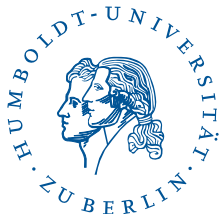


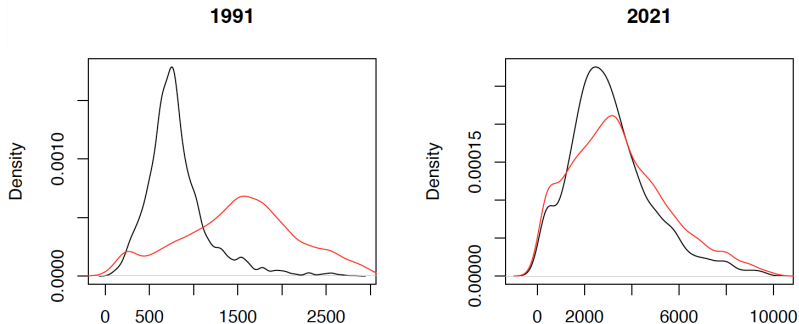
Counterfactual Density Effects and the German East-West Income Gap

Georg Keilbar
Sonja Greven

IZ2DM
Humboldt-Universität zu Berlin
July 10, 2026



East-West Income Gap in Germany



□ Monthly Income East vs. Monthly Income West



Potential Outcome Setup and Notation

- $Y \in \mathbb{R}$ is the outcome variable, $X \in \mathbb{R}^d$ are the regressors, $D \in \{0, 1\}$ is the binary treatment
- Potential outcome framework (Neyman, 1923; Rubin, 1974)
- $Y = DY_1 + (1 - D)Y_0$, where Y_1 and Y_0 are the potential outcomes for the treated and control group
- Define the conditional densities for treatment and control group as $f_{Y_1|X}(y|x)$ and $f_{Y_0|X}(y|x)$



Decomposition Methods

- $Y_k = X^\top \beta_k + \varepsilon_k$, $k = 0, 1$, with $E(\varepsilon_k | X, D = k) = 0$
- Oaxaca (1973) and Blinder (1973) introduce a decomposition for mean differences into **distribution effect** and **covariate effect**:

$$E(Y|D = 1) - E(Y|D = 0) = E(X|D = 1)^\top (\beta_1 - \beta_0) + \{E(X|D = 1) - E(X|D = 0)\}^\top \beta_0$$

- Machado and Mata (2005) and Chernozhukov, Fernandez-Val and Melly (2013) extend the OB decomposition to quantile and distributional effects
- Firpo, Fortin and Lemieux (2019) propose a decomposition method based on recentered influence function regression



Counterfactual Densities

$$f_{Y_{\langle 1,0 \rangle}}(y) := \int_{\mathcal{X}_0} f_{Y_1|X}(y|x) dF_{X_0}(x)$$

- F_{X_0} is the marginal cdf of X_0
- $f_{Y_{\langle 1,0 \rangle}}$ denotes the (unconditional) counterfactual density of the treated if they would have faced the covariate distribution of the control group
- $f_{Y_{\langle 0,1 \rangle}}$ can be defined analogously
- $f_{Y_{\langle 1,1 \rangle}}$ and $f_{Y_{\langle 0,0 \rangle}}$ are the unconditional densities of Y_1 and Y_0



Counterfactual Density Effects

$$f_{Y_{\langle 1,0 \rangle}}(y) := \int_{\mathcal{X}} f_{Y_1|X_1}(y|x) dF_{X_0}(x)$$

- ▣ Counterfactual densities have a causal interpretation under the following assumptions
- ▣ **Assumption 1 (Unconfoundedness)**, $Y_0, Y_1 \perp\!\!\!\perp D|X$
- ▣ **Assumption 2 (Overlap)**, $0 < P(D = 1|X = x) < 1$ for all $x \in \mathcal{X}$
- ▣ Counterfactual density effects:

$$\frac{f_{Y_{\langle 1,1 \rangle}}(y)}{f_{Y_{\langle 0,0 \rangle}}(y)} = \frac{f_{Y_{\langle 1,1 \rangle}}(y)}{f_{Y_{\langle 0,1 \rangle}}(y)} \times \frac{f_{Y_{\langle 0,1 \rangle}}(y)}{f_{Y_{\langle 0,0 \rangle}}(y)}$$

- ▣ Total effect = **distribution effect** $\times$ **covariate effect**



Estimation Procedure

$$\hat{f}_{Y_{\langle 1,0 \rangle}}(y) := \int_{\mathcal{X}_0} \hat{f}_{Y_1|X}(y|x) d\hat{F}_{X_0}(x)$$
$$\hat{f}_{Y_{\langle 0,1 \rangle}}(y) := \int_{\mathcal{X}_1} \hat{f}_{Y_0|X}(y|x) d\hat{F}_{X_1}(x)$$

- Plug-in estimation of counterfactual densities
- $\{(Y_{ki}, X_{ki})\}_{i=1}^{n_k}$ denotes a sample of n_k i.i.d. copies of (Y_k, X_k) , for $k = 0, 1$
- $\hat{F}_{X_k}(x) = \frac{1}{n_k} \sum_{i=1}^{n_k} 1_{\{X_{ki} \leq x\}}$ is the empirical cdf



Additive Density Regression based on Bayes Hilbert Spaces

- Maier et al. (2025, AOAS) introduce a structured additive regression model for density-valued responses
 - ▶ Observations of dependent variable are densities
 - ▶ Estimation via gradient boosting
- Maier et al. (2025, Arxiv preprint) estimate conditional densities from individual observations
 - ▶ Observations are realizations of the conditional distribution
 - ▶ (Penalized) maximum likelihood estimation
- Advantages: flexible modeling of covariate effects and approach covers continuous, discrete and mixed outcome variable case



Bayes Hilbert Space Regression

- Observations $(y_i, x_i) \in \mathcal{T} \times \mathbb{R}^d$, $i = 1, \dots, n$
- Bayes Hilbert space $B^2(\mu)$ (Egozcue et al., 2006; van den Boogaart et al., 2014) [Details in appendix](#)
- $f_i := f(\cdot | x_i) \in B^2(\mu)$ are not observed directly
- Define $\tilde{f} := \text{clr}[f] = \log(f) - \frac{1}{\mu(\mathcal{T})} \int_{\mathcal{T}} \log(f) d\mu$
- Structured additive density-on-scalar regression model:

$$f_i = \bigoplus_{j=1}^d h_j(x_i) \quad \left(\Leftrightarrow \tilde{f}_i = \sum_{j=1}^d \tilde{h}_j(x_i) \right), \quad i = 1, \dots, n$$

- Partial effects h_j mapping x_i into $B^2(\mu)$, e.g., (group-specific) intercept, linear effect, smooth effect



Types of Partial Effects

Covariate(s)	Type of effect	$h_j(\mathbf{x})$
None	Intercept	β_0
One scalar covariate x	Linear effect	$x \odot \beta$
	Flexible effect	$g(x)$
Two scalar covariates x_1, x_2	Linear interaction	$x_1 \odot (x_2 \odot \beta)$
	Functional varying coefficient	$x_1 \odot g(x_2)$
	Flexible interaction	$g(x_1, x_2)$
Grouping variable k	Group-specific intercepts	β_k
Grouping variable k and scalar x	Group-specific linear effects	$x \odot \beta_k$
	Group-specific flexible effects	$g_k(x)$



Specification of Partial Effects in $B^2(\mu)$

$$h_j(x_i) = (\mathbf{b}_j(x_i) \otimes \mathbf{b}_{\mathcal{T}})^{\top} \boldsymbol{\theta}_j = \bigoplus_{l=1}^{d_j} \bigoplus_{m=1}^{d_{\mathcal{T}}} b_{j,l}(x_i) \odot b_{\mathcal{T},m} \odot \theta_{j,l,m}$$

- $\mathbf{b}_j$ vector of basis functions describing covariate effect
(e.g., B-spline basis depending on one covariate)
- $\mathbf{b}_{\mathcal{T}}$ vector of basis functions in $B^2(\mu)$
- $\otimes$ Kronecker product for real- with $B^2(\mu)$ -valued matrix
- $\boldsymbol{\theta}_j \in \mathbb{R}^{d_j d_{\mathcal{T}}}$, vector of coefficients



Specification of Partial Effects in $L_0^2(\mu)$

$$\tilde{h}_j(x_i) = \left(\mathbf{b}_j(x_i) \otimes \tilde{\mathbf{b}}_{\mathcal{T}} \right)^{\top} \boldsymbol{\theta}_j = \sum_{l=1}^{d_j} \sum_{m=1}^{d_{\mathcal{T}}} b_{j,l}(x_i) \cdot \tilde{b}_{\mathcal{T},m} \cdot \theta_{j,l,m}$$

- $\mathbf{b}_j$ vector of basis functions describing covariate effect (e.g., B-spline basis depending on one covariate)
- $\tilde{\mathbf{b}}_{\mathcal{T}} = \text{clr}[\mathbf{b}_{\mathcal{T}}]$ vector of basis functions in L_0^2
- $\otimes$ Kronecker product
- $\boldsymbol{\theta}_j \in \mathbb{R}^{d_j d_{\mathcal{T}}}$, vector of coefficients



Bayes Hilbert Likelihood

- Estimate $\theta_1, \dots, \theta_d$ by maximizing the penalized log-likelihood

$$\ell(\theta_1, \dots, \theta_d) = \sum_{i=1}^n \left\{ \tilde{f}_i(y_i) - \log \int_{\mathcal{T}} \exp(\tilde{f}_i) \right\} + \text{pen}(\theta)$$

- $\text{pen}(\theta) = - \sum_{j=1}^d \theta_j^\top \mathbf{P}_{j, \mathcal{T}} \theta_j$
- Estimation computationally expensive due to integral term
- Solution (for mixed Bayes Hilbert spaces): approximation via multinomial / Poisson likelihood
- Discretize the continuous part of $\mathcal{T}$ by binning observations into histograms



Approximation via Multinomial Likelihood

- Partition the support of Y into G histogram bins
- The values of the histogram are $n_g^i = 1_{\{y_i \in U_g\}}$ with corresponding bin widths Δ_g for $g = 1, \dots, G$
- u_g is the bin center of bin U_g
- Then, we can view the vector $(n_1^i, \dots, n_G^i)$ as a realization of a multinomial variable with class probabilities

$$p_g^i(\theta) = \frac{\Delta_g \exp \left[\left(b(x_i) \otimes \tilde{b}_{\mathcal{T}}(u_g) \right)^{\top} \theta \right]}{\sum_{k=1}^G \Delta_k \exp \left[\left(b(x_i) \otimes \tilde{b}_{\mathcal{T}}(u_k) \right)^{\top} \theta \right]}.$$

- The multinomial log-likelihood converges to the Bayes-Hilbert likelihood if the maximal binwidth approaches 0, $\max_g \Delta_g \rightarrow 0$



Approximation via Poisson Regression

- Multinomial likelihood converges to Bayes Hilbert likelihood if histogram binwidth goes to zero
- Equivalence of multinomial and Poisson likelihood after introducing additional intercept parameters (Poisson trick, see e.g., Palmgren, 1981)
- Implementation in R available via the `mgcv` package



Application: German East-West Gap

- Burda et al. (1997), Kluge and Weber (2018), Dickey and Widmaier (2021) consider OB decomposition for conditional means / quantiles
- Socioeconomic panel dataset (SOEP) provided by DIW Berlin
- Dependent variable gross income has probability mass on zero (mixed reference measure)
- Separate estimation for Easterners, $\hat{f}_{Y_1|X}$, and Westerners, $\hat{f}_{Y_0|X}$
- Place of living is an endogenous variable, we define Easterners by place of living in 1989 / place of birth
- Covariates: sex, age, level of education, rural vs. urban, size of employer



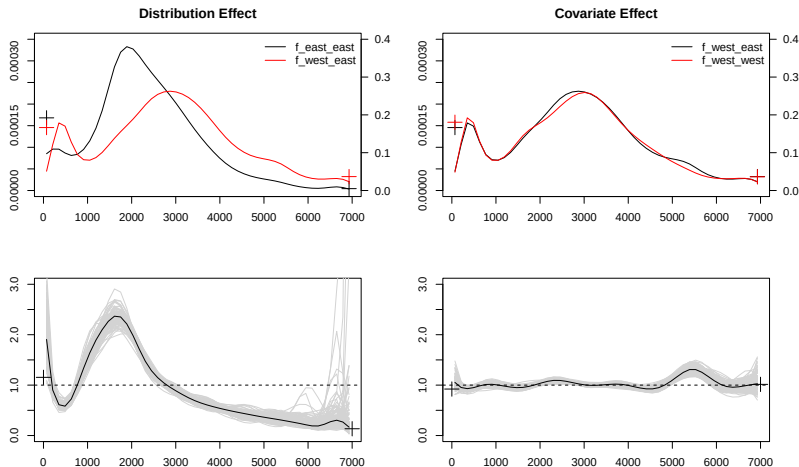
Mean- and Distance-Based Results

	DE	CE	DE	CE
1991	-1426	349	0.558	0.072
2001	-703	102	0.225	0.025
2011	-481	-71	0.149	0.041
2021	-213	-145	0.097	0.034

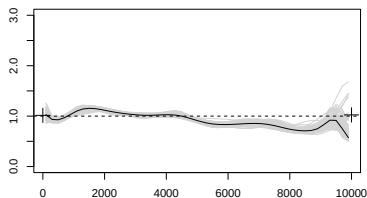
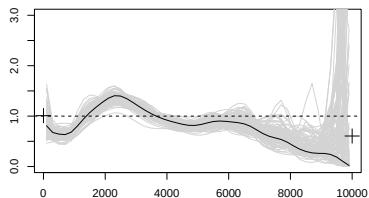
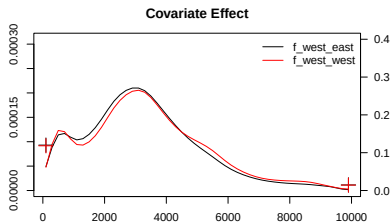
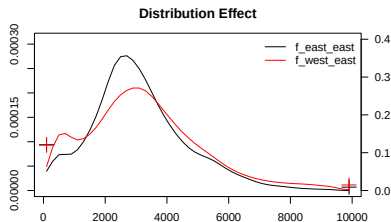
Mean-based OB decomposition (left side); OB decomposition based on scalar density effects following Kennedy et al. (2013).



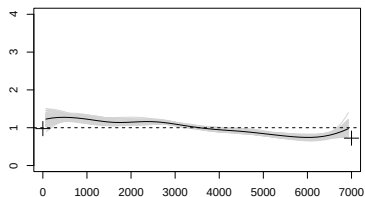
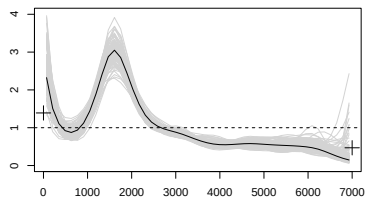
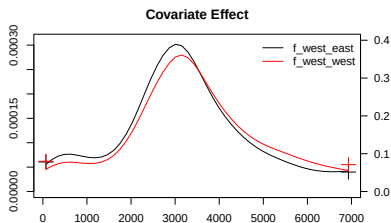
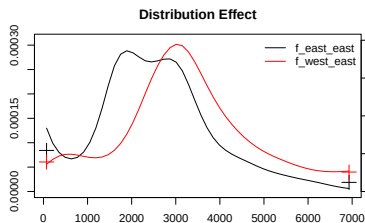
Counterfactual Density Effects (2001)



Counterfactual Density Effects (2021)



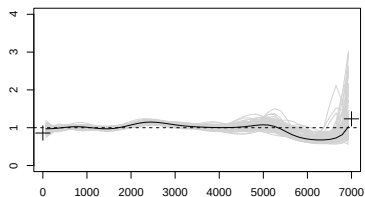
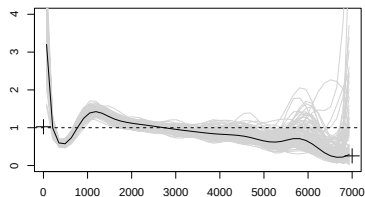
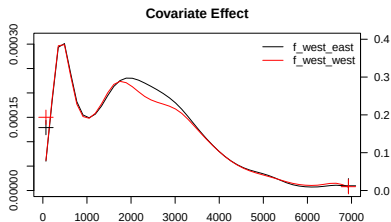
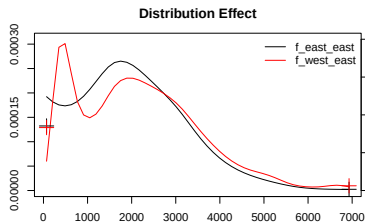
Male Counterfactual Effects (2011)



Counterfactual Density Effects —————



Female Counterfactual Effects (2011)



Counterfactual Density Effects



Conclusion

- We propose a framework for conducting causal inference based on counterfactual densities
- Extension of the Oaxaca-Blinder decomposition to density modeling
- Estimation of conditional densities using additive model in Bayes Hilbert space
- Observed differences in income distribution between east and west cannot adequately explained by differences in covariate distribution
- Preprint available: [arXiv:2603.28470](https://arxiv.org/abs/2603.28470)



Counterfactual Density Effects and the German East-West Income Gap

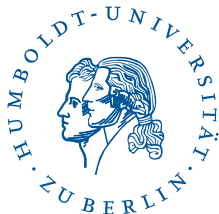
Georg Keilbar

Sonja Greven

IZ2DM

Humboldt-Universität zu Berlin

July 10, 2026



Bayes Hilbert Spaces

Bayes Hilbert Space $B^2(\mu)$ (Egozcue et al., 2006; Boogaart et al., 2014) [Go back](#)

- Let $(\mathcal{T}, \mathcal{A})$ be a measurable space with a finite (reference) measure μ on it.
- Usually $\mathcal{T} \subset \mathbb{R}$ with common cases:
 - ▶ **continuous:** $\mathcal{T} = I = [a, b]$ for $a < b$, $\mathcal{A}$ Borel σ -algebra, $\mu = \lambda$
 - ▶ **discrete:** $\mathcal{T} = \mathcal{D} = \{t_1, \dots, t_D\}$, $\mathcal{A} = \mathcal{P}(\mathcal{D})$, $\mu = \delta := \sum_{d=1}^D w_d \delta_{t_d}$ with $w_d > 0$
 - ▶ **mixed:** $\mathcal{T} = I \cup \mathcal{D}$, $\mathcal{A}$ smallest σ -algebra containing all closed sub-intervals of I and all points of $\mathcal{D}$, $\mu = \lambda + \delta$
- $B^2(\mu)$ is the set of μ -densities $f : \mathcal{T} \rightarrow \mathbb{R}$ with
 - ▶ $f > 0$ μ -almost everywhere
 - ▶ $\int_{\mathcal{T}} \log(f)^2 d\mu < \infty$, i.e. $\log(f) \in L^2(\mu)$.

