KRTK-KTI WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

# Where is the pain the most acute?

## The market segments particularly affected by gender wage discrimination in Hungary

OLGA TAKÁCS – JÁNOS VINCZE

**KRTK-KTI WP – 2023/04**

March 2023

https://kti.krtk.hu/wp-content/uploads/2023/03/KRTKKTIWP202304.pdf



INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
EÖTVÖS LORÁND RESEARCH NETWORK (ELKH)
BUDAPEST

ABSTRACT
The gender earnings gap can be attributed either to the different distribution of males and females across jobs or to within job biases in favour of men. The latter is frequently called the wage structure effect, and it may be interpreted as wage discrimination against women. In this paper we focus on this second source of the gap. In particular, we study the heterogeneity of the wage structure effect by looking for the main drivers of it. On Hungarian matched employer-employee data we identify those firm-worker profiles that exhibit extremely high gender wage differentials We apply the Causal Forest methodology, borrowed from the conditional average treatment effect (CATE) literature, which has been utilized in several observational studies, recently. Our findings show that those firms that pay relatively high wages tend to discriminate against women most strongly, and especially with respect to women who have spent a longer time in the same firm. But this tendency is moderated by regional effects; where demand side competition is strong the wage structure effect tends to be smaller. These findings are, by and large, in accordance with the view that relative bargaining power is relevant for wage-setting, or, alternatively, firms practice third degree wage discrimination.

JEL codes: J16 J31 C14
Keywords Gender pay gap, heterogeneous wage structure effects, random forest regression.

Olga Takács
Corvinus University of Budapest
e-mail address:tackyolgi@gmail.com

János Vincze
Corvinus University of Budapest and
Centre for Economic and Regional Studies, Hungary
e-mail address: vincze.janos@krtk.hu

# Hol fáj leginkább?
## A legnagyobb nemek közti bérkülönbségekkel sújtott részpiacok Magyarországon

TAKÁCS OLGA – VINCZE JÁNOS

<u>ÖSSZEFOGLALÓ</u>
A nemek átlagbérének különbsége betudható részben annak, hogy a férfiak jobban fizető pozíciókat foglalnak el, részben pedig annak, hogy adott foglalkozásokon belül is nagyobb fizetést kapnak. Ez a tanulmány az utóbbi kérdéskörre koncentrál, és azt kutatja, hogy a munkapiac mely szegmenseiben mutatható ki a legnagyobb százalékos bérkülönbség magyar (Bértarifa) adatokon. Az Oksági Erdők módszertanát használva azt találjuk, hogy leginkább azok a vállalatok diszkriminálnak, amelyek nagy bért fizetnek (feldolgozóipari, külföldi tulajdonú nagyvállalatok), és amelyek elsősorban a Közép és Nyugat Dunántúlon helyezkednek el. A diszkrimináció mértéke különösen nagy azoknál a munkavállalóknál, akik régebben (legalább 5 éve) dolgoznak ugyanazon a munkahelyen. Ezek az eredmények összhangban vannak azokkal az elméletekkel, amelyek szerint a bérek jelentős mértékben függenek a munkavállalók relatív alkuerejétől.

JEL: J16 J31 C14
Kulcsszavak: nemi bérkülönbségek, heterogén bérstruktúra hatások, véletlen erdő regresszió.

## I Introduction

The "gap" between men's and women's wages is a chronic policy problem. It reveals itself as a considerable difference in average wages, but it also exhibits substantial heterogeneity which shows up in several ways. For instance, there is ample evidence that sorting to occupations and sectors (Blau and Kahn, 2017), or even to firms (Card et al, 2016**),** can account for a large part of the average earning gap. Furthermore, it is also suggested that disparity in bargaining power or in labour supply elasticities drive a wedge between female and male wages even in the same enterprise or industry (Meng, 2004; Heinze-Wolf, 2010).

In this paper we study the heterogeneity of the expected difference between the (log) wages of men and women with the same observable characteristics on Hungarian matched employer-employee data. Observable characteristics include employee and employer specific properties. We define an elementary market as a combination of non-gender covariates, a "profile" of an employer-employee relationship. This market concept is consistent with the common finding that wages are specific to these relationships. Our primary goal is to estimate the expected (log) difference between male and female wages in each elementary market. The literature usually refers to these quantities as heterogeneous wage structure (HWS) effects. Our notion of a 'market' seems to be abusively microscopic, but, by aggregating elementary markets, we can speak, for example, of the market for the work of middle-aged women with tertiary education working in the manufacturing sector, which is meaningful. We will make statements only about aggregates of elementary markets, called submarkets, in the following.

Our main goal is to look for those submarkets that display the largest HWS effects, therefore we would like to discover those segments of the labour market where it is most likely that a potential policy intervention is sensible. Also, we hope that, by identifying these submarkets, we can learn something about the working of the labour market, in general. Thus, we look at the problem of heterogeneity from a somewhat unusual perspective. A more customary approach is to estimate mean wage structure effects and figure out the contribution of individual covariates. This usually involves estimating parametric female and male wage regressions, which constitutes an intermediate step in the traditional Oaxaca-Blinder decomposition. However, recent studies have accorded that a simple, predefined functional form cannot be a very good approximation to the theoretical conditional expectation function, and have made allowances for flexibility. Starting from a specification which contains many interactions and higher order terms employing LASSO-type estimation methodologies (Tibshirani, 1996, Fan and Li, 2001) has been an avenue followed by some researchers (Böheim-Stöllinger, 2021; Briel-Topfer, 2020; Strittmatter-Wunsch, 2021). This has the advantage of making possible the calculation of average partial derivatives with confidence bands. Still, partial derivatives should be interpreted with caution if there exists substantial correlation between covariates. Our approach tries to sidestep these problems, as we

make use of non-parametric estimates which are post-processed with a view towards identifying heterogenous submarkets.
How can we distinguish between focussing on submarkets or on variables? We want to come up with statements like this: “women of modest educational accomplishment and working for foreign-owned firms in the manufacturing sector are especially vulnerable in terms of wage discrimination”. This assertion would not imply that foreign-ownership or educational achievement have separate “causal” effects on discrimination. Its validity does not preclude either that foreign-owned firms do not discriminate in the submarket for university graduates, or that unskilled workers in the service sectors do not face substantial discrimination. Isolating the effect of foreign-ownership is practically impossible since settlement choice is non-random, manufacturing firms may prefer less densely populated and/or poorer regions, where they may have more power on the labour market (Fazekas, 2005; Manning-Petrangolo, 2017). Our covariates are not independent for more obvious reasons, too. Young workers with long tenure are obviously non-existent. The connection between age and tenure is complex, and identifying independent effects to tenure and age seems to be a futile enterprise. Still, a statement like “middle-aged women with long tenure are particularly exposed to wage discrimination” may make sense.

The emphasis on submarkets or subgroups of women can be useful from a practical policy perspective as policies are usually defined for market segments. Suppose, for instance, an intervention targets young and relatively uneducated people. An obvious precondition is that policy makers believe that age-education combinations have “heterogeneous” effects on some relevant outcome. Probably it would be impractical and unnecessary to define the target group by the presumed numerical effects of the relevant variables separately. Our point is that the estimation of individual parameters is neither sufficient nor necessary to settle on a target group. Health policy and marketing employ the concept of audience segmentation, where researchers explicitly search for highly affected (target) “audiences”, i.e. subgroups (Slater, 1996). Effectively, we propose to obtain audience segmentation on the Hungarian labour market, where the target audience contains the elementary markets, most highly affected by gender wage discrimination. We exploit several methodologies and seek the target audience as the one that robustly turns up as such in our estimates. Since we cannot exclude the existence of relevant unobserved variables (confounders) the emphasis on extreme groups can lend robustness to the diagnosis. We may miss relevant labour market segments, but, from a policy point of view, it is prudent to consider interventions only when there is strong evidence for their potential usefulness.

Our econometric procedure consists of two phases. In the first phase we estimate HWS effects for each point in the covariate space, by utilizing non-parametric Machine Learning (ML) methods. Estimating HWS effects is analogous to estimating CATE **(**Fortin et al., 2011; Jacob, 2021**).** This problem has been given substantial attention in the ML literature recently (Künzel et al., 2017, Willke et al (2012), Lipkovich et al. (2017)**)**, and also within econometrics **(**Wager-Athey, 2018; Athey-Wager, 2019; Athey et al., 2019). Basically, there exist two main approaches (Jacob, 2021): 1. applying some "base learners" for the control and treated populations and combining the estimates through some meta-learner, 2. estimating the effects directly. We explored both possibilities; finally our preferred method turned out to be the Causal Forest procedure of Athey et al, 2018 (a direct method), and we report results only thereof.

The second methodological issue concerns the best way to diagnose the groups at highest risk of being discriminated against. In the second phase we process the estimated HWS effects in order to isolate most highly affected submarkets. First, having obtained the HWS estimates we fit shallow regression tree models by the 'ctree' algorithm **(**Hothorn et al., 2006) to exhibit contiguous subsets of the covariate space that contain extremely high and low HWS effects. Regression trees have unpleasant features in terms of variance, and the step functions they deliver can be bad approximations of the true relationship. Thus, it is reasonable to complement this analysis with a possibly more robust one. We implement the CLAN methodology of Chernozhukov et al. (2018a) to find extreme groups from a different perspective. It consists in estimating the characteristics of the most and least affected subgroups, defined as the upper and lower 20 percent of the population in terms of estimated HWS effects, via a jack-knife estimator. The first approach is similar to finding those relatively large segments in a heat map where the temperature is extreme in both directions, while the second reveals the averages location of the hottest and coldest points.

Finally, we consider the market segment(s) that exhibit consistently very large gender wage gaps. The properties of these submarkets are analysed, and we interpret our findings in the light of the theoretical and empirical literature. In terms of individual characteristics, and perhaps somewhat surprisingly, the most "vulnerable" group consists of middle-aged women without university education who work for a relatively long time in the same firm. On the other hand the most "discriminating" enterprises are those that pay higher wages in general (foreign-owned, large manufacturing companies), but with a twist: they do not operate in the most affluent region, where average wages are the highest.

In the next section we relate our work to the previous literature. In Sections III and IV we describe the statistical methodologies and the data, respectively. The following section reports the findings, the penultimate analyses the extremely discriminating submarkets, and the concluding discusses the results.

**I**

## I Related literature

### 1 Gender wage gap

Blau-Kahn (2017) is a general overview on the literature on the gender wage gap. Apparently, a large amount of attention has been paid to study the gap for certain subsamples, and many papers have focussed on heterogeneity by some (for instance occupational) dimensions. However, as Bach et al. (2016) pointed out, it may be useful to address heterogeneity in an unprejudiced way, in order to find out what the main drivers of heterogeneity are, or to pinpoint the variables chiefly responsible for it. To carry out this program one has to start with a general specification, for which purpose the adoption of some ML method seems to be imperative. Bach et al., (2016) started with a high-dimensional nonlinear (in-variables) least squares regression, and uncovered substantial heterogeneity. In the same vein, several papers have applied ML techniques to the gender wage gap problem, though their focus have been either on the gap decomposition (Briel-Topfer, 2020), or on methodology (Strittmatter-Wunsch, 2021).

As we have access to matched employer-employee data our paper is intimately related to the literature on how the gap is shaped by firms' policies. Card et al. (2016) distinguished between sorting and bargaining effects on Portuguese data, and concluded that both contributed significantly to the overall gap. Their bargaining effects correspond to within firm gaps which were also identified by Heinze and Wolf (Heinze-Wolf, 2011), who, on German data, found that these are excessively variable across firms.

Many papers have been written about the Hungarian gender wage gap in the last 30 years. In the 1990s authors were mainly interested in the effect of transition (Csillag, 2006; Brainerd, 2000; Newell and Reilly, 2001; Kertesi-Köllő, 2001). Later on, studies addressed heterogeneity along the wage distribution (Newell-Reilly, 2001; Lovász, 2013**)**, and the distinction between the public and corporate sectors (Lovász, 2013). A series of papers concentrated on family roles from a human capital point of view. A general finding was that characteristically female roles in families resulted in less human capital and a poorer bargaining position for women, both increasing the gap. (Szabó-Morvai, 2018; Galasi, 2002 a, b).

**2 CATE estimation**

On the methodological side, as Fortin et al. (2011) pointed out, the wage structure effect can be regarded as a treatment effect, but without the causal interpretation. Sometimes ago the choice of an ML methodology for estimating heterogenous treatment effects seemed strange but, as Mullainathan-Spiess (2017) underlined, the „predictive" ML methods are completely reasonable choices for this purpose as well. Recently, a large literature developed in general statistics with the express purpose of creating „meta-learners" whose output is the measurement of heterogeneous treatment effects (Künzel et al.; 2019, Jacob, 2021). Several theoretical studies appeared also in the econometric literature (Athey-Imbens, 2016; Wager-Athey, 2018; Athey et al., 2019) with a similar goal. Knaus et al. (2021) is a study on the comparative performance of several methodologies.

The ML literature (Molnar, 2022) is aware of the importance to interpret or process the usually non-parametric ML estimates. There are several suggestions how to do it, and we follow Molnar (2021) by fitting shallow regression trees on the HWS effects with the 'ctree' algorithm of Hothorn et al. (2006). Also, we employ the CLAN methodology proposed in Chernozhukov et al. (2018a) whose purpose is „to discover *ex post* whether there is any relevant heterogeneity in treatment effect by covariates" (Chernozhukov et al., 2018a, p. 3). Papers methodologically similar to ours have mushroomed in recent years (Elek-Bíró, 2021; Deryugina et al., 2019; Knaus et al., 2021; Bertrand et al., 2021; Hussam et al., 2022; Davis-Heller, 2020; Knittel-Stolper, 2021; Murakami et al., 2020; Chowdhury et al., 2021; Christiansen-Weeks, 2020. Baiardi-Naghi (2021) revisits a number of previous studies by using ML techniques to estimate CATE, and shows that interesting novelties emerge from these exercises.

## III Methodologies

### 1 HWS effects estimates

We can regard our fundamental estimands (the HWS effects) as conditional average treatment effects (CATE), where we consider, arbitrarily, men as the treated sub-population. Let Y denote log wages, and X a multidimensional set of primary covariates common to both sexes, where we assume common support. We estimate

$$\mathrm{HWS(x)} = \mathrm{E(y|x, T = 1))} - \mathrm{E(y|x, T = 0)},$$

where $\mathrm{E(|)}$ denotes conditional expectations, x is an element of X, and T=1 for a man and T=0 for a woman. In the potential outcome framework with the assumption of conditional non-confoundedness one can rightfully speak of a causal effect with the interpretation that the deliberate change of the position of someone with profile x from T=0 (woman) to T=1 (man) would increase her salary by HWS(x) log points in an expected value sense. We abstain from this interpretation for obvious reasons, and, also, because the fulfilment of the non-confoundedness assumption is more than doubtful. We rather look at the estimand from an observer's , rather than an engineer's, point of view. What is the mean log difference between the wage of a man and a woman, if both have profile x? The likely existence of confounders implies that one can possibly find some explanation for the wage differential related to the different (unobservable) character of genders. For instance, if the gap is affected by the fact that there is a gender difference between human capital accumulation due to the more time married women spend on household work relative to their husbands, then a mere change of sex (tongue in cheek) may not lead to the estimated effect, and the apparent gap can be attributed to this behavioural disparity. Confounding is unavoidable with the data we have available, but we can hope that, especially as we look for extreme rather than average "effects, we find submarkets where the gap cannot be fully explained away by such factors.

## 2. The Causal Forest algorithm

As Wager-Athey (2018) emphasised, one of the possible applications of machine learning (ML) methods in economics is for problems where our main interest is in estimating a nonlinear CEF. Athey (2018) asserts that "anyplace in traditional econometrics where a kernel function might have been used, ML methods that perform better than kernels in practice may be substituted". Random Forest (RF) is a tree-based statistical learning algorithm (Breiman, 2001) that has been applied in many disciplines (Cutler, 2012). Varian (2014) proposed RF for econometricians by citing Howard and Bowles (2012)**,** who asserted that it had been one of the most successful general-purpose predictive algorithms**.** Wager-Athey (2018) described RF regression as similar to other traditional non-parametric regression methods (e.g. k-nearest-neighbour algorithms), as it delivers some weighted average of "nearby" points as the prediction. However, it has the advantage that both the weights and the proximities are determined in a data-driven way. Indeed, RF is a methodology that has been hallowed by its unexpectedly excellent predictive performance, invoking, hitherto unproven, speculations about its nature (Biau-Scornet, 2016). Though usually RF is praised because of the opportunity to deal with high dimensional problems Mullainathan-Spiess (2017) also asserts that it can do significantly better than OLS "even at moderate sample sizes and with a limited number of covariates" (Mullainathan-Spiess, 2017, p. 89).

RF was introduced in order to rectify the overfitting tendencies of single-tree based methods (Hastie et al., 2009). In an RF regression, one grows many suboptimal regression trees, and the RF prediction is calculated as an average of the individual trees' predictions. Each tree is grown on a subsample (or on a bootstrap sample), and at each node (possibly) only a random subset of explanatory variables is considered for a split. The main advantage of RF seems to be that the random and restricted manner of branch formation in individual trees achieves de-correlation among constituent trees.

With the help of an RF algorithm there are several possibilities to estimate the

$$\mathrm{HWS}(x) = \mathrm{E}(y|x, T = 1)) - \mathrm{E}(y|x, T = 0),$$

function. We opted for the Causal Forest algorithm of Athey et al. (2018). Here the splits are chosen so as to maximise the "causal" effect. In a tree-based algorithm, the general idea behind splitting is that the new nodes have to be most different in terms of the target, which in a prediction-oriented exercise is the predicted variable, while in a causality-oriented study it is the heterogeneous causal effect. As the latter is not directly observable, the CF algorithm estimates the causal effect at each node, i.e. on a series of smaller and smaller subsamples, by a simple orthogonalized regression. This algorithm is an instance of the “Generalized Random Forests” (GRF) family that implements several changes with respect to Breiman’s RF. One modification entails the penalising of unbalanced splits via two parameters to discourage the forming very different sized child nodes. The most crucial difference is "honesty". Honesty separates splitting from estimation by dividing the subsamples into two parts. One is employed for placing the split, and the other one serves for estimation in the leaves. These changes contain the tendency of the original algorithm to become excessively data-driven, leading to overfitting. It was proved that under general conditions (Athey et al., 2018) the CF algorithm produces consistent estimates of the HWS effects. Though CF was developed explicitly for cases where the non-confoundedness assumption holds, as Wager-Athey ( 2018) emphasizes it is also robust to deviations from it.

## 2 Processing the HWS estimates

By definition different profiles (covariate vectors) are assigned different HWS effects. The resulting non-parametric function must be characterized in some interpretable way. An obvious starting point is to check whether there exists significant heterogeneity, i.e. whether the function is not a mere constant. We test for heterogeneity in a way suggested in Chernozhukov et al. (2018a), by projecting an unbiased signal of the HWS effects on the estimated effects. If the projection parameter is significantly different from 0 we can be confident that our estimates capture important heterogeneity.

### 2.1 Surrogate trees

Regression-trees have been employed in many areas to carry out a group-oriented data analysis in a supervised learning context. Single "greedy" regression-trees are generally classified as well-interpretable ML algorithms with not very good predictive performance (Hastie et al., 2009). This contrasts them with RF algorithms that are based on a randomized large set of (not so greedy) regression trees, and are considered excellent predictors without straightforward interpretability. The recent literature suggests that predictive models based on RF (or other powerful) predictors can be made interpretable if we fit shallow regression-trees on the model's predictions. This is sometimes called a global surrogate analysis (Molnar, 2022), where one tries to approximate the results of a not easily interpretable model with a highly interpretable one, in our case a regression tree. There are many possibilities to choose a regression tree algorithm, and we followed the advice of Molnar **(**2022), and employed the 'ctree' algorithm of Hothorn et al, (2006). 'Ctree' improves on the CART (Breiman et al., 1984) algorithm, which suffers from overfitting and biased variable selection, the latter because variables with more values have a better chance of being selected for splitting. 'Ctree' solves both problems by using statistical hypothesis tests to determine splits. First, it tests the multiple hypotheses that all covariates are independent of the response at each node. If the null is accepted, the algorithm stops. If it is rejected, then single variable tests for independence of each covariate are computed, and the p-values compared. An attractive feature of 'ctree' is that tests are distribution free permutation tests (Hothorn et al., 2006). The variable with the lowest p-value (indicating the strongest relationship with the response variable) is chosen, avoiding thereby selection bias. After picking the splitting covariate, 'ctree' uses the same method as CART to find the split point.

For the sake of interpretability, we grow shallow conditional inference trees of depth 3. This choice results in – at most – eight distinct submarkets, of which we can select the ones with the largest and smallest average HWS effects. The splits leading to these extreme groups can give a first impression of which submarkets can be taken as good candidates for discriminative wage setting, or, alternatively, what the main drivers of heterogeneity are.

### 2.2 Classification analysis (CLAN)

Chernozhukov et al. (2018a) suggested a model agnostic analysis with the express purpose of making sense of machine learning methods by focussing on "grosser" features of the estimated heterogeneous effects. The CLAN (classification analysis) methodology consists in finding average characteristics of the most and least affected observational units as defined by the predictor of the counterfactual conditional wage structure effects. We compute estimates of several functions of the covariates belonging to the highest and lowest 20 % in terms of HWS effects. These statistics are proportions, i.e. estimates of probabilities, after redefining some of the variables.

Chernozhukov et al. (2018) offers a special methodology, variational estimation and inference (VEIN), that works even with possibly inconsistent ML estimates (a distinct possibility in our case), and produces asymptotically valid confidence intervals. This methodology is based on repeated data splitting and honest estimation.

1. Split the sample randomly into an A(uxiliary) and a M(ain) subsample.
2. Construct the CATE estimator on the A sample,
3. Predict the CATE on the M sample, then determine the extreme groups in the M sample, and compute the estimates of the proportions together with upper and lower bounds of confidence intervals,
4. Calculate the difference between the most and least affected groups, together with confidence intervals.
5. Repeat from 1 to N times to obtain N point and confidence interval estimates.
6. The final estimate is the median of the N point estimates and the medians of the lower and upper bounds of the confidence intervals.

If the elementary confidence intervals have a level 1-alpha then the implied confidence level of the final estimate is approximately 1-2*alpha, by taking into account both estimation and sample splitting uncertainty.

## IV Data

Our data come from the Wage and Earnings Survey of the National Employment Office of Hungary, and were provided by the Databank of the Centre for Economic and Regional Studies. It is a matched employer-employee database that furnishes annual information (recorded in May). Each annual sample includes all firms with more than 50 employees and a randomly selected subset of firms with 5-50 employees. However, we dropped observations with firms having less than 20 employees as former research indicated that there is probably a large divergence between reported and actual wages in that size category (Lovász, 2008; Elek et al., 2009).

We used the logarithm of gross monthly earnings, comprising the monthly base wage, overtime pay and other regular earnings paid in May of each year, as the earnings variable. As this measure is inappropriate to compare full-time and part-time employees, we restricted our sample to employees working full-time. We dropped the public sector, where wage setting is based on administrative rules. To eliminate outliers, we left out observations with a log wage above the mean plus 3 standard deviations. Table 0 contains the variables used in the analysis.

**Table 0**

| Covariate | In CF | in 'ctree' and CLAN |
|---|---|---|
| Age | numeric (years) | ordinal factor (4 levels) |
| Tenure | numeric (months) | ordinal factor (4 levels) |
| Education | ordinal factor (4 levels) | ordinal factor (4 levels) |
| Occupation | ordinal factor (5 levels) | ordinal factor (5 levels) |
| Public or private | binary | binary |
| Domestic or foreign | binary | binary |
| Collective agreement | binary | binary |
| Size | numeric (no of employees) | ordinal factor (3 levels) |
| Sector | ordinal factor (18 levels) | ordinal factor (3 levels) |
| Region | ordinal factor (7 levels) | ordinal factor (3 levels) |

**Source**: Wage and Earnings Surveys

**Notes:** Age: Age is a numeric variable measured in years. As both the sample average and median are close to 40 years, for the sake of the CLAN and other aggregate analyses we defined age as a 4-category (ordinal) factor (levels: <= 30, 31-40, 41-50, > 50). Tenure: Tenure is the length of service with the current employer, measured in months. We made use also of an aggregated version with four categories as follows. 1: Less or equal than 12 months (new-entrants), 2: 13- 60 months (short tenured workers), 3: 61-120 months (long-tenured workers, 4: more than 120 months (very long-tenured workers). The median tenure in the sample is close to 50 months, but 5 years is a more traditional definition of long tenure, therefore we used it as the cut-off point. Education: Education is a 4-level ordinal factor. It is categorized as 1: primary, 2: secondary school without degree, 3: secondary school with degree and 4: tertiary degree. Secondary school without a degree includes vocational and vocational training schools. Secondary schools with degrees are vocational high schools, grammar and technical institutions.

Occupation: The Occupational code is the Hungarian variation of the ISCO codes. We employed an ordered aggregated version of the one-digit ISCO categories. The ordering was as follows: 1. Major Group 1 (managers), 2: Major group 2 (professionals), 3: Major group 3 (associated professionals), 4: Major groups 4-8 (skilled or semi-skilled blue-collar workers), 5: Major group 9 (unskilled workers). Size: Size is measured by the number of employees. We defined a compressed version of three levels (less than 50, 51-250, and more than 250) like in the Eurostat's structure of earnings (SES) survey. State-owned: A binary factor with 1 denoting majority state ownership. Foreign-owned: A binary factor with 1 denoting majority foreign-ownership. Collective agreement: A binary factor, where 1 indicates that there exists some collective agreement in the firm in question. Region: Hungary is divided into seven NUTS regions. For the CF estimates we regrouped these regions into three, based on per capita income variations, as a relevant aspect of wage-setting is price level differences, which are strongly correlated with regional income. The regions in ascending order of per capita income: 1. Northern Great Plain, 2. Northern Hungary, 3. Southern Great Plain, 4. Southern Transdanubia, 5. Central Transdanubia, 6. Western Transdanubia, 7. Budapest and Pest county. The 3-level ordered factor for the CF was: 1. Central region (Budapest and Pest county), 2. Central and Western Transdanubia, 3: the rest of the country. Sector: In the sample the sector variable is a factor with 18 categories (NACE Rev. 2 – 1). As the CF algorithm works badly with dummy coded factors with many levels, it is recoded as an ordinal factor. We calculated each 2-digit sector's trade intensity as (imports + exports) / (imports + domestic production). Based on this measure we defined the three ordinal levels as highly tradable, medium tradable and little tradable for the 'ctree' and CLAN. The highly tradable sector is manufacturing.

## V Results

### 1 Preliminary analysis

To impose the common support assumption, we ran logit models with gender as the target variable. We defined common support as profiles with a propensity score not less than 5 or not more than 95 %. In this logit model we used the same basic covariates as in the CF regressions, with 18-level factors for sector, 7-level factors for region, and, exceptionally, 4-digit ISCO codes for occupation. Our purpose was to robustly exclude specifically male and female occupations.

Imposing common support resulted in 20-30 percent reduction in the sample sizes. In the final samples both male and female wage averages were higher than in the original samples thus mostly relatively low wage individuals were eliminated. (See Table 1.) As many more low wage male profiles were dropped the final sample had a practically balanced distribution of males and females, but a definitely larger gap.

Table 1 General features of the samples

| | 2008 | 2012 | 2016 |
|---|---|---|---|
| No of obs in original | 104 455 | 99 037 | 107 698 |
| No of obs in final | 76 250 | 77 127 | 83 524 |
| Share of  females in original | 0.394 | 0.400 | 0.389 |
| Share of females in final | 0.476 | 0.480 | 0.478 |
| Average wage in original (HUF) | 195 030 | 246 440 | 292 617 |
| Average wage in final (HUF) | 207 777 | 257 987 | 308 091 |
| Log gap (male-female)in original | 0.108 | 0.133 | 0.130 |
| Log gap (male-female) in final | 0.183 | 0.202 | 0.221 |

Source: Wage and Earnings Surveys (2008, 2012, 2016)

Note: Log gap is the difference between the logarithm of the average male wage and the logarithm of the average female wage.

In the final samples the female and male subsamples are well-balanced. Comparing the distributions in the left-out samples one can see that originally male profiles included an unusually higher share with educational level 2, whereas female profiles with educational level 3. Excluding many of these profiles led to a more balanced occupational distribution. From now on every result is based on the final samples.

For each year we ran two wage regressions by the generalized random forest algorithm, and built surrogate trees of depth 3 on the predictions. The first set of regressions used all the explanatory variables (including gender), whereas the second set only the firm specific variables. By looking at the highest and lowest average groups in the surrogate trees we can have an impression about what variables are most salient to explain differences in the level of wages. It seems that the sharpest differences in terms of wages follow from education. The unsurprising finding is that university educated people who work in management or professional jobs earn most, whereas less educated people with blue collar jobs earn the least. Inspecting the trees in more details we can detect an effect due to working for a foreign-owned company that increases wages for white collar workers in general. When we focus on firm-related covariates it appears that the highest wages are paid in foreign-owned firms operating in the Central region, which, also, have collective agreement of some sort. At the other extreme there are firms with the dissimilar characteristics (domestically-owned, outside the central region and without any collective labour agreement).

**2 Causal Forest estimates**

In this subsection we report the results of single CF estimates by using the ‘causal_forest’ command in the ‘grf’ package. The hyperparameters are the default ones. The estimated average effects and their standard errors are reported in Table 2.  There is a slight reduction in the estimated average effects over time. The package implements the heterogeneity test of Chernozhukov et al. (2018) via estimating a best linear prediction (BLP) function by regressing observed values on the CF estimates. When the beta0 estimate is close to 1 it means that the average effect is precisely estimated by the CF, whereas heterogeneity can be rejected if beta1 is not significantly different from 0. The BLP beta estimates (see Table 2) produced beta0 coefficients around 1 and beta1 coefficients well-above 1, showing that the average WS effect is well-captured, and the hypothesis of non-heterogeneity of the HWS effects can be discarded. A beta1 close to 1 would mean that also the heterogeneity is precisely estimated, which is not the case. It underlines the necessity to use robust methods for the analysis of the HWS effects.

A robust method proposed in Chernozhukov et al. (2018a) is to identify the heterogeneity groups defined by quintiles of the estimated effects (GATES). As Table 3 shows these are also well-separated at higher quintiles, especially at the 5th , while not so much at the lower ones. It is reassuring to notice that the correlations between the propensity scores and the estimated effects are close to 0 (see Table 2). We also ask whether the HWS effects correlate with the general profitability of jobs. We estimated the average wage for ISCO-4 occupations, and could not find strong correlation between these and the HWS effects (Table 2). Also, we estimated simple linear models with the basic variables, and computed OLS based HWS effects by the usual algorithm. The CF-provided HWS effects and these are positively correlated, as expected, but far from perfectly (Table 2).

Table 2 Characteristics of HWS effect estimates

| | 2008 | 2012 | 2016 |
|---|---|---|---|
| average WS effect | 0.138<br>(0.002) | 0.136<br>( 0.002) | 0.124<br>(0.002) |
| beta0 | 0.997<br>( 0.020) | 0.995<br>(0.019) | 1.000<br>(0.019) |
| beta1 | 1.320<br>(0.038) | 1.263<br>(0.034) | 1.423<br>0.048) |
| correlation with prop score | 0.002 | 0.046 | 0.044 |
| correlation with well-paid jobs | -0.146 | -0.033 | 0.087 |
| correlation with OLS-based effects | 0.587 | 0.578 | 0.380 |

Notes: When calculating the OLS-based effects we estimated linear regression on separate female and male subsamples with a set of regressors that includes: numeric age and age squared, numeric tenure, numeric size, 1-digit ISCO as an unordered factor, education as an unordered factor with 4 levels, region as an unordered factor with 7 levels, sector as an unordered factor with 18 levels, as well as the three binary variables (state ownership, foreign ownership, existence of collective agreement).

Table 3 GATES (grouped averaged treatment effects)

| | 2008 | 2012 | 2016 |
|---|---|---|---|
| regressor | 1.0072<br>(1.006 1.0075) | 1.008<br>(1.007 1.008) | 1.0079406<br>(1.007 1.008) |
| first quintile | 0.0614<br>(0.038 0.084) | 0.039<br>0.017 0.061) | 0.061<br>(0.038 0.084) |
| second quintile | 0.1110<br>(0.0879 0.1341) | 0.081<br>0.0599 0.103) | 0.080<br>(0.057 0.103) |
| third quintile | 0.1301<br>(0.1076 0.1527) | 0.130<br>0.108 0.151) | 0.142<br>(0.119 0.166) |
| fourth quintile | 0.1952<br>(0.1716 0.2188) | 0.158<br>(0.137 0.180) | 0.168<br>(0.145 0.191) |
| fifth quintile | 0.2554<br>(0.2322 0.2787) | 0.268<br>(0.246 0.290) | 0.284<br>(0.261 0.308) |

Notes: The parameters and associated 95 % confidence intervals (in parentheses) come from the linear regression defined in Chernozhukov et al. (2018), Section 2.1.

Depth-3 conditional inference trees were grown on the estimated effects. The significance level is 0.001, and the joint test uses the Bonferroni adjustment. Table 4 reports statistics for the highest and lowest effect groups.

Table 4 Conditional inference trees

| | 2008 | 2012 | 2016 |
|---|---|---|---|
| H-group average effect | 0.258 (err = 31.3) | 0.269 (err = 31.1) | 0.183 (err = 29.4) |
| H-group no of observations | 6352 | 5814 | 12422 |
| H-group splitting variables | tradable,<br>longer tenure<br>foreign ownership | tradable<br>very long tenure<br>large firm | longer tenure<br>Medium-large firm<br>tradable |
| L-group average effect | 0.055 ( err = 8.4) | 0.036 (err = 12.5) | 0.075 (err = 14.5) |
| L-group no of observations | 4537 | 3183 | 11514 |
| L-group splitting variables | small firm<br>non- or semi-<br>tradable<br>tertiary education | non- or semi-<br>tradable<br>new entrants<br>state ownership | shorter tenure<br>domestic<br>small firm |

Notes: H (L) group denotes the highest (lowest) effect groups, respectively. Longer (shorter) tenure means tenure above (below) 5 years, respectively. New entrants are those with less than 1 year tenure.

A small firm is one with less than 50 employees, whereas a medium-large firm is one with more than 50. The non- or semi-tradable sectors contain all industries with the exception of manufacturing.

**3. CLAN**

Our main tool is the CLAN, see Table 5. Whereas shallow conditional inference trees can show the place of the "big" contiguous regions with highest and lowest average effects, the CLAN calculates the difference between statistics pertaining to the highest and lowest effect elementary markets. In Table 5 we report differences of estimated probabilities of levels of covariates of the least and most affected groups, i.e. those subpopulations where the estimated HWS effects belong to the first and fifth quintile. To help assess the size of the differences Table 5 shows the respective base proportions in the final full samples. To estimate these proportions, we employed the VEIN methodology, where the number of splits was 100, and the hyperparameters and covariates were the same as in the single CF.

Table 5 CLAN and base proportions

| Feature | (1) 2008 | (2) 2008 base | (3)2012 | (4) 2012 base | (5) 2016 | (6) 2016 base |
|---|---|---|---|---|---|---|
| age <30 | -0.28 (-0.298, -0.274) | 0.256 | -0.278 (-0.290, -0.267) | 0.211 | -0.322 (-0.331, -0.313) | 0.204 |
| age 30-40 | 0.142 (0.127, 0.156) | 0.308 | 0.081 (0.066, 0.095) | 0.336 | 0.130 (0.117, 0.144) | 0.301 |
| age 40-50 | 0.117 (0.103, 0.130) | 0.234 | 0.144 (0.131, 0.158) | 0.251 | 0.202 (0.189, 0.216) | 0.282 |
| age 50 + | 0.024 (0.011, 0.037) | 0.200 | 0.055 (0.042, 0.067) | 0.199 | -0.012 (-0.024, -0.001) | 0.210 |
| tenure <12 | -0.243 (-0.255, -0.232) | 0.216 | -0.383 ( -0.395, -0.371) | 0.217 | -0.313 (-0.325, -0.302) | 0.197 |
| tenure 12-60 | -0.106 (-0.121, -0.09) | 0.371 | -0.131 (-0.145, -0.118) | 0.352 | -0.241 (-0.254, -0.228) | 0.350 |
| tenure 61-120 | 0.143 (0.130, 0.155) | 0.181 | 0.156 (0.143, 0.169) | 0.199 | 0.181 (0.169,0.193) | 0.202 |
| tenure 120 + | 0.211 | 0.229 | 0.363 | 0.230 | 0.374 | 0.249 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.197, 0.225) | | (0.349, 0.376) | | (0.360, 0.387) | |
| educ 1 | 0.099 (0.089, 0.110) | 0.152 | 0.076 (0.067, 0.085) | 0.123 | -0.007 (-0.016, 0.002) | 0.119 |
| educ 2 | 0.305 (0.293, 0.318) | 0.244 | 0.156 (0.142, 0.170) | 0.256 | 0.101 (0.089, 0.113) | 0.226 |
| educ 3 | -0.048 (-0.064, -0.033) | 0.382 | -0.049 (-0.064, -0.033) | 0.368 | 0.010 (-0.004, 0.025) | 0.375 |
| educ 4 | -0.353 (-0.366, -0.341) | 0.221 | -0.189 (-0.202, -0.176) | 0.252 | -0.099 (-0.112, -0.086) | 0.278 |
| ISCO 1 | -0.115 (-0.126, -0.105) | 0.113 | -0.060 (-0.070,-0.050) | 0.093 | -0.007 (-0.017, 0.002) | 0.107 |
| ISCO 2 | -0.145 (-0.154, -0.136) | 0.094 | -0.084 (-0.093, -0.075) | 0.112 | -0.029 (-0.039, -0.020) | 0.129 |
| ISCO 3 | -0.118 (-0.132, -0.104) | 0.237 | 0.071 (0.058, 0.085) | 0.242 | 0.052 (0.039, 0.065) | 0.226 |
| ISCO 4-8 | 0.390 (0.375, 0.404) | 0.461 | 0.078 (0.062, 0.093) | 0.439 | 0.082 (0.067, 0.097) | 0.398 |
| ISCO 9 | -0.009 (-0.017, -0.001) | 0.094 | -0.003 (-0.011, 0.005) | 0.112 | -0.098 (-0.108, -0.088) | 0.137 |
| size <50 | -0.643 (-0.654, -0.631) | 0.336 | -0.513 (-0.525, -0.501) | 0.277 | -0.593 (-0.605, -0.582) | 0.268 |
| size 51-250 | 0.007 (-0.005, 0.020) | 0.257 | 0.031 (0.018,0.044) | 0.240 | -0.028 (-0.040,-0.016) | 0.268 |
| size 250 + | 0.632 (0.620, 0.644) | 0.406 | 0.484 (0.470, 0.497) | 0.482 | 0.627 (0.615, 0.638) | 0.462 |
| private owner | 0.0555 | 0.882 | 0.071 | 0.132 | -0.054 | 0.081 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.047, 0.063) | | (0.059, 0.082) | | (-0.062, -0.046) | |
| state owner | -0.0555 (-0.063, -0.047) | 0.117 | -0.071 (-0.082,-0.059) | 0.867 | 0.054 (0.046, 0.062) | 0.918 |
| domestic owner | -0.488 (-0.501, -0.475) | 0.682 | -0.361 (-0.375,-0.347) | 0.661 | -0.337 (-0.350, -0.324) | 0.708 |
| foreign owner | 0.488 (0.475, 0.501) | 0.317 | 0.361 (0.347, 0.375) | 0.338 | 0.337 (0.324, 0.350) | 0.291 |
| no agreement | -0.312 (-0.326, -0.297) | 0.644 | -0.257 (-0.271,-0.242) | 0.638 | -0.285 (-0.297, -0.273) | 0.769 |
| agreement | 0.312 (0.297,0.326) | 0.355 | 0.257 (0.242, 0.271) | 0.361 | 0.285 (0.273, 0.297) | 0.230 |
| Central Region | -0.227 (-0.242, -0.212) | 0.411 | -0.253 (-0.268,-0.238) | 0.430 | -0.146 (-0.161,-0.131) | 0.429 |
| CW. Transdanubia | 0.166 (0.152, 0.179) | 0.229 | 0.192 (0.179, 0.205) | 0.215 | 0.123 (0.111, 0.135) | 0.202 |
| Rest of the country | 0.059 (0.044, 0.075) | 0.358 | 0.063 (0.048, 0.079) | 0.353 | 0.017 (0.002, 0.032) | 0.367 |
| non-tradable | -0.170 (-0.180, -0.160) | 0.112 | -0.107 (-0.115, -0.099) | 0.086 | -0.108 (-0.117, -0.100) | 0.088 |
| semi-tradable | -0.502 (-0.515, -0.488) | 0.517 | -0.546 (-0.559, -0.533) | 0.553 | -0.252 (-0.266, -0.237) | 0.554 |
| tradable | 0.672 (0.661, 0.684) | 0.369 | 0.652 (0.640, 0.663) | 0.360 | 0.356 (0.343, 0.370) | 0.357 |

Notes: Columns (1), (3), (5) show the differences between the proportions in the upper and lower quintile of the respective feature. The rest of the columns display the proportions in the final sample.

The CLAN confirms the lessons gained from the ‘ctree’ analysis, and also adds further details. Longer and shorter terms of tenure (below or above 60 months) seem to produce sharp edges: longer-tenured individuals have significantly higher chance to end up among the most affected. Concerning age, it appears that 30 years is the divide, though obviously those below 30 usually have lower tenures. The countervailing changes in proportions are not equally distributed, the burden falls mainly on the 30-50 age group, and after 50 the differences are small.

As concerns education the less disadvantaged group of women is that with tertiary education, but the differences seem to be trimmed over time. The pattern is not clear with respect to the character of jobs, but the most disadvantaged group consistently is of skilled or semi-skilled blue-collar workers.

Regarding firm characteristics former results are reinforced. With respect to size, smaller-firm workers seem to populate the least and large-firm workers the most affected group. Manufacturing appears again as a driver of discrimination, but, among the rest of industries, we can now see that the intermediately tradable sectors exhibit definitely less than average discrimination. A sharp distinction seems to be related to owner's nationality, as the proportion of people employed by foreign owners increases steeply from the least to the most affected group.

The 'ctree' analysis did not provide much information about the role of locality, private or state ownership and the existence of wage agreements. CLAN shows a clear regional gradient, the middle income Central and Western Trans-Danubian region accommodates significantly more highly than little affected individuals, in contrast to the high-income Central Region. More surprising perhaps is the apparent positive relationship between the existence of a collective agreement and discrimination. Concerning state ownership, we can see no clear pattern, state owned enterprises do not appear to be in any of the extremities in particular.

## VI Interpretation

The 'ctree' and CLAN seem to suggest that the following market segment exhibits exceptionally high HWS effects: 1. large 2. foreign-owned firms 3. operating in manufacturing 4. in the CWT region 5. for medium-aged and 6. longer-tenured employees 7. who have less than tertiary education. We calculated the mean HWS effects for this subpopulation, characterized by these seven features. This hypothetical highly discriminated group has average HWS effects well beyond the most affected quintile (Table 6, row 1).

To evaluate the individual role of these features we conducted a quasi-Shapley-value analysis. We considered the above average as the coalitional value of the seven features (the "grand" coalition), and computed the coalitional values for the sub-coalitions of firm and personal features, respectively. Table 6 show the results.

Both the coalitions of firm and personal characteristics produce values that fall clearly short of the value of the grand coalition, thus both types are needed to get large effects in row 1 in Table 6. In 2008 and 2012 firm features have a larger "marginal" effect, whereas in 2016 the marginal effects are roughly the same. It must be also noticed that by 2016 the gradients are smaller, it seems that extremes became less extreme, the highly affected groups are more similar to the overall average than in years 2012 or 2008. If we try to distinguish among the firm features we find that when we add the firm features one-by-one to the personal features the strongest impact is either that of foreign ownership or of manufacturing. The corresponding analysis show that, among the personal features, tenure stands out unequivocally. If we restrict the sample to those individuals with longer-tenure and a job in a foreign-owned company then the resulting average effects are as follows: in 2016 0.1705, in 2012 0.2003, and in 2008 0.2208. These are comparable or larger than the combined average effects of the firm and personal variables, respectively.

Table 6 Coalition values

| Features | 2008 | 2012 | 2016 |
|---|---|---|---|
| full | 0.2971458<br>(1046) | 0.2864221<br>(562) | 0.2103853<br>(1207) |
| firm | 0.2137833<br>(4318) | 0.2110053<br>(3336) | 0.1631979<br>(3962) |
| personal | 0.1849068<br>(15095) | 0.1876896<br>(15711) | 0.1624755<br>(17019) |
| firm+tenure | 0.2687313<br>(1964) | 0.2635134<br>(1521) | 0.1932741<br>(2262) |
| firm+age | 0.2393296<br>(2367) | 0.2284146<br>(1952) | 0.1785924<br>(2389) |
| firm+educ | 0.2264337<br>(3643) | 0.2203493<br>(2862) | 0.1685038<br>(3127) |
| personal+sector | 0.248292<br>(6407) | 0.2568588<br>(6772) | 0.1949139<br>(7376) |
| personal+size | 0.2223508<br>(7085) | 0.215643<br>(7790) | 0.1826419<br>(8802) |
| personal+foreign | 0.2547778<br>(4664) | 0.2313454<br>(5433) | 0.1958159<br>(5511) |
| personal+region | 0.2126128<br>(3804) | 0.2211041<br>(3743) | 0.1792276<br>(3893) |

Notes: Coalition values are simple averages of HWS effects over groups defined by the features indicated in the first column. The meaning of the rows is as follows: firm (large foreign owned manufacturing enterprise in the Central and West Transdanubian Region), personal: less than tertiary education with age between 30 and 50 and tenure above 5 years). Full: the features in “firm” and “personal” together. The rest of the rows shows the averages for features added to the joint features.

What do all these tell us about the Hungarian labour market? Three of the firm characteristics (size, foreign ownership, sector) are such that they suggest that firms that pay higher wages exhibit larger HWS effects. On the other hand, highest wage firms are operating in the Central Region, rather than in Western and Central Transdanubia. Thus our findings are consistent with the presumption that firms that can pay higher wages and are settled in regions where competition on the demand size is less fierce can have more monopsonistic power and can exercise more discrimination. Discrimination is, however, also dependent on supply side (personal) features. Middle-aged less educated women with longer tenure have the characteristics that may make them switching between employers more difficult relative to similar men, thus are more apt to be "exploited".

Thus, the monopsonistic version of gender discrimination (or the practically equivalent bargaining power theory) is largely in accordance with our findings (Manning, 2010**).** This theory argues that a substantial part of gender wage differences must be due to the fact that women are in a weaker position in the labour market than men, and when the competition on the supply side is not strong enough firms can exploit this advantage. There are two legs of this explanation: weaker bargaining position for women and stronger bargaining position for certain enterprises. On the supply side longer-tenured women living in less densely populated regions may be rightfully considered as more easily "indentured" (in a non-literal sense, of course) than similar men. On the other hand, firms that can pay high wages and are overwhelmingly large employers in a region face much weaker competition. So, a large and well-paying employer is relatively in a better bargaining position in face of women than of men.

The gender pay gap can be explained also by unobserved human capital differences, but we must notice that we excluded specifically "male" and "female" occupations and it seems difficult to explain why high productivity firms would hire high productivity men and low productivity women with similar features for the same kind of jobs

**VII Summary**

We estimated heterogeneous counterfactual expected wages, tried to identify market segments that are most or least affected by the GWG, and interpreted them in terms of the working of the Hungarian labour market. Econometrically speaking our diagnosis is based on estimating conditional expectation functions pointwise. We believe that this exercise can contribute indirectly to the theoretical debate on the causes of the gender wage gap. Our results point into the direction that at least some part of the gap must be understood in terms of unequal labour market power for the sexes. Finally, the question arises if some policy action needs to be taken, or whether there exists some feasible action, at all.

Narrowing the difference between the costs of switching and movement must involve providing the availability of child-care facilities, and, in general, promoting the equal social role of women and men.

It is a general agenda which a government may or may not pursue, according to its ideological basis. Nonetheless, almost free of ideological bias, there is one possible avenue to reduce the gap. Our finding that having some collective wage agreement is indeed a “predictor” of a larger gap might mean two things. Either collective agreements in their present form tend to increase the gap, or it is just an epiphenomenon, as large manufacturing companies being those that also tend to have some collective agreement. Though the first explanation is unlikely it seems that collective agreements in their present form do not do anything to reduce wage discrimination. It can be the case that paying attention to formulating fair wage rules may contribute to improving the relative earning position of women.

## VIII References

Athey S. The impact of machine learning on economics. *The economics of artificial intelligence: An agenda*. University of Chicago Press. 2018: 507-547.

Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*. 2016; 113(27): 7353-7360.

Athey S, Wager S. Estimating treatment effects with causal forests: An application. *Observational Studies*. 2019; 5(2): 37-51.

Athey S, Tibshirani J, Wager S. Generalised random forests. *The Annals of Statistics*. 2019; 47(2): 1148-1178.

Austin PC. Using the standardised difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in Statistics - Simulation and Computation*. 2009; 38(6): 1228-1234.

Bach P, Chernozhukov V, Spindler M. Closing the US gender wage gap requires understanding its heterogeneity. *arXiv preprint arXiv:1812.04345*. 2018.

Baiardi A, Naghi AA. The value added of machine learning to causal inference: Evidence from revisited studies. arXiv preprint arXiv:2101.00878. 2021.

Benmelech, E., Bergman, N.K. and Kim, H., Strong Employers and Weak Employees How Does Employer Concentration Affect Wages?. *Journal of Human Resources*, *57*(S), pp.S200-S250. 2022.

Bertrand, M., Crépon, B., Marguerie, A. and Premand, P., Do workfare programs live up to their promises? Experimental evidence from Cote D’Ivoire (No. w28664). National Bureau of Economic Research. 2021.

Hussam, R., Rigol, N. and Roth, B.N., Targeting high ability entrepreneurs using community information: Mechanism design in the field. *American Economic Review*, *112*(3), pp.861-98. 2016

Biau, G, Scornet E. A random forest guided tour. Test; 25(2): 197-227. 2022.

Elek, Péter, and Anikó Bíró. "Regional differences in diabetes across Europe–regression and causal forest analyses." *Economics & Human Biology* 40, 2021: 100948.

Blau FD, Kahn LM. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*. 2017; *55*(3): 789-865.

Blinder, AS. Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*. 1973; 8: 436-455.

Böheim, R. -Stöllinger, P., Decomposition of the gender wage gap using the LASSO estimator. *Applied Economics Letters*, *28*(10), pp.817-828. 2021.

Brainerd E. Women in transition: Changes in gender wage differentials in Eastern Europe and the former Soviet Union. *ILR Review*. 2000; 54(1), 138-162.

Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees, Chapman & Hall. 1984.

Breiman L. Random forests. Machine Learning. 2001; 45(1): 5-32.

Brieland S, Töpfer M., The gender pay gap revisited: Does machine learning offer new insights? Diskussionspapiere, No. 111. 2020.

Card D, Cardoso AR, Kline, P. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*. 2016; *131*(2): 633-686.

Chernozhukov V, Demirer M, Duflo E, Fernández-Val I. Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomised Experiments, with an Application to Immunisation in India. National Bureau of Economic Research, No. w24678. 2018a.

Chernozhukov V, Fernández-Val I, Luo Y. The sorted effects method: discovering heterogeneous effects beyond their averages. *Econometrica.* 2018b; 86.6: 1911-1938.

Chowdhury, R., Ceballos-Sierra, F. and Sulaiman, M., Grow the pie or have it? Using machine learning for impact heterogeneity in the Ultra-poor Graduation Model. 2021

Christiansen, T. and Weeks, M., Distributional Aspects of Microcredit Expansions. 2020

Csillag M. "Female work" and the gender wage gap from late socialism to today. In Galasi P, Kézdi G, editors. The Hungarian labour market review and analysis. Budapest: Institute of Economics; 2007.

Cutler A, Cutler DR, Stevens JR. Random forests. In Ensemble machine learning. Springer, Boston, MA. 2012: 157-175.

Davis, J.M. and Heller, S.B., Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*, *102*(4), pp.664-677. 2020

Deryugina, T., Heutel, G., Miller, N.H., Molitor, D. and Reif, J., The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, *109*(12), pp.4178-4219. 2019.

Earle, J.S., Telegdy, Á. and Antal, G., Foreign ownership and wages: evidence from Hungary, 1986–2008. *ILR Review*, *71*(2), pp.458-491., 2018.

Elek P, Scharle A, Szabó B, Szabó PA. A bérekhez kapcsolódó adóeltitkolás Magyarországon. Közpénzügyi Füzetek. 2009; 23.

Fan J, Li R. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*. 2001; 96(456): 1348-1360.

Fazekas K. A hazai és a külföldi tulajdonú vállalkozások területi koncentrációjának hatása a foglalkoztatás és munkanélküliség területi különbségeire. In Faluvégi A, Fazekas K, Nemes-Nagy J, Németh N. (eds) A hely és a fej: Munakpiac és regionalitás Magyarországon. MTA Közgazdaságtudományi Kutatóközpont, Budapest. 2005: 47-74.

Fortin N, Lemieux T, Firpo S. Decomposition methods in economics. *Handbook of labor economics*. Vol. 4. Elsevier. 2011: 1-102.

Galasi P. A fizetett és nem fizetett munka értéke. In Fazekas K. (eds) Munkaerőpiaci tükör 2002, MTA Közgazdaságtudományi Kutatóközpont. 2002a: 101-117.

Galasi P. A női-férfi munkakínálati különbségek tényezőkre bontása. In Fazekas K. (eds), Munkaerőpiaci tükör 2002, MTA Közgazdaságtudományi Kutatóközpont. 2002b: 127-133.

Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. New York: Springer-Verlag; 2009.

Heinze A, Wolf E. The intra-firm gender wage gap: a new view on wage differentials based on linked employer–employee data. *Journal of Population Economics*. 2010; 23(3): 851-879.

Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*. 2006; 15(3): 651-674.

Howard J, Bowles M. The two most important algorithms in predictive modeling today. In Strata Conference presentation. 2012; Santa Clara, USA. 28(2).

Jacob, Daniel. CATE meets ML-The Conditional Average Treatment Effect and Machine Learning. arXiv preprint arXiv:2104.09935. 2021.

Kertesi G, Köllő J: A gazdasági átalakulás két szakasza és az emberi tőke átértékelődése. A bérszerkezet átalakulása magyarországon 1986-1999 - iii. rész. *Közgazdasági Szemle*. 2001; 48(11): 897-919.

Knaus MC, Lechner M, Strittmatter A. Machine learning estimation of heterogeneous causal effects: Empirical monte carlo evidence. *The Econometrics Journal*. 2021; 24(1): 134-161.

Knittel, C.R. and Stolper, S., Machine Learning about Treatment Effect Heterogeneity: The Case of Household Energy Use. In *AEA Papers and Proceedings* (Vol. 111, pp. 440-44). 2021, May.

Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences.* 2019; 116.10: 4156-4165.

Lipkovich, I., Dmitrienko, A. and B D'Agostino Sr, R., Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Statistics in medicine*, *36*(1), pp.136-196. 2017

Lovász A. Competition and the Gender Wage Gap: New Evidence from Linked Employer-Employee Data in Hungary, 1986-2003. Budapest Working Papers On The Labour Market BWP. 2008 July.

Lovász A. Jobbak a nők esélyei a közszférában? A nők és férfiak bérei közötti különbség és a foglalkozási szegregáció vizsgálata a köz- és magánszférában. Közgazdasági Szemle. 2013; 60:1057-1074.

Manning A, Petrongolo B. How local are labor markets? Evidence from a spatial job search model. *American Economic Review*. 2017; 107(10): 2877-2907.

Meng X.Gender earnings gap: the role of firm specific effects. *Labour Economics*. 2004; 11(5): 555-573.

Molnar C. Interpretable Machine Learning - A Guide for Making Black Box Models Explainable. Independently published. 2022.

Mullainathan S, Spiess J. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*. 2017; 31(2): 87-106.

Murakami, K., Shimada, H., Ushifusa, Y. and Ida, T., Heterogeneous Treatment Effects of Nudge and Rebate: Causal Machine Learning in a Field Experiment on Electricity Conservation. *International Economic Review*. 2020.

Newell A, Reilly B. The gender pay gap in the transition from communism: some empirical evidence. *Economic Systems*. 2001; 25(4): 287-304.

Oaxaca, R. Male- female wage differentials in urban labor markets. *International Economic Review*. 1973; 14: 693-709.

Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (*Methodological*)*. 1996; 58(1):267-288.

Slater, Michael D. "Theory and method in health audience segmentation." *Journal of health communication* 1.3 1996: 267-284.

Strittmatter A, Wunsch C. The Gender Pay Gap Revisited with Big Data: Do Methodological Choices Matter? arXiv preprint arXiv:2102.09207. 2021.

Szabó-Morvai Á. Háztartási munkamegosztás. In Fazekas K.- Szabó-Morvai Á. (eds) Munkaerőpiaci tükör 2017, MTA Közgazdaság- és Regionális Tudományi Kutatóközpont Közgazdaságtudományi Intézet. 2018: 168-171.

Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association. 2018; 113(523): 1228-1242.

Willke, R.J., Zheng, Z., Subedi, P., Althin, R. and Mullins, C.D., 2012. From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: a primer. *BMC Medical Research Methodology*, *12*(1), pp.1-12.