GRAPE Working Paper #72

Matching it up: non-standard work and job satisfaction

Katarzyna Bech, Magdalena Smyk, Lucas van der Velde, Joanna Tyrowicz

FAME | GRAPE, 2022

Foundation of Admirers and Mavens of Economics
Group for Research in Applied Economics

Matching it up: non-standard work and job satisfaction

Katarzyna Bech
SGH and FAME|GRAPE

Magdalena Smyk
SGH and FAME|GRAPE

Joanna Tyrowicz
FAME|GRAPE, University of Regensburg and IZA

Lucas van der Velde
SGH and FAME|GRAPE

Abstract
We leverage the flexibility enactment theory to study the link between working arrangements and job satisfaction. We propose that this link is moderated by individual inclination to non-standard working arrangements. Thus, we provide novel insights on the (mis)match between preferred and actual working arrangements. We apply this approach to data from the European Working Conditions Survey and empirically characterize the extent of mismatch in working arrangements across European countries. We shed new light on several phenomena. First, the extent of mismatch is substantial and reallocating workers between jobs could substantially boost overall job satisfaction in European countries. Second, the mismatch more frequently affects women and parents. Finally, we demonstrate that the extent of mismatch differs across European countries, which hints that one-size-fits-all policies, whether they deregulate or curb non-standard arrangements, are not likely to maximize the happiness of workers.

Keywords:
non-standard working arrangements, job satisfaction, gender

JEL Classification

Corresponding author
Joanna Tyrowicz, j.tyrowicz@grape.org.pl

Acknowledgements
The authors express gratitude for insightful comments and suggestions by Arthur Alderson, Arne L. Kalleberg, Julia McQuillan, Jekaterina Navicke, Erita Narhetali, Dina Okamoto, and Linas Tarasonis. Kazimierz Czarnocki and Kamil Sekut provided wonderful research assistance. Sam Butterick provided editorial assistance. Financial support by National Science Center grant no. #2017/27/L/HS4/03219 is appreciated. The remaining errors are ours..

Published by: FAME | GRAPE
ISSN: 2544-2473

Foundation of Admirers and Mavens of Economics
ull. Koszykowa 59/7
00-660 Warszawa
Poland

W | grape.org.pl
E | grape@grape.org.pl
TT | GRAPE_ORG
FB | GRAPE.ORG
PH | +48 799 012 202

1 Introduction

We study the (mis)match between actual and preferred working arrangements. *Flexibility enactment theory* proposes viewing non-standard working arrangements through the lens of boundary management (Kossek et al. 2004). Building on Lee et al. (2002), this theory posits that non-standard working arrangements (NWAs) put workers under greater pressure to proactively manage boundaries between work and private life as compared to standard working arrangements. A general hypothesis implied by the flexibility enactment theory is that *ceteris paribus*, job satisfaction is lowered by the *mismatch* between individual ability to manage boundaries and the degree of requirement to manage these boundaries ensuing from NWAs. As physical and time barriers between work and home blur, workers are subjected to the additional mental load of maintaining separation between the two domains.

Building on the flexibility enactment theory, we theorize that individuals differ in their *ability* to manage boundaries associated with NWAs (such as varying working hours, working long hours, nights or weekends, or the ability to work from outside the official premises), that is to prevent negative spillovers across life domains. Flexibility enactment theory predicts that the ability to actively manage boundaries mediates the influence of NWAs on job satisfaction. In other words, not everyone is equally able to maintain high levels of job satisfaction under the strain of spillovers between professional and private spheres. The advantages of NWAs may well be the means to reconciling work-home conflicts for some workers. For example, it may be convenient for parents to share the care, when one of them is able to work long hours on specific days and fewer than the regular eight hours per week on other days. However, as demands from both domains materialize and responses that are adequate in one domain prove to be inappropriate in the other, they give rise to new conflicts between the domains of work and home, both in terms of time and in terms of behavior (Nippert-Eng 2008). For example, flexible hours may imply cognitive spillovers to leisure, deteriorating health and increasing the psychological strain of work (e.g. Williams et al. 2013, Cha and Weeden 2014).

Based on these theoretical foundations, we conjecture that ability to manage boundaries (and spillovers) between work and private life is a factor mediating between NWAs and job satisfaction. Workers who are better able to manage boundaries between the two might find that NWAs allow them to attend to their home needs without neglecting their work duties and vice versa. Workers who lack this ability might thrive when boundaries are rigidly established within standard working arrangements. Importantly, some workers may be *mismatched* in terms of working time arrangements and individual ability to manage NWAs. Hence emerges an ambiguous relationship between NWAs and job satisfaction. On the one hand, non-standard working arrangements may allow for greater work-life balance thus contributing to workers' happiness (Atkinson et al. 2011) and productivity (Bloom et al. 2015). On the other hand, NWAs require workers to proactively manage boundaries and not everyone is equally able to maintain high levels of job satisfaction under such strain.

We propose to empirically evaluate the conjecture that the link between job satisfaction and NWAs is heterogeneous. Admittedly, the ability to proactively manage boundaries is latent and thus not observed in the data. To uncover this latent link between job satisfaction and NWAs we use machine learning (ML) methods. We deploy the ML algorithms to obtain data-driven model of the relationship between personal characteristics, job characteristics and job satisfaction for workers employed in standard arrangements. Based on this model we obtain counterfactual levels of job satisfaction for each individual working with NWAs. The model provides the job satisfaction as if these workers were not employed in NWAs. By comparing the factual and counterfactual job satisfaction we identify the individuals matched to working conditions (their job satisfaction would have been the same even in the absence of NWAs) and individuals who are mismatched (their job satisfaction would have been higher or lower in the absence of NWAs). ML methods fully exploit the data without imposing any *ad hoc* restrictions in terms of model specification and functional forms. Indeed, we infer the counterfactuals from the deep, complex and nonlinear underlying links between individual characteristics, job characteristics and job satisfaction. Consequently, while our approach is rooted in theory, arbitrary modelling choices are avoided.

Empirical evidence to date relies on parametric approaches, and offers mixed evidence. Wheatley (2017) finds positive correlations between job satisfaction and working in non-standard arrangements for men, but negative correlations for women. Bellmann and Hübler (2020) argue that patterns for job satisfaction are generally unclear, while the correlations with work-life balance are robustly negative. Individuals with caring obligations appreciate if the employer somehow accommodates their complicated time schedules (Bainbridge and Townsend 2020). In Europe, full-time work makes mothers less happy than part-time or staying home altogether (Hamplová 2019). Moreover, flexible hours may imply cognitive spillovers to leisure, deteriorating health and increasing the psychological strain of work (e.g. Williams et al. 2013, Cha and Weeden 2014). These spillovers are highly heterogeneous across arrangement types and genders, with a multitude of patterns emerging from the data (see Lott 2015, 2020, for a cross-country comparative study, and for uniquely profound data and analysis for Germany, respectively). In addition, the workplaces which discourage flexibility are considered less attractive by all types of workers (O'Connor and Cech 2018). From the perspective of employers, workers who ask for flexible work arrangements are evaluated negatively (McCarthy et al. 2013, Cech and Blair-Loy 2014, Munsch 2016) and as poor workers (Rudman and Mescher 2013, Vandello et al. 2013), irrespective of their gender. This broad and burgeoning literature notwithstanding, empirical inquiry into the mismatch between the working arrangements and ability to manage boundaries is lacking.

We aim at three contributions to the existing literature. First, to the best of our knowledge, we are the first to operationalize the flexibility enactment theory in a quantitative, empirical context and draw conclusions implied by the theory. Taking the flexibility enactment theory as granted we characterize types of workers more able to manage boundaries. We do so by overcoming the common deficiency of the data: the ability to manage boundaries is a latent variable, unavailable in the existing data sets. We address this challenge by obtaining counterfactual levels of job satisfaction through machine learning algorithms. This innovation is our second contribution, as parametric estimates of counterfactuals could not deliver reliable results in our context. Third, our analysis is set in an international, comparative context. By using the data from the European Working Conditions Survey, we can characterize the extent and drivers of mismatches between actual and optimal NWAs across individuals and countries.

Our study has relevance for general audiences as well as policy implications. In the context of digitalization and internationalization, it becomes a policy concern if workers experience anxiety about the expectation of 24/7 availability, and about work transgressing their private spaces. These concerns have triggered policy initiatives such as *the right to disconnect*.[1] Our results may be used to evaluate firm-policies as well as government and transnational policy initiatives.

Our study is structured as follows. Section 2 provides an overview of the existing literature. In section 3, we present the data and in section 4, we introduce our empirical approach. The results of our estimations as well as the counterfactual experiments are discussed at length in section 5. Exploring heterogeneity of the mismatch between actual and optimal working arrangements, we discuss in detail the drivers of this mismatch as well as country-specific implications.

[1] In Europe, the *Working Time Directive* specifies that the time during which the worker is expected to answer phone calls or emails should count as regular working time, irrespective of whether the phone calls/emails occurred. France and Italy have introduced specific legislation: the former requires that all employment contracts specify which hours are "off grid" for the employers; in the latter the legislation concerns self-employed ICT-based workers. The US perspective is covered extensively by Secunda (2019).

2 Literature review and research questions

Early on, sociological theory provided several explanations for why individual abilities to proactively manage boundaries may be highly heterogeneous. First, *role theory*, *self-discrepancy theory* and *social identity theory* all emphasize that individuals define themselves through the lenses of how they are perceived by others and by themselves (Kahn et al. 1964, Katz and Kahn 1978). These lenses are by design individual and thus highly heterogeneous in the sense that the social norms and stereotypes exhibit differently across social classes, education achievement levels or even local communities. This combination of individual perceptions of oneself and of social norms alleviates or exacerbates work-family and family-work conflicts. This line of scholarly work was further advanced by *spillover theory* (Westman and Piotrkowski 1999) and *boundary theories* (Nippert-Eng 1996, Clark 2000). With their specific angles, these theories posit that by integrating life experiences from both personal and work environments, people automatically transmit these experiences across life spheres and thus spillovers, both positive and negative, are unavoidable.

Given that spillovers cannot be avoided, they have to be actively managed. The crux of *flexibility enactment theory* is that the ability to actively manage the spillovers depends on an individual's general boundary management abilities (Kossek et al. 2004). This is why the individual satisfaction defined from a job depends not only on the specific arrangements at work, but also on the ability to manage boundaries between work and private life, which is easy for some individuals, and troublesome for others. An individual boundary management strategy relates to a combination of boundaries (related to time, space, sense of belonging, etc.) and it is defined as a continuum over different aspects of non-standard arrangements.

Despite this rich body of theoretical work in psychology, social psychology and sociology, the empirical studies of non-standard working arrangements focus on preferences (e.g., Piasna and Drahokoupil 2021). Often they also isolate one type of NWA, e.g., part-time work or shift work. For example, it was repeatedly established that full-time work makes mothers less happy than part-time or staying home altogether (see Hamplová 2019, for evidence from Europe). This result extends to individuals with caring obligations, who appreciate if the employer somehow accommodates for complications implied by those obligations (Bainbridge and Townsend 2020, for evidence from Australia). Much less effort was devoted to which specific hours are worked, though this was found to be one of the key factors for women's engagement in so-called gig jobs (Mas and Pallais 2017, Cook et al. 2021). Recently, Lachowska et al. (2023) use a revealed preferences approach to show that the mismatch between desired and actual working hours reduces mental well-being. In addition to differences in preferences, there seem to exist profound differences in time endowments, that is the ability to occasionally work long hours (Cortes and Pan 2019, Zapf and Weber 2017) or specific hours (Duchini and van Effenterre 2018, Cubas et al. 2019).

Qualitative and quantitative evidence support the hypothesis that job pressure is felt strongly in occupations where time and location boundaries are not set. For example, a study by Schieman and Glavin (2016) corroborates this observation, exploring a nationally representative sample of workers and delineating the job pressure for individuals holding high status jobs, such as professionals and managers. In a study comparing workers from four culturally distant countries, Barney and Elias (2010) argue that, in some cultural contexts, the autonomy in setting one's own schedule actually exacerbates work-related stress. Dumas and Sanchez-Burks (2015) argue in an extensive review paper that the perspective on boundary management and spillovers in the literature alternates between treating them as a tool for handling role responsibilities on the one hand and as a tool for shaping workplace identity and relationships on the other.[2]

[2]There is also ample qualitative evidence that men less frequently request flexibility at work (Vandello et al. 2013). Requesting flexibility is stigmatizing (Rudman and Mescher 2013, Vandello et al. 2013, Munsch 2016) and stereotypically associated with different motivations across genders: leisure aspirations for men and family devotion schema for women (Albanesi and Olivetti 2009, Flabbi and Moro 2012, Williams et al. 2013). Indeed, men are less likely to receive acceptance on the request for non-standard arrangements if it is for family reasons (Brescoll et al. 2013).

In summary, the literature theorizes that working in non-standard arrangements requires individuals to manage boundaries more proactively than standard working arrangements. Systematic evidence on who is able to manage the boundaries is more scarce, both in terms of individual characteristics (such as gender and household structure) and in terms of NWA characteristics (such as the type of non-standard working arrangement). Our study contributes to the literature by empirically addressing three research questions. Note that we do not challenge the flexibility enactment theory. Instead, we test some implications of the theory. In everything what follows, we assume that flexibility enactment theory holds and all inference we provide is conditional on this assumption.

Our study builds on the flexibility enactment theory to provide new empirical results. First, we propose to think about the ability to manage boundaries as a counterfactual: for individuals who work in NWAs we obtain a hypothetical level of job satisfaction if they did not have NWAs, *ceteris paribus*. We obtain the difference between the actual and the counterfactual level of job satisfaction for workers with NWAs and study the distribution of this differential. Note that if flexibility enactment theory was not quantitatively relevant for job satisfaction, we should observe that for most individuals there is no difference between the actual and counterfactual job satisfaction.

Research Question 1 *Is the mismatch between actual and preferred working arrangements prevalent?*

We expect that some individuals work under arrangements suited for their abilities to manage boundaries, i.e. those with better abilities are more likely to enjoy flexibility, and those less able work, on average, more frequently under standard contracts. We expect these matched workers to report high levels of job satisfaction. However, some individuals are forced to work under arrangements that challenge their abilities to manage boundaries. These individuals are likely to report lower levels of job satisfaction. While there is no direct way to test that the mismatch is attributed to boundary management, our approach rests upon obtaining the most accurate counterfactual in the absence of NWAs, and thus attributing this differences in job satisfactions across working arrangements are (at least partially) driven by differences in ability to manage boundaries.

Second, we explore if the mismatch is more likely to emerge among workers with specific characteristics. If that is the case, then one could think about the mismatch in a way analogous to heterogeneity of treatment effects across different types of individuals. We consider two dimensions of heterogeneity: gender and parenthood status.

Research Question 2 *Is the mismatch between actual and preferred working arrangements systematic?*

We expect that ability to actively manage boundaries is higher for women and parents. Notably, caring for relatives is stereotypically a task expected of women. Likewise, parents need to accommodate not only their own boundaries but also those implied by the duty to care for their children. Consequently, women and parents face more spillovers between work and private life, which raises their experience in proactively managing the boundaries. When working with NWAs, they deploy this experience and thus enjoy higher job satisfaction from the granted flexibility. Note that this is not a foregone conclusion: individuals experiencing more spillover strain may well benefit mentally from clearly delineated boundaries associated with standard working arrangements. Since they may be unable to effectively commit to those arrangements due to their care obligations and the social pressures around those obligations, they may be forced to work in NWAs, to the detriment of their job satisfaction.

Having studied the individual level, we analyze the aggregate implications. After establishing a proxy of mismatch between actual and preferred working arrangements, we investigate whether policies aimed at reducing NWAs raise overall satisfaction. We ask whether the costs identified by the flexibility enactment theory outweigh the benefits that workers can derive from working in NWA, such as the possibility to reconcile job and household demands.

Research Question 3 *Can aggregate job satisfaction be raised by removing NWAs?*

The existing research finds conflicting results, admittedly studying differentiated sub-populations and different forms of NWAs. We propose to reconcile the literature by quantifying the groups that would benefit from standard rather than NWAs, and the groups that would be harmed by eliminating NWAs.

Despite new research questions and empirical strategy, our study approach draws heavily on the existing literature. Our study takes a similar starting point as Wooden et al. (2009), who study the role of mismatch between stated, preferred and actually worked hours in job satisfaction of Australian workers. They show that this mismatch is relevant for full-time workers. Studying Germany, Wunder and Heineck (2013) show that this mismatch is related to caring obligations, with strong spillovers from a partner's mismatch, not the hours *per se*. Angrave and Charlwood (2015) rely on longitudinal data from the UK and show that potential mismatch has no lasting effects on job satisfaction of workers because they either adapt or change jobs. Lee et al. (2015) show that job satisfaction may increase when the employers remedy hours mismatch reported by the workers. However, data on preferred working arrangements are rare across countries and, when available they are restricted to the number of working hours. This limitation leaves a broad range of non-standard working arrangements under-researched. Furthermore, stated preferences may be inconsistent with actions: measurement suffers from several biases.

Our study differs in three important aspects from previous research. First, we base our conjectures on flexibility enactment theory: our latent variable is the ability to proactively manage boundaries. Specifically, heterogeneity of preferences among workers is no longer required for our empirical approach to be valid. This is relevant, because axiomatically, preferences tend to be constant over time, whereas the ability to proactively manage boundaries may vary over time, jobs and living arrangements. Second, our empirical strategy does not require information about preferences towards working arrangements. Indeed, our identification relies on counterfactuals emerging from ML methods. Third, we innovate by looking at four different aspects of NWAs, whereas the existing literature typically isolated one aspect, most frequently working hours.

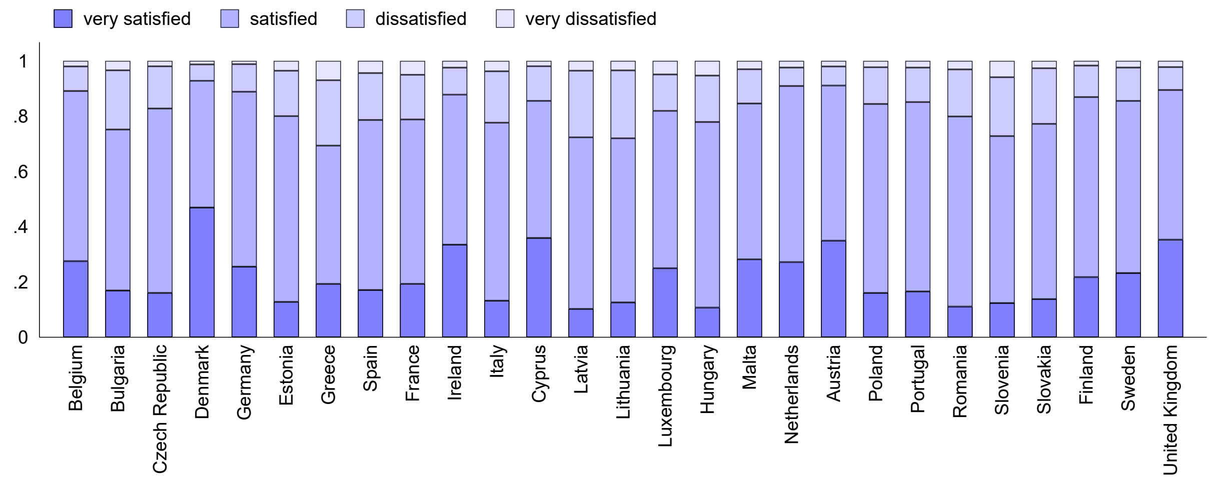
3 Data

We use data from the European Working Conditions Survey (EWCS). It is administered across an increasing number of countries, with harmonized sampling methodology and questionnaire. This survey is administered every five years. The available data span the years between 1991 and 2015. However, some of the variables necessary for our study are only available starting with the third wave, thus our study covers the period 2001-2015.

Each wave of EWCS provides extensive information about the socio-demographic individual characteristics of the workers. In addition to gender and age, individuals report the size of their household, the number and age of children within their household, their subjective health status, occupation, and industry. Individuals report also the type of contract (temporary or permanent) and the number of hours worked. Self-employment is reported separately from wage-employment. In addition, the commute time is reported. These variables jointly are used in our models as explanatory variables. The full list of variable codes and their transformation is reported in Appendix A.

Clearly, one of the most important aspects of job satisfaction are earnings individual receives. An important question is whether workers under flexible schedules are compensated more or less (depending on whether they are considered employment benefits or not) than workers under standard contracts. Also,

Figure 1: Job satisfaction across countries in waves in EWCS



Notes: The graph is based on EWCS data from waves 3-6.

different types of non-standard working arrangements should compensate workers differently. However, the dataset lacks information on wages. [3] We believe that our proposed method is to some extent robust to the exclusion of certain characteristics, and adding wages would not increase the predictive power of the model significantly, once we control for hours worked, occupation and industry.

We utilize responses for salaried workers, aged between 18 and 65 years of age. Due to specificity of public sector, such as health care, police, etc – we restrict the sample to individuals employed within private entities. Overall, the data covers twenty seven countries and roughly 58,000 workers.

3.1 Job satisfaction

All but the first wave of EWCS include the question *"On the whole, are you very satisfied, satisfied, not very satisfied or not at all satisfied with working conditions in your main paid job?"*. This variable provides our main indicator of job satisfaction (JS). The answers come on a four-levels Likert scale, ranking from very satisfied (1) to very disatisfied (4). On average, 25% of individuals are very satisfied, with an additional 57% of individuals being satisfied. The distribution of job satisfaction across countries is portrayed in Figure 1.

3.2 Information on working arrangements

EWCS provides rich information on working conditions, which makes it uniquely suitable for our analysis. The respondents are asked whether they *normally* face several dimensions of the non-standard working arrangements in their main job.

First, the workers report if their work involves **varying hours**. This arrangement implies that the worker's starting and finishing times at work vary within a week and between weeks. Prior research focused on shift work, especially evening or night shifts (e.g., Han 2008, Chait Barnett et al. 2008). By comparison, the questionnaire in EWCS has a broader meaning, as in addition to shift work and rotation, it also includes arrangements where the hours vary between days of the week or within a week in an irregular manner. The literature argues that lack of regularity in working hours is harmful to well-being (Chait Barnett et al. 2008), stability of relationships (Hertz and Charlton 1989, Florean and Engelhardt-Woelfler 2020) and parental

[3] Questions on exact/range earnings are only asked from 5th wave onwards. Hence, approximately 65% of the sample misses that information.

success (Barnett and Gareis 2007, Han 2008, Gracia and Garcia-Roman 2018). However, the discretion to adjust hours may raise well-being (Grzywacz et al. 2008), see also a review by Bolino et al. (2021). We define a dummy variable taking on the value of 0, when workers report "fixed starting and finishing times" and 1 otherwise.[4]

Second, the workers report if their work involves **nights**, that is working hours between 10pm and 5am. The effects of working at nights for physical and mental health have been researched in ergonomics (Bohle and Tilley 1989). They are detrimental for many markers of life quality, such as sleep quality (Cheng et al. 2021) and mental health (Torquati et al. 2019). However, with technological changes, especially in large cities, once mainly predominant in caring, policing and manufacturing, work during the nights is becoming increasingly normalized; see also a review by Mueller (2019). We obtain an indicator variable which takes on the value of 1 when a worker's schedule involves at least two nights a month and 0 otherwise.

Third, EWCS inquires about working **long hours** (more than 10 hours a day). This is not akin to overtime, because it signifies an arrangement in which one or more working days extend beyond the standard eight hours. Then on other working days, either the working time is shorter or there is a larger number of free days per week. Note that such working arrangements can indeed be helpful for caring obligations as well as for personal development, because shorter hours or more free days may be conducive to nurturing non-work related activities, when family or institutionalized assistance is available on the days of long working hours (Angrave and Charlwood 2015, Cortes and Pan 2019). We obtain an indicator variable, which takes on the value of 1 when a worker reports long hours at least four days a month and 0 otherwise.

Finally, fourth, the workers report if their work involves the weekends. We focus specifically on **Sundays** given special cultural role attributed to Sundays as free days across countries in our sample.[5] Working on Sundays is unequivocally forbidden by the law (except dedicated weekend working schedules, in which case it is heavily regulated), even though it may be particularly suitable for combining personal life with professional activity (Cook et al. 2021). We obtain an indicator variable, which takes on the value of 1 when a worker reports working on Sundays at least twice a month and 0 otherwise.

In the further analysis, we focus on workers who report *none or one* of those specific non-standard working arrangements (NWA). Among 58,518 workers in our sample, 28,378 report no NWAs. In addition, 14,924 workers report one of the four above NWAs. The remaining 15,216 workers report two or more NWAs. For example, some workers report that their work involves working on Sundays as well as nights. Given the special status of Sunday work across the European countries, we construct a fifth group, which is a conjunction of Sundays and nights.[6] A second large group are workers who report that their work involves long hours, but the specific start and end times vary between days in a week and/or between weeks.[7] Consequently, we construct a sixth group: a conjunction of varying hours and long hours. Overall, we have six distinct types of NWAs: varying hours, nights, long hours, Sundays, long-and-varying hours and Sunday nights. After these six distinct groups are formed, we eliminate from the analysis the few workers who report more than one NWA at the time.[8] The final sample consists of 56,107 workers.

Our approach to studying each NWA separately is consistent with the theoretical foundations stemming from (Kossek et al. 2004). Specifically, using the individual worker perspective, we focus on the types of non-standard working arrangements that require a proactive management of boundaries. We obtain seven

[4]Note that EWCS does not specify if the varying hours are set by the employer or selected autonomously by the worker. The question about autonomy to set working hours was not asked to all workers, roughly 58% of the workers declare no autonomy in setting the hours, but in roughly 24% cases the question was not asked. Thus, the autonomy question could not be used in our study.

[5]Respondents in EWCS do not report religion, thus, we cannot identify person-specific days of worship.

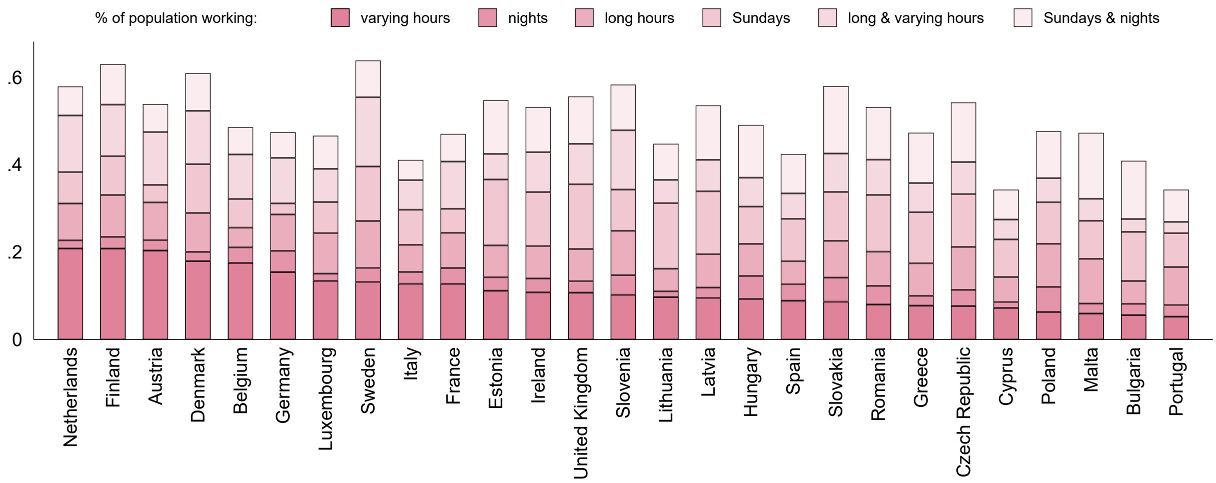
[6]This group may be working on Sunday nights or on Sundays *and* during the nights on other days of the week. EWCS does not permit to pin point if these two NWAs exist jointly or separately.

[7]Again, while EWCS is not sufficient to identify if both these features are inherent to the main job, it seems plausible that for this group of workers, a combination of these two NWAs is a designated characteristic of their jobs.

[8]For example, 636 individuals report that their schedule involves both nights work and varying hours. However, this group is too small to obtain reliable estimates.

distinct groups of workers. The reference sample consists of workers who report *none* of the four NWAs. This sample contains 28,378 workers. In addition, there are six distinct groups of workers with NWAs.

Figure 2: Non-standard working arrangements (NWAs) across countries in EWCS



Notes: The graph is based on EWCS data from waves 3-6. The countries are ordered by prevalence of work in varying hours.

Note that each of the six NWAs bring about the risk of transgressions between work and private life. In some cases, though, they may facilitate combining obligations and aspirations in the private life with the work life. Roughly 49% of the workers report NWAs (with a cross-country coefficient of variation = 1.01). We report the distribution of NWAs across countries in Figure 2. The figure reveals paramount role of varying start and finish times, as well as the prevalence of work on Sundays, with nights and long working hours contributing to a lower fraction of NWAs. The figure also reveals remarkable differences across countries, both in terms of prevalence and in terms of the composition of NWAs.

3.3 Descriptive statistics

Table 1 provides the descriptive statistics of our final sample. We report the full sample as well as sub-samples for the reference group and for each of the specific six NWAs. Approximately 40% of our sample are women. Women are less likely to work during the nights, in long hours and in arrangements combining Sundays and nights. The proportion of women in the reference group is slightly larger than in the full sample, which suggests that women are more likely to work with standard working arrangements than men. Roughly 10% of workers in our sample live in single households, while approximately 12% of individuals in our sample report having one or more children under the age of 7 in the household. Both characteristics are similar across sub-samples. The households in our sample are rarely multi-generational, only 1-2% of the individuals share a household with an elderly person. The fact that these characteristics are so similar across the sub-samples suggests that the link between caring obligations for the elderly and non-standard working arrangements is not driven by joint residence in Europe. This observation is consistent with the conflicting empirical evidence presented by Wheatley (2017).

Table 1 additionally summarizes job characteristics. Roughly 11% of workers are employed part-time, with higher figures in varying hours and Sundays sub-samples. Individuals working at nights and in long hours are rarely on part-time contracts. The proportion of workers who claim working on Saturdays is around 38%, with high correlation of working on Saturdays and Sundays. Table 1 also shows workers' subjective

Table 1: Descriptive statistics

Variable	Full sample	Reference group	Varying hours	Nights	Long hours	Sundays	Long & varying h.	Sundays & nights
% satisfied with their job	**82.9**	**85.1**	**84.6**	**75.5**	**81.0**	**78.3**	**83.7**	**76.2**
Personal characteristics:								
% of women	40.4	46.1	46.4	23.7	27.7	52.0	20.3	25.6
% of single hh	10.3	10.2	11.5	10.3	9.5	9.4	11.6	10.0
% of hh with a child aged<7 yo	12.3	12.0	12.8	13.1	12.2	11.6	14.0	12.2
% of hh with an elder member	1.5	1.5	1.0	1.8	1.1	1.6	1.8	2.2
Job characteristics:								
% working part-time	11.2	11.8	18.5	6.8	4.0	17.5	2.8	7.9
% working on Saturdays	37.6	23.1	30.9	35.8	33.8	89.7	33.5	90.7
% report hours fit schedules	81.9	89.5	86.4	74.7	78.4	71.6	71.1	58.1
% report supportive colleagues	92.7	92.5	91.1	92.4	95.1	92.6	93.2	93.2
% report enough time for tasks	92.8	94.1	93.6	93.5	89.6	92.4	88.1	91.9
% with long commute	29.7	27.0	30.6	28.1	36.2	26.3	40.6	32.0
hazardous conditions (count)	3.03	2.95	2.67	4.00	3.23	3.19	2.78	3.53
NWAs:								
% working in varying hours	11.9	0	100	0	0	0	0	0
% working nights	3.6	0	0	100	0	0	0	0
% working in long hours	8.1	0	0	0	100	0	0	0
% working on Sundays	8.4	0	0	0	0	100	0	0
% working long&varying hours	8.7	0	0	0	0	0	100	0
% working on Sunday nights	8.5	0	0	0	0	0	0	100
Observations	56 107	28 378	6 312	1 728	4 461	5 577	4 408	5 243

Notes: Table presents EWCS data from waves 3-6. We report the sample means, weighted. The weighted proportion of workers with standard working arrangements is 50.8%. In addition to the reported variables, we also use the information about age (categorized into 5-year age groups), industry (categorized into agriculture, manufacturing, market services, non-market services and others) and occupation (grouping ISCO categories into low-skilled, medium-skilled and high-skilled).

opinion on their working environment. Finally, Table 1 reports the prevalence of non-standard working arrangements in the total sample.

4 Methodology

Our starting point is that the difference between actual and preferred working arrangements arise from the fact that some individuals work in schedules inconsistent with their inner ability to manage boundaries. In order to study the difference between actual and preferred working arrangements empirically, we construct counterfactual levels of job satisfaction. We obtain a model linking individual and job characteristics to job satisfaction for all workers in standard arrangements. We then apply this model to workers actually employed in NWAs and we obtain their job satisfaction as if they worked in standard arrangements, *ceteris paribus*. Our approach resembles the well-known Oaxaca-Blinder counterfactuals (Blinder 1973, Oaxaca 1973), where parameters from a regression on one gender are used to recover counterfactual average wage for the other gender. We use machine learning (ML) methods to uncover the latent link between job satisfaction and working time arrangements. The ML algorithm predicts job satisfaction if individuals with NWAs had been working under standard arrangements.

The ML algorithm comes down to a set of decision rules, which applied to data provide classification outcomes. In our framework, the data is split according to different values of explanatory variables (here individual and job characteristics) to provide final classification into four levels of job satisfaction. There are two important advantages of this approach over a standard parametric estimation. First, we do not specify a functional form: given all the available variables, the ML algorithm recovers the deep linkages between individual (and job) characteristics on the one hand, and job satisfaction on the other hand. Second, prediction is a direct classification into job satisfaction levels rather than a parametric prediction; ML thus frees the researcher from setting arbitrary cut-off points for comparing the actual and counterfactual classifications. In other words, we eliminate the scope of arbitrary choices.

Clearly, ML methods do have drawbacks: the need for large sample sizes and commonly outlined difficulty to uncover individual marginal effects for the explanatory variables. We train the model on roughly 28 thousand observations, and our main goal is to obtain the most accurate classification.The algorithm that we use does not allow recovering individual marginal effects, nor does it distinguish the relative importance of the different characteristics.[9]

4.1 Empirical strategy

We characterize job satisfaction as a function of individual and work characteristics (x_i) and unobserved ability to manage boundaries (u_i), which is crucial if an individual works under non-standard arrangements, but is irrelevant when an individual works with standard arrangements (hence the interaction with NWAs).

$$JS_i = f(x_i, NWA_i \times u_i)$$

Denote $\mathbb{X}_0$ the set of individuals who work in standard arrangements and $\mathbb{X}_1$ the set of individuals who work in NWAs. We proceed in three steps:

Step 1. Obtain model of job satisfaction We train (estimate) the f, that is a ML model for the determinants of job satisfaction for the respondents from a reference group ($x_i \in \mathbb{X}_0$), i.e. individuals who work under standard arrangements. The ML algorithm predicts job satisfaction using workers' individual (e.g. gender, parenting status, age, etc.) and work characteristics (occupation, industry, part-time indicator, commuting time, etc.).

From a wide range of machine learning classification techniques, the random forest (RF) classifier (Breiman 2001) works best in our data.[10] We make this judgment in terms of number of cases correctly classified (i.e. accuracy) as well as the out-of-diagonal elements of the confusion matrix, that is the number of cases for which a model predicts a value different than the actual one.

Our final model has high sensitivity (accurately classified cases), and maintains high specificity. Table C.1 in Appendix C reports the actual and the model-implied levels of job satisfaction for the individuals without NWAs. For comparison, we also report analogous results from a standard parametric regression approach, demonstrating superiority of the ML. In fact, an ordered probit model utilizing the same variables assigns 95% of observations to the largest category (*Satisfied*), when in the data less than 61% of individuals reported this level. Moreover, the ordered probit model never assigns *very dissatisfied*, and less than 1% of individuals are assigned to *dissatisfied*.

We also validate our predicted job satisfaction by comparing it with an index of job satisfaction which was not used to train the ML models. This index is composite derived from several variables available solely in the sixth wave of EWCS. We estimate correlations between composite index of job satisfaction on the one hand, and actual JS as well as predicted JS on the other hand. The results are reported in Table C.2 in Appendix C. We obtain very similar estimates of correlations and R^2 for models with factual and counterfactual job satisfaction.

Step 2. Obtain counterfactual job satisfaction We use the ML trained model to predict counterfactual level of job satisfaction for respondents in $\mathbb{X}_1$, i.e. individuals who work under non-standard arrangements. We use the decision rules produced by the ML algorithm in the previous point to predict the counterfactual job satisfaction that workers with NWA would have experienced if they worked in standard work arrangements.

[9] Molnar (2022) provides middle-ground, with regression-type ML estimation, but this approach cannot be applied in our case.
[10] We elaborate on the specifics of our use of ML methods in Appendix B.

The ML algorithm recovers $\hat{f}(x_i, [NWA_i = 0] \times u_i)$ if $x_i \in \mathbb{X}_1$, that is the job satisfaction for an individual with characteristics x_i, where we counterfactually set NWA to be equal to zero. For example, we use the predictive model estimated on the reference group to predict the job satisfaction of individuals who work on Sundays, as if they worked in standard arrangements.

Step 3. Obtain the difference between factual and counterfactual job satisfaction We construct the difference between the actual and the model-derived counterfactual job satisfaction. In the remainder of this paper we refer to counterfactual job satisfaction "counterfactual JS" and to actual job satisfaction as "factual JS" . We can write this difference as:

$$\Delta JS_i = \text{Factual } JS_i - \text{Counterfactual } JS_i = f(x_i, NWA_i \times u_i) - \hat{f}(x_i, [NWA_i = 0] \times u_i) \forall x_i \in \mathbb{X}_1$$

where ΔJS_i is the difference in job satisfaction, and Counterfactual JS_i indicates the levels for worker i in working arrangements NWA. For example, imagine an individual who works long hours ($NWA = 1$) and reports low job satisfaction (4). Given this individual's personal and job characteristics, our ML model indicates that if this worker had been employed in regular working hours (i.e. taking away NWA), her/his job satisfaction would have been 1. Then, the value of the mismatch measure equals the difference in job satisfaction: $\Delta JS_i = 4 - 1 = 3$. In this example, removing NWA would lead to an improvement by three levels.

A comparison of the actual and the counterfactual levels of job satisfaction results in one of three possible outcomes for workers in $\mathbb{X}_1$. First, it can occur that the actual and the counterfactual job satisfactions are the same (a worker is just as well off with and without NWA_i). Second, it could be that job satisfaction is higher in actual than in counterfactual (a worker is better off keeping $NWA_i = 1$ and would lose from having standard working arrangements). Finally, job satisfaction could be lower in actual than in counterfactual (a worker would benefit from changing from $NWA_i = 1$ to standard working arrangements $NWA_i = 0$).

Further analyses: systematic component of mismatch Once we complete the three steps, we check whether the difference between actual and counterfactual levels of job satisfaction depends on personal characteristics.

We use ΔJS_i as the dependent variable in a standard regression model. Given that job satisfaction variables have four levels, the differences can take on seven different possible values (three positive, three negative and no change). Individuals with $\Delta JS_i = 0$ imply that no gain in job satisfaction arises from eliminating NWAs. Positive values of ΔJS_i correspond to greater improvement in job satisfaction if the counterfactual working arrangements were to become factual; analogously, negative values of ΔJS_i imply a decline in job satisfaction.

We utilize ΔJS_i variable in two ways. First, we construct a dummy, which takes on the value of 1 if there is improvement in job satisfaction and 0 otherwise. We call this outcome variable improvement. Further, we measure the intensity, i.e. we use the original seven-level categorical variable where larger positive values are greater improvement and larger negative values are greater deterioration of job satisfaction.

We estimate a probability of mismatch in working arrangements in M1 (logit) and intensity of mismatch in M2 (ordered logit), obtained through a counterfactual experiment of taking away NWA.

$$\begin{aligned} Improvement(Y/N)_i &= \beta_c + \beta_w \times woman + \beta_p \times parent + \gamma woman \times parent + \delta controls_i + \epsilon_i \quad (M1) \\ \Delta JS_i &= \alpha_c + \alpha_w \times woman + \alpha_p \times parent + \eta woman \times parent + \theta controls_i + \varepsilon_i \quad (M2) \end{aligned}$$

The parameters indexed with c denote country fixed effects. The parameters indexed with w and p reflect the own effects of gender (on non-parents) and parenthood status (on men). The interaction terms γ and η reflect the additional effect of motherhood. We focus on exploring the role of gender and parenting in the mismatch of working arrangements. We thus include the personal characteristics which are of main importance, i.e. gender and presence of child under 7 years of age in the household. All regressions adjust age and household characteristics. These models are estimated separately in six sub-samples representing six different forms of NWAs.

Further analyses: country-level analyses Aggregating the individual results to country-level, we can similarly identify three groups: those for whom non-standard arrangements are optimal, those for whom standard arrangements are optimal, those who work in non-standard arrangements but would be better off in standard working arrangements. We can compute these measures for each specific NWA, as well as for all NWAs taken together.

4.2 Assumptions

Our approach explores deep underlying patterns in the data, in a way similar to Athey et al. (2021). The key assumptions underlying our methodology are as follows.

Assumption 1 *There is common support.*

That is we are able to compute counterfactual $\hat{f}(x_i, [NWA_i = 0] \times u_i)$ $\forall x_i \in \mathbb{X}_1$ for everyone in the $NWA = 1$ sub-sample. More generally, assignment to $\mathbb{X}_0$ or $\mathbb{X}_1$ samples is probabilistic in the population. This assumption is mechanically imposed in our approach, because singletons (observations available only in either $\mathbb{X}_0$ or $\mathbb{X}_1$ sample) are dropped from the analysis before the ML model was trained.

Assumption 2 *The unobserved ability to manage boundaries (u_i) affects JS only when $NWA = 1$.*

It means that job satisfaction is independent of ability to manage borders in standard working arrangements (that is when one does not need to manage the potential spillovers). From the theory, we additionally expect that $f(.)$ is not decreasing in $NWA_i \times u_i$, i.e. more capable workers can derive higher job satisfaction from working in NWAs than workers who are not able to manage boundaries.

Assumption 3 *Individuals are heterogeneous along a single dimension.*

It means that the ability to manage boundaries is the only unobserved variable affecting job satisfaction that differs systematically between workers with NWAs and without them. Consider that there are two sources of unobserved heterogeneity, such that $JS_i = f(x_i, NWA_i \times (u_i + v_i))$, where v_i could be related to another factor systematically driven by NWAs. If that was the case, then we would not be able to tell apart the difference between u_i and v_i. For example, the additional factor v_i could be related to potential dissatisfaction of individual i with his/her pay, relative to his/her peers who work in jobs with different NWAs than individual i.

Note that the policy analysis focuses on the removal of NWAs rather than introducing them. This focus arises from our theoretical premises. The theory emphasizes that some workers will have difficulties in setting boundaries when work arrangements are non-standard. Hence, *ceteris paribus* the existence of NWAs leads to lower job satisfaction on average. For workers who work in standard arrangements we cannot construct the counterfactual job satisfaction if they worked in NWAs, because the term u_i is not defined in this population. Our empirical approach attributes NWAs critical role in determining job satisfaction through an adequate match from the same assumptions as in the standard treatment effect methodology. These rely on Assumptions 1 and 2: common support and conditional independence of treatment and potential outcomes. We provide the full derivation in Appendix D.

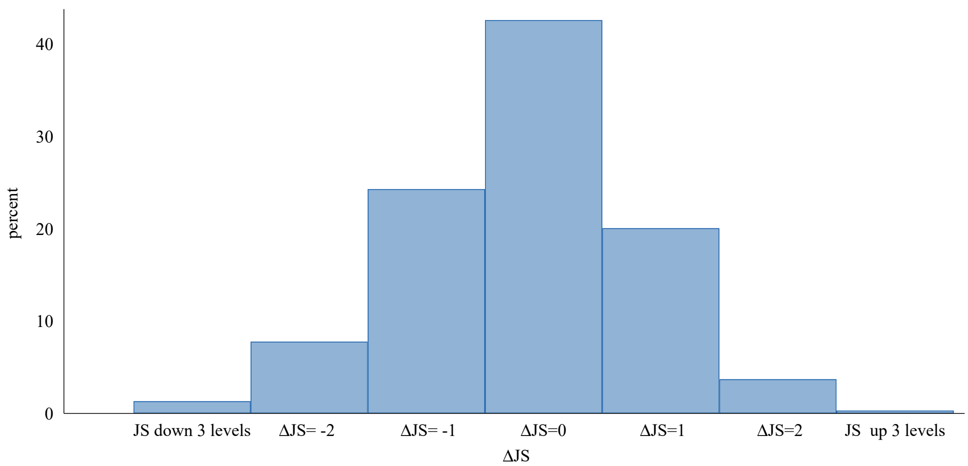
5 Results

We describe the results in three substantive parts, which match the three research questions states at the beginning. First, we briefly characterize the prevalence of mismatch in the light of our empirical approach, thus we start from Research Question 1. Next, we refer to Research Question 2 and discuss individual level drivers of the mismatch. Third, we provide aggregations of counterfactual NWA to establish the potential scope for job satisfaction gains and losses, to answer Research Question 3. We conclude this section by placing our results in the context of the flexibility enactment theory.

5.1 Prevalence of (mis)match in working arrangements

In our sample, for 43% of individuals no change in job satisfaction is implied by comparing actual job satisfaction and $\hat{f}$. Further, 33% of the sample would have observed a decline in job satisfaction if NWAs were removed, whereas the remaining 24% would have seen a rise in their job satisfaction absent NWAs. In other words, the distribution is skewed to the left. Figure 3 presents the distribution of ΔJS_i in our $x_i \in \mathbb{X}_1$ sample.

Figure 3: Changes between factual and counterfactual job satisfaction



Notes: The figures present the distribution if ΔJS_i for $x_i \in \mathbb{X}_1$. Figure E.1 in the Appendix, reports a similar graph across NWAs.

Note that positive values arise as high levels of dissatisfaction ($JS \in [3, 4]$) are replaced in the counterfactual scenarios with job satisfaction ($JS \in [1, 2]$). Conversely, negative values of ΔJS_i emerge as job satisfaction is replaced with job dissatisfaction.

The overall share of 33% of individuals whose job satisfaction would have been lower had they been deprived the opportunity to work with NWAs suggests that on average roughly a third of individuals with non-standard arrangements have larger benefits from flexibility than experience the strain associated with managing boundaries. These individuals are matched. For the 24% who would benefit from removing the NWAs, either the costs of managing boundaries are too high or the benefits of flexibility are too low for the two to balance out. These individuals are mismatched. We next inspect if this dichotomy is driven by gender and parenthood status.

5.2 Individual drivers of (mis)match in working arrangements

Research question 2 considers whether the mismatch between actual and preferred working arrangements is systematic. Specifically, women and parents are less likely to experience mismatch due to higher abilities to actively manage boundaries. Table 2 summarizes the impact of the variables of interest on the probability that removing NWAs improves the job satisfaction of a worker.

We provide two sets of estimates for the event that taking away NWAs improves job satisfaction: the probability of improvement (a *logit* specification, M1) and the seven-level categorical variable measuring the intensity of improvement (ranging from -3 to +3; a reduction of job satisfaction by 3 levels to an improvement in job satisfaction by 3 levels, respectively, an *ordered logit* specification, M2). All specifications include country fixed-effects and year fixed-effects. The former makes sure that our estimates are not affected by differences in mean values between countries. The latter addresses the same concerns for potential time trends (common across countries). Thus, our specifications isolate the effects at the individual level.

Overall, the results reported in Table 2 are consistent for the ordered logit estimators, but the logit estimators are often underpowered to reject the null hypothesis of insignificance. To address this point, we provide further evidence: we combine all NWA sub-samples into one estimation and introduce dummy variables for each NWA. The main objective of this specification is to improve the power of our analysis. The results are reported in the last panel of Table 2 and confirm that overall, both M1 and M2 models deliver similar results, both in terms of magnitude and in terms of significance.

Table 2: Does taking away NWAs improve job satisfaction?

	Varying hours		**Nights**		**Long hours**		**Sundays**	
	Logit (M1)	OLogit (M2)	Logit (M1)	OLogit (M2)	Logit (M1)	OLogit (M2)	Logit (M1)	OLogit (M2)
woman (β_w / α_w)	-0.08	-0.12**	-0.13	-0.19***	-0.29***	-0.32***	-0.01	-0.03
	(0.07)	(0.06)	(0.10)	(0.09)	(0.06)	(0.06)	(0.07)	(0.05)
parent (β_p / α_p)	-0.22**	-0.26***	-0.19*	-0.28***	-0.08	-0.12	-0.31***	-0.31***
	(0.13)	(0.11)	(0.13)	(0.11)	(0.11)	(0.09)	(0.14)	(0.09)
woman $\times$ parent (γ / η)	-0.01	0.04	-0.54	0.01	0.09	0.10	0.15	0.17
	(0.19)	(0.13)	(0.43)	(0.25)	(0.19)	(0.17)	(0.18)	(0.13)
Observations	6 312		1 728		4 461		5 577	
Mean predicted probability	0.29		0.43		0.35		0.32	
Correctly classified	71.25%		58.64%		63.91%		68.10%	
Specificity	100.00%		82.62%		99.82%		99.89%	
	Long & varying hours		**Sundays & nights**				**NWAs jointly**	
	Logit (M1)	OLogit (M2)	Logit (M1)	OLogit (M2)	(M1)	(M2)	Logit (M1)	OLogit (M2)
woman (β_w / α_w)	-0.30***	-0.31***	0.01	-0.01			-0.17***	-0.19***
	(0.08)	(0.07)	(0.05)	(0.06)			(0.03)	(0.03)
parent (β_p / α_p)	-0.41***	-0.25***	-0.21***	-0.20***			-0.24***	-0.23***
	(0.09)	(0.08)	(0.07)	(0.06)			(0.05)	(0.04)
woman $\times$ parent (γ / η)	0.41**	0.23	0.37***	0.29*			0.13***	0.13***
	(0.24)	(0.16)	(0.19)	(0.19)			(0.07)	(0.06)
Observations	4 407		5 243				27 729	
Mean predicted probability	0.35		0.34				0.33	
Correctly classified	65.31%		66.81%				66.63%	
Specificity	98.92%		98.05%				99.59%	

Notes: Table presents the results of estimating an impact of selected individual characteristics on counterfactual changes in workers' job satisfaction when taking away the non-standard working arrangements. Reported are point estimates. The dependent variable for the ordered logistic regression (OLogit) is the change in job satisfaction index (ΔJS_i). For the logistic regression (Logit) we construct a binary variable indicating an "improvement" or a positive change in job satisfaction when NWAs are eliminated. Note that $\Delta JS_i = 0$ (i.e. no change) is classified as 0 in Logit. The variable $parent$ takes on the value of 1 if there is a child younger than 7 years old in the household. Models include age, individual country indicators and wave indicators as regressors. The specification for all sub-samples jointly include additionally a sequence of dummy variables for each type of NWA. Standard errors presented in parentheses. ***, ** and * denote significance at p<0.05, p<0.1, and p<0.15, respectively.

Gender With the exception of Sundays and Sunday & nights, women's job satisfaction is less likely to improve than for men if their NWAs were replaced by a standard arrangements. The results are not robust in M2 and occasionally lose significance due to low precision in M1. The exact interpretation of the estimated ordered logit coefficients is that the ordered log-odds for a (non-parenting) woman being in a higher job satisfaction category is 0.12-0.32 lower than an equivalent (non-parenting) man. Looking at all NWAs jointly, the estimated logit coefficients suggest that the probability that (non-parenting) woman's job satisfaction increases after NWAs are removed is roughly 17% smaller than for (non-parenting) man. This result implies that women's job satisfaction is less affected by the need to manage boundaries, despite the fact that the six studied NWAs create the potential for severe transgressions from professional to private life. For Sundays, as well as Sundays & nights the effect of gender is absent not only due to low precision, but also because the estimated coefficients are close to zero.

Parenting Table 2 reveals negative coefficients for fathers of children under 7 years. This result is consistent across all the NWAs, except for long hours where the precision of the estimated coefficients makes them insignificant, but also considerably smaller. Statistical power to reject the null hypothesis is higher in the ordered logit specifications (M2), but the point estimates are similar between M1 and M2. The point estimates suggest marginal effects of roughly 7.9 percentage points on mean prevalence of roughly 33%, i.e. the probability that fathers benefit from the removal of NWAs is on average 24% smaller than for non-parenting men.

The effects of parenting appear similar for fathers and mothers: in the case of most NWAs, there is no statistically significant interaction between gender and parenting young children. The exception are two combinations of NWAs: long and varying hours as well as Sundays & nights. In these two combinations, our models suggest an improvement in job satisfaction for mothers, relative to non-mothers and the effect virtually cancels out for mothers relative to fathers. The existing literature argues that parents (or, more generally, care providers) appreciate flexibility (Bainbridge and Townsend 2020). Our analysis refines this result by showing that there are some forms of NWAs that can be neutral for mothers, but not for fathers, in terms of job satisfaction.

5.3 Mismatch in working arrangements across countries

The results in Table 2 specify who is *relatively* more likely to be matched into NWAs. To address Research Question 3, we aggregate our indicators to the national level. We use the potential outcomes implied by ΔJS_i to classify workers into three groups as described earlier (i) NWA_i reduces JS relative to standard arrangements, which we classify as mismatched; (ii) NWA_i raises JS relative to standard arrangements, which we classify as matched; and (iii) no change in JS between having and not having NWA_i, which we classify as indifferent. For Research Question 3 the indifferent workers are irrelevant,[11] thus we construct the net measure as a difference between matched NWA workers and mismatched NWA workers. The net is negative if matched share is higher than the mismatched share and positive in the opposite case. Overall, the thought experiment which we propose balances between those who would benefit from the elimination of non-standard working arrangements and those who would lose from it. We explore the comparative character of EWCS data and study if there are countries where the mismatch of NWAs is particularly pronounced.

Figure 4 portrays changes in job satisfaction if all NWAs are eliminated, aggregated over all NWAs (analogous bars for specific NWAs are portrayed in the Appendix, Figures E.5 to E.9). We present the results

[11]The share of indifferent workers is reported across NWAs and countries in Figure E.2 (for all NWAs jointly), and Figure E.3 (across specific NWAs). On average, the share of indifferent workers ranges from 37% to 47%. The share of indifferent workers may be an artefact of the way job satisfaction is measured in EWCS (a four-point Likkert scale). It is also plausible that for some individuals the trade-offs on average balance out and thus they could fluctuate between NWAs and standard arrangements without detriment to their job satisfaction.

separately for men and women (top panel), as well as parents and non-parents (bottom panel). Positive bars signify the share of mismatched workers, that is workers with NWAs who benefit from adopting standard arrangements. Conversely, the negative bars show matched workers (that is those whose job satisfaction would decline if NWAs were replaced by standard arrangements). The values on the vertical axis signify the shares of NWAs workers. For example, in Austria matched women outweigh mismatched women by roughly 8 percentage points, the relevant figure for men is close to 18 percentage points.

Figure 4 displays considerable heterogeneity across European countries, though in a majority of them we observed negative net for both men and women. Indeed, the split by genders reveals that there are few cases, where the positive bars outweigh the negative ones (the net of removing NWAs is positive). The situation is more ambiguous when we look at aggregates by parents and non-parents. There appear to be highly polar cases among parents. While these opposites may reflect country-specificity, they may also be related to a smaller sample size for parents. Overall, Figure 4 reveals that, for each country, matches outweigh mismatched NWAs workers in a sense that a higher share of workers would experience a reduction in job satisfaction if NWAs were removed than would gain from such policy change. Our method cannot help to judge if some employees in standard working arrangements would benefit from access to NWAs, thus we cannot directly evaluate the gains from reallocation of workers between NWAs and standard arrangements. However, while we cannot estimate the share of workers deprived of NWAs, our method is sufficient to grasp the maximum share of workers inadequately matched to NWAs.

Our results strongly suggest that most workers in NWA were able to individually manage the boundaries between these two spheres and their job satisfaction would not be improved if they no longer had to actively engage in setting boundaries. At the same time, both the share of indifferent NWA workers and the shares of matched and mismatched NWA workers appear to be highly heterogeneous across countries. This heterogeneity defies simple characterizations, as it does not seem to be consistent with the groupings of the European countries according to welfare state or indices based on prevalence of fixed-term employment, share of self-employment, etc.

Clearly, gender and parenting status are not the only variables affecting ability for boundary management, there can be other, direct factors (Mellner et al. 2014). For instance, the usage of multiple devices (separate phones or computers for work and private lives) is a way of creating boundaries between home and work (Fleck et al. 2015) and thus may facilitate managing boundaries. Obviously, some forms of employment (e.g. self-employment and home-based work) require better abilities to manage boundaries (Myrie and Daly 2009). Unfortunately, no such variables are available in EWCS.

5.4 Discussion

Our results show that women are substantially less likely to benefit from eliminating the NWAs. This strong gender gradient appears despite the observation that across all countries, all models and all data subsamples, the majority of individuals would experience no change in job satisfaction subsequent to the elimination of NWAs. Parenting is also relevant: fathers have a stronger decline in job satisfaction from eliminating NWAs relative to non-fathers. We find mixed results for mothers: typically they appear similar to fathers, but for some NWAs they could benefit if NWAs were replaced by regular working arrangements relative to fathers, but not-relative to non-mothers.

A potential concern related to our results stems from working with cross-sectional data. Ideally, one would study longitudinal data and measure changes in job satisfaction for individual workers as their work arrangements change between non-standard and standard, or as they change their parental status. With EWCS such analysis is not feasible, but with other data such stepwise analysis could reveal deeper and more refined policy implications. In principle, our method can be applied the same way to individual level

Figure 4: Decomposing the changes in job satisfaction if NWAs are eliminated

Notes: Taking away NWA refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of mismatched and matched individuals obtained with population weights.

data from one specific country. We have experimented with German SOEP data, but the questions related to working arrangement were only asked once every few years, which reduces the suitability of this database for our analysis.

Another source of concern might be that we treat the counterfactual levels of job satisfaction as if they were deterministic, but truly they are stochastic, i.e. our model predicts job satisfaction in the scenario when non-standard work is eliminated, but this prediction should be associated with certain probability or a confidence interval around the predicted value. As long as it does not matter on the individual level, this uncertainty starts to play a role when results are averaged (like in a regression) or aggregated (in country levels matched/mismatched groups). As a result the statistical power of tests performed on logit/ordered logit and aggregated estimates might be lower than presented. A possible way around the problem would be to recover a measure of uncertainty from the ML algorithm (in line with recovering regression type estimates) and use it to adjust standard errors in later steps. We leave it for future research.

Due to data limitations, our analysis omits wages. However, it could be that job satisfaction is driven by a wage relative to peers (especially at the same workplace). Likewise, it could be that individuals with NWAs are compensated with a wage somewhat higher than peers in standard arrangements. This would be an indirect channel of relative wages influencing job satisfaction, where NWAs appear as a context rather than a mediator. While we cannot exclude this to be the case, our specifications adjust for a number of individual characteristics relevant for wages (age, and education) as well as job and employer-specific indicators (including industry and occupation, as well as hours worked). These controls capture the systematic variation in wages, though they fail to account for any residual variation. More importantly, the omission of this variable would produce lower counterfactuals across all groups. This could affect our second stage estimates only if the rewards for women and parents for working flexibly were larger than for other groups. While these controls cannot fully capture variation in wages, they do capture the systematic component of wage dispersion.

Finally, our results would benefit from a richer measure of job satisfaction. In EWCS, the categorical job satisfaction indicator is only weakly correlated with a composite index measurement, as portrayed in Appendix C.2. Indeed, the vast majority of people report being satisfied with their job *despite* negative answers to the composite marker questions. Refined measurements of job satisfaction could permit a more complex machine learning classification than what is feasible with current EWCS data.

Our measure of job mismatch builds on a difference between the reported and counterfactual job satisfaction measures. Our measure, the change in job satisfaction, is indicative of the underlying boundary management ability, not an exact measure. We infer that this difference is informative of abilities to manage boundaries. In 2015, EWCS incorporated a more direct measure of this ability in the form of questions on family-work conflict. These questions are inspired by the works of Netemeyer et al. (1996) and Hayman (2005).[12] Table C.3 in Appendix C reveals that among those workers for whom the elimination of NWA would be beneficial, the proportion experiencing family-work and work-family conflict most of the time is higher than among those workers who are matched. Moreover, the association is stronger in the case of transgressions from work to personal life, than in transgressions in the opposite direction. These results are along our initial expectations, as flexibility enactment theory emphasizes work concerns permeating into family sphere.

Our results serve to substantiate the flexibility enactment theory. It appears that individuals employed in non-standard working arrangements have the ability to manage boundaries between work and private life and thus they do not experience dissatisfaction from them. Eliminating NWAs would not contribute to raising their job satisfaction. The policy debate about NWAs ought to account for individual abilities, as one-size-fits-all regulations are likely to leave many workers with lower job satisfaction levels. Overall, formulating policy for adequate workers' rights protection becomes a subtle challenge under flexibility enactment theory.

Our study offers several innovations relative to other strands of the existing literature. First, rather than hypothesizing about the type of workers and the types of occupations that may be more challenged by boundary management, we focus on identifying the *gradient* of job satisfaction across individuals. This gradient reflects a hypothesis long salient in the literature, that a given person may derive different levels of job satisfaction if some specific features of this job are altered. The idea of gradient builds on the earlier conceptualization of the self-discrepancy theory and social identity theory, i.e. for some individuals irregularity is a natural habitat, whereas for the others it triggers negative spillovers and transgressions,

[12] EWCS includes fewer items, and the specific formulations differ from the measures validated in previous empirical literature.

necessitating the burdensome development of a boundary management strategy.

Second, we make no *ex ante* assumptions about the drivers of the gradient. We employ machine learning rather than parametric modelling. This methodology allows us to avoid many discretionary choices. Foremost, we do not impose any restrictions on the functional form of the model nor on the statistical distribution of the data. Neither the definition of variables nor their interrelationships have to be pre-specified in a model which relies solely on classification. Moreover, given that job satisfaction is typically a categorical variable, we do not need to assume anything about the thresholds for assigning the parametric predictions to the categories: the machine learning algorithm automatically assign the level of job satisfaction to each individual.

Third, we explore data from the representative cross-country survey, thus permitting the identification of the gradient and the mismatch along all occupational groups and industries. This helps to mitigate the risks of omitting large groups of workers mismatched in their working arrangements.

6 Conclusions

In this paper, we build on flexibility enactment theory. We leverage this approach to answer novel research questions about the prevalence of mismatches between actual and optimal working arrangements. Our approach helps to reconcile two superficially opposite views in the policy debate. The first view emphasizes that flexibility at work is what helps workers reach their potential. The second view emphasizes that work tends to transgress into private life and that this transgression is easier when working arrangements leave room for abuse. Flexibility enactment theory posits that some workers are immune to such abuse, because they are able to actively manage boundaries and separate work from private life. This theory builds on role theory and social identity theory as well as spillover theory and boundary theory. From these theories jointly, we conjecture that the ability to actively manage boundaries is not merely a skill, but it actually enables deriving satisfaction from a job even if it is characterized by non-standard working arrangements.

We test implications from flexibility enactment theory using data from the European Working Conditions Survey. We deploy modern machine learning methods to uncover (potentially complex and nonlinear) relationships between individual characteristics, family characteristics, job features and job satisfaction, separately for individuals working in standard employment arrangements and in non-standard employment arrangements. The machine learning models are then used to obtain counterfactual levels of job satisfaction for workers with NWAs, as if they worked in standard arrangements. These counterfactual simulations enable the identification of who would benefit from a change in working arrangements, who would lose, and who is indifferent between the two arrangements. We characterize the extent and drivers of mismatch among 27 European countries.

Our results have important policy implications. Given substantial heterogeneity across countries as well as the systematic character of mismatches across individuals, no one-size-fits-all policy is likely to be successful in raising the job satisfaction of all workers. Neither the curtailment nor the widespread introduction of NWAs will universally improve worker welfare. If anything, the only potentially universal policy recommendation would be raising the boundary management skills (e.g. by introducing them to educational curricula). Otherwise, individuals are highly heterogeneous in terms of suitable working arrangements. Furthermore, the workers' inclination to engage in non-standard arrangements depends on personal circumstances and may change over the life-cycle.

Our results call for further theorizing on the links between job satisfaction and working conditions. While the flexibility enactment theory appears to find strong confirmation in the data, there seem to be several particularly promising areas for wider exploration. First, it appears that the notion of social identity may be more directly intertwined with flexibility enactment. Specifically, identity may be an important pillar of boundary management strategy (e.g. ethnicity-specific work ethics). While it was demonstrated that managing spillovers may be viewed as a distinguishing characteristic of some groups, it is not warranted that the extent of satisfaction derived from this ability is automatically related to this ability: feeling competent to manage some situations is not automatically equivalent to an appreciation of the situations themselves. Second, it appears that the existing theories expect the non-standard working arrangements to facilitate transgressions from work to private lives (hence the need to *manage* the spillovers). There exists a large literature on transgressions in occupations with long hours (medical professions, police force). To some individuals, however, there may be intrinsic value in the arrangements associated with their occupations, often non-standard ones (e.g. actors, musicians, television and radio producers). A better understanding of the intrinsic value of non-standard arrangements through the lens of flexibility enactment theory could help us refine the meaning of the term "transgression" and thus the mechanisms tying them to job (and life) satisfaction.

Last, but certainly not least, the gender context of the flexibility enactment theory should be revisited. Our results show that women are much less likely to benefit from eliminating the NWAs, but we find little support that this effect is driven by parenting. Indeed, for most types of mismatches, parents of both genders are equally likely to benefit from a change in arrangements. Given these findings, one cannot explain away the persistent and large gender differences in mismatches with the burdens of childbearing and rearing. These large gender differences call for thorough exploration into social identity theory and boundary theory.

References

Albanesi, S. and Olivetti, C.: 2009, Home production, market production and the gender wage gap: Incentives and expectations, *Review of Economic Dynamics* **12**(1), 80–107.

Angrave, D. and Charlwood, A.: 2015, What is the relationship between long working hours, over-employment, under-employment and the subjective well-being of workers? longitudinal evidence from the UK, *Human Relations* **68**(9), 1491–1515.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G. and Khosravi, K.: 2021, Matrix completion methods for causal panel data models, *Journal of the American Statistical Association* pp. 1–15.

Atkinson, C., Lucas, R. and Hall, L.: 2011, Flexible working and happiness in the nhs, *Employee Relations* **33**, 88–105.

Bainbridge, H. T. and Townsend, K.: 2020, The effects of offering flexible work practices to employees with unpaid caregiving responsibilities for elderly or disabled family members, *Human Resource Management* **59**(5), 483–495.

Barnett, R. C. and Gareis, K. C.: 2007, Shift work, parenting behaviors, and children's socioemotional well-being: A within-family study, *Journal of Family Issues* **28**(6), 727–748.

Barney, C. E. and Elias, S. M.: 2010, Flex-time as a moderator of the job stress-work motivation relationship: A three nation investigation, *Personnel Review* .

Bellmann, L. and Hübler, O.: 2020, Working from home, job satisfaction and work–life balance–robust or heterogeneous links?, *International Journal of Manpower* .

Blinder, A. S.: 1973, Wage discrimination: Reduced form and structural estimates, *The Journal of Human Resources* **8**(4), 436.

Bloom, N., Liang, J., Roberts, J. and Ying, Z. J.: 2015, Does working from home work? evidence from a chinese experiment, *The Quarterly Journal of Economics* **130**(1), 165–218.

Bohle, P. and Tilley, A. J.: 1989, The impact of night work on psychological well-being, *Ergonomics* **32**(9), 1089–1099.

Bolino, M. C., Kelemen, T. K. and Matthews, S. H.: 2021, Working 9-to-5? a review of research on nonstandard work schedules, *Journal of Organizational Behavior* **42**(2), 188–211.

Breiman, L.: 2001, Random forests, *Machine Learning* **45**(1), 5–32.

Brescoll, V. L., Glass, J. and Sedlovskaya, A.: 2013, Ask and ye shall receive? the dynamics of employer-provided flexible work options and the need for public policy, *Journal of Social Issues* **69**(2), 367–388.

Cech, E. A. and Blair-Loy, M.: 2014, Consequences of flexibility stigma among academic scientists and engineers, *Work and Occupations* **41**(1), 86–110.

Cha, Y. and Weeden, K. A.: 2014, Overwork and the slow convergence in the gender gap in wages, *American Sociological Review* **79**(3), 457–484.

Chait Barnett, R., Gareis, K. C. and Brennan, R. T.: 2008, Wives' shift work schedules and husbands' and wives' well-being in dual-earner couples with children: A within-couple analysis, *Journal of Family Issues* **29**(3), 396–422.

Chen, C. and Breiman, L.: 2004, Using random forest to learn imbalanced data, *University of California, Berkeley* .

Cheng, W.-J., Puttonen, S., Vanttola, P., Koskinen, A., Kivimaki, M. and Harma, M.: 2021, Association of shift work with mood disorders and sleep problems according to chronotype: a 17-year cohort study, *Chronobiology International* **38**(4), 518–525.

Clark, S. C.: 2000, Work/family border theory: A new theory of work/family balance, *Human relations* **53**(6), 747–770.

Cook, C., Diamond, R., Hall, J. V., List, J. A. and Oyer, P.: 2021, The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers, *The Review of Economic Studies* **88**(5), 2210–2238.

Cortes, P. and Pan, J.: 2019, When Time Binds: Substitutes for Household Production, Returns to Working Long Hours, and the Skilled Gender Wage Gap, *Journal of Labor Economics* **37**(2), 351–398.

Cubas, G., Juhn, C. and Silos, P.: 2019, Coordinated work schedules and the gender wage gap, *NBER WP 26548*, National Bureau of Economic Research.

Duchini, E. and van Effenterre, C.: 2018, School schedule and the gender pay gap, *Policy Brief 26*, Institute for Public Policy.

Dumas, T. L. and Sanchez-Burks, J.: 2015, The professional, the personal, and the ideal worker: Pressures and objectives shaping the boundary between life domains, *The Academy of Management Annals* **9**(1), 803–843.

Flabbi, L. and Moro, A.: 2012, The effect of job flexibility on female labor market outcomes: Estimates from a search and bargaining model, *Journal of Econometrics* **168**(1), 81–95.

Fleck, R., Cox, A. L. and Robison, R. A.: 2015, Balancing boundaries: Using multiple devices to manage work-life balance, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, Association for Computing Machinery, New York, NY, USA, pp. 3985–3988.

Florean, D. and Engelhardt-Woelfler, H.: 2020, His and her working hours and well-being in germany: A longitudinal crossover-spillover analysis, *Journal of family Research* **32**(2), 249–273.

Gracia, P. and Garcia-Roman, J.: 2018, Child and adolescent developmental activities and time use in spain: The gendered role of parents' work schedules and education levels, *European Sociological Review* **34**(5), 518–538.

Grzywacz, J. G., Carlson, D. S. and Shulkin, S.: 2008, Schedule flexibility and stress: Linking formal flexible arrangements and perceived flexibility to employee health, *Community, Work and Family* **11**(2), 199–214.

Hamplová, D.: 2019, Does work make mothers happy?, *Journal of Happiness Studies* **20**(2), 471–497.

Han, W.-J.: 2008, Shift work and child behavioral outcomes, *Work, Employment and Society* **22**(1), 67–87.

Hayman, J.: 2005, Psychometric assessment of an instrument designed to measure work life balance, *Research and practice in human resource management* **13**(1), 85–91.

Hertz, R. and Charlton, J.: 1989, Making family under a shiftwork schedule: Air force security guards and their wives, *Social Problems* **36**(5), 491–507.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D. and Rosenthal, R. A.: 1964, *Organizational stress: Studies in role conflict and ambiguity*.

Katz, D. and Kahn, R. L.: 1978, *The social psychology of organizations*, Vol. 2, Wiley New York.

Kossek, E. E., Lautsch, B. A. and Eaton, S. C.: 2004, Chapter 12: Flexibility enactment theory: Implications of flexibility type, control, and boundary management for work-family effectiveness, *Work and Life Integration: Organizational, Cultural and Individual Perspectives*, Psychology Press.

Lachowska, M., Mas, A., Saggio, R. and Woodbury, S. A.: 2023, Work hours mismatch, *Working Paper 31205*, National Bureau of Economic Research.

Lee, B. Y., Wang, J. and Weststar, J.: 2015, Work hour congruence: the effect on job satisfaction and absenteeism, *International Journal of Human Resource Management* **26**(5), 657–675.

Lee, M. D., MacDermid, S. M., Williams, M. L., Buck, M. L. and Leiba-O'Sullivan, S.: 2002, Contextual factors in the success of reduced-load work arrangements among managers and professionals, *Human Resource Management* **41**(2), 209–223.

Lott, Y.: 2015, Working-time flexibility and autonomy: A european perspective on time adequacy, *European Journal of Industrial Relations* **21**(3), 259–274.

Lott, Y.: 2020, Does flexibility help employees switch off from work? flexible working-time arrangements and cognitive work-to-home spillover for women and men in germany, *Social Indicators Research* **151**(2), 471–494.

Mas, A. and Pallais, A.: 2017, Valuing alternative work arrangements, *American Economic Review* **107**(12), 3722–3759.

McCarthy, A., Cleveland, J. N., Hunter, S., Darcy, C. and Grady, G.: 2013, Employee work–life balance outcomes in ireland: a multilevel investigation of supervisory support and perceived organizational support, *The International Journal of Human Resource Management* **24**(6), 1257–1276.

Mellner, C., Aronsson, G. and Kecklund, G.: 2014, Boundary management preferences, boundary control, and work-life balance among full-time employed professionals in knowledge-intensive, flexible work, *Nordic journal of working life studies* **4**(4).

Molnar, C.: 2022, *Interpretable Machine Learning*, 2 edn.

Mueller, M.: 2019, Night and organization studies, *Organization Studies* **40**(10), 1513–1527.

Munsch, C. L.: 2016, Flexible work, flexible penalties: The effect of gender, childcare, and type of request on the flexibility bias, *Social Forces* **94**(4), 1567–1591.

Myrie, J. and Daly, K.: 2009, The use of boundaries by self-employed, home-based workers to manage work and family: A qualitative study in canada, *Journal of Family and Economic Issues* **30**, 386–398.

Netemeyer, R. G., Boles, J. S. and McMurrian, R.: 1996, Development and validation of work–family conflict and family–work conflict scales., *Journal of Applied Psychology* **81**(4), 400–410.

Nippert-Eng, C.: 1996, Calendars and keys: The classification of "home" and "work", *Sociological Forum* **11**(3), 563–582.

Nippert-Eng, C. E.: 2008, *Home and work*, University of Chicago Press.

Oaxaca, R.: 1973, Male-female wage differentials in urban labor markets, *International Economic Review* **14**(3), 693.

O'Connor, L. T. and Cech, E. A.: 2018, Not just a mothers' problem: The consequences of perceived workplace flexibility bias for all workers, *Sociological Perspectives* **61**(5), 808–829.

Piasna, A. and Drahokoupil, J.: 2021, Flexibility unbound: understanding the heterogeneity of preferences among food delivery platform workers, *Socio-Economic Review* **19**(4), 1397–1419.

Rudman, L. A. and Mescher, K.: 2013, Penalizing men who request a family leave: Is flexibility stigma a femininity stigma?, *Journal of Social Issues* **69**(2), 322–340.

Schieman, S. and Glavin, P.: 2016, The pressure-status nexus and blurred work–family boundaries, *Work and Occupations* **43**(1), 3–37.

Secunda, P. M.: 2019, The employee right to disconnect, *Notre Dame Journal of International Comparative Law* **9**, 1.

Torquati, L., Mielke, G. I., Brown, W. J., Burton, N. W. and Kolbe-Alexander, T. L.: 2019, Shift work and poor mental health: a meta-analysis of longitudinal studies, *American Journal of Public Health* **109**(11), e13–e20.

Vandello, J. A., Hettinger, V. E., Bosson, J. K. and Siddiqi, J.: 2013, When equal isn't really equal: The masculine dilemma of seeking work flexibility, *Journal of Social Issues* **69**(2), 303–321.

Westman, M. and Piotrkowski, C. S.: 1999, Introduction to the special issue: Work–family research in occupational health psychology., *Journal of Occupational Health Psychology* **4**(4), 301.

Wheatley, D.: 2017, Employee satisfaction and use of flexible working arrangements, *Work, Employment and Society* **31**(4), 567–585.

Williams, J. C., Blair-Loy, M. and Berdahl, J. L.: 2013, Cultural schemas, social class, and the flexibility stigma, *Journal of Social Issues* **69**, 209–234.

Wooden, M., Warren, D. and Drago, R.: 2009, Working time mismatch and subjective well-being, *British Journal of Industrial Relations* **47**(1), 147–179.

Wunder, C. and Heineck, G.: 2013, Working time preferences, hours mismatch and well-being of couples: Are there spillovers?, *Labour Economics* **24**, 244–252.

Zapf, I. and Weber, E.: 2017, The role of employer, job and employee characteristics for flexible working time: An empirical analysis of overtime work and flexible working hours' arrangements, *Technical report*, IAB-Discussion Paper.

Online Appendices

A Full list of variables and their transformations

(intended for online dissemination)

The sample currently covers 36 countries. The data for the Balkans, Norway and Switzerland as well as Turkey are omitted as some of the key questions were not asked in those countries. The total EWCS dataset from wave three onward includes nearly 180 thousand working age individuals. With few exceptions, samples within countries comprise roughly 1000 individuals (representative national samples, for Germany across all waves and in selected waves for the other countries the samples are several times larger). All our specifications account for sample weights and country-fixed effects to accommodate for varying sample sizes across countries.

We utilize the following variables from European Working Conditions Survey:

- self-employment: Q7 = 2 (these observations are excluded from the analysis)
- job satisfaction: Q88 (categorical and re-coded to a dummy, with levels "satisfied" and "very satisfied" grouped to the value of 1 and "dissatisfied" and "very dissatisfied" grouped to the value 0)
- women: Q2a
- age: Q2b (in levels for regressions and in age groups for ML models; the age groups are 19-30, 31-45 and 46-60)
- children and elderly in the household: Q3c (age of household members children, and the age of the youngest child in the household, eventually a dummy taking on the value of 1 if the youngest child in the household is younger than 7 years old)
- occupation: y15_ISCO_88_1 (ISCO 88 1 digit categorized to three levels: 1 for "legislators, senior officials and managers", "professionals", "technicians and associate professionals" and "armed forces", 2 for "clerks", "service workers and shop and market sales workers" and "skilled agricultural and fishery workers" and 3 for "craft and related trades workers", "plant and machine operators and assemblers" and "elementary occupations")
- industry: y15_nace_r1_lt_4 (NACE classification of economic activities categorized to four levels: 1 for groups "A-B - Agriculture, Hunting, Forestry and Fishing", 2 for groups "C to F - Industry", 3 for "G to K - Services - excluding public administration" and 4 for "L - Public administration and defense and M to Q - other services")
- single household: Q1 = 1 (dummy coded as 1 for households with 1 person, 0 otherwise)
- enough time to finish tasks: Q61g_lt =1
- working on Saturdays: Q37c (re-coded to a dummy variable taking value 1 if an individual reports that they work on Saturday at least twice a month, 0 otherwise)
- working on Sundays: Q37b (re-coded to a dummy variable taking value 1 if an individual reports that they work on Sunday at least twice a month, 0 otherwise)
- reporting long hours: Q37d (re-coded to a dummy variable taking value 1 if an individual reports that they *normally* work more than 10 hours a day at least twice a month, 0 otherwise)
- working during nights: Q37a (re-coded to a dummy variable taking value 1 if an individual reports that they *normally* work at night (between 10.00 pm and 05.00 am) at least twice a month, 0 otherwise)
- working part-time: Q2d = 1
- working in varying hours: Q39d = 1 (answer "No" to the question: "Do you work fixed starting and finishing times?")
- long commute: Q36 (re-coded to a dummy variable taking value 1 if an individual reports that they spend more than an hour travelling from home to work and back , 0 otherwise)
- working with hazardous conditions: Q29a, Q29b, Q29c, Q29d, Q29e, Q29f, Q29g, Q29h, Q29i, Q30a,

Q30b, Q30c, Q30e and Q31 (count variable with the number of hazardous conditions an individual is exposed to at work "All of the time", "Almost all of the time", "Around 3/4 of the time", "Around half of the time", "Around 1/4 of the time" or is required to wear personal protective equipment; hazardous conditions include: vibrations, noise, high temperatures, low temperatures, smoke, fumes, powder or dust, vapours such as solvents and thinners, chemical products or substances, tobacco smoke, infectious materials, such as waste, bodily fluids, laboratory materials; more than 5 hazardous conditions is classified as 5)

- hours fit: Q44 (answering a question on whether working hours fit in with family or social commitments outside work, categorical variable with four levels for ML models, for regression re-coded to a dummy variable with levels "Well" and "Very well" grouped to the value of 1 and "Not very well" and "Not at all well" grouped to the value 0)
- supportive colleagues: $y15\_q61a\_lt$ =1

In addition, we universally perform the following steps.

- The public services workers are excluded from the analysis (based on the variable y15_nace_r1_lt_4, i.e. NACE4 classification, we eliminate codes L and M-to-Q).
- The observations from Croatia, FYROM, Turkey, Norway, Albania, Kosovo, Montenegro, Switzerland and Serbia are dropped, as key questions were not asked in those samples.

The full replication package is available at [LINK HERE, ANONYMIZED FOR REFEREEING PURPOSES].

B Machine Learning algorithms

(intended for online dissemination)

The advantage of the Machine Learning algorithms over standard statistical procedures is their extraordinary predictive power [13]. The commonly outlined drawbacks of ML methods are related to their lack of interpretation (in terms of the quantitative impact of explanatory variables on the dependent variable) and a "black-box" nature. However, in recent years many methods of extraction of regression type results from ML algorithms have been developed, see Molnar (2022) for an overview.

Nevertheless, the ML algorithms seem to be a perfect tool for our analysis, where less emphasis is put on the specific effect of individual characteristics on job satisfaction and more on the correct prediction of job satisfaction in alternative realities (where workers with flexible working arrangements are assumed to work under regular conditions). After all, our main goal is to calculate counterfactual levels of job satisfaction and to analyze the mismatch between actual and predicted levels.

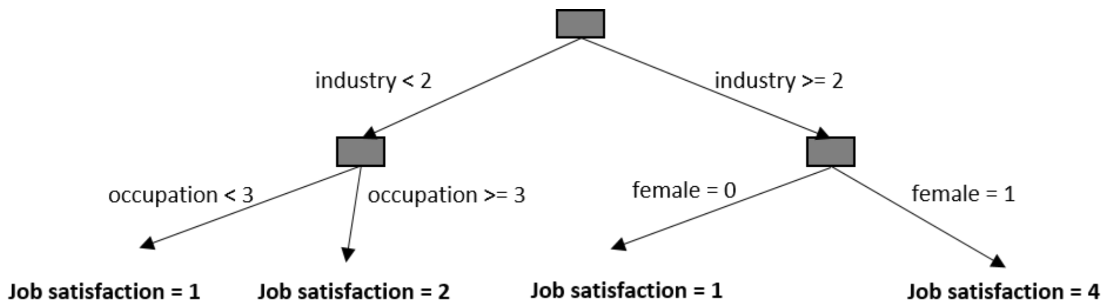
The state-of-the-art toolbox of ML algorithms commonly used in applied research offers plenty of methods appropriate for our purpose (supervised classification of workers into four job satisfaction levels based on their characteristics). We have experimented with more than 20 different ML algorithms. It turned out that Random Forest classifier outperforms other ML algorithms and provides the best prediction results in our data. The best algorithm has been selected based on the accuracy rate (i.e. number of cases correctly predicted), assuring the appropriate false positive and true negative rates.

Our model was trained in Python 3.8 (function $sklearn.ensemble.RandomForestClassifier$). The full replication package is available at [LINK HERE, ANONYMIZED FOR REFEREEING PURPOSES].

Random forest Random Forest algorithms are examples of so called ensemble methods of classification, which aggregate information from the lower level machine learning tool - a decision tree.

A decision tree is a model in a form of a tree, which splits the data top-down into smaller subsets, so that objects in these smaller groups are similar to each other and distinct from objects in other groups. An individual tree can be seen as a tool which predicts the value of a target variable (here job satisfaction) based on several input variables (here individual's and job's characteristics). At each step, the algorithm chooses an explanatory variable that best (according to Gini impurity measure, but in general yielding the largest information gain for categorical outcome) splits the data into separate groups. The trees are appropriately pruned by selecting the optimal parameter values for the trees' complexity and size. Figure B.1 visualizes an example simple decision tree. In this example, the decision rule or the model's prediction claims that e.g. "all female workers from industry category 2 or larger are predicted to have a job satisfaction level equal to 4".

Figure B.1: A decision tree visualization



[13] Deep Learning ML methods cannot be applied in our case due to insufficient number of data points.

The Random Forest algorithm consists of simultaneously building a large number of individual decision trees on various sub-samples of the dataset, while to improve the predictive accuracy, the final prediction is based on the outcome selected by most trees.

Specifically, we construct a random forest with 500 trees on bootstrapped sub-samples of the original data. We set the minimum number of samples essential to split a node to 2, and the minimum number of samples to be at a leaf to 1. The algorithm adjusts for sample weights. To make sure that unique outcomes get higher weights in the classification problem, we adjust the weights to be inversely proportional to frequencies in the data (standard approach suggested in Chen and Breiman 2004).

C Internal and external validity of ML model

(intended for online dissemination)

C.1 Internal validity

Table C.1 reports the actual and forecasted levels of job satisfaction in the predictive model for the reference group ($NWA = 0$).

We compare the results obtained by ML algorithm with standard parametric ordered probit regression model. In order to provide a point prediction from ordered probit, for each individual we calculate four distinct probabilities, i.e. estimated probability that his/her job satisfaction equals each level. The predicted job satisfaction level is the one with highest estimated probability.

Even though both methods provide similar accuracy (i.e. percentage of cases correctly predicted - slightly above 60%) it is clear that ML outperforms parametric model, which rarely predicts dissatisfaction.

Table C.1: Internal validity of the ML model

	ML - Random Forest Predicted JS				**Parametric (Ordered probit)** Predicted JS			
Actual JS	**1** Cell %	**2** Cell %	**3** Cell %	**4** Cell %	**1** Cell %	**2** Cell %	**3** Cell %	**4** Cell %
very satisfied (**1**)	**15.6**	3.9	1.9	1.2	**2.3**	20.7	0.0	0.0
satisfied (**2**)	11.1	**35.2**	10.2	6.1	1.8	**58.7**	0.2	0.0
dissatisfied (**3**)	1.2	1.4	**8.3**	1.3	0.1	13.0	**0.2**	0.0
very dissatisfied (**4**)	0.1	0.1	0.2	**2.3**	0.0	2.6	0.2	**0.0**
N	8 265	11 108	6 073	2 932	1 218	26 965	187	8

Notes: Table provides the proportion of workers correctly assigned (bold on the main diagonal) and mis-classified in both models. **N** denotes the number of observations in a given sub-sample.

The key disadvantage of using ML algorithms is that, in the case of singletons, the model cannot learn the patterns of job satisfaction. Singletons occur in small samples if a given composition of characteristics and outcomes exists solely in either the training set or in the testing set, but not in both. Since the main interest in our paper lies in the counterfactual level of job satisfaction, we abandon the split to training and testing sets and put all the available observations in the training set.

C.2 External validity #1

The consequence of putting all observations in the training set is that we cannot obtain out-of-sample predictions to study the quality of the model: we can do in-sample comparison of model with the data, but we cannot do out-of-sample comparisons.

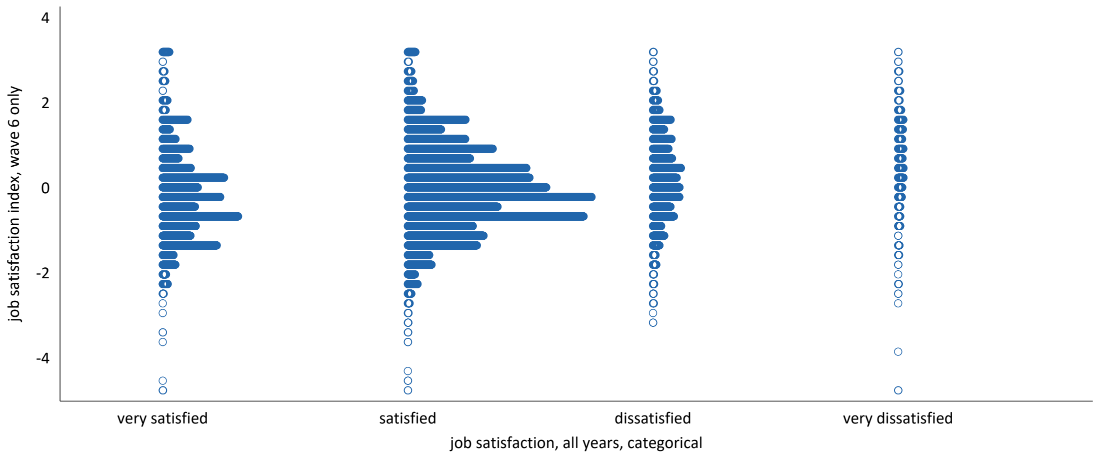
This shortcoming can be addressed using the data from the sixth wave of EWCS. This wave included a battery of questions concerning the feelings about one's job and working place. The questions fall under an umbrella *How often you feel this way?* and they include:

- At my work I feel full of wave energy
- I am enthusiastic about my job
- Time flies when I am working
- I feel exhausted at the end of the work day
- I doubt the importance of my work

These questions are answered on a five-levels Likert scale ranging from *always* through *most of the time*, *sometimes*, up to *rarely* and *never*. We revert the scales for the questions with negative statements (exhaustion and sense of work). We utilize these six variables to construct an index of job satisfaction using factors (JS_I). We use the factor analysis and establish that two factors are sufficient to construct an index: the first two factors have eigenvalues in excess of one, whereas the eigenvalue for the third factor is approximately 0.3. The Cronbach's α obtained with these six variables amounts to 0.81. We obtain as prediction after a principal component analysis.

Figure C.1 juxtaposes the index measure and self-reported job satisfaction (JS). The two measures of job satisfaction have a correlation of 0.16 ($p - value = 0.00$). This level of correlation may be indicative that not all of the items from the EWCS questionnaire are actually taken into consideration by the workers when they report their job satisfaction directly.

Figure C.1: Job satisfaction in wave 6 of EWCS: comparing categorical and index measure



Notes The graph is based on EWCS data from waves 6. Job satisfaction index (JS_I) is the factor-based measure utilizing five dedicated questions addressing various aspects of job satisfaction. Job satisfaction (JS) is a categorical self-reported variable.

We provide the correlation between the model JS and the actual JS_I, which is an adequate external validity test given that no variable used to construct JS_I was used in the modeling of JS. We estimate:

$$JS_I = \alpha_c + \alpha JS + \epsilon,$$

where the parameters α_c denote country fixed effects and α denotes the slope between categorical value of job satisfaction JS and the index JS_I constructed based on the additional variables available only in the sixth wave of EWCS. We estimate this model for actual job satisfaction for $x_i \in \mathbb{X}_0$ as well as predicted job satisfaction for $x_i \in \mathbb{X}_0$.

We report partial correlation between the original JS and JS_I in the first row of Table C.2. The correlation between the composite index and the self-reported level of job satisfaction is 0.21, highly statistically significant. The second column of Table C.2 reports the correlation coefficient for the JS implied by our machine learning model. When compared to the direct measure of raw JS, our predicted JS from the model performs well. The correlation coefficient is 0.15 and the confidence intervals between the coefficients in the first row and in the second row overlap. We infer that, in statistical terms, these estimates are not different from each other.

In parallel to the estimated correlations, the measures of R^2 are also similar between the model with raw JS and the model with predicted JS. Note that the dependent variable and the sample size are identical between rows, hence the total sum of squares is the same. While the comparison of R^2 in this context is meaningful, it also reveals that the self-reported levels of JS have low ability to explain the variation in the composite index of job satisfaction (JS_I). In summary, we establish that the correlations are remarkably

Table C.2: External validity of the model

JS_I	**Actual JS**	**Predicted JS**
α	0.206***	0.154***
	(0.0266)	(0.0148)
R^2	0.046	0.049
Observations	7 584	7 584

Notes: The model includes country-fixed effects. Data for JS_I is available only in the sixth wave of EWCS. Hence, the differences in sample sizes compared to Table C.1. Constant estimated, but not reported. Standard errors clustered at the level of country reported in parentheses, *** denotes significance at $p < 0.01$.

similar across both the original and the implied measure of JS. As a byproduct, we also show that self-reported levels of job satisfaction are not very consistent with a composite index intended to capture the same phenomenon.

C.3 External validity #2

The sixth wave of EWCS also allows comparing our measure of improvement to variables that look directly into the issue of managing boundaries, in a way similar to standard psychometric surveys (Hayman 2005, Netemeyer et al. 1996). The sixth wave included five questions concerning the work-family and family-work conflict. The questions inquired on how often a person felt that a given situation occurred. The list of questions includes:

- kept worrying about work when you were not working
- felt too tired after work to do some of the household jobs which need to be done
- found that your job prevented you from giving the time you wanted to your family
- found it difficult to concentrate on your job because of your family responsibilities
- found that your family responsibilities prevented you from giving the time you should to your job

Possible answers included "Always" , "Most of the time," "Sometimes," "Rarely", and "Never". Table C.3 presents the proportion of people who experienced a given situation "Always" or "Most of the time," thus we focus on cases where conflict is more acute. We expect that the proportion of people experiencing conflict is higher among those workers for whom abolishing NWA improves their job satisfaction.

Table C.3: External validity of the model: difficulties in managing boundaries

Removing NWA improves life satisfaction:	Yes	No	Diff.	Obs.
Situation occurs "Most of time" or "Always"				
Worry about work when not working	0.21	0.14	0.07***	7547
Felt too tired after work to do household work	0.31	0.24	0.07***	7547
Job took time from family	0.20	0.15	0.06***	7493
Cannot focus at work due to family	0.05	0.03	0.01***	7547
Family responsibilities took time from work	0.04	0.03	0.01***	7462

Notes: Data for conflict variables are available only in the sixth wave of EWCS. The sample includes only individuals who are currently working in NWA, and who do not have missing answers for conflict situations. *** denotes significance at $p < 0.01$.

All variables show a similar study. Conflict between family and work domains, and possible transgressions, are greater among those workers whom our algorithm classifies as not being well matched (for whom removal of NWA would increase job satisfaction). Moreover, differences are more pronounced in the case of work-family conflict. We notice, for example, that people who benefit from removing NWA are seven percentage points more likely to state that they worry about work when not working "most of the time" or "always". This difference represents a 50% increase with respect to workers who are matched with their current agreement. By contrast, family-work conflict are reported much less frequently. Less than 5% of workers in NWA identified those instances. And even in this situation, people who are not match are likely to experience greater conflict.

D Viewing counterfactual JS through the lens of treatment effects

(intended for online dissemination)

The main issue of causal inference is the identification of the right counterfactuals, as a simple comparison of means is likely to be biased. The ML approach followed in our empirical section requires two assumptions to hold. In particular, we have a common support assumption, i.e. $0 \leq P(NWA|X = x) \leq 1$ for all $x \in X$. Denote $JS_1 = JS(NWA = 1)$ and $JS_0 = JS(NWA = 0)$. The second assumption is that of selection on observables $NWA \perp JS_1, JS_0|x$, where JS_1 and JS_0 represent the potential outcomes for individual i with the non-standard working arrangements ($NWA = 1$) and without them ($NWA = 0$), respectively. Note that these assumptions are similar to the standard treatment effects in propensity score matching.

The individual ability to manage boundaries is latent (unobservable). This puts restrictions on the treatment effects that can actually be estimated in an unbiased way. Notably, we can estimate the effect of taking away NWAs, but we cannot estimate the effect of giving NWAs. We demonstrate it with the following thought experiment. For brevity, assume that the ability to manage boundaries is the only unobserved variable (u). Following the theory, more capable workers can derive higher satisfaction from working in NWAs than workers who are not able to manage boundaries. We further assume that the ability to manage boundaries does not affect job satisfaction when $NWA = 0$, that is when the worker does not need to manage boundaries. In other words, JS_1 increases in u, but JS_0 does not. Then, the following holds:

$$\begin{aligned} E(JS_1|NWA = 1, x, u) \quad - \quad & E(JS_0|NWA = 0, x, u) = \\ & \underbrace{E(JS_1|NWA = 1, x, u) - E(JS_0|NWA = 1, x, u)}_{\text{treatment effect on treated}} + \\ & \underbrace{E(JS_0|NWA = 1, x, u) - E(JS_0|NWA = 0, x, u).}_{\text{selection bias}} \end{aligned}$$

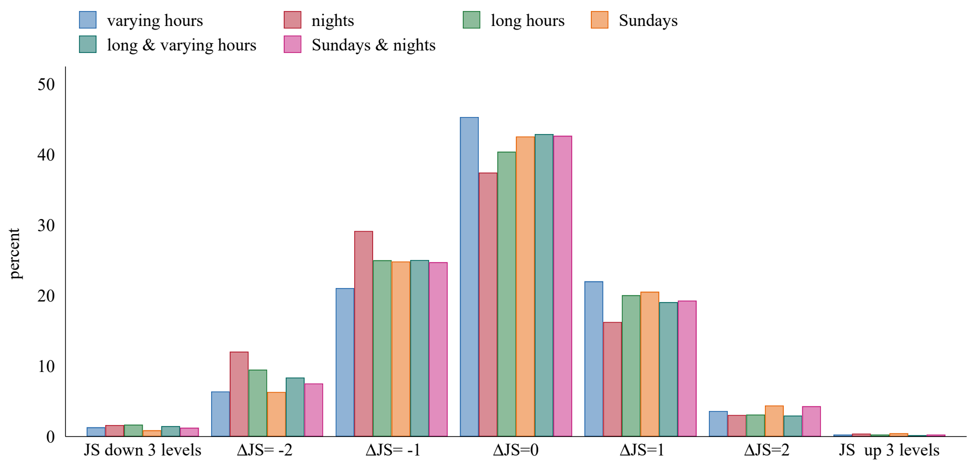
Consequently, our estimates have a causal interpretation (they are unbiased) as long as $E(JS_0|NWA = 1, x, u) = E(JS_0|NWA = 0, x, u)$. Given that job satisfaction is independent of ability to manage borders in standard employment (that is when one does not need to manage), then the assumption $E(JS_0|NWA, x, u) = E(JS_0|NWA, x)$ is plausible. Given the conditional independence assumption, we can write $E(JS_0|NWA, x) = E(JS_0|x)$. For our analysis, this means that it is possible to use observations from individuals not exposed to NWAs as counterfactuals for those workers under NWAs.

Notice that the converse need not be true. Following an analogous reasoning, computing treatment on the untreated requires that $E(JS_1|NWA = 1, x, u) = E(JS_1|NWA = 0, x, u)$. However, by definition $E(JS_1|u) \neq E(JS_1)$, and so the counterfactuals that we could potentially construct would be downward biased.

E Additional graphs and tables

(intended for online dissemination)

Figure E.1: Changes between factual and counterfactual job satisfaction across NWAs



Notes The figures present the distribution if ΔJS_i for $x_i \in \mathbb{X}_1$.

Figure E.2: Proportion of workers for whom taking away NWAs does not change JS

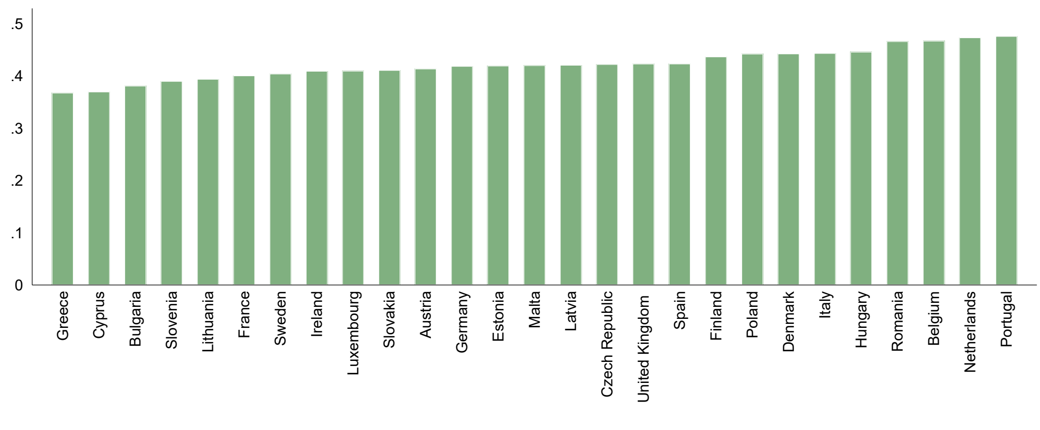
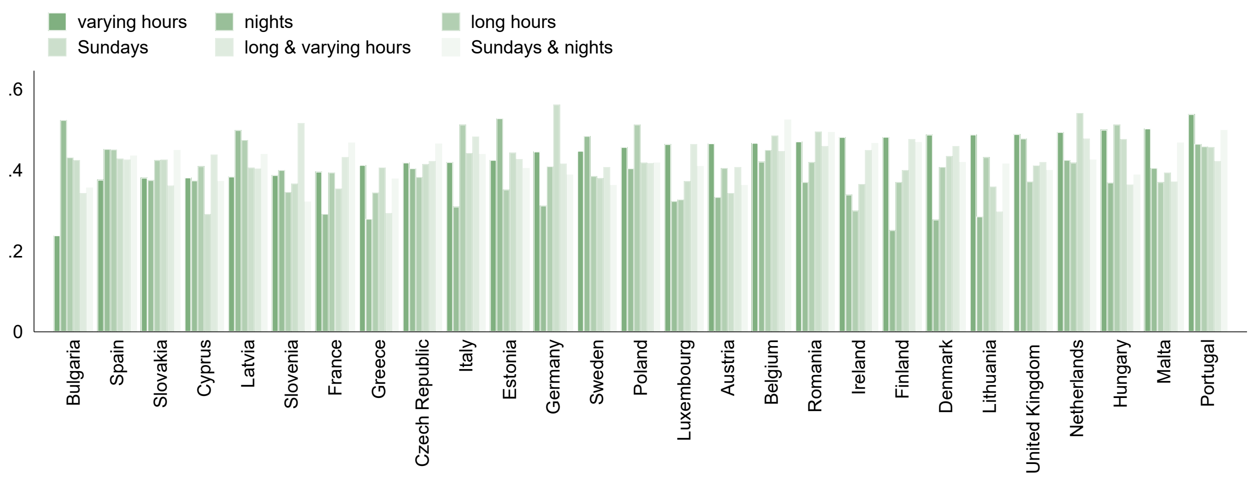
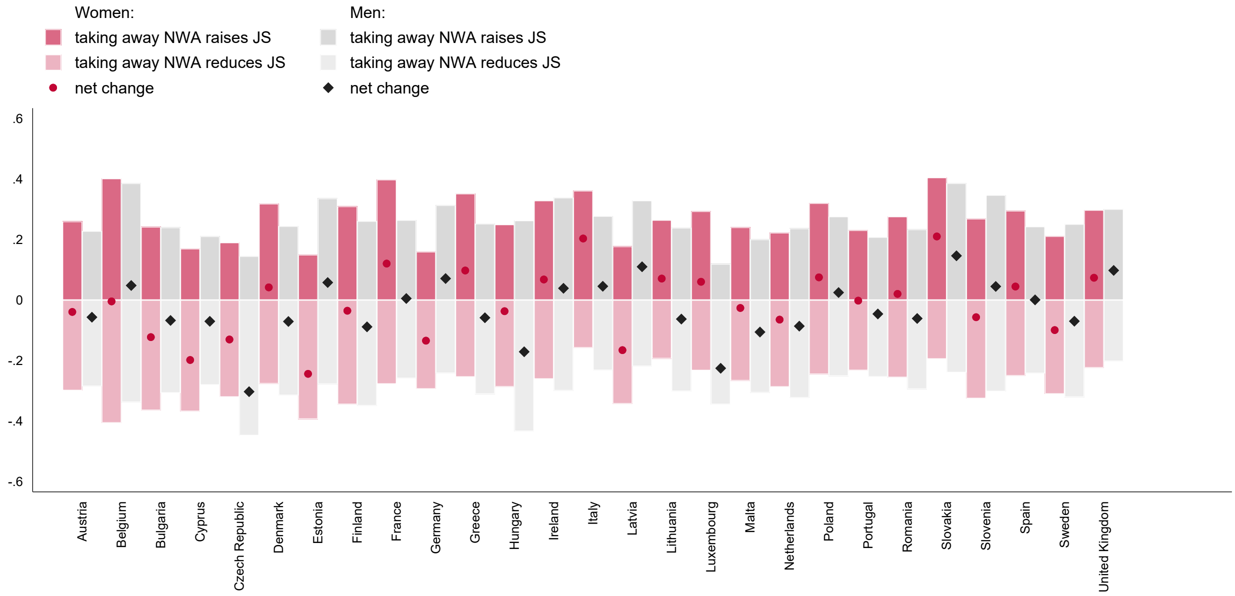


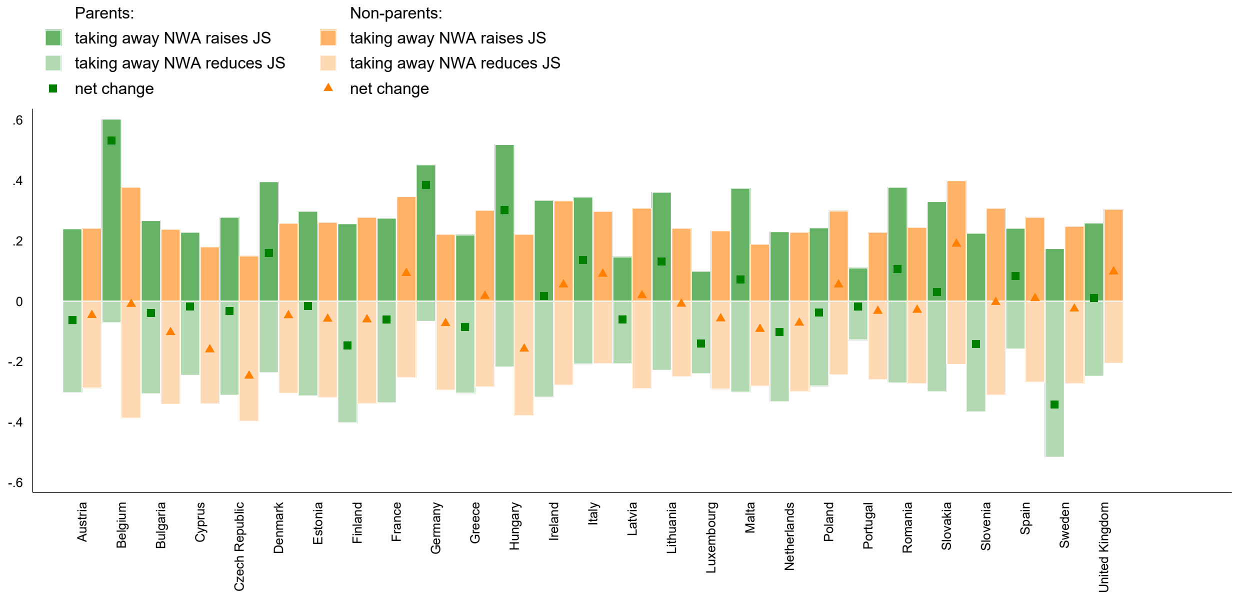
Figure E.3: Proportion of workers for whom taking away NWAs does not change JS across NWAs



Notes Taking away NWAs refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of individuals obtained with population weights.

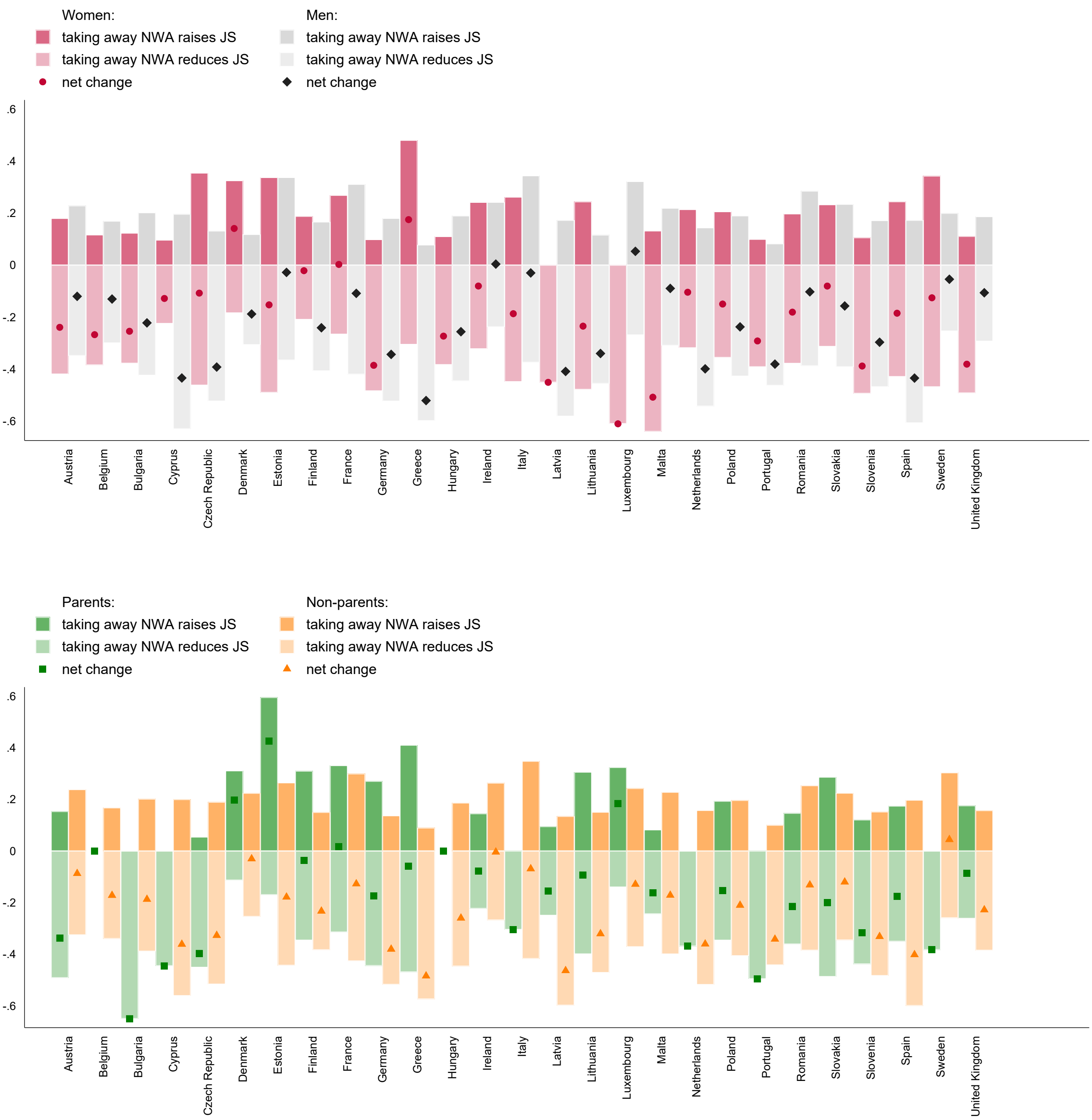
Figure E.4: Winners and losers from removing NWAs - varying hours





Notes Taking away NWA refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of mismatched and matched individuals obtained with population weights.

Figure E.5: Winners and losers from removing NWAs - nights

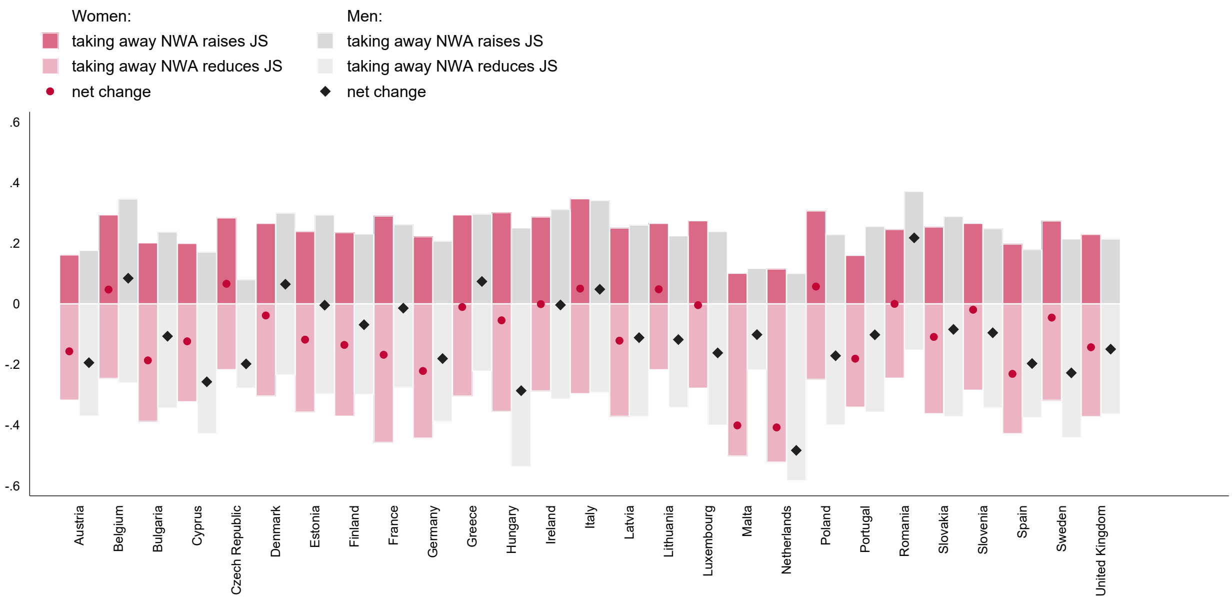


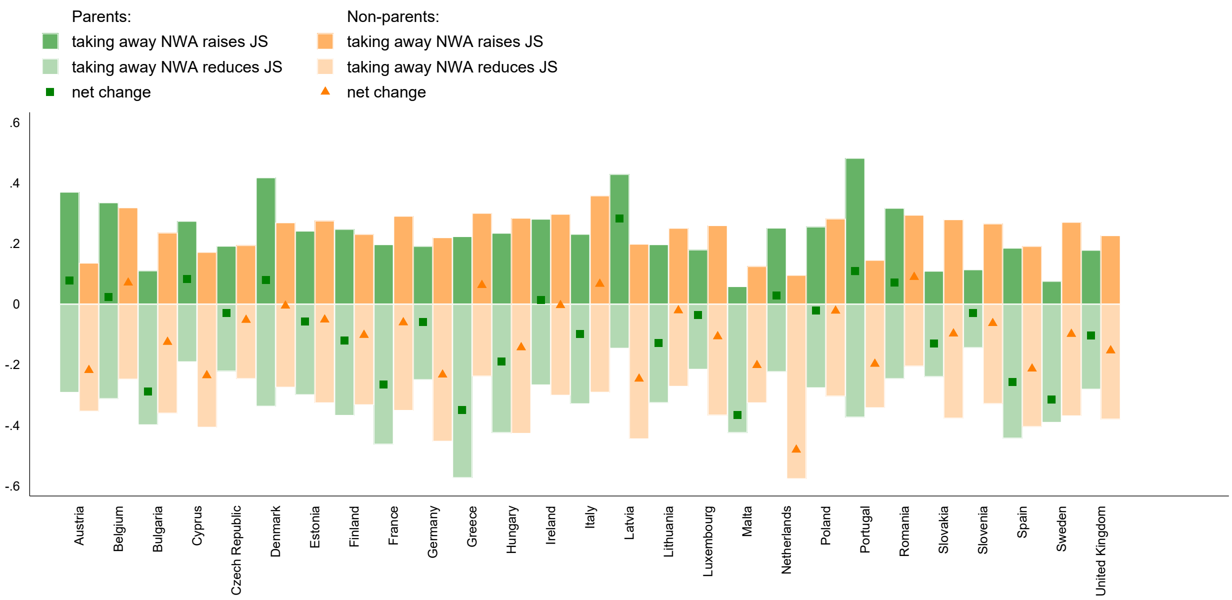
Notes Taking away NWA refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of mismatched and matched individuals obtained with population weights.

Figure E.6: Winners and losers from removing NWAs - long hours

Notes Taking away NWA refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of mismatched and matched individuals obtained with population weights.

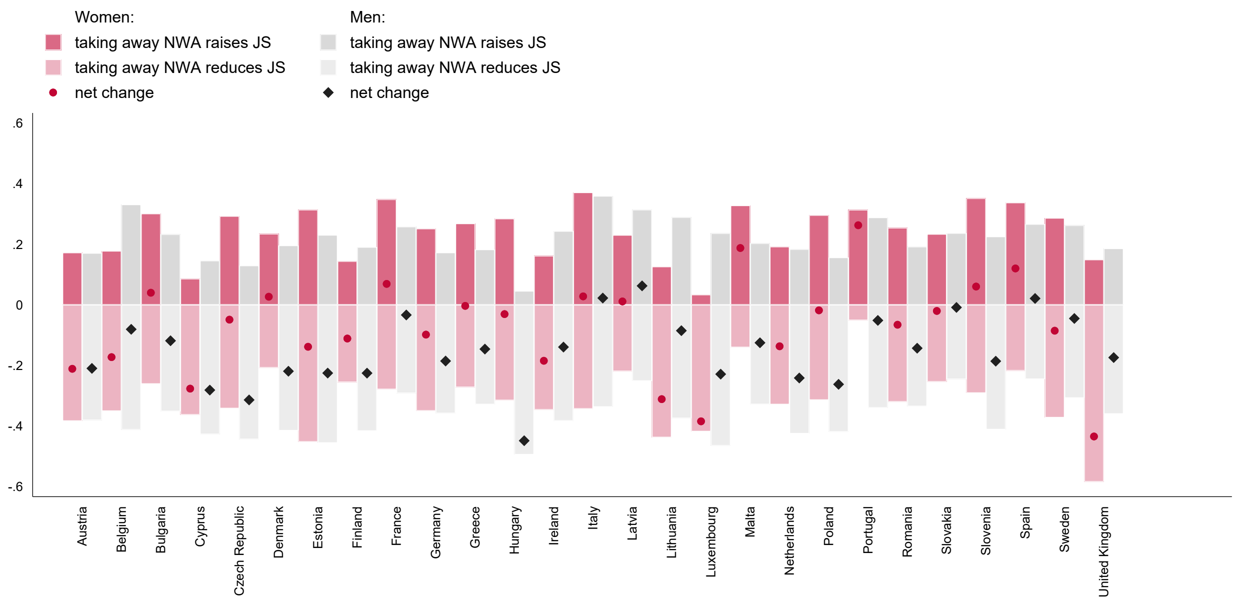
Figure E.7: Winners and losers from removing NWAs - Sundays

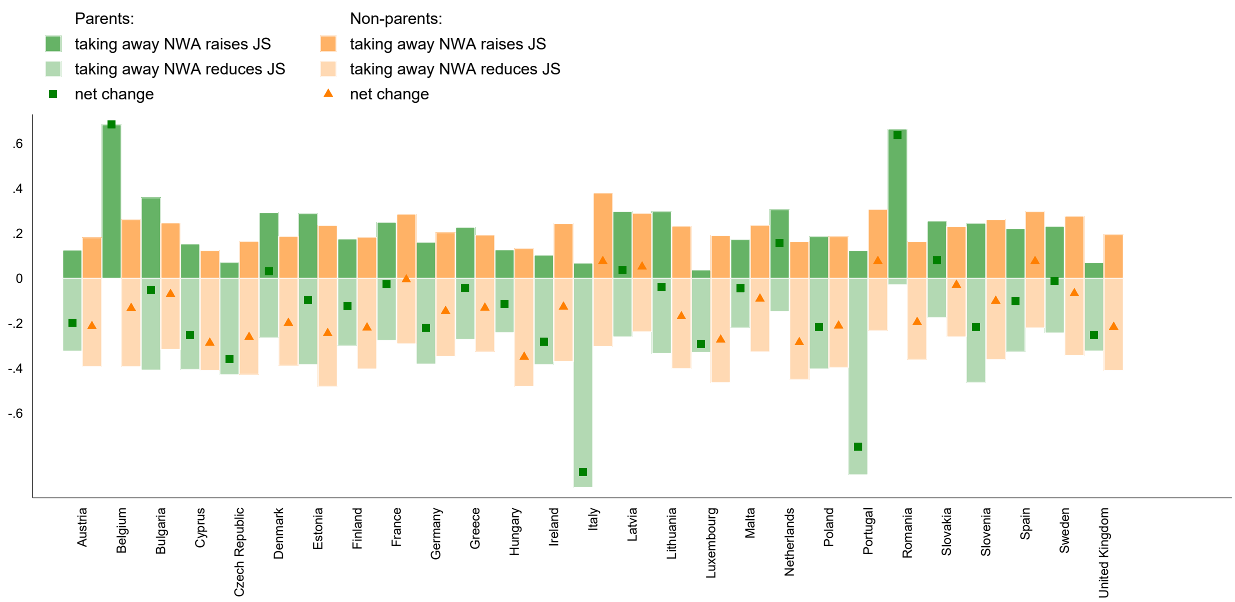




Notes Taking away NWA refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of mismatched and matched individuals obtained with population weights.

Figure E.8: Winners and losers from removing NWAs - long & varying hours



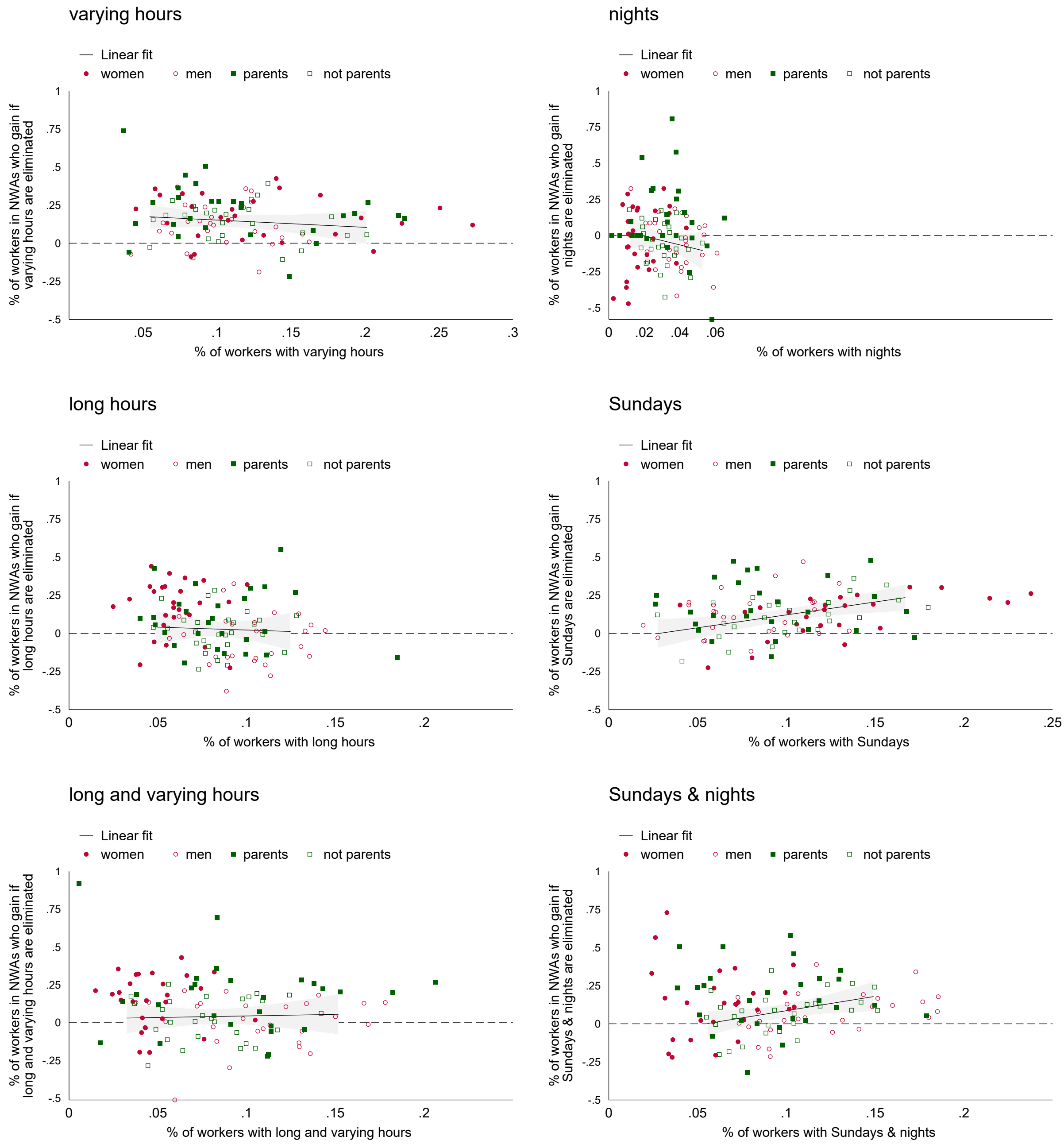


Notes Taking away NWA refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of mismatched and matched individuals obtained with population weights.

Figure E.9: Winners and losers from removing NWAs - Sundays & nights

Notes Taking away NWA refers to a counterfactual change in job satisfaction where a person is actually in a job with NWA and is assumed to change job features to fully regular working conditions. Shares of mismatched and matched individuals obtained with population weights.

Figure E.10: Prevalence of NWAs and % workers who would benefit from eliminating NWAs



Notes Optimal change in NWA involves assigning each individual an NWA such that his/her job satisfaction is maximized among options: actual JS and JS with optimal NWA. For illustrative purposes, we report shares of satisfied individuals (taking satisfied and very satisfied as one group). Shares of mismatched and matched individuals obtained with population weights. We show the results for men and women, as well as parents and non-parents. For convenience, we add a linear fit with confidence intervals for country-level aggregates in the background. We show the results for each of the NWAs separately and, for convenience, we keep the scales of the horizontal and vertical axes constant. The correlations are small and dispersion is large. After adjusting for country-fixed effects, only one slope proves to be statistically significant, all be it marginally (for working on Sundays, the effect of prevalence on net measure is statistically significant (for 1 percentage point higher prevalence, net measure is -1.27 percentage points lower, with a $p - value$ of 0.097). The results suggest that improving the match of workers to NWAs does not necessitate a change in NWA prevalence. A large part of job satisfaction gain could be achieved by workers reallocating themselves between jobs with and without non-standard arrangements. The persistence of mismatch across countries speaks of welfare costs stemming (partially) from labor market frictions preventing a better sorting of workers across working time arrangements.