DEUTSCHE BUNDESBANK
EUROSYSTEM

# Discussion:

**Cross-country differences in the size of venture capital financing rounds**

Tobias Cagala / Deutsche Bundesbank

October 22, 2019

*The presentation represents the author's personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff.*

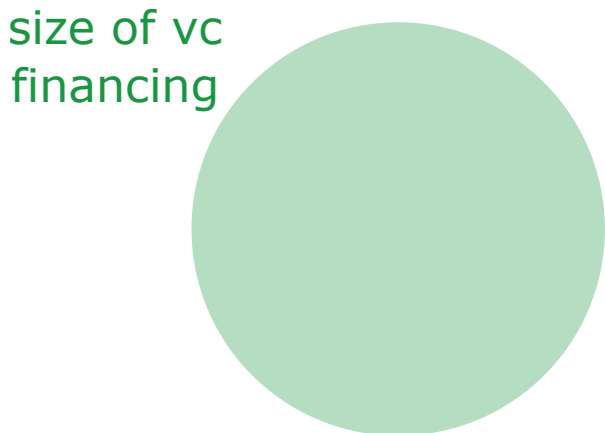
# Summary

- **Outcome:** $y :=$ Size of venture financing rounds
- **Analysis:** $y = f(\widehat{X})$
- **Model:** Predict $\widehat{X}$ with ML
- **Results:**
  1. Firm- and country characteristics $\rightarrow y$
  2. Degree of development of venture capital industry $\not\rightarrow y$

# Comments: Main Result

Degree of development of venture capital industry $\notin X$

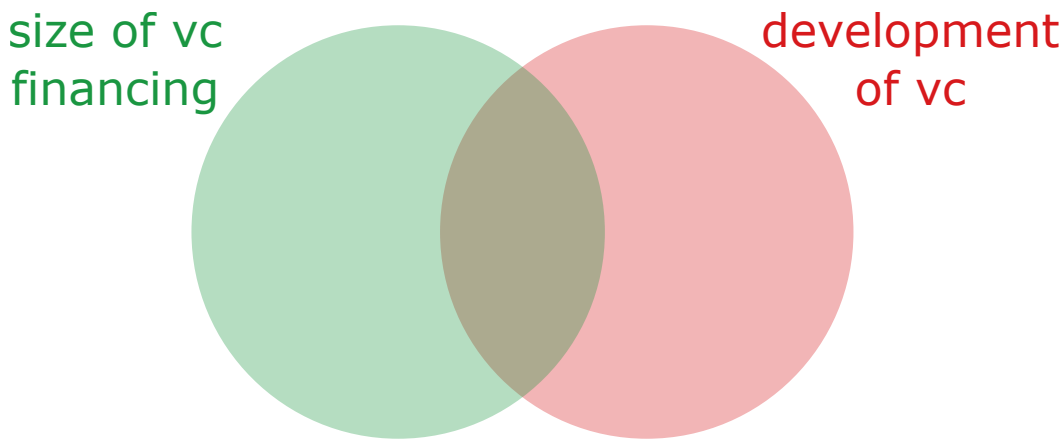
**Lack of identifying variation $\not\equiv$ Lack of causal effect**



# Comments: Main Result

Degree of development of venture capital industry $\notin X$

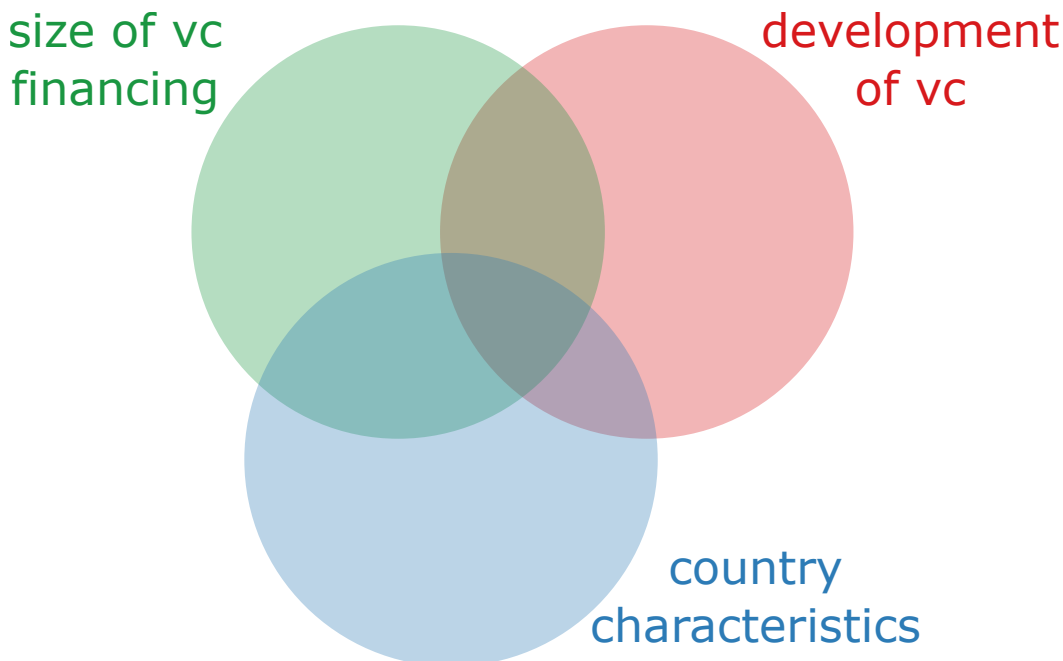
**Lack of identifying variation $\not\equiv$ Lack of causal effect**



# Comments: Main Result

Degree of development of venture capital industry $\notin X$

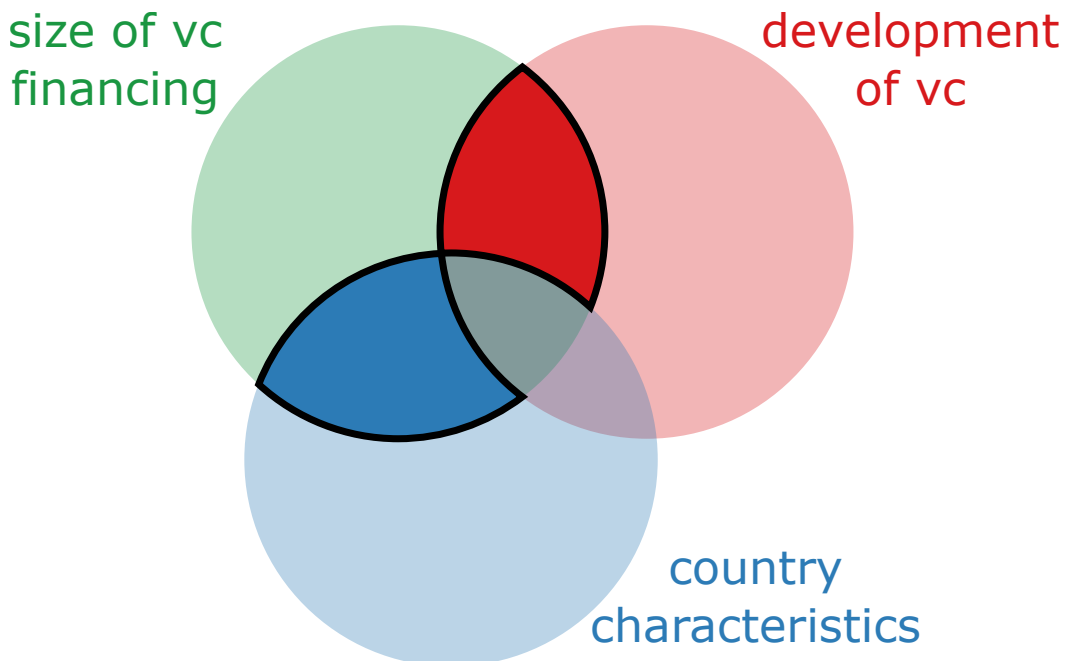
**Lack of identifying variation $\not\equiv$ Lack of causal effect**



# Comments: Main Result

Degree of development of venture capital industry $\notin X$

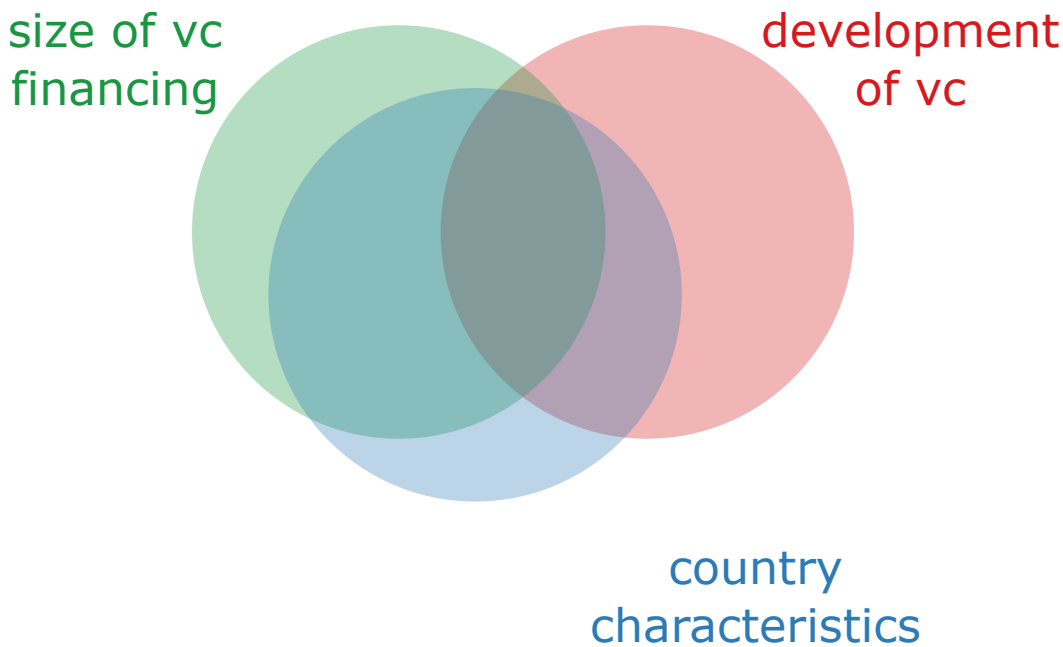
**Lack of identifying variation $\not\equiv$ Lack of causal effect**



# Comments: Main Result

Degree of development of venture capital industry $\notin X$

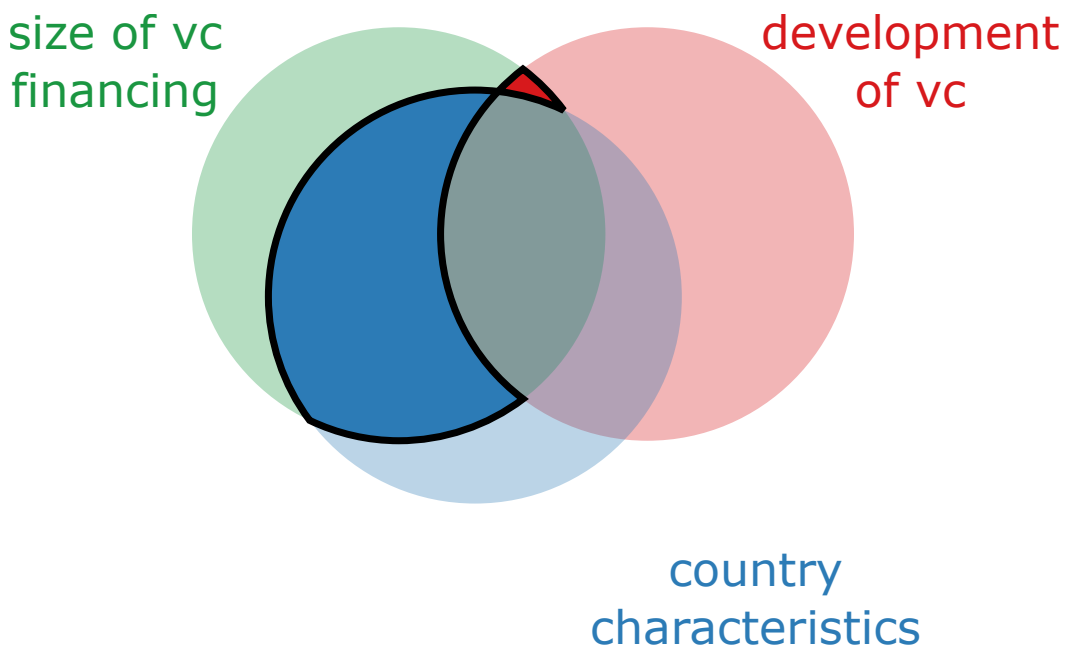
**Lack of identifying variation $\not\equiv$ Lack of causal effect**



# Comments: Main Result

Degree of development of venture capital industry $\notin X$

**Lack of identifying variation $\not\equiv$ Lack of causal effect**



# Comments: Main Result

Degree of development of venture capital industry $\notin X$

**Lack of identifying variation $\not\equiv$ Lack of causal effect**

## Possible solution

- Redundancy argument (already in the paper)
  - You show a high correlation
    ($R^2$ of 50% compared to $R^2$ of 10% in main analysis)
  - Disentangle relationship between variables by moving away from lower dimension (correlation between individual components)
- Move away from agnostic approach
  - More careful selection of control variables
  - Exploit variation over time (if existent)

# Comments: Inference

**Use of cluster-robust standard errors**

- Standard errors incorrect for two-step procedure:
  1. Generate $\widehat{X}$
  2. Estimate $y = f(\widehat{X})$ with cluster-robust standard errors

# Comments: Inference

**Use of cluster-robust standard errors**

- Standard errors incorrect for two-step procedure:
  1. Generate $\widehat{X}$
  2. Estimate $y = f(\widehat{X})$ with cluster-robust standard errors
- Ignores sampling variance in the first-stage estimates
- Similar to 2SLS in IV-estimation, we have to correct the standard errors
  - Angrist & Pischke 'Mostly Harmless Econometrics' Ch. 4
  - Wagner & Athey (2019) 'Estimation and Inference of Heterogeneous Treatment Effects using Random Forests'

# Comments: Inference

**Standard errors are incorrect for two-step procedure**

## Possible solution

- Asymptotic results likely not available
- Use block-bootstrap for entire two-step procedure

# Comments: Machine Learning

**Use of specific ML procedure is not properly motivated**

Alternative approaches for dimensionality reduction

- PCA
- Neural Networks (Autoencoder)
- Selection by domain experts

## Possible solution

- Show results with alternative approaches
- Evaluate benefits of ML procedure (gain in efficiency?)
- Compare out-of-sample predictive accuracy

# Comments: Machine Learning

**Interpretability**

- Interpretation of marginal effects of $\widehat{X}$ is very difficult
- Agnostic ML approach bears many risks

  Example: The World Bank's Starting a Business Scores of Germany and Sierra Leone are very similar

## Possible solution

- ML procedure for robustness check
- Main analysis:
  - Panel model with fixed effects; show within and between $R^2$, or
  - Careful variable selection and modelling choices (e.g., interactions with OECD dummy,...)

# Minor Points

- Low in-sample fit ($R^2 \approx 10\%$)

  Ideas on the source of unexplained variation?

- Null effect as main result

  Convince reader that null is precisely estimated (e.g., by showing confidence interval)

- Do not interpret size of insignificant coefficients

- "Boosted trees and stacked generalization allow us to construct variables that summarize all the information..."

  Dimensionality reduction always implies loss of information