The Existence of Inefficiency: LASSO+SFA

Christopher F. Parmeter[1] Artem Prokhorov[2]
Valentin Zelenyuk[3]

[1]Miami Herbert Business School
[2]University of Sydney Business School
[3]School of Economics and Centre for Efficiency and Productivity Analysis

November 7th, 2024

From the End

- Combine machine learning with stochastic frontier analysis.
- Establish moment/parameter redundancy for use of post double LASSO with MLE.
- Simple and effective step-wise estimator that preserves efficiency and valid inference.

X-inefficiency

ALLOCATIVE EFFICIENCY VS. "X-EFFICIENCY"

By HARVEY LEIBENSTEIN*

At the core of economics is the concept of efficiency. Microeconomic theory is concerned with allocative efficiency. Empirical evidence has been accumulating that suggests that the problem of allocative efficiency is trivial. Yet it is hard to escape the notion that efficiency in some broad sense is significant. In this paper I want to review the empirical evidence briefly and to consider some of the possible implications of the findings, especially as they relate to the theory of the firm and to the explanation of economic growth. The essence of the argument is that microeconomic theory focuses on allocative efficiency to the exclusion of other types of efficiencies that, in fact, are much more significant in many instances. Furthermore, improvement in "nonallocative efficiency" is an important aspect of the process of growth.

X-inefficiency?

The Xistence of X-Efficiency

By George J. Stigler*

Harvey Leibenstein called attention in an influential article (1966) to a source of economic inefficiency which was given the awful name of X-[in]efficiency. He cited studies in which misallocations of resources due to monopoly or tariffs had trifling social costs, whereas simple failure to attain the production frontier apparently led to social losses of a vastly greater magnitude. I propose to argue that this type of inefficiency can usefully be assimilated into the traditional theory of allocative inefficiency.

It is a question (to be discussed below) whether one ascribes failures to reach the ultimate limits of output from given inputs in any state of technology to inadequacy of knowledge alone, or adds also inadequate "motivation." Leibenstein (1966) separates

> [F]or the same set of human inputs purchased and the same knowledge of production techniques available to the firm, a variety of output results are possible. If individuals can choose, to some degree, the APQT bundles [choice of Activity, Pace, Quality of work, Time spent] they like, they are unlikely to choose a set of bundles that will maximize the value of output. [p. 768]

If management seeks to impose output-maximizing *APQT* bundles on the workers, indeed, these assignments of tasks would likely be " . . . less efficient than those that individuals would choose themselves under an acceptable set of [managerial] restraints" (p. 769).

In this case, and in every motivational

X-Inefficiency Xists—Reply to an Xorcist

By HARVEY LEIBENSTEIN*

Under the title "The Xistence of X-Efficiency," George J. Stigler (1976) wrote a critique of X-efficiency theory, indicated his distaste for the concept, and urged economists to abandon the idea. I will argue that this exercise in exorcism is just that. It achieves some of its results by unusual redefinitions of ordinary concepts. In the end it makes nonscientific appeals. Hence, this plea for exorcism should be ignored. At the same time, I am grateful to Stigler. As a by-product of his attack he has raised some points that others have raised orally. This provides an opportunity to clarify some issues.

Stigler makes two essential points: 1) 1975) probably available to Stigler, and in a recent book (1976), I developed such a framework. I shall refer to this larger framework as general X-efficiency theory. Table 1 contrasts the neoclassical model and general X-efficiency theory. Under the latter the neoclassical model can be included as a special case.

With the aid of the concepts developed below I shall try to show that there is nothing in the operation of an economy that is inconsistent with the existence of X-inefficiency, nor does competition necessarily lead to its elimination. Space constraints permit me to do little more than suggest the nature of the basic arguments.

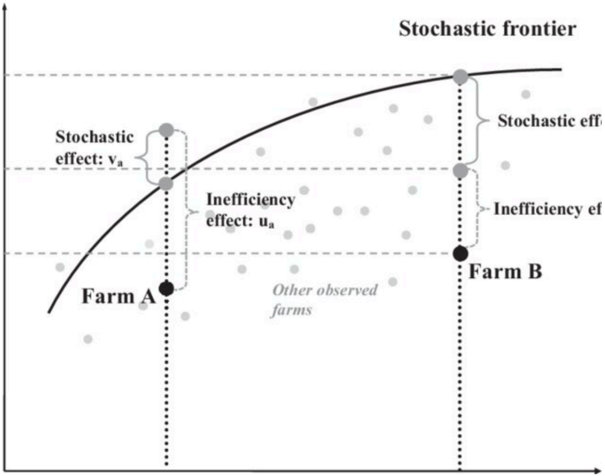
Selective Attention

LEARNING THROUGH NOTICING: THEORY AND EVIDENCE FROM A FIELD EXPERIMENT*

REMA HANNA
SENDHIL MULLAINATHAN
JOSHUA SCHWARTZSTEIN

We consider a model of technological learning under which people "learn through noticing": they choose which input dimensions to attend to and subsequently learn about from available data. Using this model, we show how people with a great deal of experience may persistently be off the production frontier because they fail to notice important features of the data they possess. We also develop predictions on when these learning failures are likely to occur, as well as on the types of interventions that can help people learn. We test the model's predictions in a field experiment with seaweed farmers. The survey data reveal that these farmers do not attend to pod size, a particular input dimension. Experimental trials suggest that farmers are particularly far from optimizing this dimension. Furthermore, consistent with the model, we find that simply having access to the experimental data does not induce learning. Instead, behavioral changes occur only after the farmers are presented with summaries that highlight previously unattended-to relationships in the data. *JEL* Codes: D03, D83, O13, O14, O30, Q16.

Stochastic Frontier Analysis (SFA)



The Stochastic Frontier Model

The stochastic frontier model we consider in this paper can be written as follows:

$$\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{v} - \boldsymbol{u} = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y}$ is an n-vector of output, $\boldsymbol{x}$ is an $p \times 1$ vector of production inputs including a constant, $\boldsymbol{\varepsilon} = \boldsymbol{v} - \boldsymbol{u}$ is the n-vector of error terms ε_i composed of a Normal part $v_i \sim N(0, \sigma_v^2)$ and a Half-Normal inefficiency component $u_i \sim N_+(0, \sigma_u^2)$.

- Aside from presence of u_i this is a trivial model to estimate.
- But we are interested in u_i.

ML Formlation of Frontier Model

$$y_i = \underbrace{\overbrace{\boldsymbol{x}_i'\boldsymbol{\beta}}^{inputs} + \overbrace{\boldsymbol{z}_i'\boldsymbol{\delta}}^{confounders} + v_i}_{\textit{stochastic frontier}} - u_i, \quad i = 1, \ldots, 2n. \tag{2}$$

- p (number of inputs) is small (and fixed).
- d (number of confounders), possibly large ($> 2n$).
- $\boldsymbol{\beta}$ can be estimated at $O(n^{-1/2})$ if $\boldsymbol{\delta}$ can be.
- Impossible to estimate $\boldsymbol{\delta}$ at this rate when d is large.

Double Machine Learning

- Consider estimation of a treatment effect (not a frontier model)

$$y_i = \underbrace{x_i\beta}_{\textit{scalar treatment}} + \underbrace{\boldsymbol{z}_i'\boldsymbol{\delta}}_{confounders} + v_i, \quad i = 1, \ldots, 2n.$$

1. Use any ML tool to predict $\mathbb{E}[y|z]$ and $\mathbb{E}[x|z]$, using half of the sample for each (hence the $2n$).
2. Obtain $\widehat{\beta}$ from the regression of $\tilde{y}$ on $\tilde{x}$ where $\tilde{w} = w - \widehat{\mathbb{E}[w|z]}$.

Double Machine Learning

- $\hat{\beta}$ is $\sqrt{n}$-consistent and asymptotically Normal even if RMSE of $\boldsymbol{z}'\boldsymbol{\delta}$ has rate $O(n^{-1/4})$ (so can estimate nonparametrically).
- The moment conditions for which $\hat{\beta}$ is constructed imply **Neyman Orthogonality**.

Neyman Orthogonality and M/P Redundancy

- In the context of asymptotically optimal testing, Neyman (1959) asked when do errors of nuisance functions **not** carry over into $\hat{\beta}$.

- Let δ denote the functional nuisance parameter and let $h_1^*(\beta, \delta)$ be the moment function implied by the FOC for $\hat{\beta}$:

$$\mathbb{E}[h_1^*(\beta, \delta)] = 0.$$

Neyman Orthogonality and M/P Redundancy

- We say $h_1^*(\cdot, \cdot)$ is Neyman orthogonal if the moment function remains valid under perturbations in δ:

$$\mathbb{D}_{12}[\delta - \delta_0] = \partial_\delta \mathbb{E}[h_1^*(\beta, \delta)][\delta - \delta_0] = 0. \tag{3}$$

- $\mathbb{D}_{12}[\delta - \delta_0]$ is the Gateaux derivative of the moment function in the direction δ around the true value δ_0.
- Neyman orthogonality is connected to and best understood in a GMM framework.

Neyman Orthogonality and M/P Redundancy

- GMM of (β, δ) is based on moment conditions assumed to hold in the population:

$$\text{[A] for } \beta: \quad \mathbb{E}[h_1(\beta, \delta)] = 0 \tag{4}$$

$$\text{[B] for } \delta: \quad \mathbb{E}[h_2(\beta, \delta)] = 0. \tag{5}$$

- We assume that [A] is enough to identify β given δ.
- Using more knowledge in the form of [B] and/or using δ_0 improves statistical efficiency asymptotically.
- Prokhorov and Schmidt (2009) asked when is it irrelevant for the estimation of β whether we know [B] and/or δ_0.

Neyman Orthogonality and M/P Redundancy

- Assume finite dimensional δ:

$$\text{[A] for } \beta: \quad \mathbb{E}[h_1(\beta, \delta)] = 0 \tag{6}$$

$$\text{[B] for } \delta: \quad \mathbb{E}[h_2(\beta, \delta)] = 0. \tag{7}$$

- When is it irrelevant for estimation of β whether we know [B] and/or δ_0?

- When asymptotic variance of GMM based on [A] with known δ is equal to asymptotic variance of GMM based on [A] and [B] with unknown δ.

Neyman Orthogonality and M/P Redundancy

$$C = \mathbb{E}\begin{bmatrix} h_1 h_1' & h_1 h_2' \\ h_2 h_1' & h_2 h_2' \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

and

$$D = \mathbb{E}\begin{bmatrix} \nabla_\beta h_1 & \nabla_\delta h_1 \\ \nabla_\beta h_2 & \nabla_\delta h_2 \end{bmatrix} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

M/P-Redundancy $\Leftrightarrow$ $C_{12} = 0$ **M**oment redundancy of [B]
$D_{12} = 0$ **P**arameter redundancy of δ

Neyman Orthogonality and M/P Redundancy

- So start by specifying

$$\text{[A] for } \beta: \quad \mathbb{E}[h_1(\beta, \delta)] = 0 \tag{8}$$
$$\text{[B] for } \delta: \quad \mathbb{E}[h_2(\beta, \delta)] = 0 \tag{9}$$

  and look for valid moment function $h_1^*(\beta, \delta)$ that is uncorrelated with $h_2(\cdot, \cdot)$ such that

$$D_{12} = \mathbb{E}\left[\nabla_\delta h_1^*(\beta, \delta)\right] = 0.$$

- The we can use any slowly converging ML tool (LASSO, GRF, etc.) to obtain $\hat{\delta}$, plug it into $h_1(\beta, \delta)$ and obtain a $\sqrt{n}$-consistent and asymptotically Normal $\hat{\beta}$.

Return to ML Formlation of Frontier Model

$$y_i = \underbrace{\overbrace{\boldsymbol{x}_i'\boldsymbol{\beta}}^{inputs} + \overbrace{\boldsymbol{z}_i'\boldsymbol{\delta}}^{confounders} + v_i}_{\textit{stochastic frontier}} - u_i, \quad i = 1, \ldots, 2n. \qquad (10)$$

- All ML tools give biased estimators.
- Inputs correlate with confounders: $x_i = m(z_i) + \eta_i$.
- Biases in $\hat{\boldsymbol{\delta}}$ and $\hat{m}(z_i)$ affect $\hat{\boldsymbol{\beta}}$ and $\hat{u}_i$.
- So what changes with the introduction of $u \geq 0$ into the model and what are the M/P redundant moments?

Return to ML Formlation of Frontier Model

$$y_i = \underbrace{\overbrace{\boldsymbol{x}_i'\boldsymbol{\beta}}^{inputs} + \overbrace{\boldsymbol{z}_i'\boldsymbol{\delta}}^{confounders} + v_i}_{\textit{stochastic frontier}} - u_i, \quad i = 1, \ldots, 2n.$$

- Conventional estimation (COLS) (assume $u \sim |N(0, \sigma_u^2)|$ and v is symmetric and $\mathbb{E}[v] = 0$):

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\delta}} \end{pmatrix} = \min_{\boldsymbol{\beta},\boldsymbol{\delta}} \sum_{i=1}^{2n} \left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{z}_i'\boldsymbol{\delta}\right)^2$$

  accounting for $\mathbb{E}[u_i] = \sqrt{\frac{2}{\pi}}\sigma_u > 0$.

- Evidence of inefficiency is captured through negative skewness of the residuals $\hat{\varepsilon}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{z}_i'\hat{\boldsymbol{\delta}}$.

Return to ML Formlation of Frontier Model

$$y_i = \underbrace{\overbrace{\boldsymbol{x}_i'\beta}^{inputs} + \overbrace{\boldsymbol{z}_i'\delta}^{confounders} + v_i}_{\textit{stochastic frontier}} - u_i, \quad i = 1, \ldots, 2n.$$

- Conventional estimation (maximum likelihood) (assume $u \sim |N(0, \sigma_u^2)|$, $v \sim N(0, \sigma_u^2)$ and $u \perp v$):

$$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} \ln \mathcal{L}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} = \left(\hat{\beta}, \hat{\delta}, \hat{\sigma}_v^2, \hat{\sigma}_u^2\right).$$

- Evidence of inefficiency: $\sigma_u^2 >> 0$.

Does Inefficiency Exist?

$$y_i = 1 + 0.3x_{1i} + 0.4x_{2i} + 0.38x_{3i} + \sum_{j=1}^{d} \delta_j z_{ij} + v_i - u_i$$

$$\text{True } \delta_j = 0, \quad z_{ij} \sim N(0,1), \quad d = cn,$$

$$v_i \sim N(0, 0.5), \quad u_i \sim |N(0, 1.2)|$$

Does Inefficiency Exist?

Average skewness of OLS residuals over 1,000 simulations

n	0	0.01	0.1	0.2	0.3	0.5	0.9
100	-0.494	-0.488	-0.420	-0.342	-0.267	-0.143	-0.001
200	-0.525	-0.517	-0.445	-0.375	-0.299	-0.177	-0.011
400	-0.536	-0.530	-0.454	-0.380	-0.308	-0.186	-0.012
800	-0.547	-0.539	-0.466	-0.391	-0.319	-0.193	-0.016
1,600	-0.549	-0.542	-0.468	-0.391	-0.319	-0.189	-0.016

Resort to ML? - Post-Single-LASSO

- LASSO $\Rightarrow$ some elements of $\hat{\delta}_{LASSO}$ are exactly 0; drop these confounders

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{LASSO} \\ \hat{\boldsymbol{\delta}}_{LASSO} \end{pmatrix} = \min_{\boldsymbol{\beta},\boldsymbol{\delta}} \sum_{i=1}^{2n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{z}_i'\boldsymbol{\delta})^2 + \lambda \sum_{j=1}^{d} |\delta_j|$$

- COLS using only confounders picked by LASSO $\Rightarrow$ **PSL-COLS**

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{PSL} \\ \hat{\boldsymbol{\delta}}_{PSL} \end{pmatrix} = \min_{\boldsymbol{\beta},\boldsymbol{\delta}} \sum_{i=1}^{2n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{z}_i'\boldsymbol{\delta})^2 ,$$

$$\text{s.t. } \delta_j = 0 \text{ for any } j \notin \text{supp}\left(\hat{\delta}_{LASSO}\right)$$

Resort to ML? - Post-Single-LASSO

- LASSO $\Rightarrow$ some elements of $\hat{\delta}_{LASSO}$ are exactly 0; drop these confounders

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{LASSO} \\ \hat{\boldsymbol{\delta}}_{LASSO} \end{pmatrix} = \min_{\boldsymbol{\beta},\boldsymbol{\delta}} \sum_{i=1}^{2n} \left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{z}_i'\boldsymbol{\delta}\right)^2 + \lambda \sum_{j=1}^{d} |\delta_j|$$

- or MLE using only confounders picked by LASSO $\Rightarrow$ **PSL-MLE**

$$\hat{\boldsymbol{\theta}}_{PSL} = \max_{\boldsymbol{\theta}} \ln \mathcal{L}(\boldsymbol{\theta}), \quad \text{s.t. } \delta_j = 0 \text{ for any } j \notin \text{supp}\left(\hat{\boldsymbol{\delta}}_{LASSO}\right).$$

Inefficiency Exists!

Average skewness of PSL-OLS residuals over 1,000 simulations

n	0	0.01	0.1	0.2	0.3	0.5	0.9
100	-0.503	-0.404	-0.386	-0.374	-0.367	-0.359	-0.350
200	-0.520	-0.452	-0.436	-0.430	-0.425	-0.420	-0.413
400	-0.536	-0.479	-0.470	-0.465	-0.463	-0.459	-0.455
800	-0.546	-0.506	-0.500	-0.498	-0.497	-0.494	-0.492
1,600	-0.552	-0.522	-0.519	-0.517	-0.516	-0.516	-0.514

Another Problem: Inference for PSL-MLE

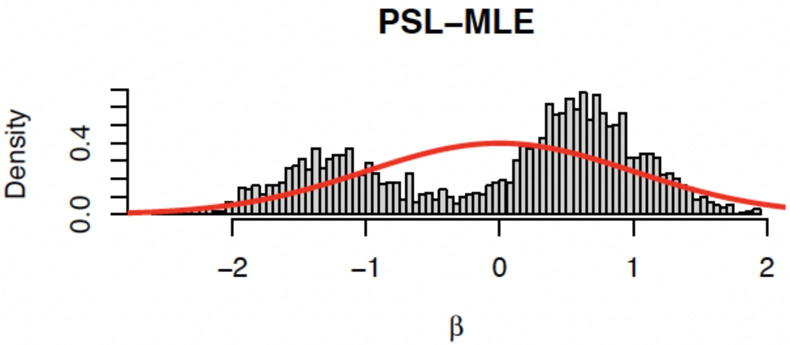
$$y_i = \beta x_i + 0.8 \sum_{j=1}^{200} \delta_j z_{ij} + v_i - u_i$$

True $\delta_j = (1/j)^2, \quad z_{ij} \sim N(0,1), \quad 2n = 100, \quad \lambda$ by CV,

$$v_i \sim N(0, 0.5), \quad u_i \sim |N(0, 1.2)|,$$

$$x_i = 0.6 \sum_{j=1}^{200} \delta_j z_{ij} + \eta_i, \quad \eta_i \sim N(0,1)$$

Sampling distribution of standardized $\hat{\beta}_{PSL}$ over 1,000 simulations



Why Does PSL Fail?

Look at MLE when $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + v_i - u_i$ for $v_i \sim N(0, \sigma_v^2) \perp u_i \sim |N(0, \sigma_u^2)|$

$$f_\varepsilon(\varepsilon_i) = \frac{2}{\sigma}\phi(\varepsilon_i/\sigma)\Phi(-\lambda\varepsilon_i/\sigma)$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u/\sigma_v$

$$\hat{\theta}_{MLE} = \max_{\boldsymbol{\theta}} \sum_{i=1}^{2n} \ln f_\varepsilon(\varepsilon_i), \quad \text{where } \varepsilon_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{z}_i'\boldsymbol{\delta}.$$

Recall: PSL zeros out some δ_js $\Rightarrow$ let δ_{LASSO} contain 0's for those j's, then

$$\xi_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta}_{LASSO} = \varepsilon_i + \boldsymbol{z}_i'(\boldsymbol{\delta} - \boldsymbol{\delta}_{LASSO}) \neq \varepsilon_i.$$

Why Does LASSO Break?

- For simplicity assume that $(\sigma, \lambda) = (1, 1)$, $d = \text{dim}(\delta) < 2n$ and define $r_i(\nu_i) = \dfrac{\phi(\nu_i)}{1 - \Phi(\nu_i)}$ (Inverse Mill's Ratio).

- Moment equations implied by FOCs from MLE:

$$\begin{aligned} [A] \text{ for } \beta: \quad & \mathbb{E}\left[\boldsymbol{x}_i' \left(\varepsilon_i + r_i(\varepsilon_i)\right)\right] = 0 \\ [B] \text{ for } \delta: \quad & \mathbb{E}\left[\boldsymbol{z}_i' \left(\varepsilon_i + r_i(\varepsilon_i)\right)\right] = 0. \end{aligned}$$

- Moment equations implied by FOCs from PSL-MLE:

$$\begin{aligned} [A] \text{ for } \beta: \quad & \mathbb{E}\left[\boldsymbol{x}_i' \left(\xi_i + r_i(\xi_i)\right)\right] = 0 \\ [B] \text{ for } \delta_{LASSO}: \quad & \mathbb{E}\left[\boldsymbol{z}_i' \left(\xi_i + r_i(\xi_i)\right)\right] = 0. \end{aligned}$$

PSL-MLE is using invalid moment conditions; LASSO regularization bias carries over to estimation of $\hat{\beta}$.

How to Conduct Valid Inference in SFA?

- Let $\tilde{\varepsilon}_i := \left(y_i - \pi_y' \boldsymbol{z}_i\right) - (\boldsymbol{x}_i - \pi_x' \boldsymbol{z}_i)' \beta - \boldsymbol{z}_i' \delta$.
- Consider the moment conditions

$$
\begin{aligned}
[A^*] \quad & \mathbb{E}\left[(\boldsymbol{x}_i - \pi_x' \boldsymbol{z}_i)' \left(\tilde{\varepsilon}_i + r_i(\tilde{\varepsilon}_i)\right)\right] = 0 \\
[B^*] \quad & \mathbb{E}\left[\boldsymbol{z}_i' \left(\tilde{\varepsilon}_i + r_i(\tilde{\varepsilon}_i)\right)\right] = 0 \\
[C] \quad & \mathbb{E}\left[\boldsymbol{z}_i' \left(\boldsymbol{x}_i - \pi_x' \boldsymbol{z}_i\right)\right] = 0 \\
[D] \quad & \mathbb{E}\left[\boldsymbol{z}_i' \left(y_i - \pi_y' \boldsymbol{z}_i\right)\right] = 0
\end{aligned}
$$

- Under homoskedasticity, $[A^*]$ satisfies Neyman orthogonality.
- Equivalently, $[B^*]$, $[C]$ and $[D]$ are M/P redundant for the estimation of β.

Sketch of the Argument

- Look at

$$[A^*] \quad \mathbb{E}\left[(\boldsymbol{x}_i - \pi_x' \boldsymbol{z}_i)' \left(\tilde{\varepsilon}_i + r_i(\tilde{\varepsilon}_i)\right)\right] = 0$$
$$[B^*] \quad \mathbb{E}\left[\boldsymbol{z}_i' \left(\tilde{\varepsilon}_i + r_i(\tilde{\varepsilon}_i)\right)\right] = 0$$

$[A^*] \perp [B^*] \Rightarrow C_{12} = 0.$

- Expected derivative:

$$\delta\colon\ \mathbb{E}\left[(\boldsymbol{x}_i - \pi_x' \boldsymbol{z}_i)' \left(\frac{\partial}{\partial \delta}\tilde{\varepsilon}_i + \frac{\partial}{\partial \delta} r_i(\tilde{\varepsilon}_i)\right)\right] =$$
$$\mathbb{E}\left[(\boldsymbol{x}_i - \pi_x' \boldsymbol{z}_i)' \left(-\boldsymbol{z}_i + \boldsymbol{z}_i r_i(\tilde{\varepsilon}_i)\left(\tilde{\varepsilon}_i + r_i(\tilde{\varepsilon}_i)\right)\right)\right] = 0.$$

Note

- Identical result holds for both π_x and π_y.
- The idea is similar to partialing out from Frisch-Waugh-Lovell.
- $[A^*]$ and $[B^*]$ correspond to running MLE where the dependent variable is the part of y_i that is orthogonal to $\boldsymbol{z}_i$ and the explanatory variables are $\boldsymbol{z}_i$ and the part of $\boldsymbol{x}_i$ that is orthogonal to $\boldsymbol{z}_i$.

Post-Double-LASSO

- LASSO of y_i on $\boldsymbol{z}_i$

$$\hat{\pi}^0_{LASSO} = \min_{\pi^0} \sum_{i=1}^{n} \left(y_i - \boldsymbol{z}_i' \pi^0\right)^2 + \lambda_0 \sum_{j=1}^{d} |\pi^0_j|.$$

- LASSO of $\boldsymbol{x}_i$ (one-by-one) on $\boldsymbol{z}_i$

$$\hat{\pi}^\ell_{LASSO} = \min_{\pi^\ell} \sum_{i=1}^{n} \left(x_{\ell i} - \boldsymbol{z}_i' \pi^\ell\right)^2 + \lambda_\ell \sum_{j=1}^{d} |\pi^\ell_j|.$$

Post-Double-LASSO

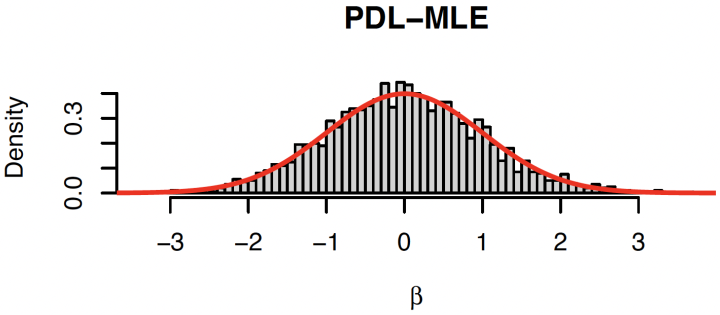
- MLE using the **union** of confounders picked by LASSO in the first two steps **PDL-MLE**

$$\hat{\boldsymbol{\theta}}_{PDL} = \max_{\boldsymbol{\theta}} \sum_{i=1}^{2n} \ln f_{\varepsilon}(\varepsilon_i),$$

$$\text{s.t. } \delta_j = 0 \text{ for any } j \notin I = \bigcup_{\ell=0}^{p} supp\left(\hat{\pi}^{\ell}_{LASSO}\right).$$

- I is called the **amelioration** set (Belloni, Chernozhukov and Hansen, 2013)

Sampling distribution of standardized $\hat{\beta}_{PDL}$ over 1,000 simulations



Empirical Example

- 137 dairy farms in Spain from 1999-2010 (Alvarez & Arias, 2004).
- y is milk production (liters).
- x is labor (man-equivalent units), cows, feed (kg), land (hectares) and roughage (expenses incurred to produce roughage: fertilizer, machines, seed, silage additives, etc.).

Empirical Example

- z is year dummies, zone dummies, land-ownership, bacteriological content of the milk, price of milk, price of feed, membership in an agricultural cooperative, milk quality indicators (fat, protein, somatic cell count), and something called AVGCOST (neither Antonio or Carlos could remember what this variable captured).
- $dim(z) = 50$ with first order terms [Cobb-Douglas]; $dim(z) = 87$ with second order terms [translog].

Empirical Example: Cobb-Douglas

	OLS	SFA	OLS Large	SFA Large	SFA-PSL	SFA-PDL
Feedstuffs	0.386	0.360	0.464	0.464	0.439	0.401
	0.012	0.013	0.011	0.011	0.011	0.013
Cows	0.595	0.642	0.467	0.467	0.546	0.560
	0.020	0.022	0.017	0.017	0.017	0.021
Land	−0.010	−0.012	0.032	0.032	0.007	0.033
	0.009	0.009	0.009	0.009	0.008	0.010
Labor	0.035	0.032	0.013	0.013	−0.015	0.005
	0.012	0.012	0.010	0.009	0.009	0.011
Roughage	0.067	0.060	0.073	0.073	0.082	0.061
	0.005	0.005	0.004	0.004	0.004	0.005
RTS	1.074	1.082	1.048	1.048	1.059	1.059
Eff	0.930	0.892	1.000	0.999	0.999	0.926

Empirical Example: Translog

	OLS	SFA	OLS Large	SFA Large	SFA-PSL	SFA-PDL
Feedstuffs	0.341	0.319	0.457	0.457	0.409	0.342
	0.014	0.014	0.013	0.013	0.013	0.014
Cows	0.633	0.676	0.454	0.454	0.574	0.618
	0.024	0.024	0.021	0.020	0.020	0.023
Land	-0.011	-0.017	-0.017	-0.017	-0.014	-0.013
	0.010	0.010	0.009	0.009	0.009	0.010
Labor	0.021	0.014	-0.007	-0.007	-0.033	0.000
	0.014	0.013	0.010	0.010	0.010	0.012
Roughage	0.093	0.088	0.122	0.122	0.126	0.079
	0.008	0.008	0.007	0.007	0.007	0.007
RTS	1.076	1.080	1.008	1.008	1.062	1.025
Eff	0.932	0.887	1.000	0.999	0.927	0.915

Concluding remarks

- Neyman orthogonality is key to ensuring valid causal inference; it is equivalent to M/P redundancy.
- Abundance of data makes it harder to establish and address inefficiency of production.
- Machine Learning tools are effective at reversing the spurious finding of full efficiency.
- Partialing out offers a way of conducting valid post-machine-learning causal inference.
- We derive and apply Neyman orthogonal moment conditions for production frontier models.