ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Unbiased instrumental variable estimation under weak instruments

Bachelor Thesis Econometrics and Operations Research
Major in Econometrics

Kyle Paul Gartner

Abstract

We show that there exists a mean unbiased instrumental variable estimation method in models with a single weak instrument and with a single endogenous regressor. We show that the unbiasedness does not come at a cost of increased estimator deviation. We compare the performance of different instrumental variable estimation methods, such as the Fuller estimator, with the unbiased estimate. The Unbiased estimator has a lower deviation than the two stage least square method, however other methods we tested outperform the Unbiased estimator. The tested methods are applied to estimate the return to schooling using college proximity as an instrument.

Keywords: Unbiased estimation, Fuller, weak instruments

Supervisor: Klooster, J
Second assessor: Naghi, AA
Date final version: July 3, 2022

1 Introduction

At present regression analysis is one of the most popular and powerful analysis tools for researchers, with a wide range of applications. One of the most common regression analysis methods is the ordinary least squares method (OLS). This method relies on seven main assumptions on the data generating process (Heij et al., 2004). Exogeneity is one of these assumptions, when explanatory variables are not exogenous parameter estimates may be biased and/or inconsistent. There are ways to circumvent this problem. Instrumental variables (IV) estimation provides us with a solution, these methods however come with their own drawbacks. In order to obtain an unbiased estimate, the two stage least square (2SLS), the conventional IV estimator for example, requires an instrument that is highly correlated with the endogenous variable (Amemiya, 1975). Such an instrument is called a strong instrument (Staiger & Stock, 1994). In practice it is however hard and not always possible to find a strong instrument (Martens et al., 2006). It is therefore useful to have an IV estimation method that (also) performs well under weak instrumental variables. Thus far not many methods have been designed that fulfil that criteria, the main methods are the Fuller estimation method proposed by Fuller (1977) and the Unbiased estimator proposed by Andrews & Armstrong (2017). A method with such properties would have a wide variety of practical applications in cases were the instruments are weak, for example when examining the effect of migrations on countries' productivity as is done by Hornung (2014) or to study the relationship between education and labor market earnings as is done by Angrist & Krueger (1995).

In order to analyse the Unbiased estimator, we aim to apply a simulation study to examine the performance of this estimator in the case of both weak and strong instrumental variables. We will therefore formulate the following central research question: *"What are the bias and deviation of the Unbiased estimator?"*. We will also analyse a few extensions to our main research question, we will formulate these extensions using different sub-questions.

First, we will compare our estimates with other estimates such as those obtained by 2SLS or the Fuller estimation approach. We will also expand our comparison with other estimation methods such as the limited information maximum likelihood (LIML) and the bias-adjusted 2SLS (B2SLS) as mentioned in Mills et al. (2014) and two estimators which we will propose ourselves. *"How do the bias and deviation of the Unbiased estimator compare to the 2SLS, Fuller, LIML, B2SLS and our two self proposed methods?"*. The unbiased estimation method relies on prior knowledge on the coefficient sign within the first stage regression, our obtained results may therefore be broadened by examining the effect on performance when assuming the wrong coefficient sign. *"How do estimates fare when the sign assumption is violated?"*. Lastly, we will conduct an empirical application of our methodology to data from Card (1993), who analyse the causal effects of education on earnings, using college proximity as instrumental variable.

1.1 Structure

In the rest of this section we will discuss the literature related to our research. Section 2 introduces the notation, the Unbiased estimator and the estimators we would like to use for our method comparison. Section 2 also discusses the theoretical consequences when the sign assumption is violated. Thereafter Section 3 presents all our simulation results and aims to answer the research questions of this paper. In Section 4 we conduct our empirical application using the data from Card (1993). Lastly, Section 5 discusses the results and presents some suggestions for future research. Auxiliary results and proofs are given in Appendix A.

1.2 Theoretical background

In practice, endogenous variables are far from rare (Martens et al., 2006). To bypass the problems caused by endogenous variables one can consider using IV estimation to obtain unbiased parameter estimates. This class of estimation methods use an instrument that is uncorrelated to the error term but is (in theory) highly correlated to the endogenous variable. In the first stage regression the relation between the instrument and the endogenous variable is estimated, thereafter in the second stage regression this relation is used to estimate a linear regression model of the dependent variable. Hirano & Porter (2015) prove that in a linear IV model with weak instruments, mean, median and quantile unbiased estimations are all impossible when the parameter space of the first stage regression is unrestricted. We thus know that in order to produce an unbiased estimate under weak instruments we must impose a restriction on the first stage regression. Staiger & Stock (1994) have proposed that instruments be valued as weak whenever the F-statistic of the first stage regression has a value less than ten, this rule of thumb was later refined and improved by Stock & Yogo (2002) but is still widely accepted and used within academia.

Andrews & Armstrong (2017) propose exploiting information on the sign of the first stage regression, i.e. whether the relationship between the instrument and the endogenous variable is positive or negative, to circumvent the previously mentioned impossibility result. Using this information they constructed unbiased estimates when instruments are weak, in both the single and multiple instrument case. In both cases the estimation methods converge towards the 2SLS estimate when the instruments increase in strength, this is a very desirable property since this implies that their method obtains asymptotically unbiased and efficient estimates when instruments are strong, since the 2SLS estimator has this property (Amemiya, 1975).

2 Estimation

In this paper we suppose that our sample contains N observations of three different variables, Y_n, X_n and Z_n where $n = 1, \ldots, N$. Y_n contains observations of our dependant variable, X_n is our endogenous independent variable and Z_n is a $m \times 1$ vector of instrumental variables. For further convenience we let Y and X be $N \times 1$ vectors where row n contains the values Y_n and X_n respectively, also let Z be a $N \times m$ matrix where row n contains the values of Z_n'. We can then capture the relationship between the variables using the following structural form of the classic linear IV model:

$$\begin{aligned} Y &= X\beta + \tilde{U} \\ X &= Z\pi + V \end{aligned} \tag{1}$$

The reduced form of Equation 1 can then be obtained by substituting the endogenous variable X by its estimated value:

$$\begin{aligned} Y &= Z\pi\beta + U \\ X &= Z\pi + V, \end{aligned} \tag{2}$$

where $\tilde{U}, U$ and V denote the error terms.

If we then denote the parameters of the OLS regression in the first- and second-stage by ξ_1 and ξ_2 respectively, we find that

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1}Z'Y \\ (Z'Z)^{-1}Z'X \end{pmatrix} \sim N\left(\begin{pmatrix} \pi\beta \\ \pi \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right), \tag{3}$$

where $\pi_i > 0$ for each $\pi_i \in \pi$. This restriction on the sign assumption can be made without any loss of generality since if $\pi_i < 0$ holds for any $i = 1, ..., m$, we can redefine our instrument Z by multiplying each value in column i by -1.

We can now focus on estimation of ξ_1 and ξ_2 since these give us sufficient information to obtain estimates for β and π in the case where the error terms U_n and V_n are independent and identically distributed (i.i.d.) over n. In the case where the regression in Equation 2 is just-identified the parameters ξ_1 and ξ_2 are scalars and we can thus write the variance-covariance matrix as follows:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = (I_2 \otimes (Z'Z)^{-1}Z')Var((U',V')')(I_2 \otimes (Z'Z)^{-1}Z')' = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

2.1 Unbiased estimation of the inverse of a normal mean

Let us assume that we can create an unbiased estimator $\hat{\tau}$ for $1/\pi$ that depends on the data only through ξ_2. We can then define the estimator $\hat{\delta}$ as

$$\hat{\delta}(\xi, \Sigma) = (\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2) \tag{4}$$

By removing the explanatory power of ξ_2 using the correlation between ξ_2 and ξ_1, $\hat{\delta}$ is constructed is such a way that it is orthogonal to $\hat{\tau}$ (Kleibergen, 2002). This estimator thus has the properties that $\mathbb{E}[\hat{\delta}] = \pi\beta - \frac{\sigma_{12}}{\sigma_2^2}\pi$ and due to orthogonality $\hat{\delta}$ is independent of $\hat{\tau}$. From this independence we get that $\mathbb{E}[\hat{\tau}\hat{\delta}] = \mathbb{E}[\hat{\tau}]\mathbb{E}[\hat{\delta}] = \beta - \frac{\sigma_{12}}{\sigma_2^2}$, we can now see that $\hat{\tau}\hat{\delta} + \frac{\sigma_{12}}{\sigma_2^2}$ will be an unbiased estimator of β. Our problem has therefore now become unbiased estimation of τ, the inverse of a normal mean.

Nikulin & Voinov (2011) have shown that an unbiased estimator of τ indeed exists if we assume the sign of π to be known. This estimator is given in Equation 5 and is a function of Φ and ϕ which denote the cumulative distribution function (c.d.f.) and probability density function (p.d.f.) of the standard normal distribution respectively.

$$\hat{\tau}(\xi_2, \sigma_2^2) = \frac{1}{\sigma_2}\frac{1 - \Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)} \tag{5}$$

We thus know that $E_\pi[\hat{\tau}(\xi_2, \sigma_2^2)] = \frac{1}{\pi}$ for all $\pi > 0$. The derivation of this non-intuitive estimator is based on the theory of bilateral Laplace transforms. Unbiasedness of this estimator was verified by Andrews & Armstrong (2017).

2.2 Unbiased estimation of β

We can now construct an unbiased estimate for our parameter of interest. As mentioned earlier $\hat{\tau}\hat{\delta} + \frac{\sigma_{12}}{\sigma_2^2}$ gives an unbiased estimate of β, if we now substitute Equation 4 and Equation 5 into our parameters $\hat{\delta}$ and $\hat{\tau}$ we obtain the following unique unbiased estimate of β:

$$\begin{aligned} \hat{\beta}_U(\xi, \Sigma) &= \hat{\tau}(\xi_2, \sigma_2^2)\hat{\delta}(\xi, \Sigma) + \frac{\sigma_{12}}{\sigma_2^2} \\ &= \frac{1}{\sigma_2}\frac{1 - \Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)}(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2) + \frac{\sigma_{12}}{\sigma_2^2} \end{aligned} \tag{6}$$

This estimator is unbiased for β if the assumption that $\pi > 0$ holds.

The 2SLS estimate can be written as:

$$\hat{\beta}_{2SLS} = \frac{\xi_1}{\xi_2} = \frac{1}{\xi_2}(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2) + \frac{\sigma_{12}}{\sigma_2^2}$$

We now find that $\hat{\beta}_U$ differs from the 2SLS estimate only in the sense that it uses the unbiased estimate $\hat{\tau}$ for $1/\pi$ instead of the plug-in estimate $1/\xi_2$. Baricz (2008) have shown that $\hat{\tau} < 1/\xi_2$ for $\xi_2 > 0$, this implies that whenever $\xi_2 > 0$, $\hat{\beta}_U$ will equal the 2SLS estimate shrunk towards σ_{12}/σ_2^2.

2.3 *k*-class estimators

An important family of estimators are the so called k-class estimators which have been constructed based on studies conducted by Nagar (1959) and Theil (1961). Mills et al. (2014) captured this family of estimators using the model specification as given in Equation 7. Different estimation methods coincide with different values for the parameter k, for example if $k = 0$ we obtain the OLS estimator and if $k = 1$ we get the 2SLS estimator. Parameter k is an arbitrary scalar which can either be stochastic or non-stochastic (Theil, 1961). It has been proven that in the general case parameter estimates using k-class estimators will be consistent if $\underset{N\to\infty}{plim}(k-1) = 0$ (Savin, 1973). We will only consider non-stochastic values for k, in order to guarantee consistency it therefore suffices to check whether $\underset{N\to\infty}{lim}(k-1) = 0$ holds.

As mentioned earlier, in our method comparison section we would like to analyse the four existing estimation methods, namely: $\hat{\beta}_{2SLS}$, $\hat{\beta}_{Full}$, $\hat{\beta}_{LIML}$ and $\hat{\beta}_{B2SLS}$ which coincide with the following values of k

$$\begin{aligned}
&\text{2SLS}: && k = 1,\\
&\text{LIML}: && k = k_{LIML} = \text{the smallest root } \iota \text{ of } \det((Y'P_zY/N + \Sigma) - \iota\Sigma) = 0,\\
&\text{B2SLS}: && k = 1 + (m-2)/N,\\
&\text{Fuller}: && k = k_{LIML} - 1/N.
\end{aligned}$$

$$\hat{\beta}(k) = \frac{X'P_ZY + N(1-k)\sigma_{12}}{X'P_ZX + N(1-k)\sigma_2^2}, \tag{7}$$

where $P_A = A(A'A)^{-1}A'$ for any full column matrix A.

Note that the Fuller method is part of a larger class of estimates. Fuller (1977) define the modified limited information estimator class as the class for which it holds that $k =$

$k_{LIML} - a/N$ for $a \in \mathbb{R}$. The LIML and Fuller methods are both part of this class and coincide with a value of a equal to 0 and 1 respectively. The different methods within this class place different weights on the variance and covariance terms in Equation 7. Fuller (1977) have argued that "If one desires estimates that are nearly unbiased a is set equal to 1. [...] If one wishes to minimize the mean square error of the estimators an a of 4 is appropriate."(p. 951). This result seems to indicate the existence of a bias-variance trade-off between different parameter values a.

We would like to further investigate this potential trade-off and will therefore propose two new estimators which are contained within the modified limited information estimator class. To our knowledge these methods have not yet been tested/used within other academic papers. First, a value of a equal to -1 which we will refer to as the Mirrorfuller. Second, we will use a value of a equal to 2 which we will refer to as the Doublefuller. Further method comparison within the modified limited information estimator class is an interesting topic for future research. The Mirrorfuller and Doublefuller methods coincide with the following values of k

$$\begin{aligned} \text{MirrorF}: \quad & k = k_{LIML} + 1/N, \\ \text{DoubleF}: \quad & k = k_{LIML} - 2/N. \end{aligned}$$

Anderson & Sawa (1979) have proven that in the just-identified case $\hat{\beta}_{LIML}$ is equivalent to the 2SLS estimator, this also implies that $\hat{\beta}_{B2SLS}$ is equivalent to the Fuller estimator since $k_{LIML} = 1$ and $m = 1$ by definition. Due to this result we must reduce our analysis to the comparison of $\hat{\beta}_U$ with $\hat{\beta}_{2SLS}$, $\hat{\beta}_{Full}$, $\hat{\beta}_{MirrorF}$ and $\hat{\beta}_{DoubleF}$. If we substitute the corresponding values of k into Equation 7 we obtain the following four estimators:

$$\begin{aligned} \hat{\beta}_{2SLS} &= \hat{\beta}(1) = \frac{\xi_1}{\xi_2}, \\ \hat{\beta}_{Full} &= \hat{\beta}(1 - 1/N) = \frac{\xi_2\xi_1 + \sigma_{12}}{\xi_2^2 + \sigma_2^2}, \\ \hat{\beta}_{MirrorF} &= \hat{\beta}(1 + 1/N) = \frac{\xi_2\xi_1 - \sigma_{12}}{\xi_2^2 - \sigma_2^2}, \\ \hat{\beta}_{DoubleF} &= \hat{\beta}(1 - 2/N) = \frac{\xi_2\xi_1 + 2\sigma_{12}}{\xi_2^2 + 2\sigma_2^2} \end{aligned} \tag{8}$$

These methods are all consistent since $\lim_{N\to\infty} (k - 1) = 0$ clearly holds for all four values of k.

2.4 Asymptotic behaviour: $\hat{\beta}_U$ under strong instruments

In Section 2.2 we have shown that $\hat{\beta}_U$ is unbiased. Even though this is a very desirable property we must also analyse the performance of this estimator when instruments become more informative. Desirably we would want the Unbiased estimator to be asymptotically efficient, i.e. to be efficient as instruments become stronger. This efficiency is achieved if the estimator is asymptotically equivalent to an efficient (IV) estimator, such as 2SLS (Amemiya, 1975). We will investigate the asymptotic properties by letting the parameter π become increasingly larger. Conventionally one would keep π fixed and take the sample size $N \to \infty$, this would result in $\Sigma \to 0$. Andrews & Armstrong (2017) have shown however that results stay the same when one takes π to infinity, we will therefore focus on the latter to simplify notation.

In Section 2.2 we also showed that the 2SLS estimator only differs from the Unbiased estimator in the fact that it replaces $\hat{\tau}(\xi_2, \sigma_2^2)$ in Equation 6 by $1/\xi_2$. Intuitively one can see that these two estimators for $1/\pi$ both converge toward zero as the value of ξ_2 increases:

$$\lim_{\xi_2 \to \infty} \hat{\tau}(\xi_2, \sigma_2^2) = \lim_{\xi_2 \to \infty} 1/\xi_2 = 0$$

Small (2010) more formally defined the conversion of these to estimators using the following inequality:

$$\sigma_2 \big| \hat{\tau}(\xi_2, \sigma_2^2) - 1/\xi_2 \big| \leq \left| \frac{\sigma_2^3}{\xi_2^3} \right|$$

Here we see that the conversion happens rapidly since the right side of the inequality grows as a cube of ξ_2, therefore the two estimators coincide with high precision for large values of ξ_2. Next we know that parameter ξ_2 has a mean equal to π, thus as $\pi \to \infty$ we will also find that $\xi_2 \to \infty$. We can then conclude that as $\pi \to \infty$ the difference between $\hat{\tau}(\xi_2, \sigma_2^2)$ and $1/\xi_2$ will converge toward zero rapidly, hence the estimator $\hat{\beta}_U$ has the same limiting distribution as $\hat{\beta}_{2SLS}$ as instruments become more informative.

2.5 Violation of restriction on first stage sign

Given that the construction of the Unbiased estimator $\hat{\beta}_U$ relies on the assumption that the sign of π in Equation 2 is known, namely $\pi > 0$, it is interesting to analyse the performance of $\hat{\beta}_U$ when this assumption is violated. These results can for example be used to check whether the right sign has been assumed within an empirical application or to improve the robustness of the Unbiased estimator.

Let us assume that we calculate the Unbiased estimator using an instrument Z^*, where $Z^* = -1 \cdot Z$, for which it thus holds that Z^* has a negative effect on X, i.e. $\pi^* < 0$ in $X = Z^*\pi^* + V^*$. Within this first stage regression we obtain parameter ξ_2^* which (in expected value) is equal to $-1 \cdot \xi_2$, where ξ_2 is the correct parameter in the sense that $\hat{\xi}_2$ would be obtained in the first stage regression if we chose Z as instrumental variable. If we now use ξ_2^* to calculate $\hat{\tau}$ in Equation 5 we obtain the following estimate for $1/\pi$:

$$\hat{\tau}^* = \hat{\tau}(\xi_2^*, \sigma_2^2) = \frac{1}{\sigma_2}\frac{1 - \Phi(\xi_2^*/\sigma_2)}{\phi(\xi_2^*/\sigma_2)} = \frac{1}{\sigma_2}\frac{\Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)}, \tag{9}$$

where we use the fact that $\Phi(A) = 1 - \Phi(-A)$ and that $\phi(A) = \phi(-A)$, $\forall A \in \mathbb{R}$.

We will now derive the first moment of this new estimator $\hat{\tau}^*$ in order to analyse its properties. From Equation 3 we know that $\xi_2/\sigma_2 \sim N(\pi/\sigma_2, 1)$, let $x = \xi_2/\sigma_2$ we then have:

$$\begin{aligned}
E_\pi(\hat{\tau}(\xi_2^*, \sigma_2^2)) =& \frac{1}{\sigma_2}\int \frac{\Phi(x)}{\phi(x)}\phi(x - \pi/\sigma_2)dx \\
=& \frac{1}{\sigma_2}\int \Phi(x)\exp((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2)dx \\
=& \frac{1}{\sigma_2}\exp(-(\pi/\sigma_2)^2/2)\Bigg\{\Big[\Phi(x)(\sigma_2/\pi)\exp((\pi/\sigma_2)x)\Big]_{x=-\infty}^{\infty} \\
& - \int (\sigma_2/\pi)\exp((\pi/\sigma_2)x)\phi(x)dx\Bigg\}
\end{aligned}$$

We made use of integration by parts in the previous step of the derivation; $\int udv = uv - \int vdu$, where $u = \Phi(x)$ and $v = \exp(\pi/\sigma_2)x$.

$$\begin{aligned}
=& \frac{1}{\sigma_2}\exp(-(\pi/\sigma_2)^2/2)\Big[\Phi(x)(\sigma_2/\pi)\exp((\pi/\sigma_2)x)\Big]_{x=-\infty}^{\infty} \\
& - \frac{1}{\sigma_2}\int (\sigma_2\pi)\exp((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2)\phi(x)dx \\
=& \frac{1}{\sigma_2}\exp(-(\pi/\sigma_2)^2/2)\lim_{t\to\infty}\Big[\Phi(t)(\sigma_2/\pi)\exp(t(\pi/\sigma_2)) \\
& - \Phi(-t)(\sigma_2/\pi)\exp(-t(\pi/\sigma_2))\Big] - \frac{1}{\pi}\int \phi(x - \pi/\sigma_2)dx
\end{aligned}$$

It can now clearly be seen that the limit within the square brackets in the last expression will go towards infinity as $t \to \infty$. We can thus conclude that $E(\hat{\tau}(\xi_2^*, \sigma_2^2))$ has an infinite first moment for all π. Therefore if the sign assumption of π is violated $\hat{\beta}_U$ will no longer be an unbiased estimator for β, once noticed this problem can however easily be solved since the instrumental variable can be redefined in order to obtain an unbiased estimate of β.

3 Simulation studies

In this section we will present the results we obtained from our simulation study and aim to compare the performance measures of the Unbiased estimator with the different estimation methods proposed in Section 2.3. The Unbiased estimator $\hat{\beta}_U$ as defined in Equation 6 relies on five parameters $(\beta, \pi, \sigma_1^2, \sigma_{12}, \sigma_2^2)$. This five dimensional parameter space is too large to fully explore via simulation, we will therefore reduce the parameter space using the equivariance argument used by Andrews & Armstrong (2017). See Appendix A.2 for details. We are able to set $\beta = 0$ and $\sigma_1 = \sigma_2 = 1$ without loss of generality, we thus only need to explore the two dimensional parameter space $(\pi, \sigma_{12}) \in (0, \infty) \times [0, 1)$, which can be fully explored via simulation. Since the reduced parameter space is quite small we were able to obtain comprehensive simulation results.

3.1 Method comparison

3.1.1 Estimator mean

In the left column of Figure 1 we plot the bias comparison for $\hat{\beta}_U, \hat{\beta}_{Full}$ and $\hat{\beta}_{DoubleF}$ (we omit $\hat{\beta}_{2SLS}$ and $\hat{\beta}_{MirrorF}$ from our comparison since the first moments of these two estimators do not exist in the just-identified case). Here we analyse a wide range of values for $\pi > 0$ but limit our analysis to $\sigma_{12} \in \{0.1, 0.5, 0.95\}$. We find that, as we would expect, the bias of the Unbiased estimator is equal to zero for all values within our two dimensional parameter space. Next we observe that the Doublefuller estimation method has a uniformly higher bias for all three values of σ_{12} compared to the Fuller and Unbiased estimator.

If we, instead of analysing the mean bias, would consider the median bias it is then possible to include $\hat{\beta}_{2SLS}$ and $\hat{\beta}_{MirrorF}$ to our analysis. Due to the nature of the Mirrorfuller estimator it is possible to obtain a negative median bias. To make the comparison more easily visible we will therefore plot the absolute median bias. Results are given in Appendix A.1. We see that the relationship found when comparing the mean bias of the three previously mentioned estimators still holds when analysing median bias. We find that the Doublefuller estimator has the largest median bias for (nearly) the entire parameter space. For small values of π, $\mathbb{E}[F] \lessapprox 2$, the Unbiased estimator outperforms all other estimation methods. When the instrument becomes stronger however the 2SLS estimate has the overall lowest absolute median bias. As mentioned earlier we see that the absolute median bias of the Mirrorfuller estimator is not uniformly decreasing over π, this is caused by the fact that we are subtracting σ_{12} and σ_2^2 in Equation 8 this may lead to a negative median bias thus increasing its absolute median bias. This does imply that the median bias of this estimator crosses a value of zero, the Mirrorfuller has the smallest absolute median bias for $\mathbb{E}[F]$ around a value of five and for a value of $\mathbb{E}[F]$ between approximately four and nine $\hat{\beta}_{MirrorF}$ beats all other methods except for the 2SLS estimator.

3.1.2 Estimator deviation

In the right column of Figure 1 we plot the logarithm of the median absolute deviation (or equivalently, 50th percentile log absolute deviation) from the true parameter value β for $\sigma_{12} \in \{0.1, 0.5, 0.95\}$. We plot the log quantiles as this makes the plots better visible. Results have also been obtained for 10th and 90th percentile log absolute deviation and are reported in Appendix A.1. Our results for $\hat{\beta}_U, \hat{\beta}_{2SLS}$ and $\hat{\beta}_{Full}$ are the same to those reported in Section 4.1.2 of Andrews & Armstrong (2017), namely that $\hat{\beta}_U$ has a uniformly lower 50th and 90th percentile absolute deviation compared to $\hat{\beta}_{2SLS}$, also for the 10th percentiles we find that $\hat{\beta}_U$ outperforms $\hat{\beta}_{2SLS}$ across nearly the entire parameter space except for cases where σ_{12} is very high and π is very small. Similarly, we also find that $\hat{\beta}_{Full}$ outperforms $\hat{\beta}_U$ except for the cases where σ_{12} is very high and π is fairly small.

If we now analyse the performance of $\hat{\beta}_{MirrorF}$ and $\hat{\beta}_{DoubleF}$ we find that $\hat{\beta}_{MirrorF}$ has the largest absolute deviation across all methods and across nearly the entire parameter space with an exception for very small values of π. Conversely, $\hat{\beta}_{DoubleF}$ exhibits the lowest absolute deviation from the true parameter β across all methods except for cases where σ_{12} is very high and π is fairly small. These results seem to indicate that there exists a bias-variance trade-off between $\hat{\beta}_{MirrorF}$ and $\hat{\beta}_{DoubleF}$ since the method with the best performance in the bias analysis has the worst performance in the absolute deviation analysis, and visa versa.

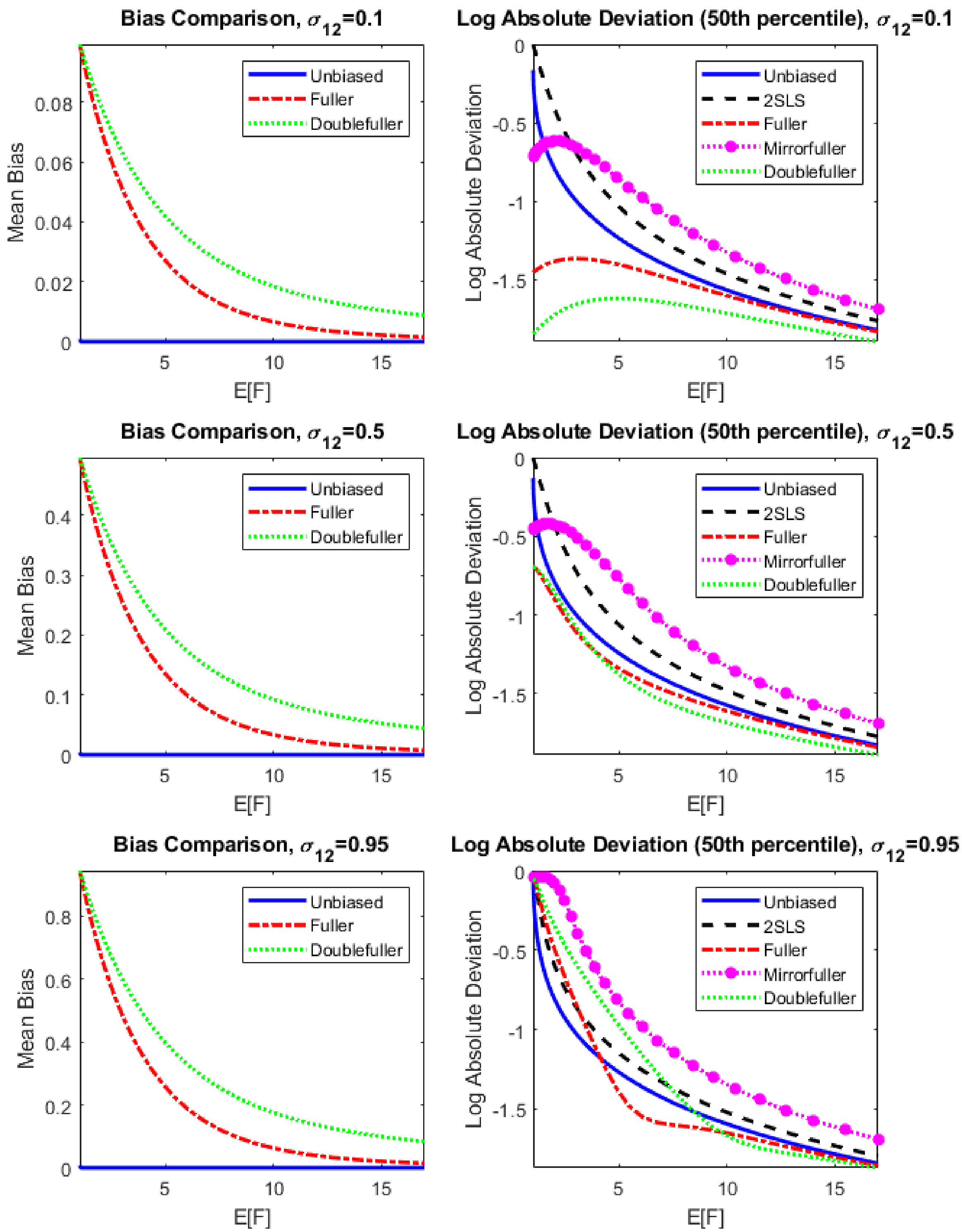

Figure 1: The three panels in the left column plot the mean bias of three different estimators against the mean of the F-statistic of the first stage regression, this is done for $\sigma_{12} \in \{0.1, 0.5, 0.95\}$. The three panels in the right column plot the median (50th percentile) log absolute deviation from the true parameter value β for five different estimators, this is done for the same three values of σ_{12}. Based on 10 million simulation draws.

3.2 Violation of restriction on first stage sign

A crucial assumption made when constructing the Unbiased estimator is the sign assumption on the relationship between the instrument and your endogenous variable, $\pi > 0$ in Equation 2. If this inequality does not hold we need to redefine our instrumental variable. Theoretically one needs to argue why this first stage sign is known using prior knowledge on the relationship between the variables, in practice however it could be that one assumes the wrong sign for parameter π. It is thus useful to analyse the behaviour of $\hat{\beta}_U$ when the restriction on the first stage sign is violated, these results may be used to verify the sign assumption or to improve the robustness of the Unbiased estimator.

In Figure 2 we have again plotted the median bias and the logarithm of the median absolute deviation in the left and right columns respectively, we have now however wrongly assumed the sign of π, i.e. we assume $\pi > 0$ even though in reality $\pi < 0$ holds. It is clear from the left column that the bias of the Unbiased estimator increases exponentially as $\mathbb{E}[F]$ increases, or equivalently as π decreases (becomes more negative). The median bias of this estimator spans several orders of magnitude. Even for relatively low values of $\mathbb{E}[F]$ the Unbiased estimator still has a significantly higher bias that the other estimators. For a value of $\mathbb{E}[F] = 1.5$ the median bias of $\hat{\beta}_U$ is already more than twice as large as the bias of $\hat{\beta}_{DoubleF}$, which has the largest bias out of the remaining estimators.

On inspection of the right column of Figure 2 we find that the median log absolute deviation of the Unbiased estimator from the true parameter β seems to increase linearly with $\mathbb{E}[F]$. Similar results are found for both the 10th and 90th percentile log absolute deviation and are reported in Appendix A.1. This indicates that both the bias and the absolute deviation increase exponentially with $\mathbb{E}[F]$. These results imply that one can quite easily detect a wrongly assumed sign of the parameter π by comparing the result with other estimation methods, such as the standard 2SLS estimator.

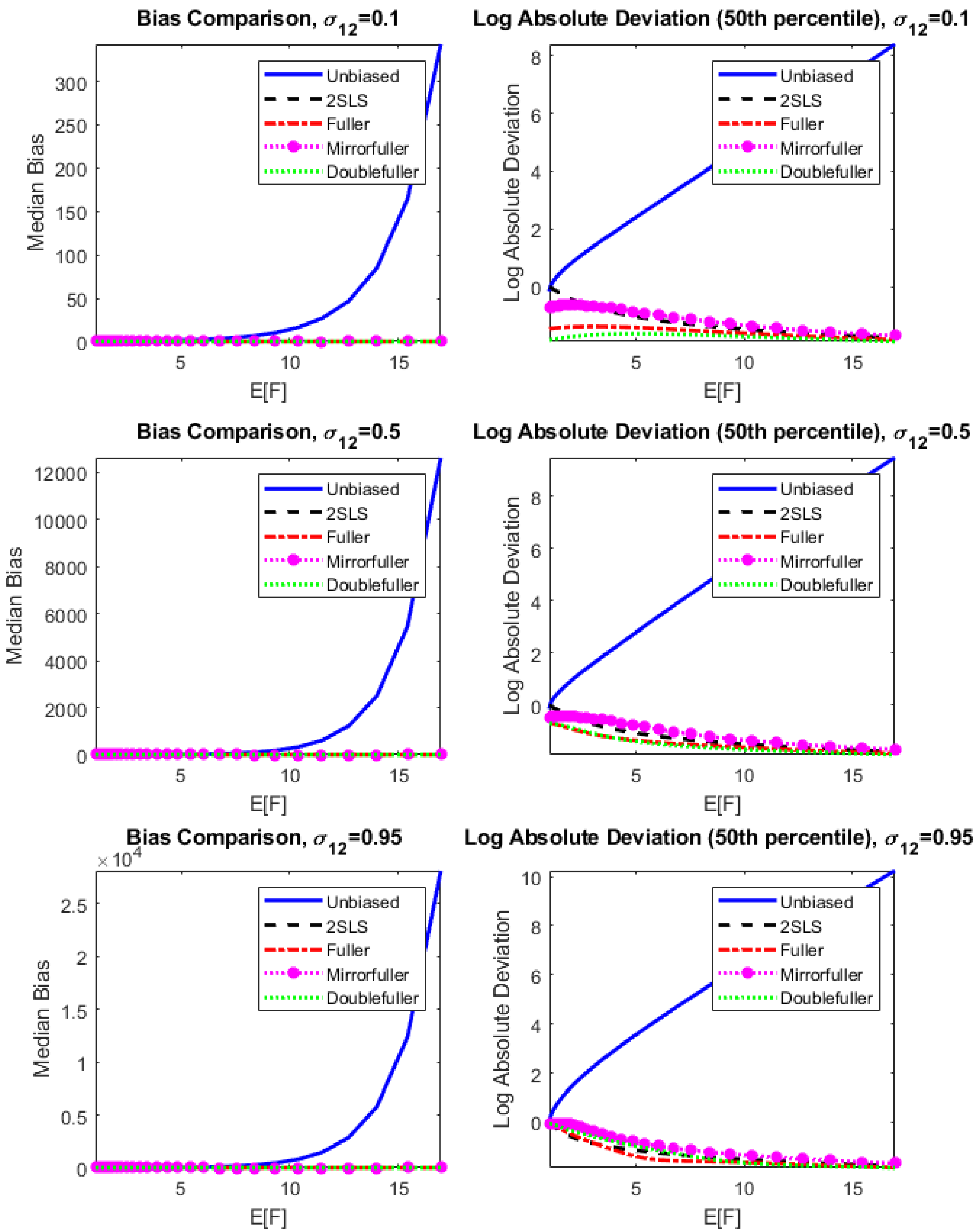

Figure 2: The three panels in the left column plot the median bias of five different estimators against the mean of the F-statistic of the first stage regression, this is done for $\sigma_{12} \in \{0.1, 0.5, 0.95\}$. The three panels in the right column plot the median (50th percentile) log absolute deviation from the true parameter value β for five different estimators, this is done for the same three values of σ_{12}. Here the assumption on the sign of π has been violated. Based on 100.000 simulation draws.

4 Empirical application: Card (1993)

In order to apply the unbiased estimation method to our empirical application we must first redefine our notation in order to allow for extra exogenous variables in our model. As in the general setup our sample contains N observations, we now however have four different variables, $\tilde{Y}_n$, $\tilde{X}_n$, $\tilde{Z}_n$ and W_n where $n = 1, \ldots, N$. $\tilde{Y}_n$ contains observations of our dependant variable, $\tilde{X}_n$ is our endogenous independent variable, $\tilde{Z}_n$ is a $m \times 1$ vector of instrumental variables and W_n is a vector of additional exogenous variables. Let $\tilde{Y} = (\tilde{Y}_1, ..., \tilde{Y}_n)'$, $\tilde{X} = (\tilde{X}_1, ..., \tilde{X}_n)'$, $\tilde{Z} = (\tilde{Z}_1, ..., \tilde{Z}_n)'$ and $W = (W_1, ..., W_n)'$. We thereafter define the residual maker matrix as $M_W = I - W(W'W)^{-1}W'$, lastly we let $Y = M_W\tilde{Y}$, $X = M_W\tilde{X}$ and $Z = M_W\tilde{Z}$ denote the residuals from regressing $\tilde{Y}$,$\tilde{X}$ and $\tilde{Z}$ on W respectively. We can now continue with the estimation procedure using the model as described in Section 2. First, we estimate the variance-covariance matrix Σ using the following hetroskedasticity-robust estimate in order to account for possible hetroskedastic data.

$$\hat{\Sigma} = N \cdot (I_2 \otimes (Z'Z)^{-1})\mathbb{E}\begin{pmatrix} \hat{U}^2 ZZ' & \hat{U}\hat{V}ZZ' \\ \hat{U}\hat{V}ZZ' & \hat{V}^2 ZZ' \end{pmatrix}(I_2 \otimes (Z'Z)^{-1})$$

Where $\hat{U}$ and $\hat{V}$ are the residuals obtained when estimating the models in Section 4.2, Table 2.

Finally, we use $\hat{\Sigma}$ to calculate $\hat{\beta}_U(\xi, \hat{\Sigma})$ as defined in Equation 6.

4.1 Data

Card (1993) study the effect of education on future earnings using data from the Young Men Cohort of the National Longitudinal Survey (NLSYM), which is a project that follows the lives of a sample of American youth. They argue that education is an endogenous variable in the earnings equation and address this by taking the proximity to a 4-year college as an exogenous instrument. Within their motivation they argue that students who live in an area without a college face higher costs of college education since the option of living at home is precluded. They believe that this higher cost reduces investments in higher education, especially among low income families. For further information and motivation of this instrument, see Card (1993). The above argument implies that the first stage sign is known, we can thus apply the Unbiased estimation method.

We will also analyse data from the NLSYM obtained from the Erasmus University Rotterdam. We use the following notation: $\tilde{Y}$ = wage, $\tilde{X}$ = years of education, $\tilde{Z}$ = residence near a 4-year college and W = [a quadratic function of experience, residence in the South, residence in a metropolitan area (SMSA), race], where residence near a 4-year college, residence in the South, residence in SMSA and race are given as binary variables. In Table 1 some of the key characteristics of our data are given. In our sample of 3010 respondents approximately 40% lived in the South, 70% lived in a metropolitan area, 70% lived near a 4-year college and 23% of the respondents are black.

Table 1: Sample characteristics of National Longitudinal Survey of Young Men.

	Average	St.Dev.		Percent (%)
Years of education	13.3	2.7	Lived in the south	40.4
Years of experience	8.9	4.1	Lived in SMSA	71.3
Wage	6.3	0.4	Lived near 4-year college	68.2
			Race: Black	23.4
Sample size (N)	3010			

4.2 Results

To capture the relationship between education and earnings we use the following structural- and reduced form models:

Table 2

Structural model:	Reduced form models:
$\mathrm{Y} = \mathrm{X}\beta + \tilde{U}$	$\mathrm{X} = \mathrm{Z}\pi + V$
	$\mathrm{Y} = \mathrm{Z}\delta + R$
	$\mathrm{Y} = \mathrm{Z}\pi\beta + U$

Table 3 presents the estimation results of our reduced form models and our structural model using college proximity as an instrumental variable for completed education. Within our reduced form models we used OLS to regress education and earnings on college proximity, parameter estimates with corresponding F-statistics are given in columns 2 and 3. Growing up near a college has a strong positive effect on both education (0.34 years of education) and earnings (4.5 percent). We see that college proximity is a strong instrument as we find a first stage F-statistic of 17.55, we would therefore expect parameter estimates of our structural model to be similar for the different IV estimation methods.

In the first column we find the estimated parameters for our structural model using different estimation methods, with the standard errors for estimators with known second moment given in parentheses. As expected the different IV estimation methods produce very similar results with all results falling within a tenth of a standard deviation of the 2SLS estimate (0.1274 - 0.1372). The obtained OLS and 2SLS estimates are the same as those found by Card (1993). With the use of college proximity as an instrument for schooling, return to education is estimated to be approximately 0.13, this estimate implies an earnings gain per year of additional schooling of approximately 13%. However, due to the high standard deviation of the 2SLS estimate one cannot reject the hypothesis that OLS parameter estimates are inconsistent. The consistency of the OLS parameter estimate is examined using the Durbin–Wu–Hausman test (Hausman, 1978).

Table 3: OLS parameter estimates for reduced form models and parameter estimates for structural model of earnings using different estimation methods. The standard errors, for estimators with known second moment, are given in parentheses.

	Structural model:	Reduced form models:	
	Earnings [β]	Education [π]	Earnings [δ]
OLS	0.0740 (0.0035)	0.3373 (0.0824)	0.0446 (0.0170)
2SLS	0.1323 (0.0492)	-	-
Unbiased	0.1290	-	-
Fuller	0.1287	-	-
Mirrorfuller	0.1363	-	-
Doublefuller	0.1287	-	-
F-stat.	-	17.55	7.439

5 Discussion

We have shown that there exists an unbiased estimator in the case of weak instrumental variables, the unbiasedness does not come at the cost of increased estimator deviation. However, the Unbiased estimator does come at the cost of a restricted first stage parameter space, namely a sign restriction on the parameter in the first stage regression. We compared the performance of this estimator with members of the k-class estimator family such as the Fuller estimator. Although the Unbiased estimator has the overall lowest absolute bias it is outperformed by other estimators when analysing absolute median bias or absolute estimator deviation. Our results suggest some areas of future research. First, we analysed two estimators within the modified limited information estimator class, we found evidence for the existence of a bias-variance trade-off. Therefore, it seems interesting to analyse the performance of other estimators captured in this class in order to better understand the relationship between estimator bias and variance. Second, Mills et al. (2014) finds certain conditional t-tests that perform well in an instrumental variable regression model, it may thus be interesting to examine possible tests based on the Unbiased estimator. Lastly, we know that in order to produce an unbiased estimate under weak instruments we must restrict the first stage parameter space, it may therefore be useful study other ways to make use of our knowledge on the first stage sign for both estimation and testing purposes.

References

Amemiya, T. (1975). The nonlinear limited-information maximum-likelihood estimator and the modified nonlinear two-stage least-squares estimator. *Journal of Econometrics*, *3*(4), 375–386.

Anderson, T. W., & Sawa, T. (1979). Evaluation of the distribution function of the two-stage least squares estimate. *Econometrica: Journal of the Econometric Society*, 163–182.

Andrews, I., & Armstrong, T. B. (2017). Unbiased instrumental variables estimation under known first-stage sign. *Quantitative Economics*, *8*(2), 479–503.

Angrist, J. D., & Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, *13*(2), 225–235.

Baricz, Á. (2008). Mills' ratio: Monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications*, *340*(2), 1362–1370.

Card, D. (1993). *Using geographic variation in college proximity to estimate the return to schooling.* National Bureau of Economic Research Cambridge, Mass., USA.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica: Journal of the Econometric Society*, 939–953.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251–1271.

Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K., et al. (2004). *Econometric methods with applications in business and economics.* Oxford University Press.

Hirano, K., & Porter, J. R. (2015). Location properties of point estimators in linear instrumental variables and related models. *Econometric Reviews*, *34*(6-10), 720–733.

Hornung, E. (2014). Immigration and the diffusion of technology: The huguenot diaspora in prussia. *American Economic Review*, *104*(1), 84–122.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, *70*(5), 1781–1803.

Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V., & Klungel, O. H. (2006). Instrumental variables: application and limitations. *Epidemiology*, 260–267.

Mills, B., Moreira, M. J., & Vilela, L. P. (2014). Tests based on t-statistics for iv regression with weak instruments. *Journal of Econometrics*, *182*(2), 351–363.

Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica: Journal of the Econometric Society*, 575–595.

Nikulin, M., & Voinov, V. (2011). Unbiased estimators and their applications. In *International encyclopedia of statistical science* (pp. 1619–1621). Springer.

Savin, N. E. (1973). Systems k-class estimators. *Econometrica: Journal of the Econometric Society*, 1125–1136.

Small, C. G. (2010). *Expansions and asymptotics for statistics*. Chapman and Hall/CRC.

Staiger, D. O., & Stock, J. H. (1994). *Instrumental variables regression with weak instruments.* National Bureau of Economic Research Cambridge, Mass., USA.

Stock, J. H., & Yogo, M. (2002). *Testing for weak instruments in linear iv regression.* National Bureau of Economic Research Cambridge, Mass., USA.

Theil, H. (1961). Economic forecasts and policy.

A Appendix

A.1 Method comparison

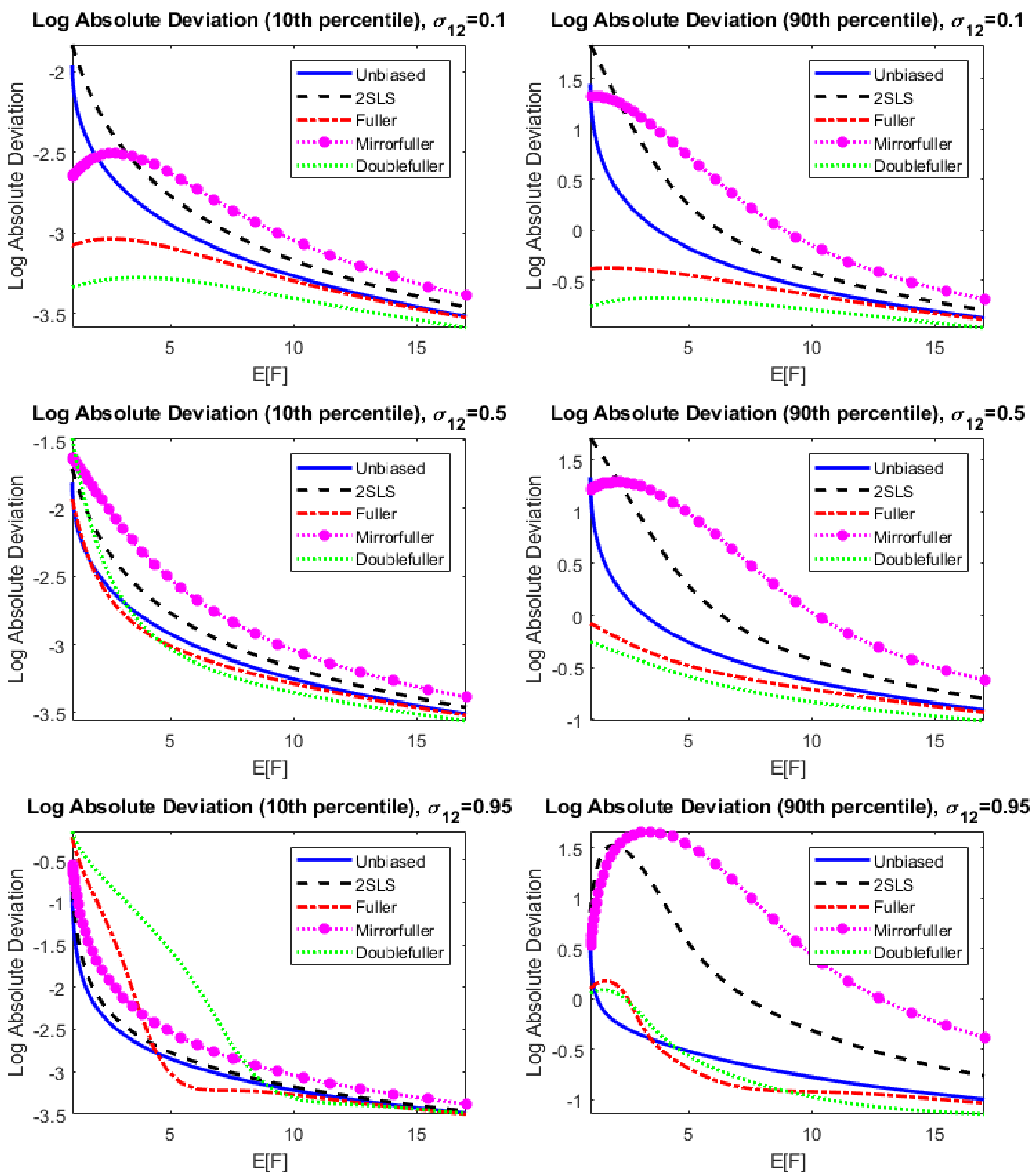

Figure 3: The three panels in the left and right column plot the 10th and 90th percentile log absolute deviation from the true parameter value β for five different estimators respectively, this is done for $\sigma_{12} \in \{0.1, 0.5, 0.95\}$. Based on 10 million simulation draws.

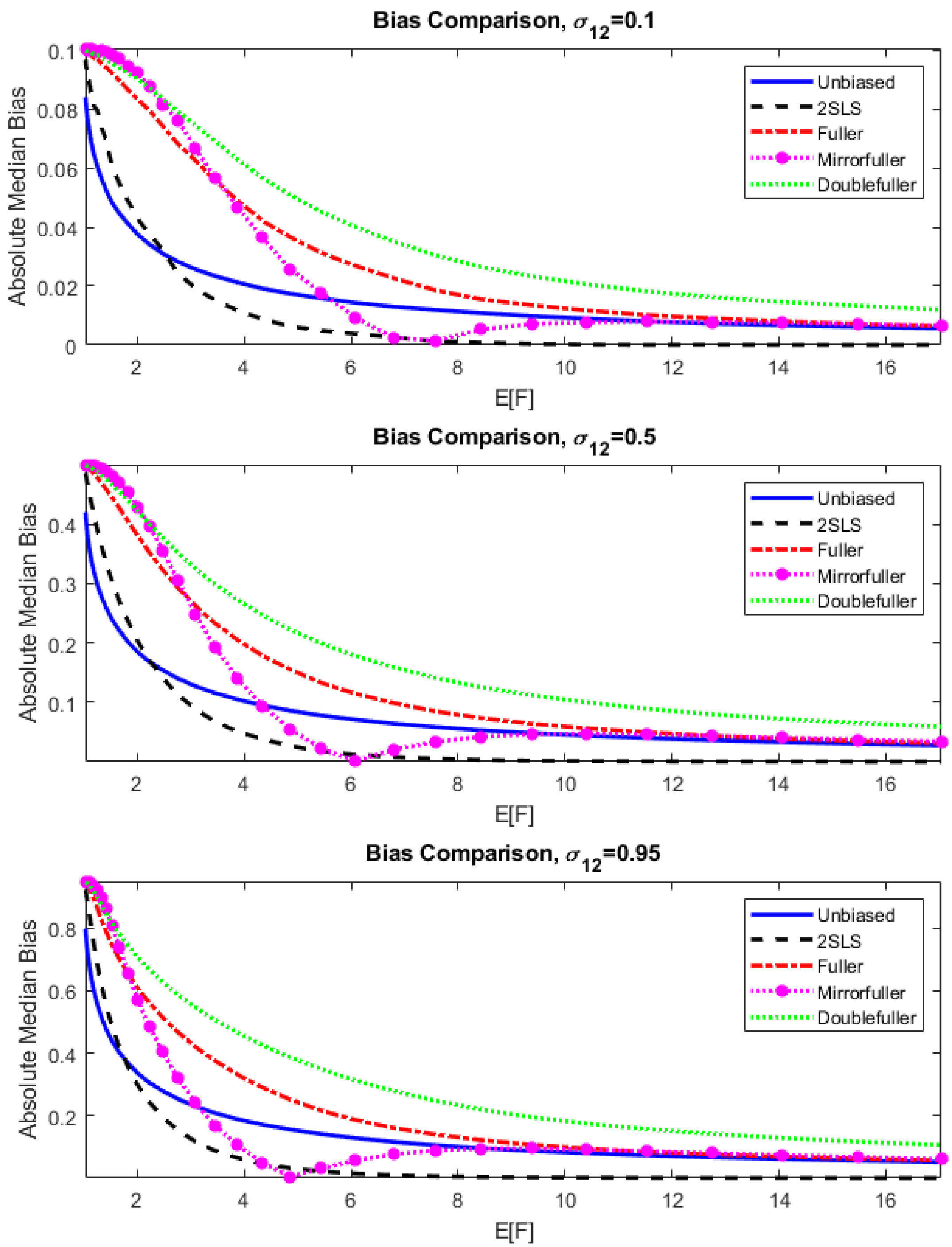

Figure 4: Absolute median bias for five different estimators, this is done for $\sigma_{12} \in \{0.1, 0.5, 0.95\}$. Based on 10 million simulation draws.

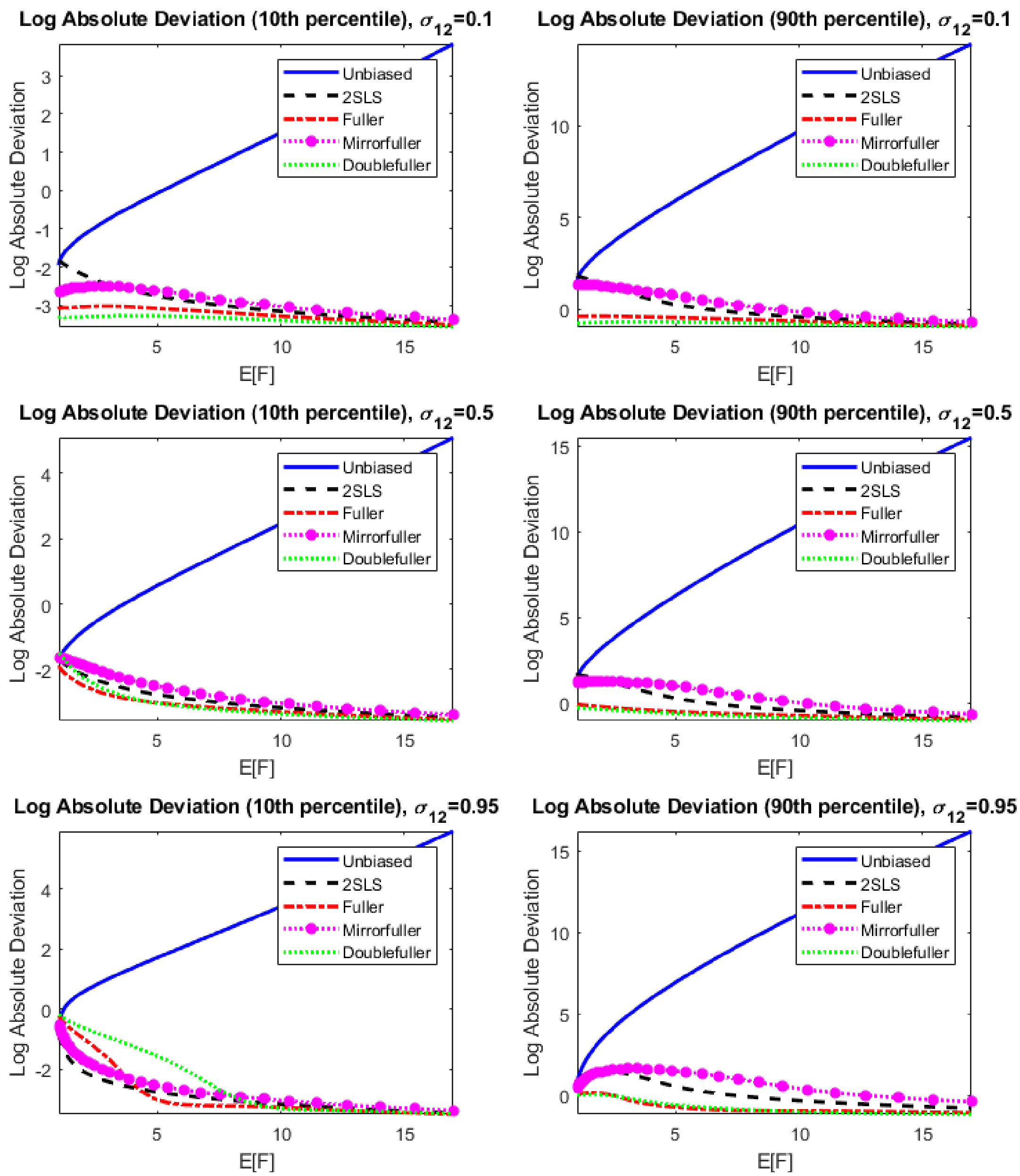

Figure 5: The three panels in the left and right column plot the 10th and 90th percentile log absolute deviation from the true parameter value β for five different estimators respectively, this is done for $\sigma_{12} \in \{0.1, 0.5, 0.95\}$. Here the assumption on the sign of π has been violated. Based on 100.000 simulation draws.

A.2 Reduction of parameter space

The following proof is an exact citation of Appendix E.1 in Andrews & Armstrong (2017), which discusses the dimension reduction of the parameter space discussed in Section 3 using an equivariance argument:

"*For comparisons between* $(\hat{\beta}_U, \hat{\beta}_{2SLS}, \hat{\beta}_{FULL})$ *in the just-identified case, it suffices to consider a two-dimensional parameter space. To see that this is the case, let* $\theta = (\beta, \pi, \sigma_1^2, \sigma_{12}, \sigma_2^2)$ *be the vector of model parameters and let* $A = \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix}$, $a_1 \neq 0, a_3 > 0$, *be the transformation*

$$g_A\xi = \tilde{\xi} = A\begin{pmatrix} \xi 1 \\ \xi 2 \end{pmatrix} = \begin{pmatrix} a_1\xi_1 + a_2\xi_2 \\ a_3\xi_2 \end{pmatrix},$$

which leads to $\tilde{\xi}$ *being distributed according to the parameters*

$$\tilde{\theta} = (\tilde{\beta}, \tilde{\pi}, \tilde{\sigma}_1^2, \tilde{\sigma}_{12}, \tilde{\sigma}_2^2),$$

where

$$\begin{aligned} \tilde{\beta} &= \frac{a_1\beta + a_2}{a_3}, \\ \tilde{\pi} &= a_3\pi, \\ \tilde{\sigma}_1^2 &= a_1^2\sigma_1^2 + a_1a_2\sigma_{12} + a_2^2\sigma_2^2, \\ \tilde{\sigma}_{12} &= a_1a_3\sigma_{12} + a_2a_3\sigma_2^2, \end{aligned}$$

and

$$\tilde{\sigma}_2^2 = a_3^2\sigma_2^2.$$

Define $\mathcal{G}$ *as the set of all transformations* g_A *of the form above. Note that the sign restriction on* π *is preserved under* $g_A \in \mathcal{G}$, *and that for each* g_A, *there exists another transformation* $g_A^{-1} \in \mathcal{G}$ *such that* $g_A g_A^{-1}$ *is the identity transformation. We can see that the model (2) is invariant under the transformation* g_A. *Note further that the estimators* $\hat{\beta}_U, \hat{\beta}_{2SLS}$, *and* $\hat{\beta}_{FULL}$ *are all equivariant under* g_A, *in the sense that*

$$\hat{\beta}(g_A\xi) = \frac{a_1\hat{\beta}(\xi) + a_2}{a_3}.$$

Thus, for any properties of these estimators (e.g., relative mean and median bias, relative dispersion) that are preserved under the transformations g_A, *it suffices to study these properties on the reduced parameter space obtained by equivariance. By choosing A appropriately,* *we can always obtain*

$$\begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \tilde{\pi} \end{pmatrix}, \begin{pmatrix} 1 & \tilde{\sigma}_{12} \\ \sigma_{12} & 1 \end{pmatrix}\right)$$

for $\tilde{\pi} > 0, \sigma_{12} \geq 0$, *and thus reduce to a two-dimensional parameter* (π, σ_{12}) *with* $\sigma_{12} \in [0, 1), \pi > 0$"