} - \beta] \approx (E[(\pi'\mathbf{Z}' + \eta')P_Z(\mathbf{Z}\pi + \eta)])^{-1}E[\eta'P_Z\varepsilon] \quad (\text{A.5})$$

We have that  $E[\pi'\mathbf{Z}'\eta] = 0$ , so

$$E[\hat{\beta}_{2SLS} - \beta] \approx [E(\pi'\mathbf{Z}'\mathbf{Z}\pi) + E(\eta'P_Z\eta)]^{-1}E(\eta'P_Z\varepsilon) \quad (\text{A.6})$$

Notice that  $\eta'P_Z\eta$  is a scalar and is equal to its trace.  $P_Z$  is an idempotent matrix, so its trace is equal to its rank,  $Q$ . So

$$E(\eta'P_Z\eta) = E[\text{tr}(\eta'P_Z\eta)] = E[\text{tr}(P_Z\eta\eta')] = \text{tr}(P_ZE[\eta\eta']) = \text{tr}(P_Z\sigma_\eta^2 I) = \sigma_\eta^2 \text{tr}(P_Z) = \sigma_\eta^2 Q \quad (\text{A.7})$$

---

With a similar technique we can show that  $E(\eta' P_Z \varepsilon)$  is equal to  $\sigma_{\varepsilon\eta} Q$ . Substituting these results in equation (A.6) we have

$$E[\hat{\beta}_{2SLS} - \beta] \approx \sigma_{\varepsilon\eta} Q [E(\pi' \mathbf{Z}' \mathbf{Z} \pi) + \sigma_\eta^2 Q]^{-1} = \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2} \left[ \frac{E(\pi' \mathbf{Z}' \mathbf{Z} \pi)/Q}{\sigma_\eta^2} + 1 \right]^{-1} \quad (\text{A.8})$$

The population F-statistic for the first stage regression is the following

$$F = \frac{E(\pi' \mathbf{Z}' \mathbf{Z} \pi)/Q}{\sigma_\eta^2} \quad (\text{A.9})$$

So (A.8) can be expressed as

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2} \frac{1}{F + 1} \quad (\text{A.10})$$

Assume that the  $\pi$  coefficients are zero and  $F = 0$ . In this case  $\sigma_X^2 = \sigma_\eta^2$  and

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\varepsilon\eta}}{\sigma_X^2} \quad (\text{A.11})$$

Thus, when  $\pi \neq 0$  and F is small, then 2SLS will be biased towards OLS.

Note that  $\frac{\sigma_{\varepsilon\eta}}{\sigma_X^2}$  is also the bias of OLS estimator, because when  $\pi = 0$ ,  $\text{cov}(\varepsilon_i, X_i) = \sigma_{\varepsilon\eta}$  and

$$\beta_{OLS} = \frac{\text{cov}(Y_i, X_i)}{\text{var}(X_i)} = \frac{\text{cov}(\beta X_i + \varepsilon_i, X_i)}{\text{var}(X_i)} = \beta + \frac{\text{cov}(\varepsilon_i, X_i)}{\text{var}(X_i)} = \frac{\sigma_{\varepsilon\eta}}{\sigma_X^2} \quad (\text{A.12})$$

Hence in this case OLS and 2SLS estimators on average are the same. And if  $\pi$  is different from zero,  $\hat{\beta}_{2SLS}$  will be biased in the direction of OLS estimator. If we add weak instruments to the regression bias of 2SLS will only increase.



## A.2 Deriving the LIML Estimator

LIML Estimator Derivation Hansen [2019] Davidson et al. [203]:

Structural Equation:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon = \mathbf{X}_0\beta_0 + \mathbf{X}_1\beta_1 + \varepsilon \quad (\text{A.13})$$

Reduced-form Equation:

$$\mathbf{X}_1 = \mathbf{Z}\pi + \eta = \mathbf{Z}_0\pi_0 + \mathbf{Z}_1\pi_1 + \eta \quad (\text{A.14})$$

Here,  $\mathbf{X}_0$  is the matrix of exogenous variables and  $\mathbf{X}_1$  of endogenous variables.  $\mathbf{Z}_0 = \mathbf{X}_0$  is the matrix of included exogenous variables,  $\mathbf{Z}_1$  is the matrix of excluded exogenous variables.

The LIML estimate of  $\beta_1$  in (A.13) is given by minimize the ratio:

$$\kappa \equiv \frac{(\mathbf{Y} - \mathbf{X}_1\beta_1)' \mathbf{M}_{\mathbf{X}_0} (\mathbf{Y} - \mathbf{X}_1\beta_1)}{(\mathbf{Y} - \mathbf{X}_1\beta_1)' \mathbf{M}_{\mathbf{Z}} (\mathbf{Y} - \mathbf{X}_1\beta_1)} \equiv \frac{\gamma' \mathbf{Y}'_* \mathbf{M}_{\mathbf{X}_0} \mathbf{Y}_* \gamma}{\gamma' \mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_* \gamma} \quad (\text{A.15})$$

where  $\mathbf{Y}_* \equiv [\mathbf{Y} \ \mathbf{X}_1]$ ,  $\gamma = [1 \ \vdots \ -\beta_1]$ ,  $\mathbf{M}_{\mathbf{Z}} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ ,  $\mathbf{M}_{\mathbf{X}_0} = \mathbf{I} - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'$ .

The first order conditions obtained by differentiating the rightmost expression in (A.15) with respect to  $\gamma$  are

$$2\mathbf{Y}'_* \mathbf{M}_{\mathbf{X}_0} \mathbf{Y}_* \gamma (\gamma' \mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_* \gamma) - 2\mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_* \gamma (\gamma' \mathbf{Y}'_* \mathbf{M}_{\mathbf{X}_0} \mathbf{Y}_* \gamma) = 0 \quad (\text{A.16})$$

Premultiplying (A.4) by  $(\mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_*)^{-1/2}$  and inserting that factor multiplied by its inverse before  $\gamma$ , we yield

$$((\mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_*)^{-1/2} \mathbf{Y}'_* \mathbf{M}_{\mathbf{X}_0} \mathbf{Y}_* (\mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_*)^{-1/2} - \kappa \mathbf{I}) \gamma_* = 0 \quad (\text{A.17})$$

where  $\gamma_* \equiv (\mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_*)^{1/2} \gamma$  and  $\gamma$  is an eigenvalue of  $(\mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_*)^{-1/2} \mathbf{Y}'_* \mathbf{M}_{\mathbf{X}_0} \mathbf{Y}_* (\mathbf{Y}'_* \mathbf{M}_{\mathbf{Z}} \mathbf{Y}_*)^{-1/2}$ .

$\hat{\kappa}$  is the smallest eigenvalue, because  $\hat{\kappa}$  is the minimum value of (A.15).

The LIML estimator of  $\beta$  is defined by estimating equation

$$\mathbf{X}'(\mathbf{I} - \hat{\kappa} \mathbf{M}_{\mathbf{Z}})(\mathbf{Y} - \mathbf{X}\hat{\beta}_{LIML}) = 0 \quad (\text{A.18})$$

Once  $\hat{\kappa}$  has been computed, we find that

$$\hat{\beta}_{LIML} = (\mathbf{X}'(\mathbf{I} - \hat{\kappa} \mathbf{M}_{\mathbf{Z}})\mathbf{X})^{-1} \mathbf{X}'(\mathbf{I} - \hat{\kappa} \mathbf{M}_{\mathbf{Z}})\mathbf{Y} \quad (\text{A.19})$$

## **Appendix B**

### **Tables & Figures**

Table B.1: Selected Critical Values for Weak Instrument Tests for TSLS Based on the First-stage F statistic. Source: Stock and Yogo [2002], table 1.

number of IV (K)	Relative bias > 10%	
	Threshold $\mu^2/K$	F statistic critical value 5%
3	3.71	9.08
5	5.82	10.83
10	7.41	11.49
15	7.94	11.51

Table B.2: Performance of the estimators under Just-identification with strong IV, Just-identification with weak IV, Over-identification with strong IVs, Over-identification with weak IVs

Model 1	Quantiles around $\beta_1$					Median Bias	Coverage Prob. <i>C.I.</i> <sub>.95%</sub>
	0%	25%	50%	75%	100%		
OLS	0.43	0.56	0.59	0.62	0.75	0.587	0.000
2SLS	-0.58	-0.08	0.00	0.07	0.37	0.000	0.536
LIML	-1.39	-0.09	0.00	0.07	0.31	0.000	0.539
JIVE	-0.74	-0.11	-0.02	0.06	0.36	-0.023	0.999
Model 2	Quantiles around $\beta_1$					Median Bias	Coverage Prob. <i>C.I.</i> <sub>.95%</sub>
	0%	25%	50%	75%	100%		
OLS	0.74	0.84	0.86	0.89	0.98	0.864	0.000
2SLS	-242.30	-0.30	0.00	0.20	80.93	0.000	0.141
LIML	-404.49	-0.23	0.06	0.21	1531.56	0.059	0.145
JIVE	-628.81	-0.74	-0.16	0.15	2182.60	-0.160	0.846
Model 3	Quantiles around $\beta_1$					Median Bias	Coverage Prob. <i>C.I.</i> <sub>.95%</sub>
	0%	25%	50%	75%	100%		
OLS	-0.05	0.04	0.06	0.08	0.16	0.063	0.370
2SLS	-0.10	-0.01	0.00	0.03	0.11	0.005	0.925
LIML	-0.11	-0.02	0.00	0.02	0.11	0.000	0.921
JIVE	-0.11	-0.02	0.00	0.02	0.11	0.000	0.999
Model 4	Quantiles around $\beta_1$					Median Bias	Coverage Prob. <i>C.I.</i> <sub>.95%</sub>
	0%	25%	50%	75%	100%		
OLS	0.33	0.47	0.50	0.53	0.66	0.500	0.000
2SLS	-0.24	0.03	0.08	0.14	0.32	0.085	0.477
LIML	-0.61	-0.07	0.00	0.02	0.11	0.000	0.921
JIVE	-0.61	-0.08	-0.01	0.05	0.28	-0.014	0.996

Table B.3: First stage regression results(dependent variable: Years of schooling

Regressor	(1)	(2)	(3)	(4)	(5)
quarter 2		0.057 (0.016)		0.056 (0.016)	0.045 (0.016)
quarter 3		0.117 (0.016)		0.113 (0.016)	0.11 (0.016)
quarter 4	0.092 (0.013)	0.151 (0.016)	0.09 (0.013)	0.148 (0.016)	0.156 (0.038)
9 year of birth dummies			✓	✓	✓
50 state of birth dummies					✓

Table B.4: 2SLS, LIML and JIVE estimates of the economic returns to schooling

	(1)	(2)	(3)
2SLS	0.105 (0.020)	0.089 (0.016)	0.093 (0.009)
LIML	0.106 (0.020)	0.093 (0.018)	0.106 (0.012)
JIVE	0.467 (0.001)	0.472 (0.001)	0.499 (0.003)
F-statistic (first stage)	32.27	4.91	2.58
$R^2$ (first stage, $\times 100$ )	0.029	0.044	0.14
Adjusted $R^2$ (first stage, $\times 100$ )	0.028	0.036	0.086
<i>Controls</i>			
Year of birth	✓	✓	✓
State of birth			✓
<i>Excluded instruments</i>			
Quarter-of-birth dummies	✓		
Quarter of birth*year of birth		✓	✓
Quarter of birth*state of birth			✓
Number of instruments	3	30	180

Figure B.1:  $R_{\infty}^2 = 0.1$ ,  $K = 7$

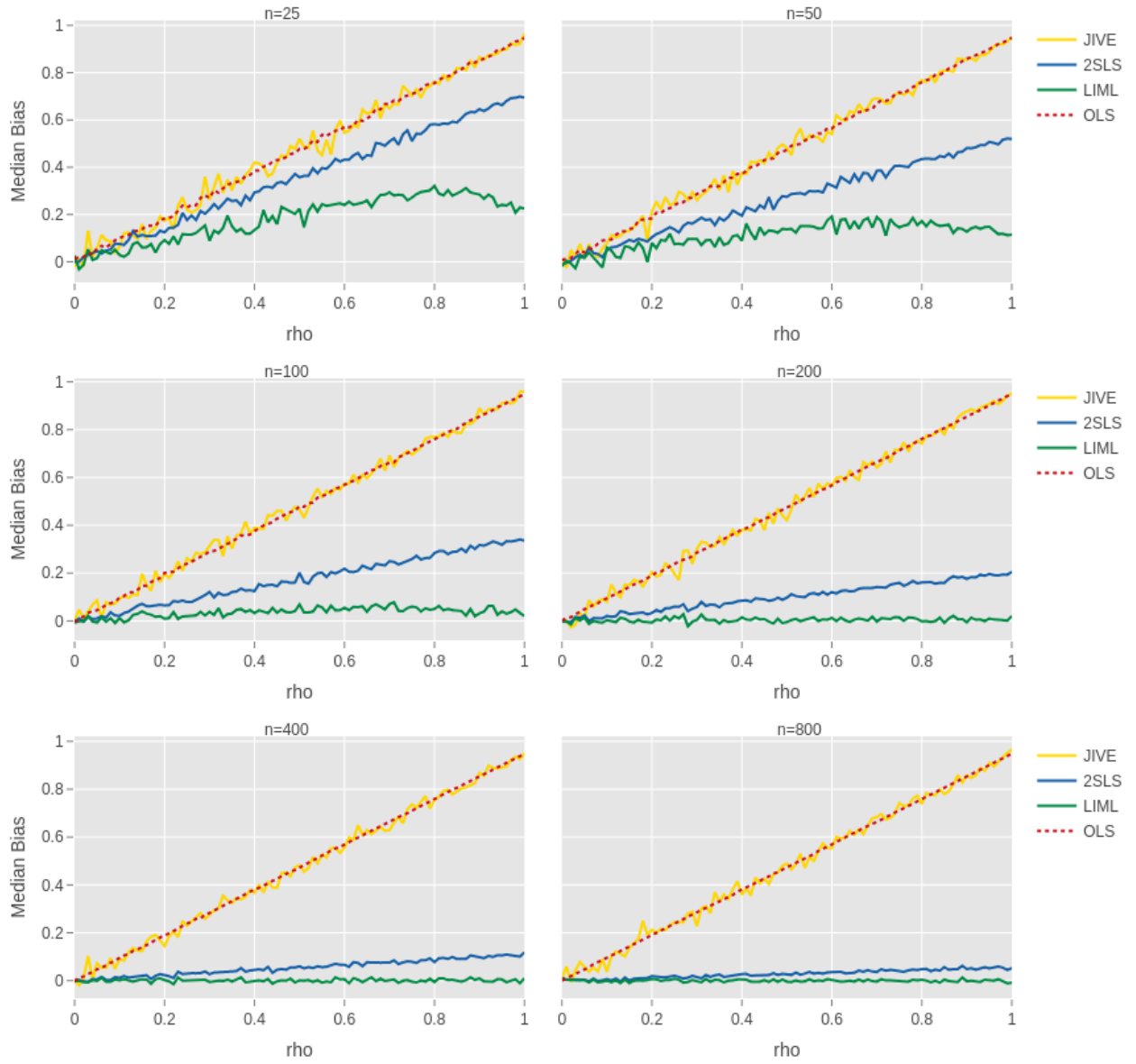


Figure B.2:  $\rho = 0.9$ ,  $K = 7$

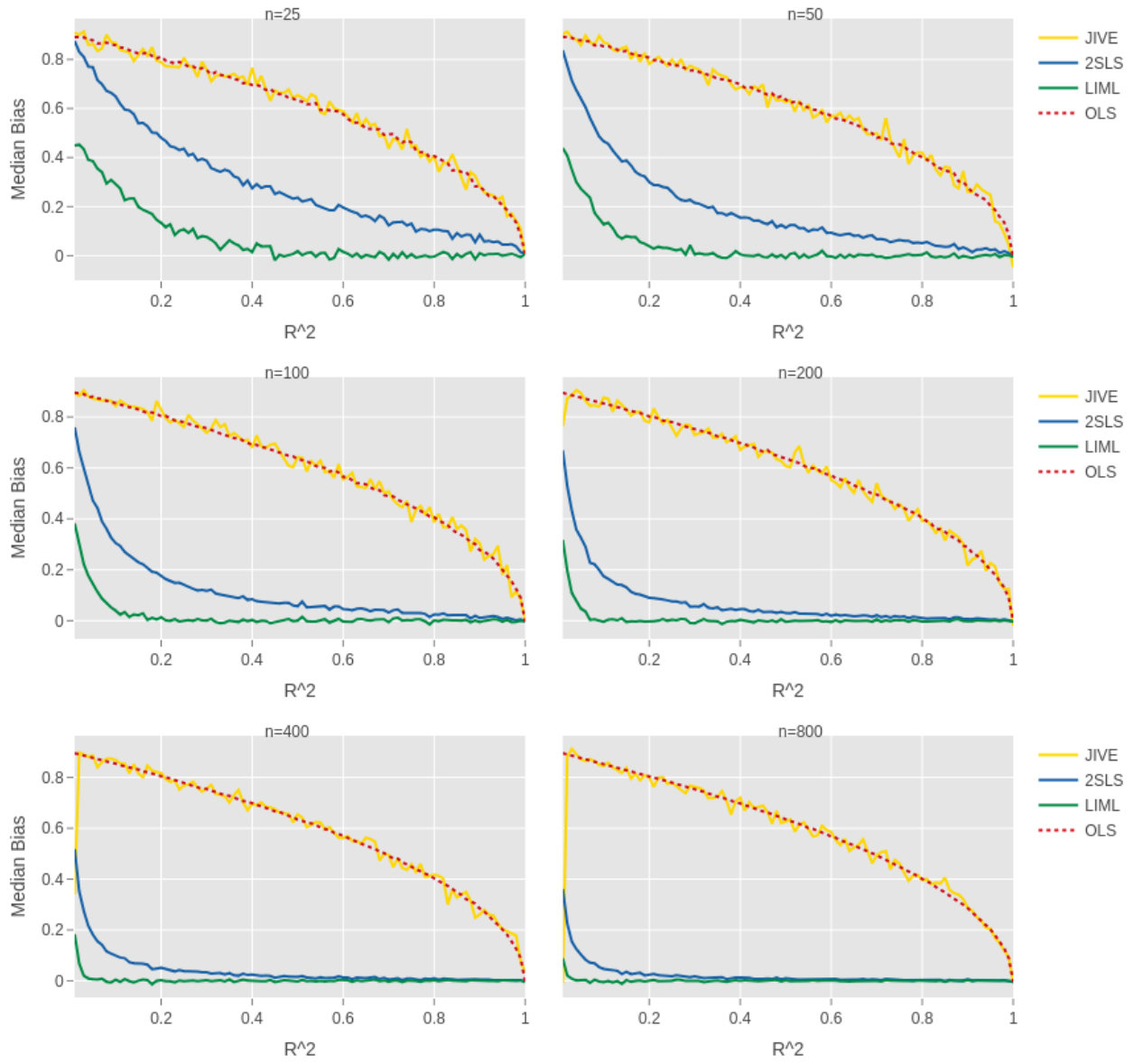


Figure B.3:  $R_{\infty}^2 = 0.1, \rho = 0.9$

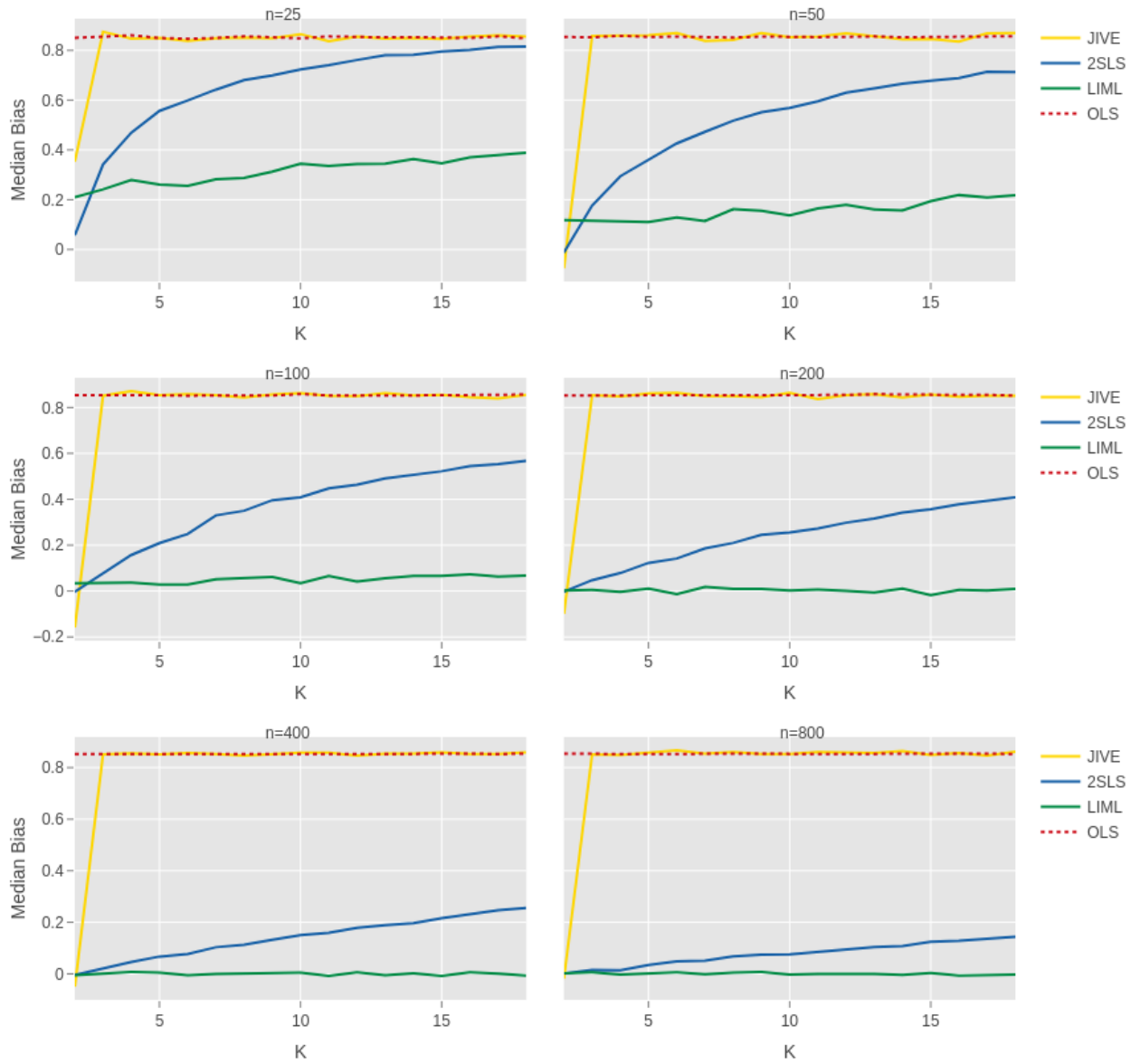


Figure B.4: Joint distribution of  $\beta$  in Model 1.

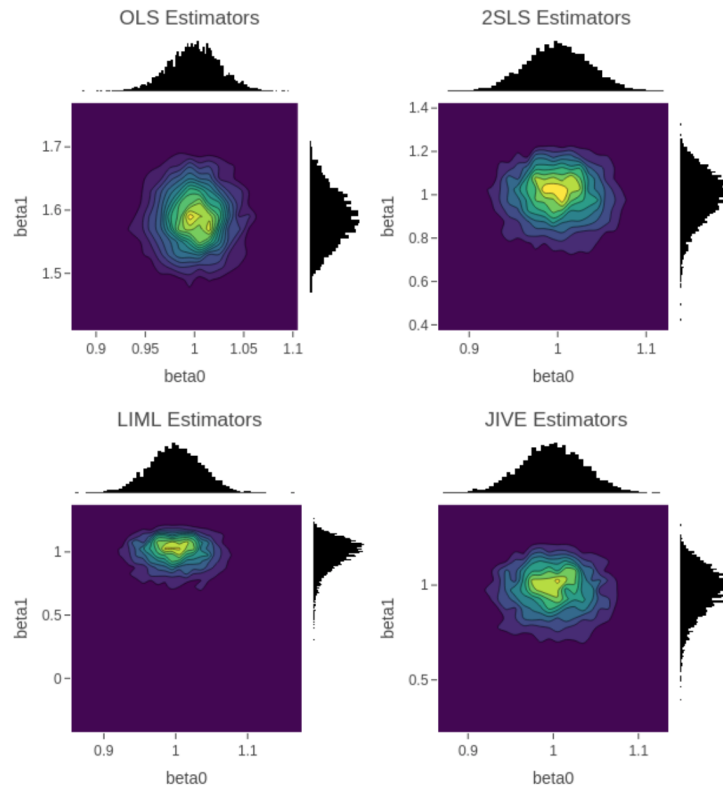


Figure B.5: Joint distribution of  $\beta$  in Model 2.

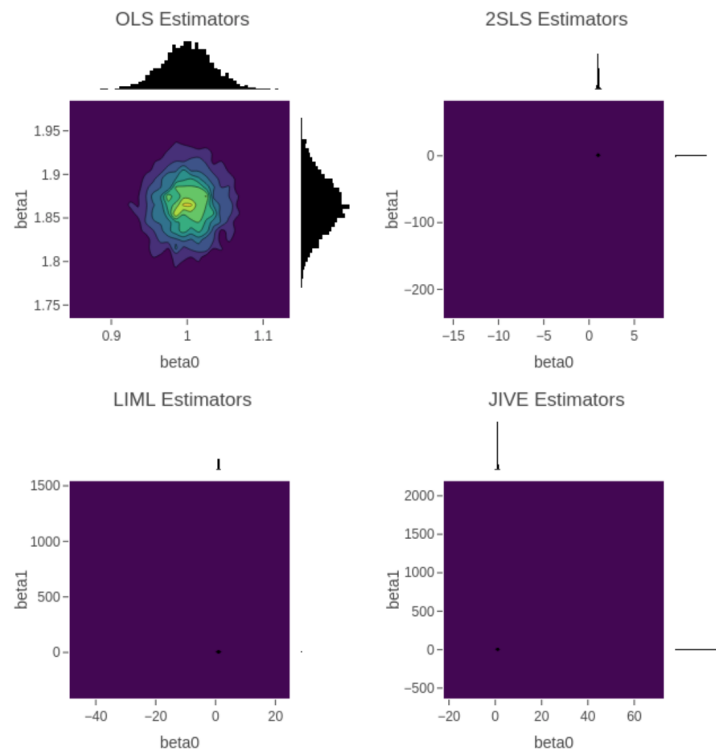




Figure B.6: Joint distribution of  $\beta$  in Model 3.

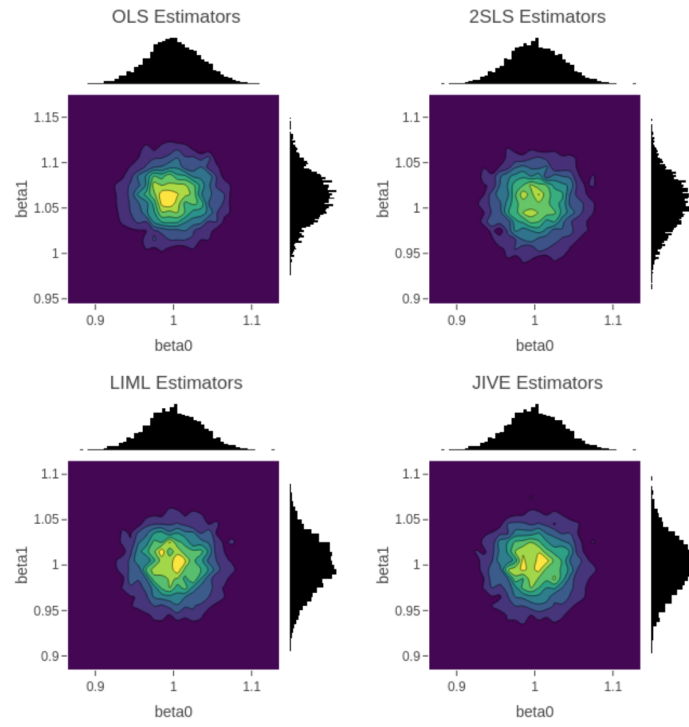


Figure B.7: Joint distribution of  $\beta$  in Model 4.

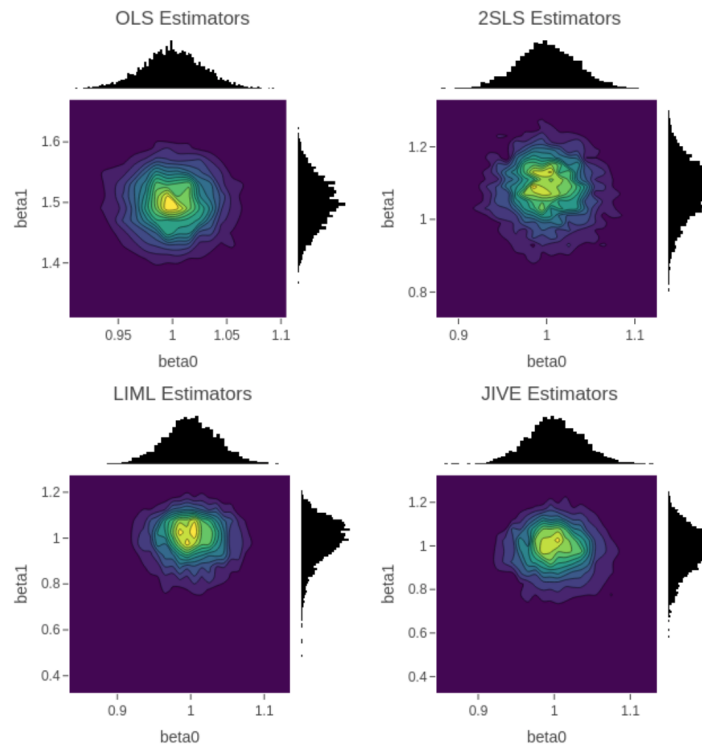


Figure B.8: Distribution of  $\beta_1$  in Model 1.

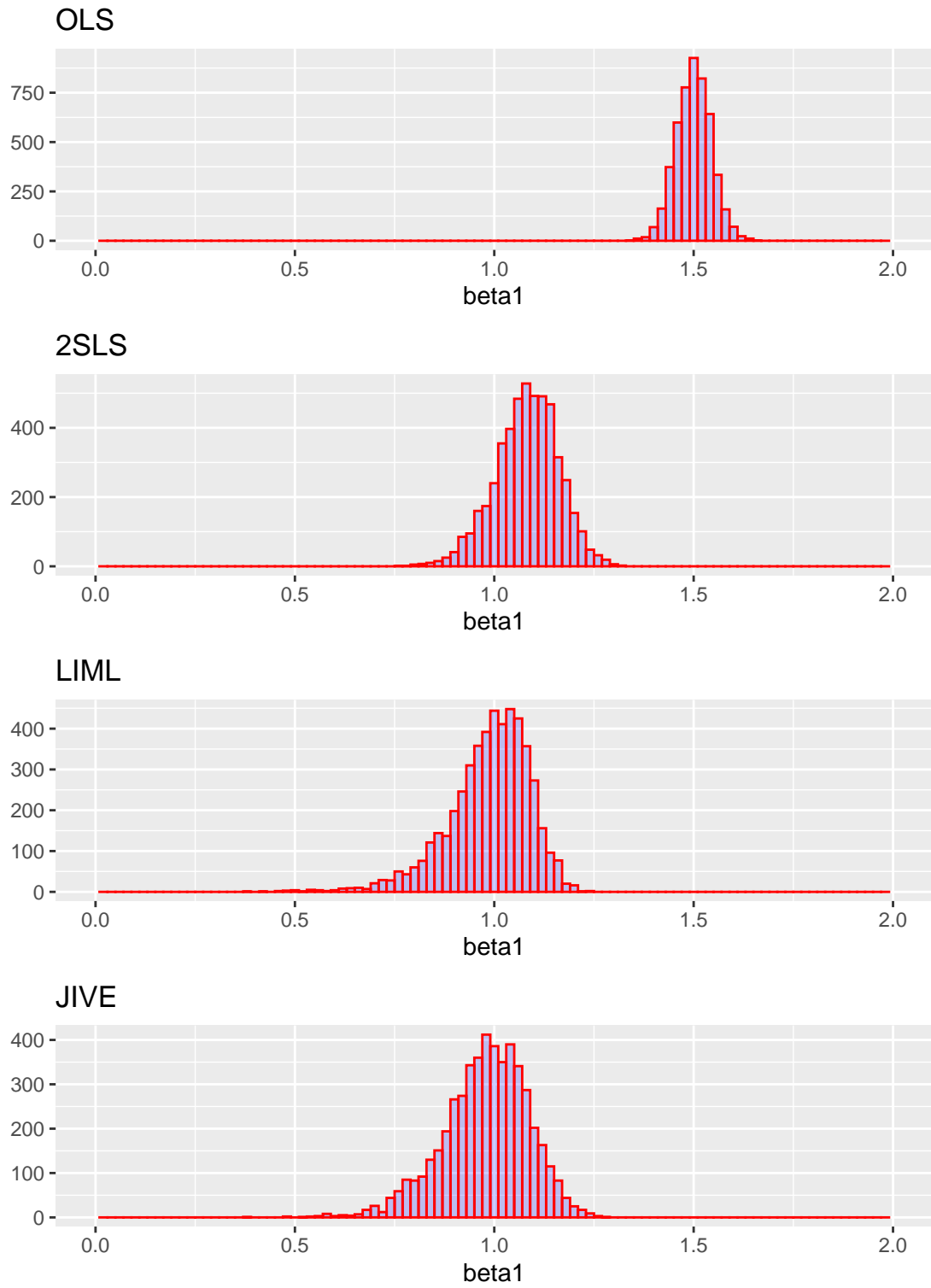


Figure B.9: Distribution of  $\beta_1$  in Model 2.

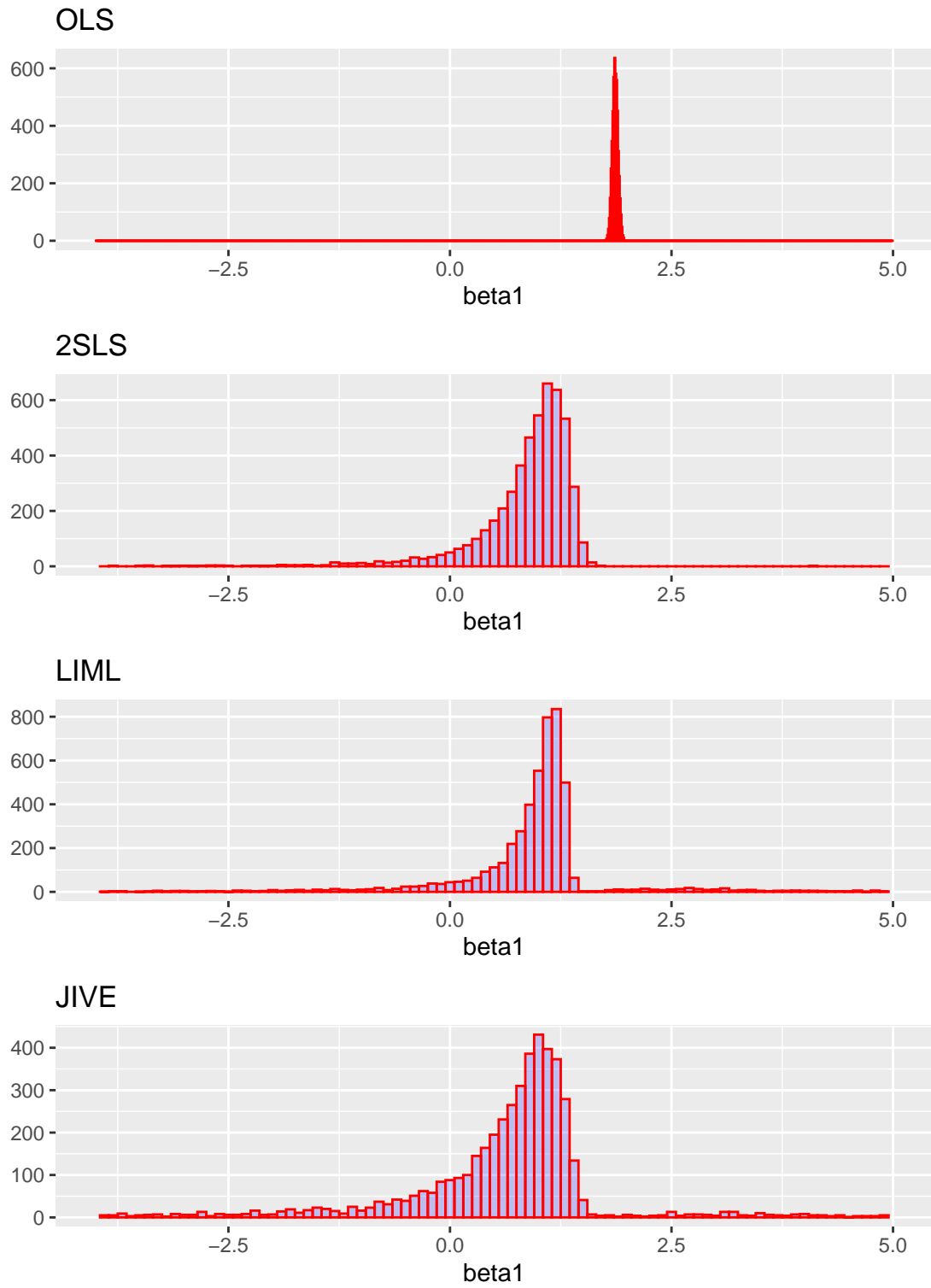


Figure B.10: Distribution of  $\beta_1$  in Model 3.

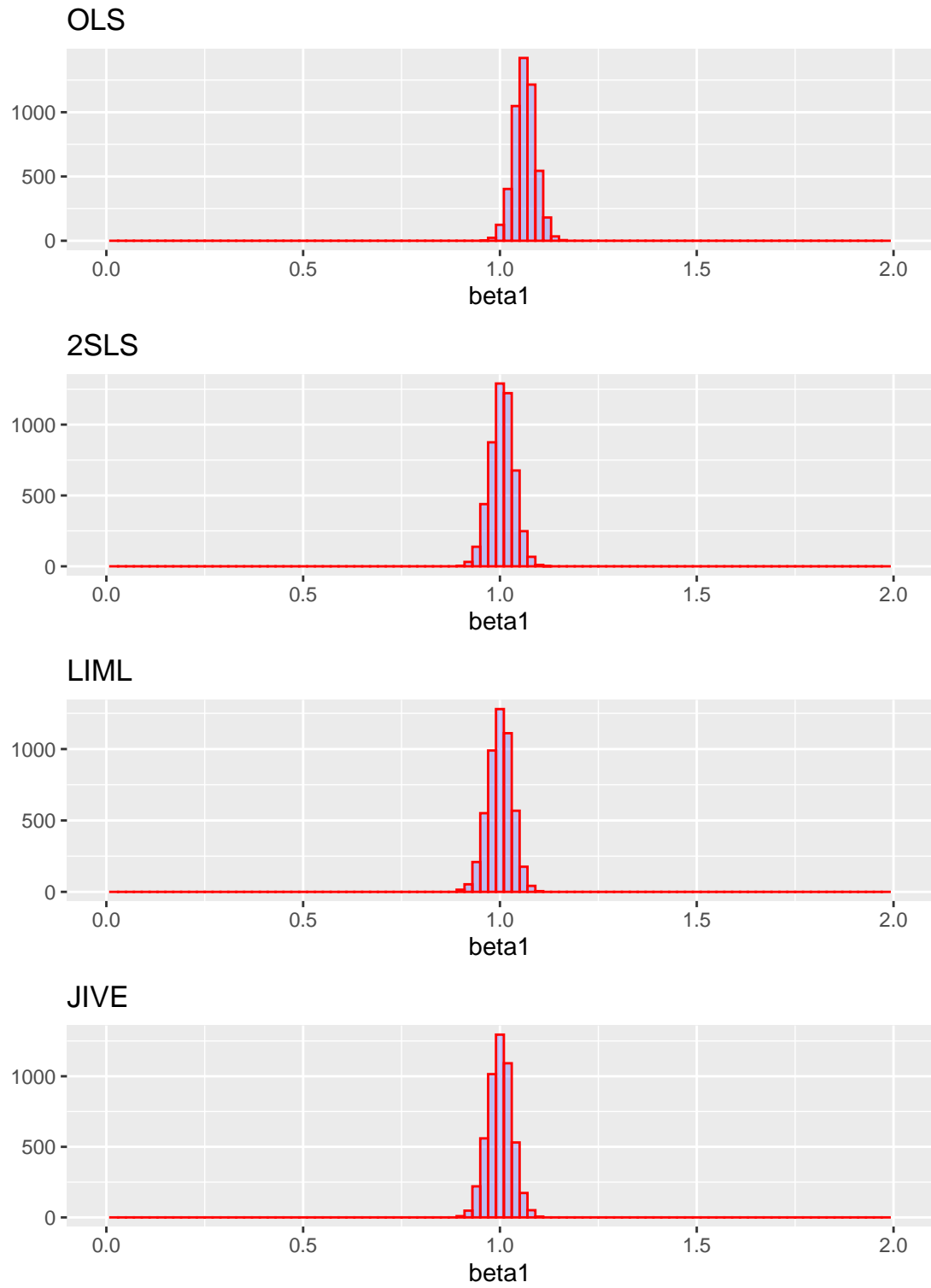


Figure B.11: Distribution of  $\beta_1$  in Model 4.

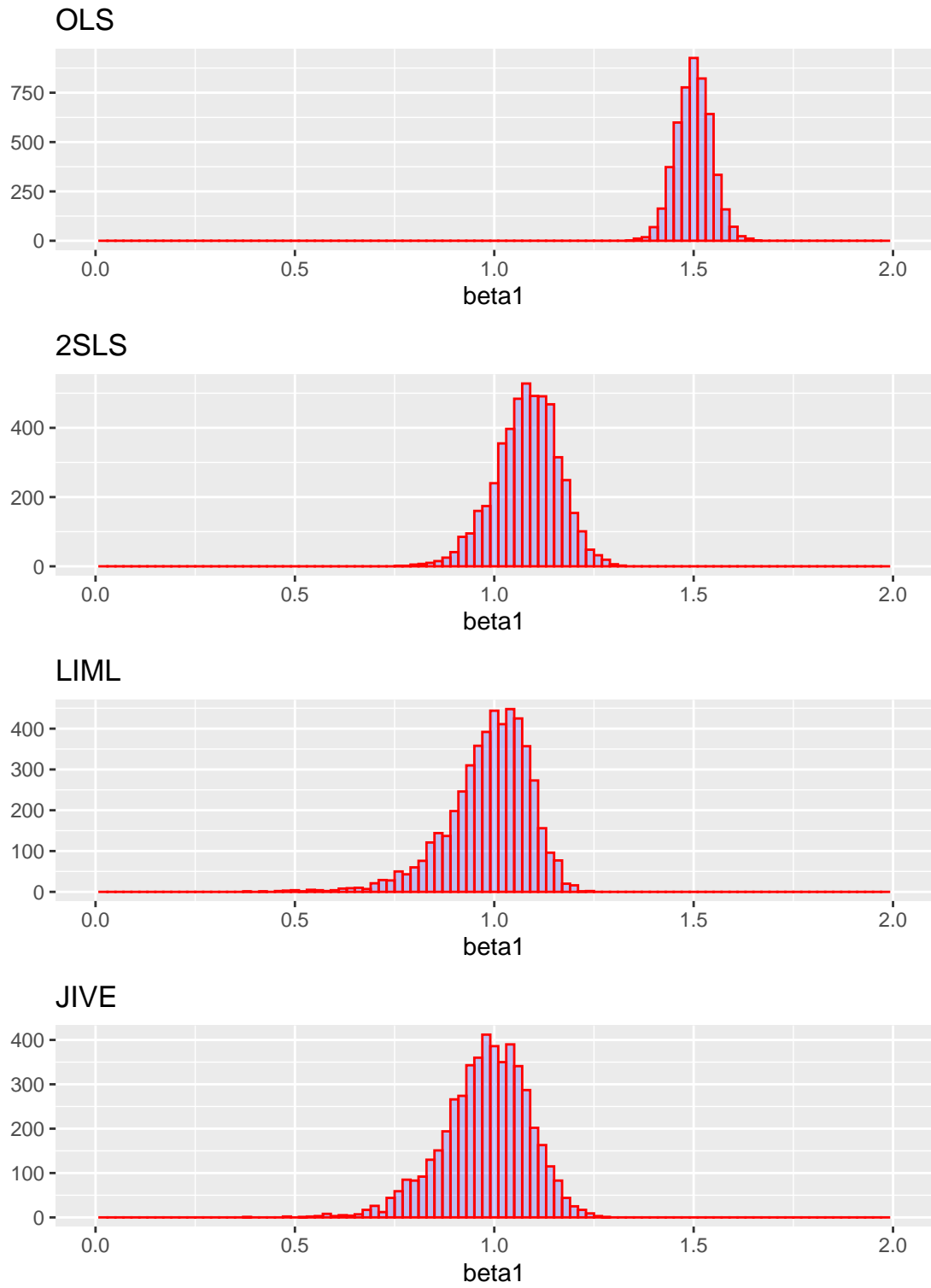


Figure B.12: Years of Completed Education and Quarter of Birth

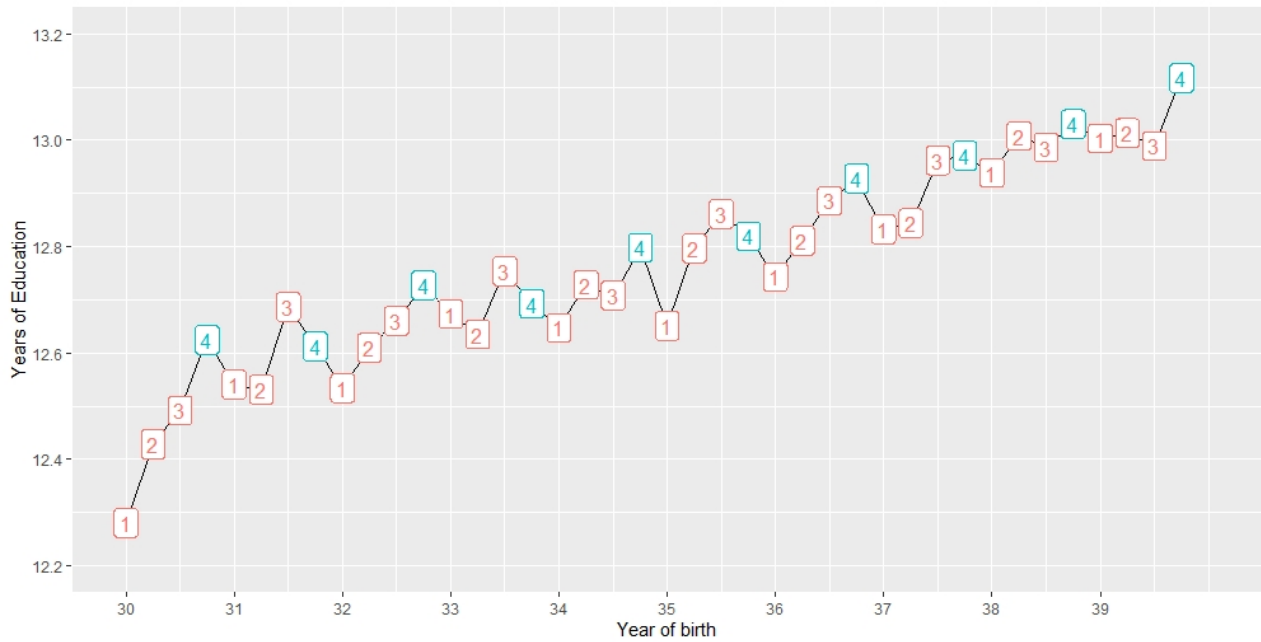
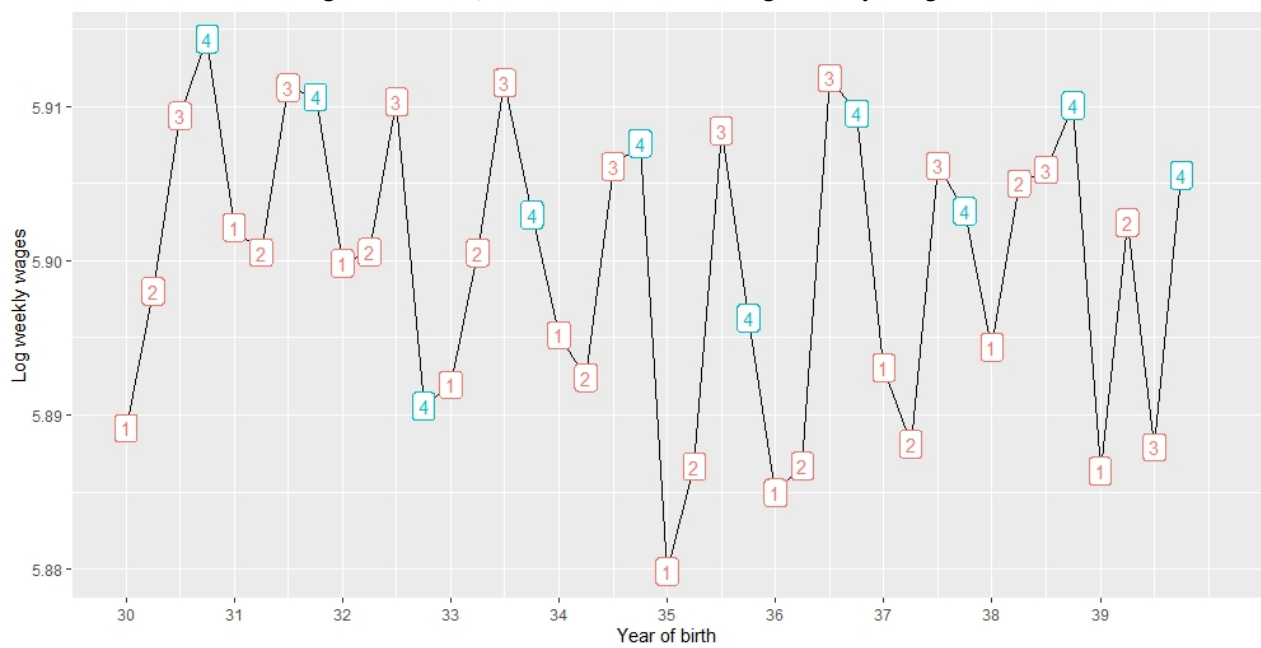


Figure B.13: Quarter of Birth and Log Weekly Wages



# Bibliography

- Joshua D Angrist and Alan B Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.
- John Bound, David A Jaeger, and Regina M Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450, 1995.
- Anirudh L Nagar. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica: Journal of the Econometric Society*, pages 575–595, 1959.
- Robert L Basman. On the asymptotic distribution of generalized linear estimators. *Econometrica, Journal of the Econometric Society*, pages 97–107, 1960.
- David H Richardson. The exact distribution of a structural coefficient estimator. *Journal of the American Statistical Association*, 63(324):1214–1226, 1968.
- Takamitsu Sawa. The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *Journal of the American Statistical association*, 64(327):923–937, 1969.
- Douglas Staiger and James Stock. Stock (1997). instrumental variables with weak instruments. *Econometrica*, 65(3):557–586, 1997.
- James H Stock, Jonathan H Wright, and Motohiro Yogo. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4): 518–529, 2002.
- Joshua D Angrist and Alan B Krueger. Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, 13(2):225–235, 1995.
- Wayne A Fuller. Some properties of a modification of the limited information estimator. *Econometrica: Journal of the Econometric Society*, pages 939–953, 1977.
- Stephen G Donald and Whitney K Newey. Choosing the number of instruments. *Econometrica*, 69(5): 1161–1191, 2001.
- Joshua David Angrist, Guido W Imbens, and Alan B Krueger. Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67, 1999.

- 
- Sören Blomquist and Matz Dahlberg. Small sample properties of liml and jackknife iv estimators: experiments with weak instruments. *Journal of Applied Econometrics*, 14(1):69–88, 1999.
- Theodore W Anderson, Herman Rubin, et al. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63, 1949.
- Thomas J Rothenberg. Approximating the distributions of econometric estimators and test statistics. *Handbook of econometrics*, 2:881–935, 1984.
- Bruce E. Hansen. *ECONOMETRICS*, chapter 11, pages 403–405. University of Wisconsin Department of Economics, 2019.
- Adolf Buse. The bias of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, pages 173–180, 1992.
- Paul A Bekker. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, pages 657–681, 1994.
- James H Stock and Motohiro Yogo. Testing for weak instruments in linear iv regression. Technical report, National Bureau of Economic Research, 2002.
- TW Anderson, Naoto Kunitomo, and Yukitoshi Matsushita. On the asymptotic optimality of the liml estimator with possibly many instruments. *Journal of Econometrics*, 157(2):191–204, 2010.
- Naoto Kunitomo. A third order optimum property of the ml estimator in a linear functional relationship model and simultaneous equation system in econometrics. *Annals of the Institute of Statistical Mathematics*, 39(3):575–591, 1987.
- Russell Davidson and James G MacKinnon. The case against jive. *Journal of Applied Econometrics*, 21(6):827–833, 2006.
- Pischke. Weak instruments. URL: <http://econ.lse.ac.uk/staff/spischke/ec533/Weak%20IV.pdf>, 10 2018.
- Russell Davidson, James G MacKinnon, et al. *Econometric theory and methods*, chapter 12, pages 535–540. Oxford University Press New York, 203.