Automatic debiased machine learning and sensitivity analysis for sample selection models

Jakob Bjelac, Victor Chernozhukov, Phil-Adrian Klotz, Jannis Kueck, **Theresa M. A. Schmitz**

European Causal Inference Meeting

Oxford, UK — April 16, 2026

Massachusetts Institute of Technology

Heinrich Heine Universität Düsseldorf

Outline

Introduction

Causal Framework and Sample Selection Problem

Riesz Representers and Debiased Machine Learning

Sensitivity Analysis and Estimation Approach

Simulation Study and Application

Conclusion

Introduction

Motivation

- Researchers often face **two selection problems** in empirical studies.
  - **Treatment assignment** is non-random ('selection into treatment').
  - **Outcomes** are only observed for part of the population ('selection into outcome observability'; e. g., job training evaluations miss earnings for the unemployed).

- **Standard methods** for confounding adjustment **fail** when outcomes are selectively missing (e. g., regression or propensity score weighting).
- **Machine learning** offers powerful tools for **high-dimensional covariate adjustment**, e. g., by deriving Neyman-orthogonal score functions for treatment effects under sample selection (see Bia, Huber, and Lafférs, 2024).

- An **alternative perspective** comes from the Riesz representation theorem, where parameters can be characterized through unique weighting functions called '**Riesz representers**'.

Contribution

- We derive a **Riesz representer** for causal inference under **sample selection**. It enables **stable estimation** and a **decomposition** of **omitted variable bias** (OVB) into a data-identified scale factor, outcome confounding, and selection confounding, yielding **sharp bounds**.

- Leveraging **automatic debiased machine learning**, we avoid unstable direct propensity score inversion of the traditional plug-in approach by jointly learning the outcome regression and the Riesz representer with the **ForestRiesz** estimator (Chernozhukov, Newey, and Singh, 2022).

- **Simulation:** Standard double machine learning (**DML**) approaches can be **sensitive to tuning** parameters, while **ForestRiesz** provides **more stable** average treatment effect (**ATE**) estimates.

- **Application:** U. S. gender wage gap
  - **ForestRiesz** yields **larger ATE estimates** than conventional DML.
    $\rightarrow$ Ignoring sample selection underestimates the gender wage gap.
  - **Robustness:** Only implausibly strong unobserved confounding would overturn the results.

Related Literature

DML for Sample Selection Models

- DML for sample selection: Bia, Huber, and Lafférs, 2024.
- DML for multivariate sample selection: Dolgikh and Potanin, 2025.

Riesz Representation & Sensitivity Analysis

- Automatic debiased ML via Riesz regression: Chernozhukov et al., 2021.
- Sensitivity analysis for causal ML: Chernozhukov et al., 2022a.
- OVB-based sensitivity analysis: Cinelli and Hazlett, 2020.

Causal Framework and Sample Selection Problem

Causal Framework and Target Parameter

- Based on the **Potential Outcomes framework** (Rubin, 1974; Rubin, 1977), and with $D \in \{0, 1\}$ indicating treatment status, let us for each unit i define
  - $\mathbf{Y_i(1)}$ Outcome under treatment ($D = 1$).
  - $\mathbf{Y_i(0)}$ Outcome under control ($D = 0$).

- **SUTVA** — Each unit's potential outcomes are unaffected by the other units' treatment assignments, and both potential outcomes are well-defined (Rubin, 1980).

Target parameter

$$\textbf{ATE} = \mathbb{E}\big[Y(1) - Y(0)\big] = \mathbb{E}\big[Y(1)\big] - \mathbb{E}\big[Y(0)\big].$$

The Sample Selection Problem — Intuition

- In many practical applications, the outcome variable Y is **not** observed for all units in the sample ('sample selection'). Let S be a binary indicator variable such that $S_i = 1$ if the outcome Y_i is observed for unit i, and $S_i = 0$ otherwise.

⚡ If the **mechanism** determining whether the outcome is observed S is **related to** the potential **outcomes** $Y(d)$ themselves, then the subsample for whom we observe the outcome ($S = 1$) is no longer representative of the full population ('sample selection bias').

! When **non-random treatment assignment** and **non-random sample selection** occur simultaneously, researchers face a 'double selection problem' (Bia, Huber, and Lafférs, 2024).

➲ **Assumptions** addressing **both sources of potential bias** are needed.

The Sample Selection Problem — Assumption I

- No unobservables jointly affect the treatment and potential outcomes conditional on covariates X:

Assumption 1 (Conditional Independence of the Treatment)
$Y(d) \perp D \mid X = x$ for all $d \in \{0, 1\}$ and x in the support of X.

- Within groups defined by a treatment status d and covariate values x, whether an outcome $Y(d)$ is observed ($S = 1$) or missing ($S = 0$) does not depend on the potential outcome's value itself:

Assumption 2* (Conditional Independence of Selection)
$Y(d) \perp S \mid D = d, X = x$ for all $d \in \{0, 1\}$ and x in the support of X.

The Sample Selection Problem — Bia, Huber, and Lafférs (2024)

Bia, Huber, and Lafférs (2024) show that under **Assumption 1** (Conditional Exogeneity), **Assumption 2*** (Conditional Independence of Selection), and **Common Support for Selection** $\Big[P(S=1 \mid D=d, X=x) > 0 \text{ for all } d \text{ and } x\Big]$, the **ATE** is **identified** by

$$\theta_0 = \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[\phi_1 - \phi_0]$$

with the **efficient score function** ϕ_d and **conditional mean outcome** $\mu(D, S, X)$:

$$\phi_d = \frac{I\{D=d\} \cdot S \cdot [Y - \mu(d,1,X)]}{p_d(X) \cdot \pi(d,X)} + \mu(d,1,X), \qquad \mu(D,S,X) := \mathbb{E}[Y|D,S,X].$$

Assumption 2* might be violated in many real-world scenarios. **Selection could depend on unobserved factors** A **that also influence** the **outcome**, even after controlling on D and X ('non-ignorable nonresponse').

The Sample Selection Problem — Assumption II

We investigate the **potential bias** when the **sample selection** process (S) is only ignorable if we **condition** on both observed covariates ($\mathbf{X}$) and unobserved covariates ($\mathbf{A}$).

- We consider a **weaker condition** as selection independence might still hold if we could additionally **condition on** the relevant **unobserved factors** A:

Assumption 2 (Conditional Independence with Unobservables)
$Y(d) \perp S \mid D = d, X = x, A = a \quad$ for all $d \in \{0, 1\}$ and x, a in the support of X and A.

While we cannot observe A, we can use the **Riesz Representer framework** to derive **sharp bounds** on the **size of OVB** that results from not observing it (Chernozhukov et al., 2022a).

The Sample Selection Problem — Assumption III

- We restrict **unobservables** to **not affect treatment assignment** beyond X:

  Assumption 3 (No Unobserved Confounding in the Treatment Assignment)
  $A \perp D \mid X = x \quad$ for all x in the support of X.

Note: For a setting in which A also affects D, we would require additional sensitivity components for treatment confounding.

The Sample Selection Problem — Assumption IV

- Additionally, with the **propensity scores** defined as
  - $p_d(X) := \mathbb{P}(D = d \mid X)$ for $d \in \{0, 1\}$
  - $\pi_0(d, X, A) := \mathbb{P}(S = 1 \mid D = d, X, A)$,

  we assume that the **treatment assignment** is **non-degenerate**, the **selection probability** is **non-zero** for each conditioning value and the **Riesz representer** is **square-integrable** (see here):

Assumption 4 (Common Support and Weak Overlap)

(i) $p_d(X) > 0$ **and** $\pi_0(d, X, A) > 0$ **almost surely** for $d \in \{0, 1\}$

(ii) **inverse-propensity moments** satisfy $\mathbb{E}\left[\frac{1}{p_1(X)\pi_0(1,X,A)} + \frac{1}{p_0(X)\pi_0(0,X,A)}\right] < \infty$.

Riesz Representers and Debiased Machine Learning

Double Machine Learing Framework

Single/naive ML estimation leads to bias in treatment effect estimation.

- To enable valid inference using ML estimators, the recent literature on *DML* focuses on constructing **Neyman-orthogonal scores** (Chernozhukov et al., 2018).
- The target parameter θ_0 is identified by a **moment condition**

$$\mathbb{E}[\psi(W, \theta_0, g_0)] = 0,$$

and the **score** $\psi(W, \theta, g, \dots)$ is **orthogonal** if it is *insensitive* to estimation errors in g_0.

Under $n^{-1/4}$-consistency of estimators of g_0, the **estimator** $\hat{\theta}_0$ which solves

$$\mathbb{E}_n[\psi(W, \theta, \hat{g}_0)] = 0$$

is $\sqrt{\mathbf{n}}$**-consistent** and **asymptotically normally** distributed.

Orthogonal Scores and the Riesz Representer

- For many parameters of interest, with a linear and continuous suitable function class $g \to \mathbb{E}[m(W, g)]$ and $g_0(x) = \mathbb{E}[Y|X = x]$, it holds that:

$$\theta_0 = \mathbb{E}[m(W, g_0)]$$

- The **Riesz Representation Theorem** guarantees the existence of a unique function, α_0, known as the **'Riesz Representer'**:

$$\theta_0 = \mathbb{E}[m(W, g_0)] = \mathbb{E}[\alpha_0(Z)\, g_0(Z)],$$

- Given a Riesz representer, an **orthogonal score** is given by:

$$\psi(W, \theta, g, \alpha) = m(W, g) - \theta + \alpha(Z)(Y - g(Z)).$$

The Riesz representer appears naturally in the structure of many causal and structural parameters (e. g. ATE). **We derive** the **Riesz representer for sample selection models**.

Riesz Representation Approach I

- Our **goal** is to estimate the **Average Treatment Effect** (ATE), defined as:

$$\theta_0 = \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[m(W, g)] = [g_0(1, X, A) - g_0(0, X, A)],$$

with $W := (Y, D, S, X, A)$ and $g_0(d, x, a) := [Y \mid D = d, S = 1, X = x, A = a]$ as conditional expectation of Y given treatment, selection, observed covariates, and unobserved confounders.

Since A is not observed in practice, we can only **identify** the **'short parameter'**

$$\theta_s = \mathbb{E}[m(W_s, g_s)] = \mathbb{E}[g_s(1, X) - g_s(0, X)],$$

with $W_s := (Y, D, S, X)$ and $g_s(d, S, X) := \mathbb{E}[Y \mid D = d, S = 1, X]$.

- By the **Riesz Representation Theorem** there **exist** unique functions α_0 and α_s which satisfy

$$\theta_0 = \mathbb{E}[\alpha_0(Z)\, g_0(Z)] \quad \textbf{and} \quad \theta_s = \mathbb{E}[\alpha_s(Z_s)\, g_s(Z_s)].$$

Riesz Representation Approach II

Theorem

In **sample selection models** *the* **Riesz Representers** *for the long parameter* θ_0 *and short parameter* θ_s *under the* **Assumptions 1 – 4**, *are given by*

$$\alpha_0(W) = \frac{\mathbf{1}\{D=1\} \cdot S}{p_1(X)\,\pi(1,X,A)} - \frac{\mathbf{1}\{D=0\} \cdot S}{p_0(X)\,\pi(0,X,A)},$$

and

$$\alpha_s(W_s) = \frac{\mathbf{1}\{D=1\} \cdot S}{p_1(X)\,\pi(1,X)} - \frac{\mathbf{1}\{D=0\} \cdot S}{p_0(X)\,\pi(0,X)}.$$

- The key **difference** between the long and short parameters lies in whether we account for **unobserved confounders** A in the selection process.
- Weighting by $\frac{1}{\pi(d,X,A)}$ in the long parameter, or $\frac{1}{\pi(d,X)}$ in the short parameter, gives **more weight to units** that were **less likely to be selected** into our observed sample.

Sensitivity Analysis and Estimation Approach

Omitted Variable Bias in Sample Selection Models

- The OVB can be interpreted as the **covariance** between the **error parts** of g and α (Chernozhukov et al., 2022a):

$$\theta_0 - \theta_s \;=\; \mathbb{E}[(g_0 - g_s)\,(\alpha_0 - \alpha_s)].$$

The **squared bias** satisfies a **sharp bound** $|\theta_0 - \theta_s|^2 \;=\; \rho^2\, B^2 \quad \leq\; B^2$

- $B^2 \;=\; \mathbb{E}[(g_0 - g_s)^2]\;\mathbb{E}[(\alpha_0 - \alpha_s)^2]$ captures the additional variation generated by A.
- ρ^2 represents the squared correlation between the two sources of variation.

- The **squared bias bound** B^2 can be decomposed as:

$$B^2 \;=\; \tilde{S}^2\, C_Y^2\, C_S^2.$$

Scaling factor $\tilde{S}^2$ is identifiable from **observed** data. C_Y^2 and C_S^2 (bounded between 0 and 1) are unknown. **Informed assumptions** about **impact of unobserved confounding** are needed.

Benchmarking Sensitivity to Unobserved Confounding

- **Outcome confounding strength** | Proportion of residual outcome variation explained by A.

$$C_Y^2 = \frac{\mathbb{E}[(g_0 - g_s)^2]}{\mathbb{E}[(Y - g_s)^2]} = R^2_{Y-g_s \sim g_0 - g_s} = \eta^2_{Y \sim A | D, X, S=1}.$$

- **Selection confounding strength** | Proportion of variation in α_0 explained by A.

$$C_S^2 = \frac{\mathbb{E}[\alpha_0^2] - \mathbb{E}[\alpha_s^2]}{\mathbb{E}[\alpha_s^2]} = \frac{1 - R^2_{\alpha_0 \sim \alpha_s}}{R^2_{\alpha_0 \sim \alpha_s}}, \quad \text{with} \quad R^2_{\alpha_0 \sim \alpha_s} = \frac{\mathbb{E}[\alpha_s^2]}{\mathbb{E}[\alpha_0^2]}.$$

Quasi-Gaussian Selection Sensitivity | Let $S = \mathbf{1}\{S^* > 0\}$ with $S^* = h(D, X) - U$ and $U \mid D, X \sim N(0, 1)$. Model confounding independent of (D, X) as $U = \mu_S A + \sqrt{1 - \mu_S^2}\, \varepsilon_S$, with $A, \varepsilon_S \overset{\text{i.i.d.}}{\sim} N(0, 1)$ and $\mu_S^2 = R^2_{U \sim A} = \eta^2_{S^* \sim A | D, X}$ (for more details see here).

Compare the **hypothetical impact** of the unobserved A with the **measured impact** of observed covariates X that we suspect might act as confounders by influencing both the outcome Y and the selection process S (see benchmarking/sensitivity analysis).

Estimating the Riesz Representer

- Since α_0 is generally unknown, one needs to construct a feasible estimator $\hat{\alpha}$, traditionally by using the **plug-in approach**.

However, the plug-in approach suffers from several drawbacks, particularly in high-dimensional or complex settings (complexity of derivation, potential instability due to inverse weighting).

Recent research has focused on methods that estimate the Riesz representer α_0 directly (**Automatic debiased machine learning**)

- **Variational Methods (Riesz Regression)**: Leverage the characterization of α_0 as the solution to a specific minimization problem ($\arg\min E[a^2 - 2m(W; a)]$, see here).
- **Adversarial (Minimax) Methods**: Minimax game seeking an α that satisfies the defining Riesz property $E[m(W, g)] = E[\alpha(W)g(W)]$ uniformly well against a worst-case choice of test function g from a specified class.

Estimating the Riesz Representer: ForestRiesz

- **ForestRiesz** (Chernozhukov, Newey, and Singh, 2022) adapts the **random forest** methodology to estimate the Riesz representer by **solving a regularized Riesz-regression problem**.
- The **Riesz representer** is modeled as **locally linear** with respect to a **pre-specified feature map**. The algorithm **constrains splits** to covariates X **exclusively** to preserve **sufficient variation** in the treatment variable D within each node.
- Defined via solutions to **moment equations** $m(\cdot) = 0$ (Chernozhukov et al., 2022b), thus we can apply the **Generalized Random Forests** framework (Athey, Tibshirani, and Wager, 2019).

It **simultaneously learns** the **regression** function $\hat{g}$ and the **Riesz representer** $\hat{\alpha}$ by augmenting the node-splitting criteria with regression-based objectives and yields the **final estimate**

$$\hat{\theta}_{\mathrm{DR}} = \mathbb{E}_n\big[\, m(W; \hat{g}) + \hat{\alpha}(Z)\big(Y - \hat{g}(Z)\big)\,\big].$$

$\Rightarrow$ Use **cross-fitted form** to avoid overfitting (more details on ForestRiesz here).

Simulation Study and Application

Simulation — Data-generative process (DGP)

- The **DGP** follows the **conditional missing-at-random (MAR) design** outlined in Appendix E of Bia, Huber, and Lafférs, 2024:

$$
\begin{aligned}
D_i &= \mathbf{1}\{X_i'\beta_0 + w_i > 0\}, \\
S_i &= \mathbf{1}\{D_i + X_i'\beta_0 + v_i > 0\}, \\
Y_i &= \theta_0 D_i + X_i'\beta_0 + u_i.
\end{aligned}
$$

  - With Y_i **only observed if** $S = 1$, and

$$
X_i \sim N(0, \sigma_X^2), \quad (u_i, v_i) \sim N(0, \sigma_{u,v}^2), \quad w_i \sim N(0, 1).
$$

  - For conditional **MAR** to hold, $\sigma_{u,v}^2$ is specified as an **identity matrix**.

Simulation - Results

	IRM			SSM			FR		
N	ATE	SE	MAE	ATE	SE	MAE	ATE	SE	MAE
1000	0.8017	0.0564	0.1983	1.1046	0.0451	0.1165	1.1306	0.0944	0.1365
4000	0.7457	0.0280	0.2543	1.0863	0.0222	0.0874	1.0677	0.0461	0.0703
16000	0.7046	0.0139	0.2954	1.0621	0.0110	0.0622	1.0349	0.0230	0.0357

Note: Average simulation results based on $\theta_0 = 1$ and 200 Monte Carlo iterations. Comparison of ForestRiesz (FR) to the interactive regression model (IRM) (Chernozhukov et al., 2018) and sample selection model (SSM) (Bia, Huber, and Lafférs, 2024). Benchmark models use random forests for estimating nuisance functions and three-fold cross-fitting. For each sample size (N), the table presents average treatment effects (ATE), standard errors (SE), and mean absolute errors (MAE).

- Across all sample sizes, IRM underestimates θ_0.
- SSM and FR converge to the simulated $\theta_0 = 1$ as sample size increases.
- Quadrupling sample size roughly halves the standard errors of all estimators.

Application – Gender Wage Gap in the U. S.

Using the **2016 American Community Survey (ACS)** data, we quantify the **gender wage gap** among respondents with a high school or college degree (c.f., Bach, Chernozhukov, and Spindler (2024)).

- **Sample** Representative 1 % sample of the U. S. population with mandatory participation.
- **Outcome** (Log) weekly wages (in USD).
- **Treatment** Gender ($D = 1$: female).
- **Covariates** 158 variables for socio-economic characteristics at individual and household level (e. g. referring to education, industry, and occupation).

Issue

Some respondents do **not report wages** even though they are employed.

Thus, the gender wage gap analysis is subject to a **sample-selection problem**.

Gender Wage Gap — Estimation Results

	IRM		SSM		FR	
	College	High school	College	High school	College	High school
Estimate	-0.0989***	-0.141***	-0.153***	-0.198***	-0.128***	-0.172***
SE	0.003	0.003	0.001	0.001	0.002	0.002
P-value	0.000	0.000	0.000	0.000	0.000	0.000

Note: Estimation results for the gender wage gap. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

- **Significant** gender wage gap in college and high school sub-populations.
- **Not controlling for non-reporting** respondents **underestimates** the gender wage **gap**.
- **Detailed analysis**: **Female** respondents with **higher experience** and **university degree** have a **higher probability to report** their income compared to male respondents.

Gender Wage Gap — Sensitivity Analysis I

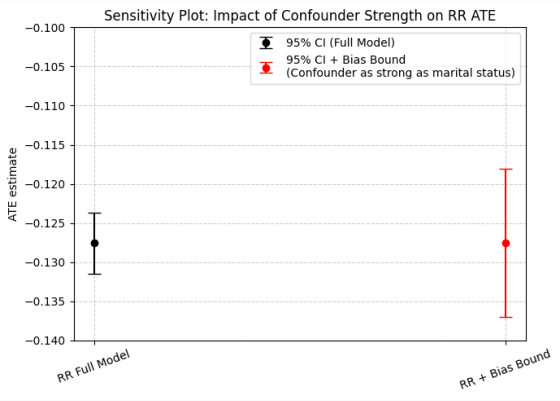
Benchmarking Analysis for the Gender Wage Gap.

Group	k	θ_{full}	θ_{-j}	$\Delta\theta$	$\lvert G_{Y,j}\rvert$	$\lvert G_{S,j}\rvert$	$\lvert\rho_j\rvert$
Marital status	5	-0.1276	-0.1331	-0.00553	0.00177	0.00097	1.000
Region	8	-0.1276	-0.1298	-0.00225	0.00110	0.00174	1.000
Race	3	-0.1276	-0.1287	-0.00116	0.00115	0.00264	0.540
Children	1	-0.1276	-0.1278	-0.00027	0.00079	0.00138	0.210
Education	3	-0.1276	-0.1275	0.00007	0.00435	0.01603	0.007
Experience	2	-0.1276	-0.1275	0.00003	0.00003	0.00020	0.284

Note: As estimated sensitivity parameters are approximately unbiased for true shares, they can be negative when true shares are close to zero (see also here).

k	Number of covariates dropped from group$_j$.
θ_{full}	ATE with full covariate set.
θ_{-j}	ATE when group$_j$ is removed.
$\Delta\theta$	Sensitivity of ATE to group$_j$, with $\Delta\theta = \theta_{-j} - \theta_{\text{full}}$.
$G_{Y,j}$	Share of outcome variation uniquely explained by group$_j$.
$G_{S,j}$	Share of selection variation uniquely explained by group$_j$.
ρ_j	Minimal confounding correlation needed for group$_j$ to explain $\Delta\theta$.

Gender Wage Gap — Sensitivity Analysis II



Notes: Sensitivity of the estimated gender wage gap to potential omitted confounding. Full model Riesz Representer ATE estimate vs. counterfactual confidence interval that would arise if an unobserved confounder were as influential as marital status.

- **Confidence interval widens** under hypothetical confounding, while the estimated **ATE** remains **negative**.
- Suggests that the **estimated gender wage gap** is **robust** to confounding of realistic magnitude.

Examples for **latent confounders**

- disclosure preferences
- compensation features
- personality traits

Conclusion

Conclusion

- We extend the **Riesz representation framework** of Chernozhukov et al. (2022a) to **sample selection models**, providing a unified framework for identification, estimation, and sensitivity analysis when unobserved confounders A affect selection but not treatment conditional on X.
- We derive a formula for **omitted variable bias** in sample selection models depending on three **interpretable parts** (scale factor | outcome confounding | selection confounding). This yields **sharp bounds** on the bias magnitude without requiring distributional assumptions.
- By using **ForestRiesz** we **learn the Riesz representer directly**, avoiding unstable inverse probability weighting. We find that **conventional double machine learning approaches** can be highly **sensitive to tuning parameters**, while ForestRiesz delivers a more stable performance.
- Our application shows the **practical relevance** of the Riesz Representer, as ignoring sample selection underestimates the U. S. **gender wage gap**, with the **sensitivity analysis** indicating that only implausibly strong unobserved confounding would overturn the results.

Thank you for your attention!

Questions?

Contact

Theresa M. A. Schmitz
Doctoral Researcher
Chair of Statistics and Econometrics
Heinrich Heine University Düsseldorf
theresa.schmitz@hhu.de

hhu Heinrich Heine Universität Düsseldorf

Working Paper

arXiv 2601.08643

References

Altonji, Joseph G, Todd E Elder, and Christopher R Taber (2005). **"Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools".** In: *Journal of political economy* 113.1, pp. 151–184.

Athey, Susan, Julie Tibshirani, and Stefan Wager (2019). **"Generalized random forests".** In: *The Annals of Statistics* 47.2, pp. 1148 –1178. DOI: 10.1214/18-AOS1709. URL: https://doi.org/10.1214/18-AOS1709.

Bach, Philipp, Victor Chernozhukov, Sven Klaassen, Malte S. Kurz, and Martin Spindler (n.d.). **DoubleML - Double Machine Learning in Python.** URL: https://github.com/DoubleML/doubleml-for-py.

Bach, Philipp, Victor Chernozhukov, and Martin Spindler (2024). **"Heterogeneity in the US gender wage gap".** In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 187.1, pp. 209–230.

Bia, Michela, Martin Huber, and Lukáš Lafférs (2024). **"Double Machine Learning for Sample Selection Models".** In: *Journal of Business & Economic Statistics* 42.3, pp. 958–969.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018). **Double/debiased machine learning for treatment and structural parameters.**

References

Chernozhukov, Victor, Carlos Cinelli, Whitney Newey, Amit Sharma, and Vasilis Syrgkanis (2022a). **Long story short: Omitted variable bias in causal machine learning.** Tech. rep. National Bureau of Economic Research.

Chernozhukov, Victor, Whitney Newey, Vıctor M Quintas-Martınez, and Vasilis Syrgkanis (2022b). **"Riesznet and forestriesz: Automatic debiased machine learning with neural nets and random forests".** In: *International Conference on Machine Learning*. PMLR, pp. 3901–3914.

Chernozhukov, Victor, Whitney K Newey, Victor Quintas-Martinez, and Vasilis Syrgkanis (2021). **"Automatic debiased machine learning via Riesz regression".** In: *arXiv preprint arXiv:2104.14737*.

Chernozhukov, Victor, Whitney K Newey, and Rahul Singh (2022). **"Automatic debiased machine learning of causal and structural effects".** In: *Econometrica* 90.3, pp. 967–1027.

Cinelli, Carlos and Chad Hazlett (2020). **"Making sense of sensitivity: Extending omitted variable bias".** In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.1, pp. 39–67.

References

Dolgikh, Sofiia and Bodan Potanin (2025). **"Double machine learning for causal inference in a multivariate sample selection model".** In: *arXiv preprint arXiv:2511.12640*. URL: https://arxiv.org/abs/2511.12640.

Imbens, Guido W (2003). **"Sensitivity to exogeneity assumptions in program evaluation".** In: *American Economic Review* 93.2, pp. 126–132.

Newey, Whitney K (1994). **"The asymptotic variance of semiparametric estimators".** In: *Econometrica: Journal of the Econometric Society*, pp. 1349–1382.

Oster, Emily (2019). **"Unobservable selection and coefficient stability: Theory and evidence".** In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. (2011). **"Scikit-learn: Machine learning in Python".** In: *Journal of Machine Learning Research* 12, pp. 2825–2830. URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Rubin, Donald B (1974). **"Estimating causal effects of treatments in randomized and nonrandomized studies.".** In: *Journal of educational Psychology* 66.5, p. 688.

References

Rubin, Donald B (1977). **"Assignment to treatment group on the basis of a covariate".** In: *Journal of educational Statistics* 2.1, pp. 1–26.

— (1980). **"Randomization analysis of experimental data: The Fisher randomization test comment".** In: *Journal of the American statistical association* 75.371, pp. 591–593.

Appendix

A.1. Other Potential Uses Cases

Our setting covers, e. g.,

- **Stratified randomized controlled trials** with attrition or nonresponse.
- **Observational studies** where rich covariates address treatment selection, but outcomes are selectively missing due to latent factors that also determine Y.

We do **not cover** cases, where latent factors jointly **affect** both **treatment take-up and outcome observability** (e. g., in program evaluations due to unobserved motivation).

Allowing **unobservables** to **jointly affect** D and S would require **additional sensitivity components** for treatment confounding; we leave this important extension to future work.

A.2. Confounding in Observational Studies I

- In an **RCT**, treatment D is assigned randomly, independent of any unit characteristics (implying $Y(d) \perp D$ for $d \in \{0, 1\}$), so that:

$$E[Y \mid D = 1] - E[Y \mid D = 0] = E[Y(1) \mid D = 1] - E[Y(0) \mid D = 0]$$
$$\xrightarrow{\text{Randomization}} E[Y(1)] - E[Y(0)] = \text{ATE}.$$

- To identify the ATE in **observational studies**, researchers often invoke the selection-on-observables assumption, also known as unconfoundedness or the Conditional Independence Assumption (CIA):

Assumption (Conditional Independence of Treatment):
$Y(d) \perp D \mid X = x \quad$ for all $d \in \{0, 1, \ldots, Q\}$ and x.

$\Rightarrow$ This assumption states that **conditional on covariates** X, **potential outcomes** are **independent of treatment status**, yielding $E[Y(d)|D = d, X] = E[Y(d)|X]$.

A.2. Confounding in Observational Studies II

- **Identification** also requires a **common support** (or overlap) condition, ensuring that for all relevant values of the covariates X, there is a positive probability of observing units under each treatment status:

  Assumption (Common Support for Treatment):
  $P(D = d \mid X = x) > 0 \quad$ for all d and x.

Under these assumptions, the **ATE** can be **identified** from observational data **using standard methods** like regression adjustment or propensity score methods (matching, stratification, inverse probability weighting).

A.3. Weak Overlap and Square-Integrability of the Riesz Representer I

- The Riesz approach **requires square-integrability** of the representer, $E[\alpha^2] < \infty$, rather than uniform positivity bounds. As our representer includes the **selection indicator** S, its **second moment** collapses to an *average inverse-probability scale*. Concretely, for $d \in \{0, 1\}$,

$$E\left[\left(\frac{1\{D = d\}S}{p_d(X)\pi_0(d, X, A)}\right)^2 \middle| X, A\right] = \frac{E[1\{D = d\}S \mid X, A]}{p_d(X)^2\pi_0(d, X, A)^2} = \frac{1}{p_d(X)\pi_0(d, X, A)}.$$

$\Rightarrow$ Thus, **Assumption 4 (ii)** is **sufficient**, see also Appendix B in our paper.

(Aligns with the **weak overlap condition** implied by **semi-parametric efficiency theory**, cf. Newey (1994))

A.3. Weak Overlap and Square-Integrability of the Riesz Representer II

- Further, we denote the **conditional mean outcome** by $\mu_d(X) = \mathbb{E}[Y|D = d, S = 1, X]$. Under **Assumption 1**, **Assumption 4**, and **conditional independence of selection**, the ATE is identified by: $\theta_0 = \mathbb{E}[\phi_1 - \phi_0]$ with

$$\phi_d = \frac{\mathbf{1}\{D = d\} \cdot S \cdot [Y - \mu_d(X)]}{p_d(X) \cdot \pi_s(d, X)} + \mu_d(X)$$

being the efficient score function derived by Bia, Huber, and Lafférs (2024).

- Hence, the **ATE is identified** using outcomes Y from the selected sample ($S = 1$) and selection indicators S for all units.

Intuitively, identification involves modeling the **conditional outcome mean** within the **selected sample**, $\mathbb{E}[Y|D = d, S = 1, X]$, and then **appropriately adjusting or re-weighting** based on estimates of the **propensity scores** $p_d(X)$ and the $\pi_s(d, X)$.

B.1. Riesz Representer Estimation

- As shown in Chernozhukov et al., 2021, α_0 is defined as the function in a given set $\mathcal{A}$ that can be obtained by **solving** the following **optimization problem**:

$$\alpha_0 \;=\; \arg\min_{a \in \mathcal{A}} \mathbb{E}[a^2(Z) - 2m(W; a)],$$

  where $\mathcal{A}$ is the **space of candidate functions** and $m(W; a)$ is a m**odel-specific function** of the data W and the candidate a.

- When we find the **minimum** by **setting the derivative to zero**, we obtain exactly the defining **property of the Riesz representer**:

$$\mathbb{E}[\alpha_0\, g_0] = \mathbb{E}[m(W;\, g_0)].$$

- For **practical applications** with data, the expectation is **replaced** by a **sample average**:

$$\hat{\alpha}_0 \;=\; \arg\min_{a \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \Big[a^2(Z_i) - 2\, m(W_i; a)\Big].$$

- This can be **estimated** by several **ML approaches**, e.g., neural networks (RieszNet), random forests (ForestRiesz), or Lasso-type methods.

B.1. Riesz Representer Estimation - Proof

$$
\begin{aligned}
\alpha_0 &= \arg\min_a \mathbb{E}\big[(\alpha_0(X) - a(X))^2\big] \\
&= \arg\min_a \mathbb{E}\big[\alpha_0(X)^2 - 2\alpha_0(X)a(X) + a(X)^2\big] \\
&= \arg\min_a \big\{-2\,\mathbb{E}[v_m(X)a(X)] + \mathbb{E}[a(X)^2]\big\} \\
&= \arg\min_a \mathbb{E}\big[-2m(W,a) + a^2\big].
\end{aligned}
$$

with $v_m(X) = \alpha_0$.

C.1. Benchmarking Sensitivity Analysis

- The **squared bias bound** B^2 has an intuitive decomposition:

$$B^2 = \tilde{S}^2 \, C_Y^2 \, C_S^2.$$

  - Scaling factor $\tilde{S}^2$ is identifiable from **observed** data.
  - C_Y^2 measures **confounding strength** in the **outcome** equation.
  - C_S^2 measures **confounding strength** in the **selection** equation.

While bounded between 0 and 1 both C_Y^2 and C_S^2 are unknown. **Sensitivity analysis maps assumptions** about A to potential bias in θ_s, but does **not** tell us **how plausible** those are.

Researchers **need to make informed decisions** about **how strongly** A **predicts** Y **and** S. A **benchmarking** approach, which uses the **observed influence** of **observed covariates** X_j as a **reference** point for the potential influence of A, can help guiding these decisions (see Imbens (2003), Altonji, Elder, and Taber (2005), Oster (2019), Cinelli and Hazlett (2020), and Chernozhukov et al. (2022a)).

C.2. Quasi-Gaussian Selection Sensitivity I

- **Represent selection indicator S as latent index S^* with Gaussian shocks** crossing a threshold: Let $S = \mathbf{1}\{S^* > 0\}$, with $S^* = h(D, X) - U$ and $U \mid D, X \sim N(0, 1)$.
- **Does not entail loss of generality** | Gaussian parameterization only for interpretation/calibration: Given $\pi_s(D, X) = P(S = 1 \mid D, X)$ with $\pi_s(D, X) \in (0, 1)$, we can take $U \sim N(0, 1)$ independent of (D, X) and set $h(D, X) = \Phi^{-1}(\pi_s(D, X))$, so that S has the same conditional distribution as $h(D, X) - U > 0$.
- Model **confounding** independent of (D, X) as

$$U = \mu_S A + \sqrt{1 - \mu_S^2}\, \varepsilon_S \qquad \text{with} \qquad A, \varepsilon_S \overset{\text{i.i.d.}}{\sim} N(0, 1).$$

$\Rightarrow$ μ_S^2 is the R^2 in the regression of the Gaussian shock U on the latent confounder A. By definition, it is equal to $\eta^2_{S^* \sim A \mid D, X}$, the nonparametric partial R^2 in the regression of the latent index S^* on A, after nonparametrically partialling out (D, X).

C.2. Quasi-Gaussian Selection Sensitivity II

One can also map μ_S^2 to the technical sensitivity parameter:

- Compute short and long selection probability as
  $\pi_s(d,x) = \mathbb{P}(S=1 \mid D=d, X=x) = \Phi(h(d,x))$, so $h(d,x) = \Phi^{-1}(\pi_s(d,x))$, and
  $\pi_0(d,x,a) = \mathbb{P}(S=1 \mid D=d, X=x, A=a) = \Phi\left((h(d,x) - \mu_S a)/\sqrt{1-\mu_S^2}\right)$.

$\Rightarrow$ **Appendix C** in our paper shows that $\mathbb{E}[\alpha_0^2]$ and $\mathbb{E}[\alpha_s^2]$ can be expressed in terms of these probabilities and can therefore be seen as functions of μ_S^2.

- We then derive the **maps** from **interpretable** to **technical sensitivity parameters**:
  $\mu_S^2 \mapsto 1 - R^2_{\alpha_0 \sim \alpha_s}(\mu_S^2)$.

$\Rightarrow$ **Results in** a **one-parameter, probit-scale calibration of selection confounding** that is directly **compatible** with the **Riesz-based bias bounds** and does not impose any assumptions on the data (solely interpretation device).

C.3. Estimating the Riesz Representer: ForestRiesz

- For **each node** in the forest, compute a **Jacobian matrix** and a **local moment vector**

$$J(\text{node}) = \frac{1}{|\text{node}|} \sum_{i \in \text{node}} r(Z_i) r(Z_i)^\top \quad \text{and} \quad M(\text{node}) = \frac{1}{|\text{node}|} \sum_{i \in \text{node}} m(W_i; r).$$

- **Optimal coefficient vector** within each node: $\beta(\text{node}) = J(\text{node})^{-1} M(\text{node})$.

- Grows the forest by **recursively splitting nodes** based on the covariates. For **each candidate split**, the two resulting **child nodes** are **evaluated** by computing their J and M.

- Seeks to **maximize** the **stability-adjusted signal** by **minimizing** the **aggregate local Riesz loss**:

$$- \sum_{\text{child} \in \{1,2\}} |\text{child}| \, \beta(\text{child})^\top J(\text{child}) \, \beta(\text{child}).$$

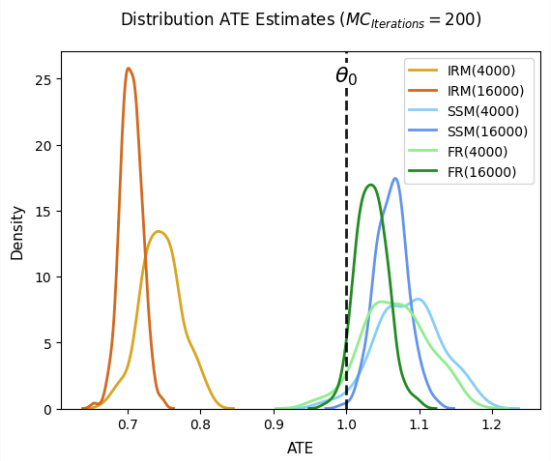
- ForestRiesz **favors splits** that yield **child nodes** where the **local moment** M is **strong and well-supported**, **penalizes** splits that produce nodes with **nearly singular** J.

D.1. Computational Details - Main Simulation

Parameters - scikit-learn	**Parameters - doubleML**
RandomForest classes	IRM and SSM
n_estimators = 500	n_folds= 3, n_rep= 1
max_depth = 20	SSM
min_samples_leaf = 5	score = 'missing-at-random'
max_features ='sqrt'	normalize_ipw = True

Notes: Final hyperparameter set up used for the *RandomForestRegressor* and *RandomForestClassifier* classes from *scikit-learn* (Pedregosa et al., 2011); and settings for the *DoubleMLIRM* and *DoubleMLSSM* estimator classes from the *doubleML* (Bach et al., n.d.) Python package. Parameters not reported are kept at their default values.

D.1. Simulation — Results: Distribution ATE estimates

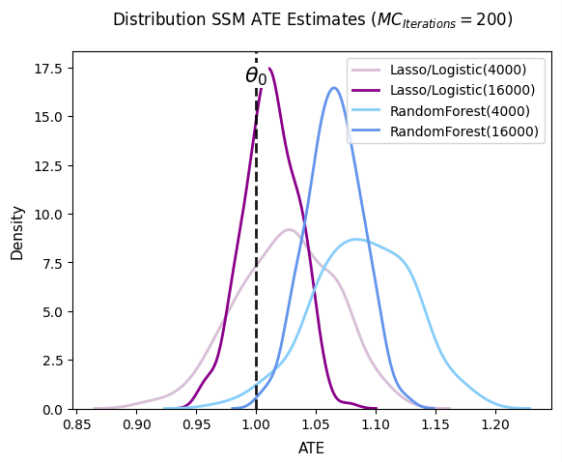


Note: Distribution of ATE estimates based on $\theta_0 = 1$ and 200 Monte Carlo iterations. Comparison of ForestRiesz (FR) to the interactive regression model (IRM) and sample selection model (SSM). Benchmark models use random forests for estimating nuisance functions and three-fold cross-fitting.

With increasing sample size

- **IRM**
  Increasing downward bias.
- **SSM** & **FR**
  Converge to the simulated ATE.

D.2. Simulation — Results: Distribution SSM ATE estimates



Note: Distribution of SSM ATE estimates based on $\theta_0 = 1$ and 200 Monte Carlo iterations. Comparison of sample selection model (SSM) with Lasso/Logistic specification of Bia, Huber, and Lafférs, 2024 and Random Forest specification with $max\_depth = 20$.

	Lasso/Logistic			RandomForest		
N	ATE	SE	MAE	ATE	SE	MAE
1000	1.0511	0.0460	0.0863	1.1228	0.0450	0.1287
4000	1.0254	0.0222	0.0393	1.0895	0.0222	0.0901
16000	1.0123	0.0111	0.0205	1.0653	0.0110	0.0653

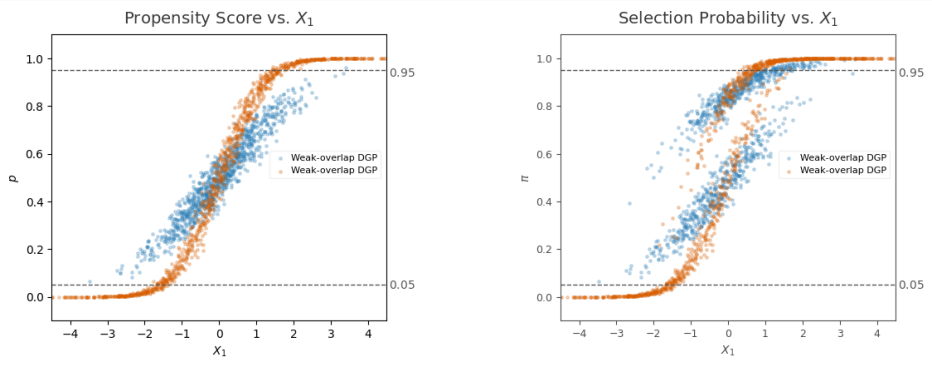
Note: Average SSM simulation results based on $\theta_0 = 1$ and 200 Monte Carlo replications.

- **SSM Lasso/Logistic** converges faster to the simulated ATE than the SSM Random Forest specification.
- Highlights **importance** of proper **hyperparameter tuning** when applying **random forest learners** to estimate the true ATE within the **SSM framework**.

D.3. Weak Overlap — Simulation Results I

- We build on the data-generative process (**DGP**) described **before** and induce **weak overlap** through a **perturbation** of the **covariate distribution**.
- We **increase** the **coefficient** of the **first covariate** to $\beta_1 = 1$ and **inflate its variance** by rescaling it as $2X_{i,1}$ after drawing $X_i \sim N(0, \sigma_X^2)$ during data generation (in the original DGP $\beta_j = \frac{0.4}{j^2}$, $\forall\, j \in \{1, \dots, p\}$). This increases the variance of X_1 by a factor of four, while preserving the correlation structure of the other covariates.
- Compared to the original DGP, the weak design **increases** the **dispersion** of $X_i'\beta_0$ through an amplified contribution of $X_{i,1}$, driven by both its larger coefficient and its higher variance.
- Thus, as propensity scores approach zero or one for large absolute values of X_1, **treatment assignment** and **selection** become **nearly deterministic** in the **tails of the covariate distribution**.

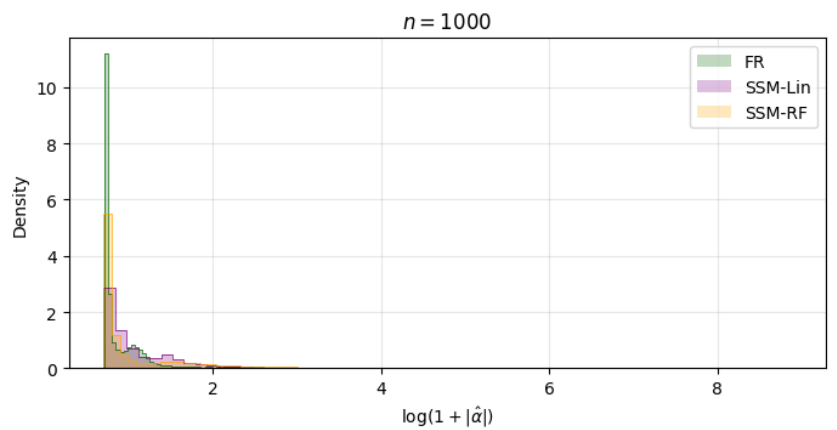
D.3. Weak Overlap — Simulation Results II



Note: Treatment propensity score and selection probability under original and weak-overlap design. Oracle treatment propensity scores (left panel) and selection probabilities (right panel) as functions of the covariate X_1 for a sample of size $n = 1\,000$ from the first Monte Carlo iteration in the $p = 100$ covariate setting. Dashed lines indicate regions with near-deterministic treatment assignment and selection.

D.3. Weak Overlap — Simulation Results III

Histogram of $\log(\mathbf{1} + |\hat{\alpha}|)$ | $p = 100$ | $MC_{\text{Iterations}} = \mathbf{200}$



D.3. Weak Overlap — Simulation Results IV

Tail behavior and second moments of estimated Riesz representers under weak overlap

$p = 400 \quad | \quad \text{MC}_{\text{Iteration}} = 50$

Model	$n_{\hat{\alpha}}$	Mean($\hat{\alpha}^2$)	Max($\lvert\hat{\alpha}\rvert$)	Q.900($\lvert\hat{\alpha}\rvert$)	Q.950($\lvert\hat{\alpha}\rvert$)	Q.990($\lvert\hat{\alpha}\rvert$)	Q.995($\lvert\hat{\alpha}\rvert$)	Q.999($\lvert\hat{\alpha}\rvert$)
				n = 1 000				
FR	28191	6.659	21.739	2.502	4.062	11.815	13.983	17.191
Oracle plug-in	28191	298.273	1744.652	5.233	7.864	24.322	38.988	126.625
SSM-Lin	28191	141186.889	39584.227	4.530	6.910	25.544	53.141	408.258
SSM-RF	28191	50.308	195.261	6.943	12.242	28.384	39.336	70.464

Note: Tail behavior and second moments of estimated Riesz representers under weak overlap for a setting with $p = 400$ covariates, based on 50 Monte Carlo iterations. Total number of nonzero $\hat{\alpha}$ (selected observations), the mean of $\hat{\alpha}^2$ and the maximum and upper quintiles of $|\hat{\alpha}|$ across estimation methods and sample sizes.

D.4. Application — Determinants of Reporting Wages

Dependent variable: S (reported wage indicator)

Interaction	Coef.	SE	p
Experience × Female	-0.0242	0.010	0.011**
Experience^2 × Female	0.0007	0.0002	0.001***
Household size × Female	0.0993	0.021	0.000***
Children < 5 × Female	0.0827	0.063	0.190
Master degree × Female	-0.0173	0.058	0.764
Professional degree × Female	0.3203	0.071	0.000***
Doctoral degree × Female	-0.1325	0.113	0.240
Married (absent spouse) × Female	-0.0862	0.174	0.621
Married (present spouse) × Female	-0.1838	0.072	0.011**
Never married × Female	0.1465	0.085	0.083*
Separated × Female	-0.2638	0.217	0.225
Widowed × Female	0.1148	0.229	0.616
Chinese × Female	-0.3191	0.189	0.091*
Other Asian × Female	-0.2551	0.132	0.053*
White × Female	-0.2321	0.091	0.011**
Not well English × Female	0.2016	0.234	0.388
English only × Female	0.4320	0.147	0.003***
English very well × Female	0.4255	0.159	0.007***
English well × Female	0.4326	0.190	0.023**
Hispanic × Female	0.0152	0.111	0.890
Veteran × Female	-0.1159	0.180	0.519
East South Central × Female	0.3031	0.126	0.016**
Middle Atlantic × Female	0.1689	0.086	0.048**
Mountain × Female	0.1045	0.111	0.348
New England × Female	-0.0324	0.104	0.756
Pacific × Female	-0.0938	0.079	0.238
South Atlantic × Female	-0.0907	0.082	0.270
West North Central × Female	0.1876	0.114	0.099*
West South Central × Female	0.0895	0.093	0.335

Notes: Logit estimates for the probability of reporting wages ($S = 1$). Interaction terms between key socio-economic characteristics and female. Positive coefficients indicate characteristics that increase women's reporting probability relative to men.

Results indicate **substantial gender heterogeneity in wage reporting**, suggesting that non-random selection into observed wages varies systematically across demographic groups.

D.4. Application — Wage Determination

Dependent variable: log(wages)

Variable	**Interpretation**
Age	Life-cycle earnings growth
Experience	Linear experience premium
Experience^2	Concavity of returns to experience
College Degree	Returns to education
Married, spouse present	Household stability effect
Professional degree	Very high skill premium
Household size	Family composition
Never married	Labor supply differences
Pacific Division	Regional wage differences
Doctoral degree	Advanced education returns

Notes: Most frequently used covariates in splitting rules of the ForestRiesz regression learner when predicting Y. Variables with higher split frequency are interpreted as having stronger predictive power for log(weekly wages).

D.4. Benchmarking Sensitivity to Unobserved Confounding I

- Let g_s be the outcome model and α_s the Riesz representer using all observed covariates X.
- Let $g_{s,-j}$ and $\alpha_{s,-j}$ be the versions omitting X_j.

$\Rightarrow$ Then, we can measure the **impact** of the **omitted** X_j with:

- **Outcome Prediction**
  - X_j's impact on predicting the outcome Y (within the selected sample, $S = 1$)
  - $\uparrow$ R-squared when X_j is added to the model.
  - How much does X_j improve outcome prediction beyond other variables?
- **Selection Weights**
  - X_j's impact on the statistical weights α_s used for correction
  - Relative change in the weights' overall size when X_j is included.
  - How much does X_j change the necessary adjustment for selection and treatment assignment?
- **ATE Estimate**
  - X_j's direct impact on the final result
  - Change in the ATE estimate when X_j is included versus excluded as a control variable.
  - How sensitive is the estimated ATE is to controlling for X_j?
- **Alignment of Effects**
  - Relation between X_j's effects on outcome & selection weights
  - Correlation between the changes they cause when X_j is removed.
  - Are X_j's effects on outcome & selection weights work together or against each other?

D.4. Benchmarking Sensitivity to Unobserved Confounding II

This yields the following metrics as **benchmark values** for **sensitivity parameters**:

- **Outcome Gain Metric ($G_{Y,j}$)**

$$G_{Y,j} := \frac{\Delta\eta^2_{Y\sim X_j|D,X_{-j},S=1}}{1-\eta^2_{Y\sim D,X,S=1}} \approx C^2_Y = \eta^2_{Y\sim A|D,X,S=1}.$$

  - **Benchmark:** How much A might explain the remaining variance in Y (after accounting for $D, S = 1, X$)?
  - **Proxy** for the sensitivity parameter $\mathbf{C^2_Y}$.
  - **Assumption:** A's relative contribution to explaining residual outcome variance is similar to X_j's.

- **Selection / Representer Gain Metric ($G_{S,j}$)**

$$G_{S,j} := 1 - R^2_{\alpha_s\sim\alpha_{s,-j}} \approx 1 - R^2_{\alpha_0\sim\alpha_s}.$$

  - **Benchmark:** A's association with the selection mechanism.
  - **Proxy** for the sensitivity parameter $\mathbf{C^2_S}$.
  - **Assumption:** We can link the relative change in the Riesz representer due to A to the change in the Riesz representer due to the observed X_j.

D.4. Benchmarking Sensitivity to Unobserved Confounding III

- **Correlation / Degree of Adversity Metric (ρ_j)**

$$\rho_j := \text{Cor}(g_{s,-j} - g_s, \alpha_s - \alpha_{s,-j}).$$

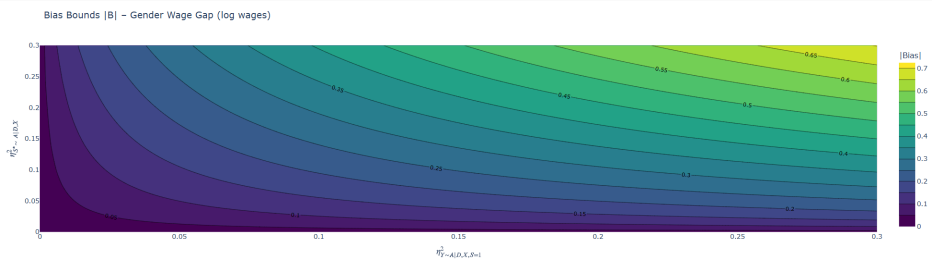
  - **Benchmark:** A's association with the selection mechanism. Measures the correlation between the change in the outcome model g_s and the change in the Riesz representer α_s when X_j is removed.
  - **Proxy** for ρ
  - **Assumption:** How aligned the confounding effects of X_j are on the outcome and selection mechanism (via the RR) are a good indicator for ρ?

$G_{Y,j}$, $G_{S,j}$, and ρ_j provide **concrete reference points**, that **correspond directly** to **values** used in the **sensitivity analysis** (C_Y^2, C_S^2 and ρ).

$\Rightarrow$ Helps to **assess robustness** of study's main **conclusions**. We can check if A would be required to be substantially more influential (in terms of outcome variance explained, impact on the selection mechanism's RR structure, or correlation/adversity) than key observed covariates like X_j.

D.4. Application — Gender Wage Gap - Sensitivity Analysis

Contour plot of bias bounds as a function of outcome sensitivity and selection sensitivity.



- Shows **how large** an **omitted confounder must be** in terms of explanatory power for both wages and selection into observed wages **to overturn** the observed **gender wage gap**.

- **Only confounders** with **combined sensitivity above** the robustness threshold (**RV** $= 0.063$) could eliminate the estimated effect, implying **strong robustness to selection and outcome confounding**.