Statistics and Probability Letters 115 (2016) 54–59

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Regressor and disturbance have moments of all orders, least squares estimator has none[☆]

Kenneth D. West[a], Zifeng Zhao[b,*]

[a] Department of Economics, University of Wisconsin-Madison, United States

[b] Department of Statistics, University of Wisconsin-Madison, United States

ARTICLE INFO

Article history:
Received 8 January 2016
Received in revised form 26 March 2016
Accepted 30 March 2016
Available online 11 April 2016

Keywords:
Least squares bias
Mean squared error
Negative moments

ABSTRACT

We construct an example in which the least squares estimator has unbounded bias and no moments, even though the regressor and disturbance have moments of all orders and are independently distributed across observations.

1. Introduction

In this note, we construct a univariate example in which the least squares estimator has unbounded bias and no moments, even though the regressor and disturbance have moments of all orders and are independently distributed across observations. As well, the regressor is continuously distributed, so our example does not turn on a possible inability to invert the moment matrix of the regressor.

Our example has two motivations. The first comes from consideration of bias or mean squared prediction error in least squares regressions. Let X be the $T \times k$ design matrix, where T is the sample size and k the number of regressors. Then $X'X$ is the $k \times k$ moment matrix of the regressors. To insure integrability of the least squares estimator, literature on least squares bias or mean squared prediction error sometimes makes a high level assumption on the existence of moments of $(X'X)^{-1}$ (e.g., Shaman, 2010) or implicitly assumes integrability without making a specific assumption (e.g., Bao and Ullah, 2010). Our paper contributes by illustrating that innocuous assumptions about dependence and moments do not suffice to insure integrability. The second motivation relates to the existence of negative moments. Research has given a necessary and sufficient condition for the existence of a negative first moment for a given random variable (Khuri and Casella, 2002) and, in the least squares context, a sufficient condition for existence of moments of $(X'X)^{-1}$ (Findley and Wei, 2002). Our paper contributes by supplying a weaker sufficient condition than in Findley and Wei (2002), although ours, unlike Findley and Wei's, is only applicable for a univariate regression setup.

In Section 2, we give a necessary and sufficient condition for the existence of moments of $(X'X)^{-1}$ for a univariate regression model. We then present a sufficient condition for the existence of moments of the least square estimator. In Section 3, we construct a counterexample in which the least squares estimator has unbounded bias and no moments, even though the regressor and disturbance have moments of all orders. Section 4 has some comments.

☆ We thank two referees for helpful comments. West thanks the National Science Foundation grant SES-1127527 for financial support.

* Correspondence to: 1300 University Avenue, Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, United States.
E-mail addresses: kdwest@wisc.edu (K.D. West), zifeng.zhao@wisc.edu (Z. Zhao).

http://dx.doi.org/10.1016/j.spl.2016.03.021
0167-7152/

2. Lemmas on the existence of moments

Write the regression as

$$y_t = x_t\beta + \eta_t, \quad t = 1, \dots T. \tag{1}$$

where all variables are scalars; y_t and x_t are observed, η_t is the unobservable disturbance, and β is the parameter of interest. Thus the least squares estimator is $\hat{\beta}_T = (\sum_{t=1}^T x_t^2)^{-1}(\sum_{t=1}^T x_t y_t) = \beta + (\sum_{t=1}^T x_t^2)^{-1}(\sum_{t=1}^T x_t\eta_t)$. (We omit the usual $\frac{1}{T}$ normalization of each of the sums to keep notation uncluttered.) In the notation of the introduction, $k = 1$ and $\sum_{t=1}^T x_t^2$ corresponds to $X'X$.

From Holder's inequality, for any $\varepsilon > 0$,

$$\left|E(\hat{\beta}_T - \beta)\right| \leq \left[E\left|\left(\sum_{t=1}^T x_t^2\right)^{-1}\right|^{(1+\varepsilon)}\right]^{\frac{1}{1+\varepsilon}} \times \left[E\left|\left(\sum_{t=1}^T x_t\eta_t\right)\right|^{\frac{1+\varepsilon}{\varepsilon}}\right]^{\frac{\varepsilon}{1+\varepsilon}}.$$

Hence if x_t and η_t have moments of all orders (as we shall shortly assume), a necessary condition for $E\left|\hat{\beta}_T - \beta\right| = \infty$ is that $(\sum_{t=1}^T x_t^2)^{-1}$ has no moment of order higher than one, and a sufficient condition for $E\left|\hat{\beta}_T - \beta\right| < \infty$ is that there exists $\varepsilon > 0$ s.t. $E(\sum_{t=1}^T x_t^2)^{-(1+\varepsilon)} < \infty$.

Intuitively, in order for $E(\sum_{t=1}^T x_t^2)^{-1} = \infty$, one wants $(\sum_{t=1}^T x_t^2)^{-1}$ to have a fat tail, i.e., one wants $\sum_{t=1}^T x_t^2$ to have high probability of being near zero. The following Lemma 1 is central to our construction of x_t that satisfies this criterion, delivering a sufficient and necessary condition for $(\sum_{t=1}^T x_t^2)^{-1}$ to have moments of order s.

Lemma 1. *Let δ_t be a sequence of parameters, $0 < \delta_t < 1, t = 1, 2, \dots, T$. For a sequence of independent random variables $z_t, t = 1, \dots, T$, let the cumulative distribution function (cdf) of z_t be $P(z_t \leq z) = z^{\delta_t}, 0 < z \leq 1$. Then for any $s > 0$, a sufficient and necessary condition for $E(\sum_{t=1}^T z_t)^{-s} = \infty$ is $\sum_{t=1}^T \delta_t \leq s$ for some $s > 0$.*

Proof. To show $\sum_{t=1}^T \delta_t \leq s \Rightarrow E(\sum_{t=1}^T z_t)^{-s} = \infty$, let $c > 0$ be an arbitrary positive constant. By Markov's inequality,

$$E\left(\frac{1}{\sum_{t=1}^T z_t}\right)^s \geq E\left(\frac{1}{\left(\sum_{t=1}^T z_t\right)^s} \cdot I\left(\frac{1}{\left(\sum_{t=1}^T z_t\right)^s} > c\right)\right) \geq P\left[\frac{1}{\left(\sum_{t=1}^T z_t\right)^s} \geq c\right] \cdot c.$$

Further,

$$\begin{aligned}
P\left[\frac{1}{\left(\sum_{t=1}^T z_t\right)^s} \geq c\right] \cdot c &= P\left[\sum_{t=1}^T z_t \leq \frac{1}{c^{1/s}}\right] \cdot c \\
&\geq c \times P\left(z_1 \leq \frac{1}{2c^{1/s}}, z_2 \leq \frac{1}{2^2c^{1/s}}, \dots, z_T \leq \frac{1}{2^Tc^{1/s}}\right) \\
&= c \times P\left(z_1 \leq \frac{1}{2c^{1/s}}\right) P\left(z_2 \leq \frac{1}{4c^{1/s}}\right) \cdots P\left(z_T \leq \frac{1}{2^Tc^{1/s}}\right) \\
&= c\left(\frac{1}{2c^{1/s}}\right)^{\delta_1}\left(\frac{1}{2^2c^{1/s}}\right)^{\delta_2} \cdots \left(\frac{1}{2^Tc^{1/s}}\right)^{\delta_T} = 2^{-\sum_{t=1}^T t\delta_t} \cdot c^{1-\sum_{t=1}^T \delta_t/s}.
\end{aligned}$$

The equality on the third line follows from independence of the z_t's.

For $\sum_{t=1}^T \delta_t < s$, since c can be made arbitrarily large, the desired result follows. For $\sum_{t=1}^T \delta_t = s$, by the above equation, we have

$$E\left(\frac{1}{\left(\sum_{t=1}^T z_t\right)^s} \cdot I\left(\frac{1}{\left(\sum_{t=1}^T z_t\right)^s} > c\right)\right) \geq 2^{-\sum_{t=1}^T t\delta_t}, \quad \forall c > 0,$$

which is obviously not true if $E(\sum_{t=1}^T z_t)^{-s} < \infty$, so we have $E(\sum_{t=1}^T z_t)^{-s} = \infty$.

The converse ($\sum_{t=1}^{T}\delta_t > s \Rightarrow E(\sum_{t=1}^{T} z_t)^{-s} < \infty$) follows from an argument similar to one made in Sriram and Bose (1988) for constant δ. Details are in the Appendix. □

The key to Lemma 1 is that the cdf of z_t is z^{δ_t} in a neighborhood of the origin (equivalently, the pdf (probability density function) of z_t is z^{δ_t-1} in that neighborhood); the result follows even if we relax the condition to only require $P(z_t \le z) = O(z^{\delta_t})$ as $z \to 0$.

Using Lemma 1, we give a sufficient condition for $E\left|\hat{\beta}_T - \beta\right| < \infty$:

Lemma 2. *Let $\{(x_t, y_t)\}_{t=1}^{\infty}$ come from model (1) with observations are independent over time t. Assume that x_t and η_t have moments of all orders and $\forall t,\ P(x_t^2 \le x) = O(x^{\delta_t})$ as $x \to 0$, then we have $E\left|\hat{\beta}_T - \beta\right| < \infty$ for all T such that $\sum_{t=1}^{T}\delta_t > 1$.*

Proof. As mentioned above, since $\sum_{t=1}^{T} x_t\eta_t$ has moments of all orders, by Holder's inequality, the sufficient condition for $E\left|\hat{\beta}_T - \beta\right| < \infty$ is that there exists an $\varepsilon > 0$ s.t. $E(\sum_{t=1}^{T} x_t^2)^{-(1+\varepsilon)} < \infty$.

Since $\sum_{t=1}^{T}\delta_t > 1$, we have that there exists an $\varepsilon > 0$ s.t. $\sum_{t=1}^{T}\delta_t > 1+\varepsilon$. By assumption, $\forall t,\ P(x_t^2 \le x) = O(x^{\delta_t})$ as $x \to 0$, so by applying Lemma 1 at $s = 1+\varepsilon$, we have $E(\sum_{t=1}^{T} x_t^2)^{-(1+\varepsilon)} < \infty$. □

3. A counterexample

In this section, we shall construct an example in which least squares has unbounded bias, i.e., $E(\hat{\beta}_T - \beta) = \infty$, which implies as well that $\hat{\beta}_T$ has no moments.

Our process for x_t in (1) will behave like z_t when $x_t > 0$, but when $x_t < 0$ will have no mass in a neighborhood of zero. To simplify notation, our example sets $\theta_t = 1 - \delta_t$.

The joint distribution of (x_t, η_t) in our example is as follows. Suppose $\{x_t, \eta_t\}_{t=1}^{\infty}$ are independent and not identically distributed (i.n.i.d.) random variables, where the pdf of x_t's is defined as:

$$f_t(x) = \begin{cases} x^{-\theta_t} & \text{if } 0 < x \le c(\theta_t) \\ \dfrac{1}{2K(\theta_t)} & \text{if } -\dfrac{5}{2}K(\theta_t) < x \le -\dfrac{3}{2}K(\theta_t), \\ 0 & \text{otherwise} \end{cases}$$

where $0 < \theta_t < 1$, and

$$\int_0^{c(\theta_t)} x^{-\theta_t} = \frac{1}{2}, \quad \text{i.e. } c(\theta_t) = \left(\frac{1-\theta_t}{2}\right)^{\frac{1}{1-\theta_t}},$$

and

$$K(\theta_t) = \int_0^{c(\theta_t)} x \cdot f_t(x)dx = E(x_t \cdot I(x_t > 0)).$$

Thus, by construction $E(x_t \cdot I(x_t < 0)) = -K(\theta_t)$. So we have $E(x_t) = 0$ and median of x_t is 0.

Based on x_t, the conditional distribution of η_t is defined as

$$\begin{cases} \text{if } x_t > 0, \quad P(\eta_t = \eta | x_t) = \begin{cases} p_t & \text{if } \eta = 1 \\ 1 - p_t & \text{if } \eta = -x_t, \end{cases} \\ \text{if } x_t < 0, \quad P(\eta_t = \eta | x_t) = \begin{cases} \dfrac{1}{2} & \text{if } \eta = c_t \\ \dfrac{1}{2} & \text{if } \eta = d_t \end{cases} \end{cases}$$

where non-stochastic p_t, c_t and d_t will be defined so that $E(\eta_t) = 0$ and $E(\eta_t x_t) = 0$. That is,

$$\begin{aligned} E(\eta_t) &= E(E(\eta_t|x_t)) \\ &= E((p_t - (1-p_t)x_t) \cdot I(x_t > 0) + \frac{c_t + d_t}{2} \cdot I(x_t < 0)) \\ &= p_t - 2(1-p_t)K(\theta_t) + \frac{c_t + d_t}{2} = 0, \end{aligned}$$

and

$$E(\eta_t x_t) = E(x_t E(\eta_t | x_t))$$
$$= E((p_t - (1-p_t)x_t) \cdot x_t \cdot I(x_t > 0) + \frac{c_t + d_t}{2} \cdot x_t \cdot I(x_t < 0))$$
$$= p_t K(\theta_t) - (1-p_t)K_2(\theta_t) - \frac{c_t + d_t}{2} K(\theta_t) = 0,$$

where $K_2(\theta_t) = E(x_t^2 \cdot I(x_t > 0))$.

Solving the above equations, we can easily get

$$p_t = \frac{2K(\theta_t)^2 + K_2(\theta_t)}{2K(\theta_t)^2 + K_2(\theta_t) + 2K(\theta_t)},$$

and

$$c_t + d_t = 4(1-p_t)K(\theta_t) - 2p_t,$$

and we choose arbitrary c_t and d_t such that $c_t + d_t$ satisfies this last equation.

In other words, for any sequence $\{\theta_t\}_{t=1}^{\infty}$ where $0 < \theta_t < 1\ \forall t$, there exists $\{p_t, c_t, d_t\}_{t=1}^{\infty}$ such that $\{x_t, \eta_t\}_{t=1}^{\infty}$ has the distributions described above with $E(x_t) = 0$, $E(\eta_t) = 0$ and $E(\eta_t x_t) = 0$ and median of x_t is 0.

Suppose we have (1), with $\hat{\beta}_T - \beta = (\sum_{t=1}^{T} x_t^2)^{-1}(\sum_{t=1}^{T} x_t \eta_t)$. In the following, for simplicity we set $\beta = 0$.

Theorem 1. *Assume that* $\sum_{t=1}^{T} \delta_t = \sum_{t=1}^{T}(1-\theta_t) \leq 1$, *then we have*

$$E\left[(\hat{\beta}_T - \beta)^+\right] = E(\hat{\beta}_T^+) = \infty \quad \textit{and} \quad E\left[(\hat{\beta}_T - \beta)^-\right] = E(\hat{\beta}_T^-) < \infty,$$

so,

$$E(\hat{\beta}_T - \beta) = E(\hat{\beta}_T) = E(\hat{\beta}_T^+) - E(\hat{\beta}_T^-) = \infty,$$

where $(\hat{\beta}_T - \beta)^+ = \max\{\hat{\beta}_T - \beta, 0\}$ *and* $(\hat{\beta}_T - \beta)^- = \max\{-(\hat{\beta}_T - \beta), 0\}$.

Proof. For the positive part,

$$E(\hat{\beta}_T^+) = E\left(\frac{\sum_{t=1}^{T} \eta_t x_t}{\sum_{t=1}^{T} x_t^2}\right)^+ \geq E\left(\frac{\left(\sum_{t=1}^{T} \eta_t x_t\right)^+}{\sum_{t=1}^{T} x_t^2} \cdot \prod_{t=1}^{T} I(x_t > 0)\right)$$
$$\geq E\left(\frac{\left(\sum_{t=1}^{T} \eta_t x_t\right)^+}{\sum_{t=1}^{T} x_t^2} \cdot \prod_{t=1}^{T} I(x_t > 0) \cdot \prod_{t=1}^{T} I(\eta_t = 1)\right) = E\left(\frac{\left(\sum_{t=1}^{T} \eta_t x_t\right)^+}{\sum_{t=1}^{T} x_t^2} \cdot \prod_{t=1}^{T} I(x_t > 0, \eta_t = 1)\right)$$
$$= \int_0^{c(\theta_T)} \cdots \int_0^{c(\theta_1)} \frac{\sum_{t=1}^{T} x_t}{\sum_{t=1}^{T} x_t^2} \cdot \prod_{t=1}^{T} (p_t \cdot f_t(x_t))\, dx_1 \cdots dx_T$$
$$= \prod_{t=1}^{T} p_t \cdot E\left(\frac{\sum_{t=1}^{T} x_t}{\sum_{t=1}^{T} x_t^2} \cdot \prod_{t=1}^{T} I(x_t > 0)\right) \geq \prod_{t=1}^{T} p_t \cdot E\left(\frac{1}{\sum_{t=1}^{T} x_t} \cdot \prod_{t=1}^{T} I(x_t > 0)\right) = \infty.$$

The last inequality follows because $(\sum_{t=1}^{T} x_t)^2 \cdot \prod_{t=1}^{T} I(x_t > 0) \geq (\sum_{t=1}^{T} x_t^2) \cdot \prod_{t=1}^{T} I(x_t > 0)$. The second to last equality follows from the definition of joint distribution of (x_t, η_t) and the last equality follows from Lemma 1.

For the negative part,

$$E(\hat{\beta}_T^-) = E\left(\frac{\sum_{t=1}^T \eta_t x_t}{\sum_{t=1}^T x_t^2}\right)^- = E\left(\frac{\left(\sum_{t=1}^T \eta_t x_t\right)^-}{\sum_{t=1}^T x_t^2}\right) \le E\left(\frac{\sum_{t=1}^T (\eta_t x_t)^-}{\sum_{t=1}^T x_t^2}\right)$$

$$= \sum_{t=1}^T E\left(\frac{(\eta_t x_t)^-}{\sum_{t=1}^T x_t^2}\right) \le \sum_{t=1}^T E\left(\frac{(\eta_t x_t)^-}{x_t^2}\right) = \sum_{t=1}^T E\left(\frac{\eta_t}{x_t}\right)^- < \infty.$$

The last equality follows by explicitly writing the expectation $E(\cdot)$ out. □

4. Further discussion

Some comments:

1. In our example, since x_t and η_t each has compact support, each has moments of all orders.
2. The assumption that the median of x_t is zero is to simplify the algebra. We could instead choose a value $c(\theta_t)$ such that a given quantile of x_t is zero.
3. Similarly, it was for simplicity that we used discreteness in part of the conditional distribution of η_t given x_t. Our result still holds if η_t is continuous around each of the discrete values used above.
4. Much of the complication of the example is because we set ourselves the goals of $E(\hat{\beta}_T - \beta) = \infty$, which requires $E(\hat{\beta}_T - \beta)^+ = \infty$ and $E(\hat{\beta}_T - \beta)^- < \infty$. Had we aimed solely for $E|\hat{\beta}_T - \beta| = \infty$, we could have accomplished this with the same process for x_t but with η_t simply an i.i.d. Bernoulli process.
5. In our example, as $x \to 0$, $P(x_t^2 \le x) = P(|x_t| \le \sqrt{x}) = \frac{1}{1-\theta_t} x^{(1-\theta_t)/2} = O(x^{-\delta_t/2})$. By construction, we have $\sum_{t=1}^T \delta_t \le 1$, so $\sum_{t=1}^T \delta_t/2 \le \frac{1}{2}$. It obviously violates the sufficient condition in Lemma 2, which requires $\sum_{t=1}^T \delta_t/2 > 1$.
6. In the notation of Lemma 1, Sriram and Bose (1988) considered $E(\sum_{t=1}^T z_t)^{-1}$ with constant δ (i.e., $\delta_t = \delta < 1$ for all t). They showed $E(\sum_{t=1}^T z_t)^{-1} = \infty$ for T sufficiently small, $E(\sum_{t=1}^T z_t)^{-1} < \infty$ for T sufficiently large. As in Stuttgen (2014), our Lemma 1 allows the density of z_t to shift with t, with δ_t becoming smaller and the density becoming steeper as t increases. The first part of the proof of Lemma 1 is completely different from the proof in Sriram and Bose (1988).
7. As noted below Lemma 1, we only require a symmetric Lipschitz condition of $P(x_t^2 \le x)$ at origin. This contrasts with the stronger assumption in Findley and Wei (2002), which, when specialized to the univariate case, imposes a uniform Lipschitz condition on $\mathbb{R}$.

Appendix

Details on proof of Lemma 1:
Here is the proof that $\sum_{t=1}^T \delta_t > s \Rightarrow E(\sum_{t=1}^T z_t)^{-s} < \infty$.
Assume $\sum_{t=1}^T \delta_t > s$. We mimic the argument made in Sriram and Bose (1988) for constant δ.
Let $S_T = \sum_{t=1}^T z_t$. Then

$$E(1/S_T^s) = \int_0^\infty P[1/S_T^s \ge x]dx = \int_0^\infty P[e^{-xS_T^s} \ge e^{-1}]dx \le e\int_0^\infty E[e^{-x^{1/s}S_T}]dx$$

$$= e\left(\int_0^1 + \int_1^\infty\right)\prod_{t=1}^T E[e^{-x^{1/s}z_t}]dx,$$

where the inequality follows from Markov inequality and the last equality follows from the independence of z_t's. Clearly $\int_0^1 \prod_{t=1}^T E[e^{-x^{1/s}z_t}]dx < \infty$. For $1 < x < \infty$, we have,

$$Ee^{-x^{1/s}z_t} = \int_0^1 P[e^{-x^{1/s}z_t} \ge y]dy = \int_0^1 P[z_t \le -(\log y)/x^{1/s}]dy$$

$$= x^{1/s}\int_0^\infty P[z_t \le z]e^{-zx^{1/s}}dz \le x^{1/s}\int_0^\infty z^{\delta_t}e^{-zx^{1/s}}dz = C \cdot x^{-\delta_t/s},$$

where the third and last equality follow for change of variable and the inequality follows from the fact $P(z_t \le z) = z^{\delta_t} < 1$. Here C is a generic constant.

All together, we have

$$E\left(\sum_{t=1}^{T} z_t\right)^{-s} \le e\left(\int_0^1 + \int_1^\infty\right)\prod_{t=1}^{T} E[e^{-x^{1/s} z_t}]dx \le C_1 + C\int_1^\infty x^{-\sum_{t=1}^{T}\delta_t/s}dx < \infty,$$

since $\sum_{t=1}^{T}\delta_t > s$. □

References

Bao, Yong, Ullah, Aman, 2010. Expectation of quadratic forms in normal and nonnormal variables with applications. J. Statist. Plann. Inference 140, 1193–1205.
Findley, David F., Wei, Ching-Zong, 2002. AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. J. Multivariate Anal. 83, 415–450.
Khuri, André, Casella, George, 2002. The existence of the first negative moment revisited. Amer. Statist. 56, 44–47.
Shaman, Paul, 2010. Generalized Levinson–Durbin sequences, binomial coefficients and autoregressive estimation. J. Multivariate Anal. 101, 1263–1273.
Sriram, T.N., Bose, Arup, 1988. Sequential shrinkage estimation in the general linear model. Sequential Anal. 7, 149–163.
Stuttgen, Luke, 2014. Considerations of Unbounded Bias in Least Squares Estimators (senior honors thesis), University of Wisconsin, (unpublished).