BETTER POLICIES FOR BETTER LIVES

The impact of structural reforms on employment: a country-centric assessment

Eleonora Mavroeidi, Elia Pérennès, Nicolas Woloszko

Bank of England & King's College conference on Modelling with Big Data and Machine Learning
4-5 Nov 2019

Outline

1. Introduction
2. Method
3. Data
4. Empirical Monte-Carlo simulations
5. Estimation results
6. Conclusion

OECD

1. INTRODUCTION

Modeling structural policies with machine learning

- **Estimating structural policies from macro panels**.
  - A large body of literature, including within the OECD Economics Department (Egert and Gal 2016, among others)
  - Relevant for helping governments gauge the return on possible reforms
  - Issue: **country heterogeneity** constrains the application of linear models.
- This paper introduces a new approach that relies on **machine learning** to address causal heterogeneity.

OECD

A country-centric assessment of reforms

Variable generation

- Generating non-linearities (interaction terms)
- ***Ex ante* specification**

Variable selection

- Using Double Post Lasso
- ***Ex post* specification**

OECD

Five benefits

1. Provides **country-centric estimates**, by taking into account both the observed policy conditions and the unobserved country idyosyncratic characteristics.

2. Introduces **cross-country variance** by removing some country fixed-effects using a data-driven criteria.

3. Uncovers new patterns of **policy interactions**.

4. Flexibly captures **short-term effects**.

5. Enhances the **estimates accuracy** of policy effects assessed in Monte Carlo simulations.

II. METHOD

1. Variable generation
2. Variable selection
   - 2.1 Lasso
   - 2.2 **Double Post Lasso**
   - 2.3 Chosing λ

1. Variable generation

<u>Baseline</u>: $y_{i,t+1} = Z_{i,t}\beta + X_{i,t}\alpha + cfe_i + \epsilon_{i,t}$

policies controls Fixed effects

+ **square terms**: $\sum_k \gamma_k z_{k,i,t}^2$

+ **idyosyncratic effects**: $\sum_k \sum_i \gamma_{k,i} * cfe_i * z_{k,i,t}$

+ **policy interactions**: $\sum_k \sum_{k' \neq k} \gamma_{k,\,k'} * z_{k,i,t} * z_{k',i,t}$

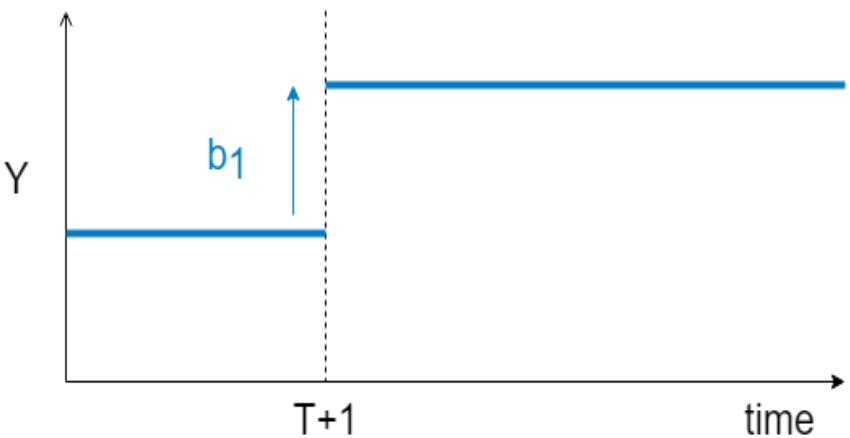
+ **multiple lags***

➔ 1000+ terms in the *ex ante* specification

Timing of policy effects

A. Model with 1 lag

$$y_t = \beta_0 + \beta_1 x_{t-1} + \sigma$$

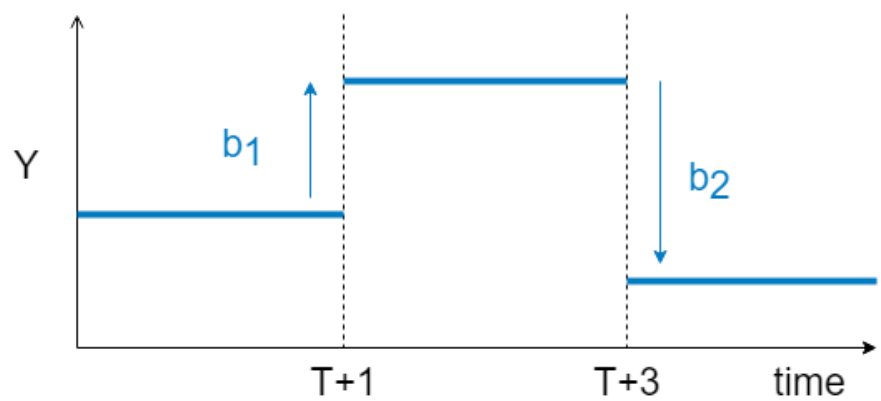


Short-term effect = β_1

Long-term effect = β_1

B. Model with 2 lags

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-3} + \sigma$$



Short-term effect = β_1

Long-term effect = $\beta_1 + \beta_2$

OECD

2. Variable selection

- Variable selection is performed on the generated covariates using **Double Post Lasso** (Chernozhukov, Hansen, Belloni, 2014, JEP)

- DPL uses **Lasso**, a widely-used penalised regression algorithm. A naïve use of Lasso for inference can result in both *regularization bias* and *omitted variable bias*.

- DPL ensures variable selection and **unbiased estimation** with appealing asymptotic properties

OECD

2.1. Lasso

- Lasso minimises the following objective function:

$$\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j})^2 + \lambda \sum_{j=1}^{p} | \beta_j |$$

- It results in a sparse solution (=many zeros in the vector of coefficients), thus performing **variable selection.** Variables with non-zero Lasso coefficient = <u>the active set</u>.

- Lasso is imperfect. Variables with small but significant correlation to the target can be wrongly dropped.

- **Naive inference with Lasso is not valid**:
  - **Regularization bias ➔** Using Lasso for selecting the model, and estimating the mode with OLS: Post Lasso
  - **Omitted variable bias ➔** Double Post Lasso

OECD

2.2. Double Post Lasso

Let y: outcome Z: policy X: set of covariates P(X,Z): generated covariates

STEP 1 – FIRST SELECTION **Lasso regression:** $y \sim P(X,Z) \setminus Z$

Let $\boldsymbol{I_1}$= set of *active variables*

STEP 2 – SECOND SELECTION **Lasso regression:** $Z \sim X$

Let $\boldsymbol{I_2}$= set of *active variables*

STEP 3 - INFERENCE **OLS regression:** $y \sim \{Z\} \cup I_1 \cup I_2$

Unbiased estimation of the effect of the policy Z on y using OLS and standard asymptotic tests

*Active variables**: variables with non-zero coefficient in a Lasso regression

2.2. Double Post Lasso (ctnd)

Assumptions :

- Identifying assumption: all potential confounders are observed

- The true model is **sparse** : only a small number of covariates actually affect the mean outcome

Inference :

- Valid confidence intervals for the DPL's final step OLS coefficients

OECD

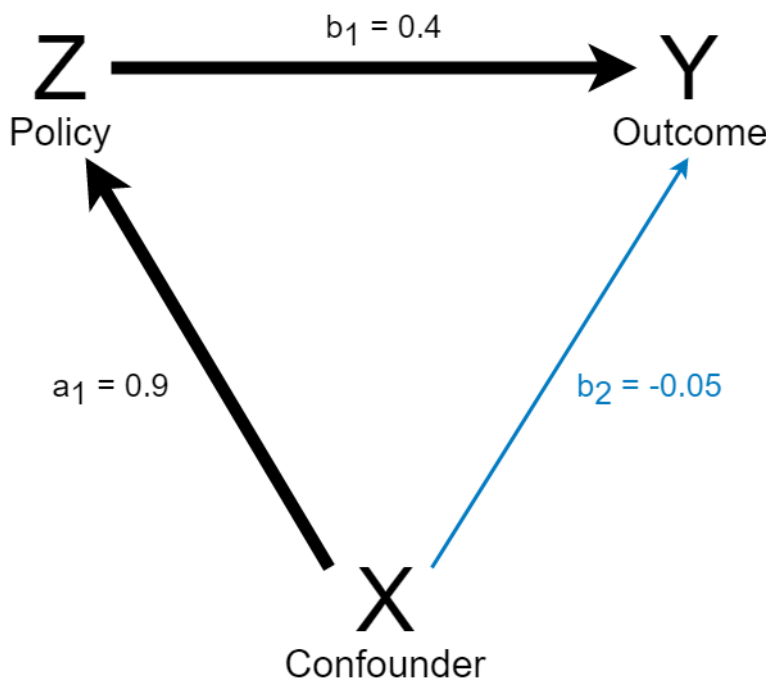
Motivation for *Double* Post Lasso

Small sample issue: weak signals hard to distinguish from the noise.

A confounder (X) weakly correlated to the outcome Y could be dropped, thus resulting in significant bias in the policy (Z) effect estimate.

True model: $Y = b_0 + b_1 Z + b_2 X + \epsilon$

And $Z = a_0 + a_1 X + \varepsilon$



Omitting X in the regression of Y on Z would result in omitted variable bias. The bias is equal to $a_1 b_2$.

Even with b_2 small (b_2 = -0.05), the bias is significant :

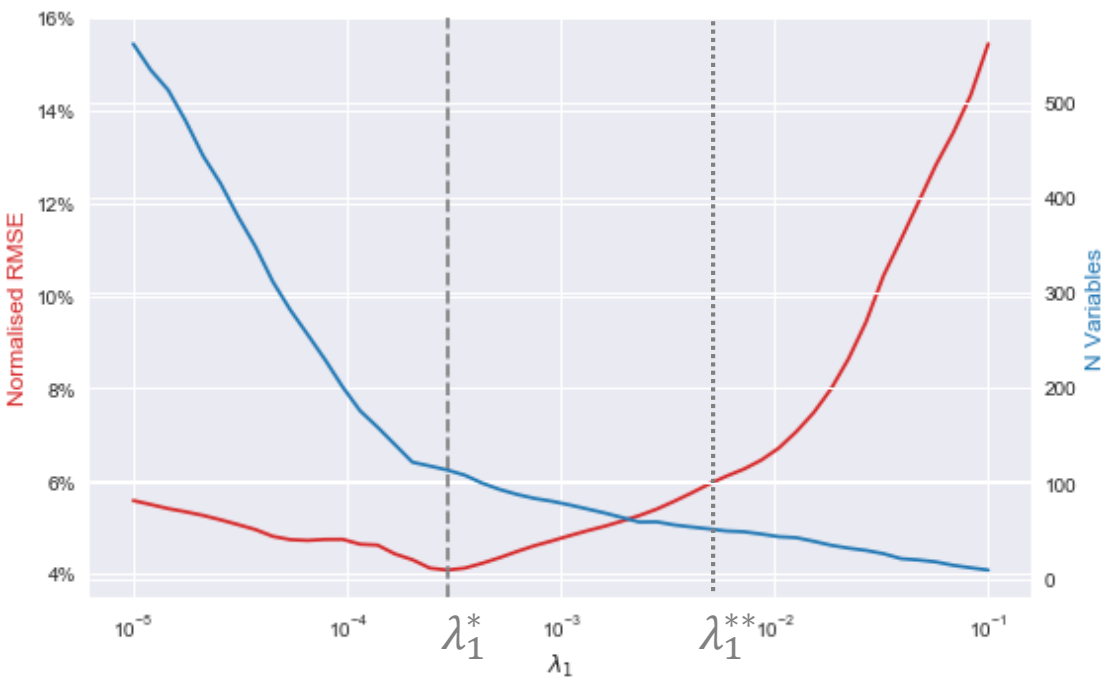
Estimate when omitting X = $0.4 - 0.05 * 0.9 = 0.35$

Bias = 12.5% of true coefficient

Bias can add up in case of several omitted confounders…

2.3. Chosing λ

- Lasso regressions require to determine the **level of regularization λ**.
- High λ => higher regularization => lower number of variables in the active set.
- λ^* is determined using "**forward looking cross-validation**", that is splitting in a training set including all observations earlier than t and a test set including observations in t for 10 values of t.



- Here, λ is chosen in the first DPL Lasso regression in order to yield a given number of active variables (n=50).
- Slightly increasing λ from its optimal value λ^* may greatly reduce the number of variables and thus **enhance interpretability at the expense of a small loss in accuracy**.

3. DATA

OECD

Data and variables

- **Data :**
  - OECD's Structural Policy Indicators Database for Economic Research (SPIDER)
  - Panel data
  - 22 countries, yearly coverage varying with policy variable

- **Variables used in the specifications :**
  - Target variable (Y) : Employment rate (15-64 yo)
  - Policy variables (Z) : EPL, Unemployment benefits, ALMP, Minimum wage, ETCR, Pension age, Family benefits, Weeks of maternity leave
  - Vector of covariates (X_j) : Structural policies, macroeconomic control variables
  - Country fixed effects

OECD

4. EMPIRICAL MONTE-CARLO SIMULATIONS

1. Simulation setting
2. Results: linear data generating process
3. Results: non-linear data generating process

Empirical Monte Carlo Simulations

Monte Carlo simulations where the data generating process (DGP) involves empirical data.

Actual outcome

Simulated policy

$$y' = y + f(Z)$$

Simulated outcome

Simulated policy effect

The policy effect $\hat{f}$ is estimated using the proposed method. The **policy estimated marginal effect** is computed as $\hat{\rho} = \hat{f}(Z+1) - \hat{f}(Z)$. It is compared to its true value $\rho = f(Z+1) - f(Z)$, that is unknown in the nature, using RMSE$(\rho, \hat{\rho})$.

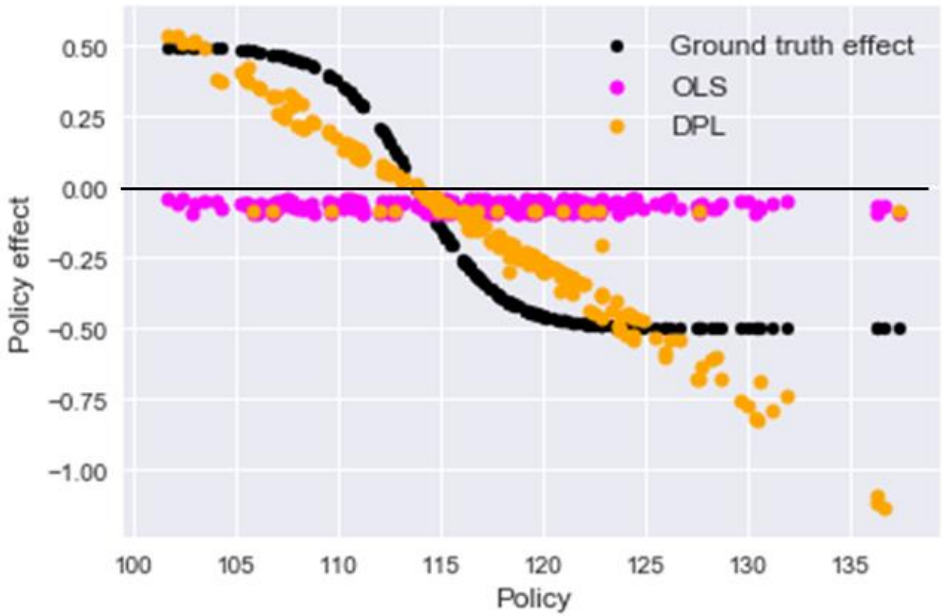
Simulations are performed with both a linear and non-linear simulated policy effect f. Simulations are replicated 100 times.

Results

		Out-of-sample RMSE on marginal effect
Linear DGP	Within	0.061
	DPL	*0.025*
Non-linear DGP	Within	0.39
	DPL	*0.16*

The causal accuracy measures how well the effect of a reform (of magnitude 1) is estimated. With both linear and non-linear DGPs, **DPL provides sizeably more accurate estimates of the reform effect** than a standard within estimator.

Non linear DGP: $f(Z)$ has a Bell curve shape so that the *policy marginal effect* $f(Z+1) - f(Z)$ is positive below a threshold Z^* and negative after



5. ESTIMATION RESULTS

Results

The estimation results yield:

- Heterogenous country effects
- Consistency with OECD estimates
- Some interactions between policies
- Interaction with business cycle is non significant

OECD

Model

- **One model for each policy**. Defined by:
  - *Ex ante* specification including multiple interactions and lags
  - *Ex post* specification resulting from the two selection steps

- Model: $Y = \beta_0 + \beta_1 * Z + \beta_2 * X * Z + \sigma$

- In a simple model without interactions, β_1 is interpreted as the marginal effect of Z. Here, **the marginal effect** of policy Z is :

$$\frac{dY}{dZ} = \beta_1 + X * \beta_2$$

- As a consequence, our betas are not comparable with betas in the literature.

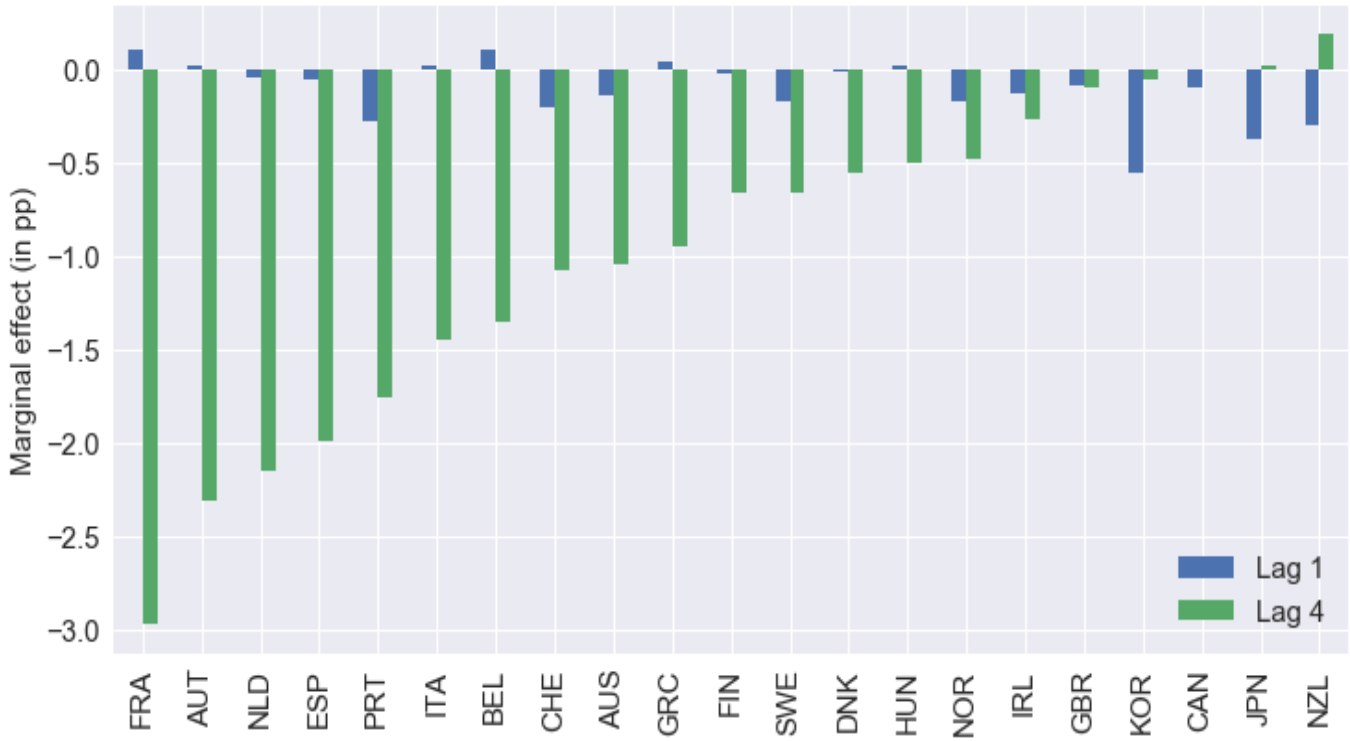
➔ This section shows the **marginal effects at the various lags of Z**: $\frac{dY_t}{dZ_{t-1}}$, $\frac{dY_t}{dZ_{t-6}}$, <u>that can be compared with the betas found in the literature</u>.

- Short term effect: $\frac{dY_t}{dZ_{t-1}}$, Long term effect: $\frac{dY_t}{dZ_{t-1}} + \frac{dY_t}{dZ_{t-6}}$

OECD

Product Market Reforms

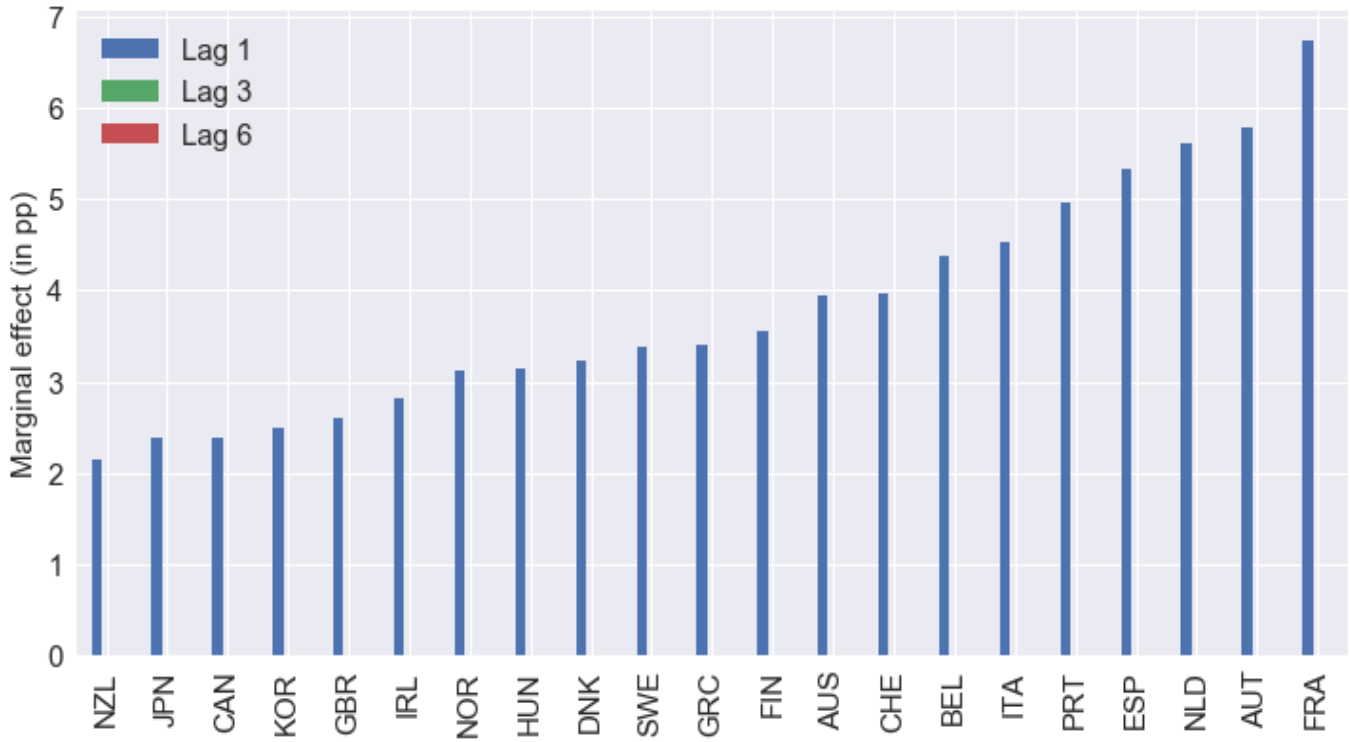
Marginal Effect of ETCR on employment



	Lag 1	Lag 4
ETCR	3.36	
ETCR * Age of retirement	-0.0546	
ETCR * Excess coverage		-0.0329***

Active Labour Market Policies

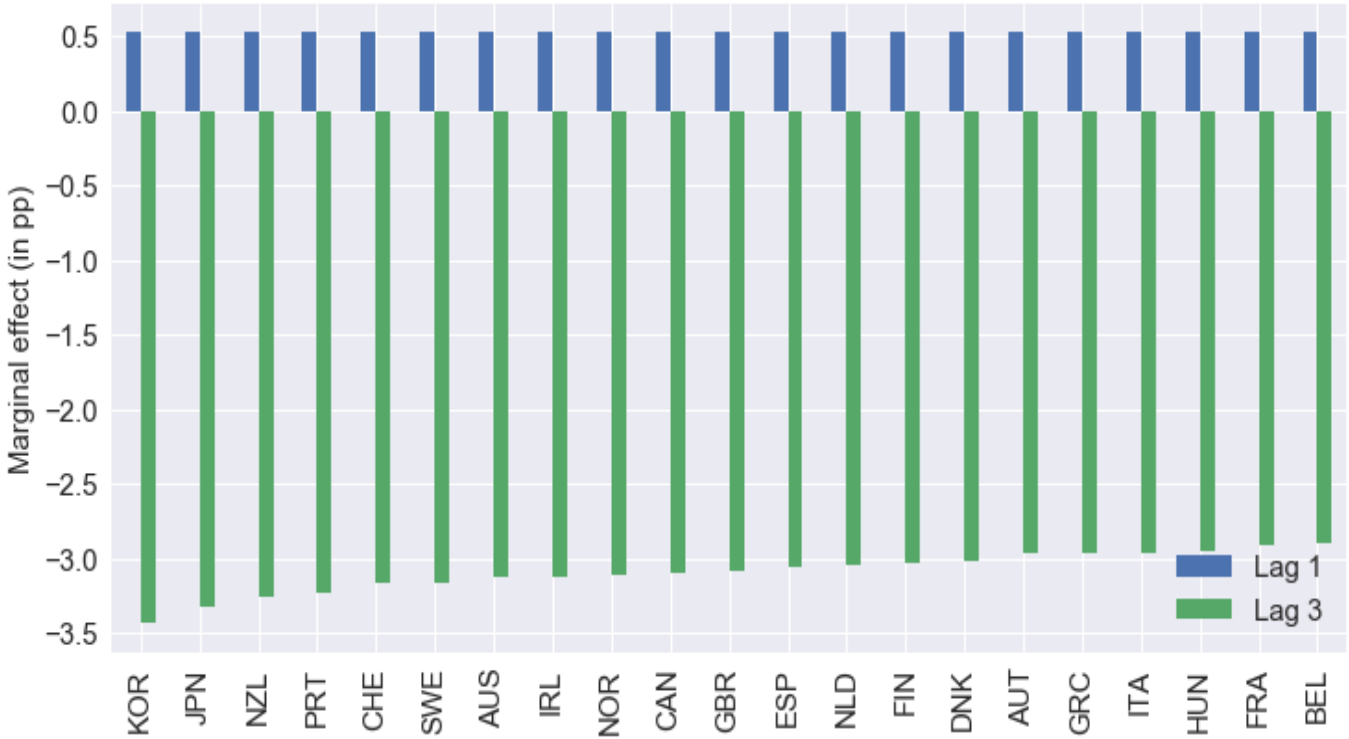
Marginal Effect of ALMP on employment



	Lag 1	Lag 3	Lag 6
ALMP	2.41**		
ALMP * Excess coverage	0.0479*		

Employment Protection Legislation

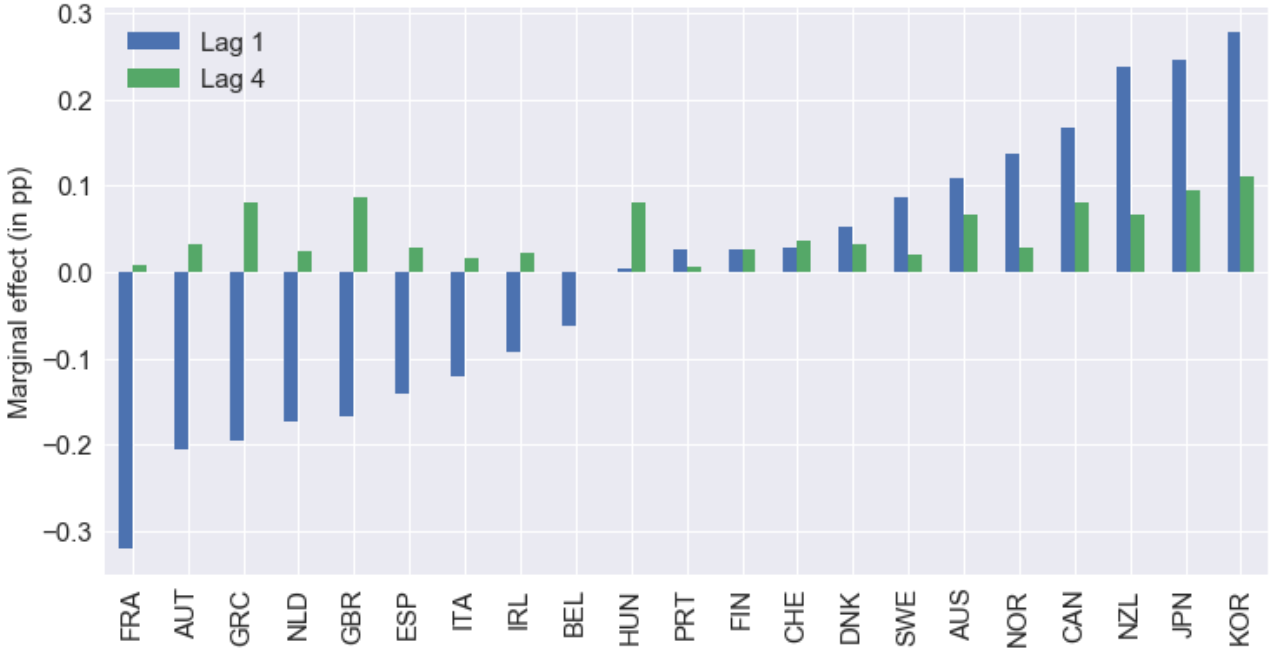
Marginal Effect of EPL on employment



	Lag 1	Lag 3
Age of retirement * EPL		-0.0469***
EPL	0.554	

Unemployment Rate

Marginal Effect of UE on employment



	Lag 1	Lag 4
Age of retirement * Unemployment benefit replacement rate	0.0102	0.00188
Unemployment benefit replacement rate	-0.335	
Unemployment benefit replacement rate * Excess coverage	-0.00504***	
Unemployment benefit replacement rate * Weeks of maternity leave	-0.00885	0.000246
Unemployment benefit replacement rate^2		-0.00138

Conclusion

- A new approach that relies on **machine learning** to address causal heterogeneity
- Consistent with current OECD estimates

Next steps:

- Extend to analysis for other outcome variables (capital, MFP)
- And to more granular data

Thank you very much

OECD

ADDITIONAL MATERIAL

OECD

Variables

POLICIES
Tax-benefit and activation
Unemployment benefit replacement rate
ALMP spending on unemployed as % of GDP/capita (HP-trend)
Tax wedge (couple, 2 children)
Wage setting institutions
Minimum wage (% of median)
Adjusted bargaining coverage rate: proportion of all wage
Excess coverage
Labour and product market
EPL regular workers
ETC regulation
Pension system
Legal age for pensions (total)
Other policies
Family benefits in cash (% of GDP)
Number of weeks of maternity leave
CONTROLS
Average Years of Schooling (Morrisson and Martin) - interpolated
Central government debt, total (% of GDP)
GDP, PPP (constant 2011 international $)
Government primary balance (% GDP)
Population, total
BUSINESS CYCLE INDICATOR
Output gap (smooth transition function)
OUTCOME VARIABLE
Employment-population rate (15-64)