A Model-Free Theory of Parametric Regression

Andreas Buja, Lawrence Brown, Arun Kuchibhotla,
Richard Berk, Ed George, Linda Zhao

Department of Statistics, The Wharton School
University of Pennsylvania

Larry's Memorial — 2018/12/01

Ten Years of Joy with Larry

Ten Years of Joy with Larry

Larry's Group: The Path to Misspecification

- Starting Point: Post-Selection Inference ("PoSI", Berk et al. 2013)
- Does model selection produce "true models"?

  Certainly not!
- The PoSI method was in a framework of
  - normal, homoskedastic errors and
  - fixed regressors,

  but
  - we allowed misspecified conditional response means,
  - if an estimate of σ was available.
- Then:
  - We got interested in treating regressors as random.
  - We discovered discrepancies between standard errors from linear models theory and from the *x*-*y* bootstrap.

Parametric Regressor Ancillarity

A "justification" for conditioning on the regressors / treating them as fixed:

- Parametric Regression: $q(y, \vec{\boldsymbol{x}};\, \boldsymbol{\theta}) \;=\; q(y|\vec{\boldsymbol{x}};\, \boldsymbol{\theta})\, q(\vec{\boldsymbol{x}})$
  The regressor distribution $q(\vec{\boldsymbol{x}})$ is a non-parametric nuisance parameter.
- Fact: The set of pairwise probability ratios $\left(q(y, \vec{\boldsymbol{x}};\, \boldsymbol{\theta}_1) / q(y, \vec{\boldsymbol{x}};\, \boldsymbol{\theta}_2) \right)_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}$ forms a universally sufficient statistic.
- Ancillarity of the regressor distributions:

$$\frac{q(y, \vec{\boldsymbol{x}};\, \boldsymbol{\theta}_1)}{q(y, \vec{\boldsymbol{x}};\, \boldsymbol{\theta}_2)} \;=\; \frac{q(y \,|\, \vec{\boldsymbol{x}};\, \boldsymbol{\theta}_1)}{q(y \,|\, \vec{\boldsymbol{x}};\, \boldsymbol{\theta}_2)}$$

  Removes $q(\vec{\boldsymbol{x}})$ from inference about the parameter $\boldsymbol{\theta}$.
- **Problem:** Regressor Ancillarity holds only if the regression model $q(y|\vec{\boldsymbol{x}}; \boldsymbol{\theta})$ is correctly specified.

Discrepancies between Linear Models and Bootstrap

	$\hat{\beta}_j$	SE_{lin}	SE_{boot}	$\frac{\mathrm{SE}_{boot}}{\mathrm{SE}_{lin}}$	t_{lin}	t_{boot}
Intercept	0.760	22.767	16.505	**0.726**	0.033	0.046
MedianIncome ($K)	-0.183	0.187	0.114	**0.610**	-0.977	-1.601
PercVacant	4.629	0.901	1.385	**1.531**	5.140	3.341
PercMinority	0.123	0.176	0.165	0.937	0.701	0.748
PercResidential	-0.050	0.171	0.112	**0.653**	-0.292	-0.446
PercCommercial	0.737	0.273	0.390	**1.438**	2.700	1.892
PercIndustrial	0.905	0.321	0.577	**1.801**	2.818	1.570

- We used the *x-y* bootstrap, not the residual bootstrap.
- The *x-y* bootstrap is obviously valid without model assumptions.

Why the discrepancies?

Discrepancies between Linear Models and Bootstrap

	$\hat{\beta}_j$	SE_{lin}	SE_{boot}	$\frac{\mathrm{SE}_{boot}}{\mathrm{SE}_{lin}}$	t_{lin}	t_{boot}
Intercept	0.760	22.767	16.505	**0.726**	0.033	0.046
MedianIncome ($K)	-0.183	0.187	0.114	**0.610**	-0.977	-1.601
PercVacant	4.629	0.901	1.385	**1.531**	5.140	3.341
PercMinority	0.123	0.176	0.165	0.937	0.701	0.748
PercResidential	-0.050	0.171	0.112	**0.653**	-0.292	-0.446
PercCommercial	0.737	0.273	0.390	**1.438**	2.700	1.892
PercIndustrial	0.905	0.321	0.577	**1.801**	2.818	1.570

- We used the *x*-*y* bootstrap, not the residual bootstrap.
- The *x*-*y* bootstrap is obviously valid without model assumptions.

Why the discrepancies? **Misspecification!**

Regression and Statistical Functionals

First steps of emancipation from model assumptions:

- “Non-Model Assumption”: $(Y_i, \vec{\boldsymbol{X}}_i) \sim \mathbf{P} = \mathbf{P}_{Y,\vec{\boldsymbol{X}}}$ iid
- Compute whatever, but think of it operationally: OLS, GLMs, ...
- Q: What is the target of estimation, making “no” assumptions?
  A: A statistical functional $\theta(\mathbf{P})$, same computation, but at $\mathbf{P}$.
- Example: Linear OLS, $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E_P}[(Y - \vec{\boldsymbol{X}}'\theta)^2]$
- Example: GLMs, $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E_P}\left[\, b(\vec{\boldsymbol{X}}'\theta) - Y\vec{\boldsymbol{X}}'\theta\,\right]$
- General optimizations: $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E_P}[\mathcal{L}(\theta; Y, \vec{\boldsymbol{X}})]$
- Solving Estimating Equations (EE): $\mathbf{E_P}[\psi(\theta; Y, \vec{\boldsymbol{X}})] \;=\; \mathbf{0}$
- Ad hoc simple lin. reg. (X univariate): $(X', Y'), (X'', Y'') \sim \mathbf{P}$ iid
  $\theta(\mathbf{P}) \;=\; \mathbf{E_P}[(Y' - Y'')/(X' - X'')\,|\,|X' - X''| > \delta]$

Regression and Statistical Functionals

First steps of emancipation from model assumptions:

- "Non-Model Assumption": $(Y_i, \vec{\boldsymbol{X}}_i) \sim \mathbf{P} = \mathbf{P}_{Y,\vec{\boldsymbol{X}}}$ iid
- Compute whatever, but think of it operationally: OLS, GLMs, ...
- Q: What is the target of estimation, making "no" assumptions?
  A: A statistical functional $\theta(\mathbf{P})$, same computation, but at $\mathbf{P}$.
- Example: Linear OLS, $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E}_{\mathbf{P}}[(Y - \vec{\boldsymbol{X}}'\theta)^2]$
- Example: GLMs, $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E}_{\mathbf{P}}\left[\, b(\vec{\boldsymbol{X}}'\theta) - Y\vec{\boldsymbol{X}}'\theta \,\right]$
- General optimizations: $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E}_{\mathbf{P}}[\mathcal{L}(\theta; Y, \vec{\boldsymbol{X}})]$
- Solving Estimating Equations (EE): $\mathbf{E}_{\mathbf{P}}[\psi(\theta; Y, \vec{\boldsymbol{X}})] \;=\; \mathbf{0}$
- Ad hoc simple lin. reg. (X univariate): $(X', Y'), (X'', Y'') \sim \mathbf{P}$ iid
  $\theta(\mathbf{P}) \;=\; \mathbf{E}_{\mathbf{P}}[(Y' - Y'')/(X' - X'')\,|\,|X' - X''| > \delta]$

$\Longrightarrow$ **"Regression Functionals"** $\theta(\mathbf{P}) = \theta(\mathbf{P}_{Y,\vec{\boldsymbol{X}}})$

Just What is Regression?

Is the previous slide all there is to regression?

Just What is Regression?

Is the previous slide all there is to regression? There's more:

- Regression is the asymmetric analysis of association between Y and $\vec{\boldsymbol{X}}$.
- Motivation: Causation and Prediction
- Interest focuses on the conditional response distribution:

$$Y|\vec{\boldsymbol{X}} \sim \mathbf{P}_{Y|\vec{\boldsymbol{X}}}$$

- The goal/hope is that $\theta(\mathbf{P})$ is a property of $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$ alone, not of $\mathbf{P}_{\vec{\boldsymbol{X}}}$.
- Useful notation: $\mathbf{P}_{Y,\vec{\boldsymbol{X}}} = \mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}}$

Just What is Regression?

Is the previous slide all there is to regression? There's more:

- Regression is the asymmetric analysis of association between Y and $\vec{\boldsymbol{X}}$.
- Motivation: Causation and Prediction
- Interest focuses on the conditional response distribution:

$$Y|\vec{\boldsymbol{X}} \sim \mathbf{P}_{Y|\vec{\boldsymbol{X}}}$$

- The goal/hope is that $\theta(\mathbf{P})$ is a property of $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$ alone, not of $\mathbf{P}_{\vec{\boldsymbol{X}}}$.
- Useful notation: $\mathbf{P}_{Y,\vec{\boldsymbol{X}}} = \mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}}$

Definition: $\theta(\mathbf{P})$ **is "well-specified" for** $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$ **if** "$\theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}}) = \theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}})$."

(a form of ancillarity requirement for $\mathbf{P}_{\vec{\boldsymbol{X}}}$)

Well-Specification of Regression Functionals

Well-specification ...

- ... means in mathy terms: $\theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}}{}') = \theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}}{}'') \quad \forall \mathbf{P}_{\vec{\boldsymbol{X}}}{}', \mathbf{P}_{\vec{\boldsymbol{X}}}{}''$;
- ... means in practice: the quantity of interest, $\theta(\cdot)$, does not depend on where the data fall in regressor space;
- ... is a joint property of $\theta(\cdot)$ *and* $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$;
  $\theta(\cdot)$ will be well-specified for some $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$ but not for others;
- ... of ML functionals, $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E}_{\mathbf{P}}[-\log(q(Y|\vec{\boldsymbol{X}};\theta))]$?

Well-Specification of Regression Functionals

Well-specification ...

- ... means in mathy terms: $\theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}}{}') = \theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}}{}'') \quad \forall \mathbf{P}_{\vec{\boldsymbol{X}}}{}', \mathbf{P}_{\vec{\boldsymbol{X}}}{}''$;
- ... means in practice: the quantity of interest, $\theta(\cdot)$, does not depend on where the data fall in regressor space;
- ... is a joint property of $\theta(\cdot)$ *and* $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$;
  $\theta(\cdot)$ will be well-specified for some $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$ but not for others;
- ... of ML functionals, $\theta(\mathbf{P}) = \mathrm{argmin}_{\boldsymbol{\theta}}\, \mathbf{E}_{\mathbf{P}}[-\log(q(Y|\vec{\boldsymbol{X}};\theta))]$?
  Yes, if the model is correctly specified: $\mathbf{P}_{Y|\vec{\boldsymbol{X}}} = q(y|\vec{\boldsymbol{x}};\theta_0)$ for some θ_0; in which case $\theta(\mathbf{P}) = \theta_0$.
- ...: Can an ML functional be well-specified for $\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \notin \{q(y|\vec{\boldsymbol{x}};\theta)\}_{\boldsymbol{\theta}}$?
  Indeed, see next.

Well-Specification: Some Finger Practice

Let $\quad \mu(\vec{\boldsymbol{X}}) \;=\; \mathbf{E}_{\mathbf{P}}[\, Y \,|\, \vec{\boldsymbol{X}}\,]$.

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) = \mathbf{E}_{\mathbf{P}}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E}_{\mathbf{P}}[Y]$ well-specified?

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) \;=\; \mathbf{E}_{\mathbf{P}}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E}_{\mathbf{P}}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E}_{\mathbf{P}}[Y] = \mathbf{E}_{\mathbf{P}}[\mu(\vec{\boldsymbol{X}})]$.

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) \;=\; \mathbf{E_P}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E_P}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E_P}[Y] = \mathbf{E_P}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E_P}[YX]/\mathbf{E_P}[X^2]$ well-specified (X univariate)?

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) = \mathbf{E_P}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E_P}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E_P}[Y] = \mathbf{E_P}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E_P}[YX]/\mathbf{E_P}[X^2]$ well-specified (X univariate)?
  A: Iff $\mu(X) = \theta_0 X$.

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) = \mathbf{E}_{\mathbf{P}}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E}_{\mathbf{P}}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E}_{\mathbf{P}}[Y] = \mathbf{E}_{\mathbf{P}}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E}_{\mathbf{P}}[YX]/\mathbf{E}_{\mathbf{P}}[X^2]$ well-specified (X univariate)?
  A: Iff $\mu(X) = \theta_0 X$.
- Q: When is $\theta(\mathbf{P}) = \mathrm{cor}(Y, X) \neq 0$ well-specified?

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) = \mathbf{E}_{\mathbf{P}}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E}_{\mathbf{P}}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E}_{\mathbf{P}}[Y] = \mathbf{E}_{\mathbf{P}}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E}_{\mathbf{P}}[YX]/\mathbf{E}_{\mathbf{P}}[X^2]$ well-specified (X univariate)?
  A: Iff $\mu(X) = \theta_0 X$.
- Q: When is $\theta(\mathbf{P}) = \mathrm{cor}(Y, X) \neq 0$ well-specified?
  A: Never

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) \;=\; \mathbf{E}_{\mathbf{P}}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E}_{\mathbf{P}}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E}_{\mathbf{P}}[Y] = \mathbf{E}_{\mathbf{P}}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E}_{\mathbf{P}}[YX]/\mathbf{E}_{\mathbf{P}}[X^2]$ well-specified (X univariate)?
  A: Iff $\mu(X) = \theta_0 X$.
- Q: When is $\theta(\mathbf{P}) = \mathrm{cor}(Y, X) \neq 0$ well-specified?
  A: Never
- Q: When is linear OLS well-specified? $\theta(\mathbf{P}) = \mathbf{E}_{\mathbf{P}}[\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}']^{-1}\mathbf{E}_{\mathbf{P}}[\vec{\boldsymbol{X}}\mu(\vec{\boldsymbol{X}})]$

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) = \mathbf{E_P}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E_P}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E_P}[Y] = \mathbf{E_P}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E_P}[YX]/\mathbf{E_P}[X^2]$ well-specified (X univariate)?
  A: Iff $\mu(X) = \theta_0 X$.
- Q: When is $\theta(\mathbf{P}) = \mathrm{cor}(Y, X) \neq 0$ well-specified?
  A: Never
- Q: When is linear OLS well-specified? $\theta(\mathbf{P}) = \mathbf{E_P}[\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}']^{-1}\mathbf{E_P}[\vec{\boldsymbol{X}}\mu(\vec{\boldsymbol{X}})]$
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \theta_0'\vec{\boldsymbol{X}}$, regardless of heteroskedasticity, non-Gaussianity.
  Well-specification of OLS slopes is weaker than the full linear model!

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) = \mathbf{E_P}[\, Y \,|\, \vec{\boldsymbol{X}} \,]$.

- Q: When is $\mathbf{E_P}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E_P}[Y] = \mathbf{E_P}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E_P}[YX]/\mathbf{E_P}[X^2]$ well-specified (X univariate)?
  A: Iff $\mu(X) = \theta_0 X$.
- Q: When is $\theta(\mathbf{P}) = \mathrm{cor}(Y, X) \neq 0$ well-specified?
  A: Never
- Q: When is linear OLS well-specified? $\theta(\mathbf{P}) = \mathbf{E_P}[\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}']^{-1}\mathbf{E_P}[\vec{\boldsymbol{X}}\mu(\vec{\boldsymbol{X}})]$
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \theta_0'\vec{\boldsymbol{X}}$, regardless of heteroskedasticity, non-Gaussianity.
  Well-specification of OLS slopes is weaker than the full linear model!
- Q: When is ridge well-specified? $\theta(P) = (\mathbf{E_P}[\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}'] + \boldsymbol{\Omega})^{-1}\mathbf{E_P}[\vec{\boldsymbol{X}}\mu(\vec{\boldsymbol{X}})]$

Well-Specification: Some Finger Practice

Let $\mu(\vec{\boldsymbol{X}}) \;=\; \mathbf{E_P}[\,Y \,|\, \vec{\boldsymbol{X}}\,]$.

- Q: When is $\mathbf{E_P}[Y]$ well-specified?
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \mathrm{const}$ because $\mathbf{E_P}[Y] = \mathbf{E_P}[\mu(\vec{\boldsymbol{X}})]$.
- Q: When is $\theta(\mathbf{P}) = \mathbf{E_P}[YX]/\mathbf{E_P}[X^2]$ well-specified (X univariate)?
  A: Iff $\mu(X) = \theta_0 X$.
- Q: When is $\theta(\mathbf{P}) = \mathrm{cor}(Y, X) \neq 0$ well-specified?
  A: Never
- Q: When is linear OLS well-specified? $\theta(\mathbf{P}) = \mathbf{E_P}[\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}']^{-1}\mathbf{E_P}[\vec{\boldsymbol{X}}\mu(\vec{\boldsymbol{X}})]$
  A: Iff $\mu(\vec{\boldsymbol{X}}) = \boldsymbol{\theta}_0'\vec{\boldsymbol{X}}$, regardless of heteroskedasticity, non-Gaussianity.
  Well-specification of OLS slopes is weaker than the full linear model!
- Q: When is ridge well-specified? $\theta(P) = (\mathbf{E_P}[\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}'] + \boldsymbol{\Omega})^{-1}\mathbf{E_P}[\vec{\boldsymbol{X}}\mu(\vec{\boldsymbol{X}})]$
  A: $\boldsymbol{\Omega} = c\,\mathbf{E}[\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}']$ and $\mu(\vec{\boldsymbol{X}}) = \boldsymbol{\theta}_0'\vec{\boldsymbol{X}}$, hence $\boldsymbol{\theta}(P) = \boldsymbol{\theta}_0/(1 + c)$.

Well-Specification and Causality

- DAG theory of causality provides criteria to select regressors $\vec{\boldsymbol{X}}$ for an outcome Y to correctly infer causal effects.
- The causal mechanism $\vec{\boldsymbol{X}} \rightarrow Y$ is described by $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$.
- For $\vec{\boldsymbol{X}}$ to be causal, it must not matter how $\vec{\boldsymbol{X}}$ is manipulated or sampled. The transmission of the causal effect is always through $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$.
- Now we want to describe properties of $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$, regardless of $\mathbf{P}_{\vec{\boldsymbol{X}}}$.
  $\Rightarrow$ Use regression functionals $\theta(\cdot)$ that are well-specified for $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$.
- Peters, Bühlmann, Meinshausen (2016) propose a scheme for finding causal associations from multiple data sources with same variables.
  Idea: If $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$ is causal for some $\vec{\boldsymbol{X}}$, it will be shared across data sources.
  Interpretation: They are selecting for well-specification, not causation.

A Well-Specification Diagnostic: Reweighting

- Q: Can we detect misspecification of regression functionals empirically?
  A: Yes, with reweighting.

- If $w(\vec{\boldsymbol{x}}) > 0$ is a weight function, of $\vec{\boldsymbol{x}}$ alone, define (density notation)
  $$p^{(w)}(y, \vec{\boldsymbol{x}}) \;=\; w(\vec{\boldsymbol{x}})\, p(y, \vec{\boldsymbol{x}}) \qquad\qquad (\mathbf{E}_{\mathbf{P}} w(\vec{\boldsymbol{X}}) = 1)$$

- Facts: $p^{(w)}(y|\vec{\boldsymbol{x}}) = p(y|\vec{\boldsymbol{x}})$ and $p^{(w)}(\vec{\boldsymbol{x}}) = w(\vec{\boldsymbol{x}})p(\vec{\boldsymbol{x}})$

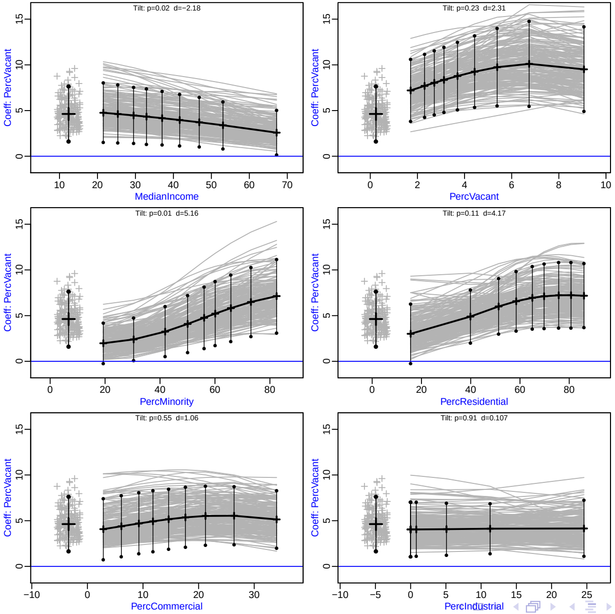
- Corollary: If $\theta(\cdot)$ is well-specified for $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$, then $\theta(\mathbf{P}^{(w)}) = \theta(\mathbf{P})$

- Conversely: If $\theta(\mathbf{P}^{(w)}) \neq \theta(\mathbf{P})$, then $\theta(\cdot)$ is misspecified for $\mathbf{P}_{Y|\vec{\boldsymbol{X}}}$.

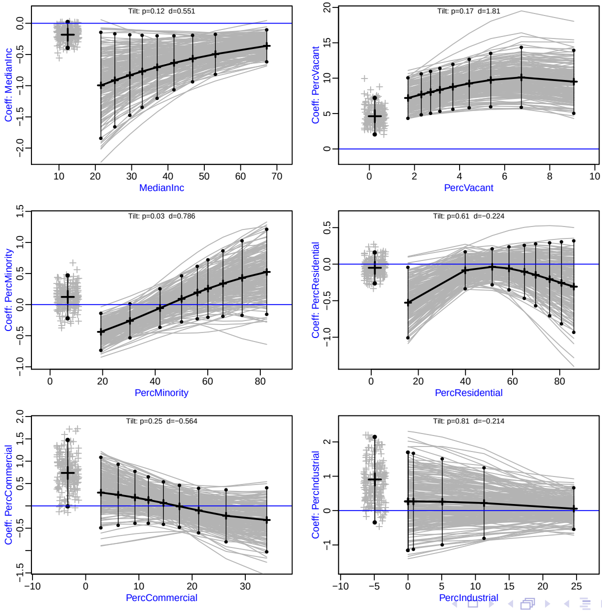
- Methodology: Let $w_\xi(\vec{\boldsymbol{x}}) = w(x_j - \xi)$ be a weight function of x_j alone,
  centered at ξ. Plot $\xi \mapsto \theta(\mathbf{P}^{(w_\xi)})$.

- LA Homeless Data: $\theta(\mathbf{P}) = \beta_{PercVacant}(\mathbf{P})$ OLS slope of interest

LA Homeless Data, $\theta(\cdot)$ = Slope of PercVacant



Reweighting of Slopes on Own Regressors



Misspecification and Random $\vec{\boldsymbol{X}} \to$ Sampling Variabilty

- If $\theta(\mathbf{P}) = \theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \mathbf{P}_{\vec{\boldsymbol{X}}})$ is misspecified, it depends on $\mathbf{P}_{\vec{\boldsymbol{X}}}$.
  Hence the following object has sampling variability:

$$\theta(\mathbf{P}_{Y|\vec{\boldsymbol{X}}} \otimes \widehat{\boldsymbol{P}}_{\vec{\boldsymbol{X}}})$$

- Compare: Linear OLS, $\quad \mathbf{V}[\,\hat{\theta}\,] \;=\; \mathbf{E}[\,\mathbf{V}[\,\hat{\theta}\,|\,\vec{\boldsymbol{X}}\,]\,] + \mathbf{V}[\,\mathbf{E}[\,\hat{\theta}\,|\,\vec{\boldsymbol{X}}\,]\,]$

- The “Conspiracy Movie”: Misspecification and random $\vec{\boldsymbol{X}}$ conspiring

```
source("http://stat.wharton.upenn.edu/~buja/src-conspiracy-animation2.R")
```

Thank you, Larry !