# Discussion Paper Series

IZA DP No. 18432

March 2026

# Who Shirks at Work? An Application of Machine Learning to Time Use Data

**José Ignacio Giménez-Nadal**
IEDIS, University of Zaragoza

**José Alberto Molina**
IEDIS, University of Zaragoza and IZA@LISER

**Jorge Velilla**
IEDIS, University of Zaragoza



IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18432 March 2026

# Who Shirks at Work? An Application of Machine Learning to Time Use Data*

## Abstract

Worker productivity depends not only on hours worked, but also on how work time is actually used, and time-use evidence shows that non-work at work is non-trivial. This paper provides a data-driven characterization of shirking, and studies which observable characteristics best predict shirking behavior using American Time Use Survey data over 2003–2024. We implement a machine-learning forward selection procedure based on out-of-sample predictive performance. Our results suggest that shirking strongly depends on stochastic or unobserved factors, and that the determinants of the extensive and intensive margins are different. Moreover, the most informative predictors are predominantly job-related and time-allocation variables, whereas macro and labor-market indicators seem less relevant. This suggests that policies or managerial approaches to improve worker efficiency relying on observables face important limitations.

## JEL classification

J22, C53

## Keywords

shirking, non-work at work, ATUS data, prediction

## Corresponding author

Jorge Velilla
jvelilla@unizar.es

* *Funding*: This work was supported by the Government of Aragón [Project S32 23R]. This paper was partially written while José Alberto Molina was Visiting Scholar at the Department of Economics of Boston College (US), to which he would like to express his thanks for the hospitality and facilities provided.

Declaration of Competing Interest: None.

# 1 Introduction

Improving worker productivity remains a central concern in labor economics, as it shapes firm performance, wage determination, and aggregate economic growth. A growing body of time-use evidence shows that productivity depends not only on hours worked but also on how paid work time is actually allocated, and that non-work at work is quantitatively non-trivial (Hamermesh, 1990; Burda et al., 2016, 2020; Giménez-Nadal et al., 2018; Giménez-Nadal and Sevilla, 2022; Giménez-Nadal et al., 2023). This misallocation of work time is persistent, difficult to address, and largely unobserved in standard labor market data, limiting the effectiveness of policies and managerial practices aimed at improving efficiency. The literature has long emphasized the role of effort, discipline, and incentives in shaping productivity (Shapiro and Stiglitz, 1984; Lazear, 1999), but empirical evidence on the determinants of within-job time allocation remains scarce. By systematically characterizing shirking at work and identifying which observable characteristics predict it, this paper advances understanding of how productivity can be more accurately measured and improved.

Economic theory predicts that shirking (often referred to as loafing or non-work at work) arises endogenously from incentive constraints, monitoring costs, and equilibrium trade-offs between leisure and effort. Efficiency wage models and urban labor market frameworks imply that firms optimally tolerate some degree of non-work at work because eliminating it completely is costly or counterproductive (Shapiro and Stiglitz, 1984; Ross and Zenou, 2008; Zenou, 2009). These models also predict substitution between leisure at home and leisure at work, as well as heterogeneous and margin-specific responses of shirking to wages, commuting, and supervision (Ross and Zenou, 2008; Giménez-Nadal et al., 2018). Empirical studies provide support for these mechanisms, documenting negative relationships between leisure and shirking, positive relationships between commuting and shirking, and heterogeneity across occupations and monitoring regimes (Ross and Zenou, 2008; Giménez-Nadal et al., 2018; Vogelsang, 2024). This paper builds on these theoretical insights by assessing, in a data-driven way, which observable characteristics meaningfully predict shirking behavior.

Institutional features of labor markets shape shirking by influencing worker discipline, monitoring technologies, and outside options. Prior work shows that shirking varies with supervision intensity, occupation-specific monitoring technologies, and institutional arrangements such as unemployment insurance and wage-setting practices (Ross and Zenou, 2008; Burda et al., 2016, 2020; Belloc et al., 2025). More generous unemployment insurance and favorable labor market conditions have been linked to increased shirking in the extensive margin and altered incentives to supply effort (Burda et al., 2020). Differences between employees and the self-employed further reflect institutional variation in monitoring and incentive structures (Hamermesh, 1990). This institutional heterogeneity suggests that observables related to job design and labor market institutions may be relevant predictors of shirking, a hypothesis we evaluate directly in this paper.

We use the American Time Use Survey (ATUS) over the period 2003–2024, which provides nationally representative diary data on individuals' activities, locations, and timing throughout the day. Time-use diaries have been widely used to measure shirking at work by combining information on workplace location with non-work activities (Hamermesh, 1990; Burda et al., 2016; Giménez-Nadal et al., 2018; Giménez-Nadal and Sevilla, 2022). The ATUS allows direct observation of within-job time allocation rather than relying on self-reported effort or firm-level proxies.[1] The long time span further enables us to capture variation across business cycles, institutional regimes, and technological change, providing a comprehensive empirical foundation for analyzing the determinants of shirking.

Our empirical approach departs from the existing literature by implementing a machine-learning forward selection procedure based on out-of-sample predictive performance. Prior studies have primarily relied on reduced-form regressions with theory-driven covariate sets (Ross and Zenou, 2008; Burda et al., 2020; Giménez-Nadal et al., 2018; Hamermesh et al., 2021), leaving open the question of how well shirking can be predicted using observables more broadly. We estimate separate models for the extensive margin (whether a worker shirks) and the intensive margin (how much time is spent shirking conditional on doing so), consistent with theoretical and empirical evidence that these margins respond differently to incentives (Burda et al., 2020). This data-driven framework allows us to assess the practical limits of prediction and to evaluate the relative importance of different classes of observables without imposing strong functional form assumptions.

Our results show that shirking is only weakly predictable using observable characteristics, indicating that stochastic or unobserved factors play a dominant role in shaping within-work time allocation. While previous studies document statistically significant relationships between shirking and demographics, occupation, supervision, wages, and labor market conditions (Ross and Zenou, 2008; Burda et al., 2020; Hamermesh et al., 2021; Belloc et al., 2025), our predictive framework reveals that these variables contribute relatively little out-of-sample explanatory power. We also find that the determinants of the extensive and intensive margins differ markedly, consistent with the margin-specific responses documented by Burda et al. (2020). Moreover, job-related and time-allocation variables emerge as the most informative predictors, whereas macroeconomic and institutional indicators add little

[1] Alternative measures of productivity shirking behavior include worker absenteeism and reduced attendance (see, e.g., Van Ommeren and Gutiérrez-i Puigarnau, 2011; Brunello et al., 2025; Martín-Román et al., 2026).

predictive value.

Several potential threats to credibility warrant consideration, including measurement error in time-use diaries, misclassification of work versus non-work activities, and unobserved heterogeneity in job tasks or workplace norms. Prior research has shown that time-use diaries provide reliable and internally consistent measures of daily activity patterns and their use in measuring shirking is well established (Hamermesh, 1990; Burda et al., 2016; Giménez-Nadal et al., 2018). Our machine-learning framework further mitigates concerns about spurious correlations by evaluating models exclusively on out-of-sample performance. Although our analysis is not causal, the consistency of results across margins, time periods, and model specifications strengthens confidence in the conclusion that observables have limited predictive power for shirking.

This paper adds value by answering the research question of which observable characteristics predict shirking more comprehensively and rigorously than the existing literature. Prior work has documented correlates of shirking using theory-driven covariate sets and reduced-form regressions (Ross and Zenou, 2008; Burda et al., 2020; Hamermesh et al., 2021; Belloc et al., 2025), but has not systematically assessed the predictive content of observables nor made a clear distinction between extensive and intensive margins in a data-driven framework. This gap exists because traditional econometric approaches prioritize causal interpretation over predictive performance, and because large, high-dimensional time-use datasets have only recently become widely accessible. By leveraging two decades of ATUS data and machine-learning methods, we fill this gap and show that most variation in shirking remains unexplained by observables. In doing so, we clarify the structural limits of policies and managerial interventions based solely on observable worker and job characteristics, thereby advancing understanding of how productivity can be more effectively measured and improved.

The remainder of the paper is structured as follows. Section 2 reviews the literature on shirking using time use diaries and derives a series of testable hypotheses. Section 3 presents the ATUS data and describes the relevant variables, while Section 4 shows empirical strategy that identifies variables to model shirking time in terms of their predictive performance. Section 5 shows the main results. Finally, Section 6 concludes.

# 2 Literature review and hypotheses

On-the-job non-work, often referred to as shirking or loafing, is quantitatively non-trivial, and has been measured and analyzed in time use diaries, mostly in the US and the UK. For instance, time use diaries allow to measure shirking by combining the location of activities (e.g., at workplace) and the type of activity (non-work activities). Existing research has quantified shirking time using time use diaries, ranging from about 30 minutes per day (e.g., Burda et al., 2016, 2020; Giménez-Nadal et al., 2018) to more than 1 hour per day in the US and the UK in the recent years (Giménez-Nadal and Sevilla, 2022; Giménez-Nadal et al., 2023). Furthermore, Hamermesh (1990) concluded that, even though shirking may relate to decreased output per paid hour, eliminating it completely would be counterproductive for firms.

**Hypothesis 1: Existence.** *Shirking is quantitatively non-trivial, and it exists because eliminating job breaks and shirking is not optimal for firms.*

Moreover, existing research has also examined which profiles of workers are more related to shirking in the US. For instance, Hamermesh et al. (2021) find that there are racial and ethnic differences that, though quantitatively small, are statistically significant even controlling for observed heterogeneity (e.g., worker demographics, industry, occupation, geographic and urban controls, etc.). Similarly, Belloc et al. (2025) report gender differences in shirking in the US among workers in non-supervised occupations.

**Hypothesis 2: Gender and racial differences.** *There are gender (Belloc et al., 2025) and racial (Hamermesh et al., 2021) differences in shirking, which are only partially driven by occupational segregation (Hamermesh et al., 2021).*

Worker occupation and industry have been analyzed in relation to shirking, as the ability to shirk may differ depending on the monitoring and supervision technology across occupations and industries. For example, Ross and Zenou (2008) separately analyze workers in heavily supervised occupations and in slightly supervised occupations, as the cost of supervision is likely related to the incentives of workers to shirk. For example, in some occupations shirking is a large problem, and firms operate through supervision and monitoring, while in other occupations firms prefer to pay higher wages or provide alternative regards.[2] Ross and Zenou (2008) conclude that, in the US, shirking is more costly in occupations with high levels of supervision and among blue-collar workers than in slightly supervised occupations and among white collar workers. This result is in line with previous analyses focusing on occupation and industry unexplained differences in wages.[3]

Based on Ross and Zenou (2008), recent empirical work using the time use data in the US by Giménez-Nadal et al. (2018) found differences across occupations in shirking time. For instance, shirking ranked from about 36 minutes in occupations related to supervision and monitoring, to about 23 minutes in unsupervised occupations, although they also report

[2] We return to wages and shirking below.

[3] See, e.g., Krueger and Summers (1988), Gibbons and Katz (1992), Neal (1993) or Chen and Edin (2002).

large heterogeneity across occupations (e.g., more than 40 minutes of shirking per day in production occupations, but less than 20 in management and business occupations). Relatedly, Giménez-Nadal and Sevilla (2022) analyze changes in the structure of work among workers in routine task-intensive occupations due to technological change and automation. They find that technological change is linked to an increase of effort and a decrease of shirking time in the UK, which is especially relevant for routine task-intensive occupations.[4]

**Hypothesis 3: Monitoring.** *Occupations or industries with greater monitoring relate to decreased shirking.*

**Hypothesis 4: Routine task.** *Workers in more routine task intensive occupations exhibit lower frequency of shirking (Giménez-Nadal and Sevilla, 2022).*

In addition to demographic and occupation differences, existing research has also examined differences between employees and self-employed workers. For instance, self-employed workers are not supervised, so they lack monitoring mechanisms to prevent non-work at work. On the other hand, Hamermesh (1990) find that self-employed workers spend half as much time on non-work at work than employees, as they aim to maximize productivity and determine their shirking time accordingly.

**Hypothesis 5: Self-employment.** *Shirking is different among the self-employed: they are not monitored or supervised, but their non-work at work directly affects their output, and they typically organize their time better, which relates to decreased shirking.*

Shirking also plays a central role in urban efficiency wage models. These models assume a complementarity relationship between leisure at home and performance at work, which turns out to be a substitution relationship between leisure and shirking (Ross and Zenou, 2008; Zenou, 2009). For instance, Ross and Zenou (2008) and Giménez-Nadal et al. (2018) analyze the predictions of these models and conclude that there is a negative relationship between leisure time and shirking time, which is especially relevant among blue collar workers who are more likely to be monitored and supervised. In addition to these models, Vogelsang (2024) finds that leisure time influences effort at work and workplace time allocation, so that it may incentive productivity and relate to decreased shirking at work.

Another central element of these urban efficiency wage models is commuting. Ross and Zenou (2008) assume that commuting is a shock to worker time allocation that reduces the time available for leisure, ultimately increasing the incentives to shirk at work. Thus,

[4] A relatively recent strand of the literature has analyzed "cyberloafing", i.e., the use of Internet-enabled devices such as smartphones or computers for personal and/or non-work activities during work hours (see, e.g., Lim, 2002). For instance, Elciyar and Şimşek (2021) link cyberloafing and monitoring and find that the perception of penalties and benefits affects the propensity to do shirking and cyberloafing, while Zhang et al. (2025) report heterogeneity in the propensity to do cyberloafing in terms of worker behavior and monitoring.

they establish a positive relationship between commuting time on one hand, and shirking time on the other hand. This hypothesis of the model was empirically confirmed in the US by Giménez-Nadal et al. (2018). Another central element of these urban models is wages, as these models assume firms are willing to pay workers more than expected to promote efficiency, compensate the cost of longer commutes, and discourage shirking (Shapiro and Stiglitz, 1984; Ross and Zenou, 2008; Giménez-Nadal et al., 2018).

**Hypothesis 6: Leisure.** *There is a substitution relationship between leisure and shirking, i.e., increased leisure relates to decreased shirking.*

**Hypothesis 7: Wages.** *Wages relate to decreased shirking as firms pay higher wages to promote efficiency.*

**Hypothesis 8: Commuting.** *Commuting is a shock to time allocation that increases the incentives to shirk, so longer commutes relate to increased shirking.*

Shirking time has also been linked to the timing of work. For instance, (Brodsky and Amabile, 2018) analyze "idle" time at work, finding that workers slow down and work with less intensity if they expect idle time at work, likely early in the morning or late in the evening. Giménez-Nadal and Sevilla (2022) also studied the timing of non-work at the workplace in the UK, reporting that about 20% of workers were doing some non-work while at the workplace at 10am, though up to 39% were doing leisure while at the workplace between 12:30pm and 1pm. However, a significant part of this time was due to meals at work, which is arguably different to shirking.

**Hypothesis 9: Timing.** *Shirking time is concentrated at specific work hours with low demand or when idle time is anticipated. Working early in the morning or late in the evening relates to increased shirking.*

The literature has also established some links between worker shirking behavior on one hand, and the macroeconomic context, labor market conditions, and institutions on the other hand. For instance, Burda et al. (2016, 2020) and Belloc et al. (2025) argue that a better macroeconomic context relates to increased shirking, as workers face a lower risk and cost of job loss and have more bargaining power, which reduces the incentive to supply effort, making shirking at work more likely. Specifically, Burda et al. (2020) document a cyclical pattern whereby higher local unemployment relates to a lower probability of engaging in shirking, but also to a higher amount of shirking among those who shirk. This indicates that, when unemployment is high, some workers avoid the risky behavior of engaging in shirking while at work. However, at the same time, this only leaves workers with a strong preference for shirking actually doing some shirking, thus increasing shirking in the intensive margin. In other words, those shirkers who would shirk only a little prefer not to shirk at all if local unemployment is high, i.e., if the cost of job loss is high. Moreover, Burda et al. (2020) also relate shirking and unemployment insurance (UI) generosity. They argue that more generous UI raises the outside option of workers and weakens discipline, increasing the likelihood that workers with a lower preference for shirking do some shirking at work.

**Hypothesis 10: Macro context.** *A favorable macroeconomic context relates to increased shirking.*

**Hypothesis 11: Unemployment discipline.** *Local unemployment relates to decreased shirking in the extensive margin, but to increased shirking among those who shirk.*

**Hypothesis 12: UI generosity.** *State unemployment generosity relates to increased shirking in the extensive margin, but to decreased shirking in the intensive margin.*

# 3 Data

## 3.1 The American Time Use Survey

We use data from the American Time Use Survey (ATUS) for the period 2003-2024. The ATUS is an ongoing survey on time use in the US. It started in 2003 and is conducted as part of the Current Population Survey (CPS) by the US Census Bureau and administered by the Bureau of Labor Statistics. It is often considered the official time use survey of the US (Flood et al., 2025). Moreover, it is part of the Integrated Public Use Microdata Series of the Institute for Social Research and Data Innovation at the University of Minnesota.[5] The ATUS is a general-purpose survey that provides information about the activities in which respondents engage based on time use diaries, together with questionnaires including information about the characteristics of those respondents.

The ATUS data comprise repeated cross-sections, and the samples are nationally representative, supported by the sample weights provided in the database (Flood et al., 2025). The respondents in the ATUS are selected randomly from households that participate in the CPS so that they represent the population of the US, and only one randomly selected person aged 15 or older in each household is asked to complete a time use diary during a so-called diary day. The diary days are distributed across the days of the week so that 50 percent correspond to weekdays (and then evenly distributed across weekdays), 25 percent to Saturdays, and 25 percent to Sundays, as the allocation of time is relatively similar across the five weekdays, but differs from that on weekends (Horrigan and Herz, 2004).

The ATUS data provide us with questionnaires about respondents, including several

[5] See https://www.atusdata.org/atus/.

socioeconomic variables, but also with information on individual time use based on diaries, where those respondents report their activities during the 24 hours of the day, from 4 am to 4 am of the next day. Several authors have reported the advantage of self-reported diary data over other types of surveys using stylized questionnaires (Bonke, 2005; Kan, 2008), such as more reliable and accurate estimates. Thus, time use surveys have become the gold standard in the study of worker daily behaviors (Aguiar and Hurst, 2007; Guryan et al., 2008; Harms et al., 2019).

## 3.2 Sample requirements

We use data from the ATUS for the years 2003 to 2024. The initial sample consists of 252,808 individuals, but we omit 4,081 individuals who filled in diaries during their holidays. In addition, to mitigate the impact of time-allocation choices across the life cycle, we drop 85,234 individuals younger than 21 or older than 60 years, consistent with Mazzocco (2007) and Giménez-Nadal and Sevilla (2012). Next, we retain employed individuals only, dropping 34,804 individuals who are unemployed or not in the labor force. We drop 26,209 individuals with censored information on work hours and income, and missing information on the key variables.

We exploit the information in diaries to define non-work days as those days in which respondents spend less than 60 minutes in paid work activities, consistent with Giménez-Nadal et al. (2018), and we omit 46,217 individuals who, satisfying the previous requirements, filled in the diaries during non-work days. This is important to avoid including in the sample employed individuals who did not work during the diary day, or who reported very short work spells that reflect marginal work activities (e.g., answering emails or phone calls) rather than regular work days. Finally, we drop 8,564 individuals who do not start or finish their diaries at 4 am with sleeping activity entries, to avoid including unusual work schedules in the sample.

These restrictions leave a sample of 47,699 employed individuals with complete information who worked during their respective diary days. Conditional on working on the diary day, we consider respondents interviewed on both weekdays and weekends, and 37,247 observations (i.e., 78.1 percent of the sample) correspond to weekdays, while the remaining 10,452 observations correspond to individuals who filled in the diaries on weekends. Excluding weekends would omit a non-negligible set of work days, and would bias the analysis toward standard Monday-to-Friday employment.

## 3.3 Variables

### Time use variables

We exploit the time use diaries to define several time use variables. First, the main variable in the analysis is the time spent shirking by workers, also referred to as "loafing" or "non-work at work" (Burda et al., 2016, 2020). To define shirking time, we identify all the time use episodes done at the workplace by workers, and measure the minutes spent doing non-work activities (excluding compulsory work breaks). Thus, our measure of shirking time comprises activities such as waiting, socializing, relaxing and thinking, eating and drinking, tobacco and drug use, reading for personal interest, watching television or movies, listening to the radio or music, or computer use for leisure, consistent with Giménez-Nadal et al. (2018).

Second, we define paid work time, measured in minutes, as the time spent working (in the main job, other jobs, or not elsewhere classified), in security or security procedures as part of the job, and in work-related activities. We also define the start time of the first episode of paid work, and the end time of the last episode of paid work, to account for the start and end of the work journey of workers, which may relate to shirking time due to accumulated fatigue or tiredness.

Third, we identify those episodes that correspond to work breaks while working that could be considered compulsory work breaks, and thus should not be classified as shirking, but may help workers disconnect and rest. These are activities such as waiting associated with working, leisure as part of the job, eating as part of the job, exercise as part of the job, and waiting as part of the job. As existing research has documented that the number of work breaks may also be relevant for explaining work performance (Giménez-Nadal et al., 2023), we also define the number of separate work breaks during the work journey.

We define other time use variables to analyze how they correlate with shirking time, all of them measured in daily minutes. These include commuting time, leisure time, childcare time, and unpaid work time. Leisure time is defined following Aguiar and Hurst (2007) and Giménez-Nadal and Sevilla (2012), and includes activities such as watching television, socializing, attending parties or social events, relaxing, or playing games, hobbies, and sports. Commuting includes travel related to working or work-related activities (Giménez-Nadal et al., 2018). Childcare time refers to primary childcare, i.e., those episodes that involve active childcare such as caring for, reading to, playing with, or helping/teaching children. Unpaid work then refers to episodes of chores (e.g., cleaning, washing, laundry, food preparation and serving), repairing or maintaining, or home security, organization, or management (Aguiar and Hurst, 2007; Giménez-Nadal and Sevilla, 2012).

Finally, the time use diaries allow us to identify the day on which the diary was filled in. This is relevant, as existing research has documented differences between weekdays and weekends. However, as weekdays often display similar work time use allocation (Horrigan and Herz, 2004), we define a dummy that takes the value 1 for weekdays, and the value 0 for weekends.

### Demographic, employment and other variables

In addition to the time use variables, the ATUS questionnaires allow us to define several interviewees' characteristics. We define a dummy that identifies males (value 1) and females (value 0) and workers' age (measured in years). We also consider workers' education level, and we define a factor variable that identifies workers with primary education only, secondary but not university education, or university education. The ATUS also includes information on workers' race and native status, and we recode them as factor variables that identify White workers, Black workers, and other races, and native workers, Hispanic origin workers, Asian origin workers, and other non-natives, respectively.

The ATUS also includes variables related to the household composition of respondents, and we define the number of family unit members, the marital status (a dummy that takes value 1 for those married or cohabiting with a partner, value 0 otherwise), the number of children, and the age of the youngest child. We also define the time spent in the presence of children.[6]

ATUS questionnaires also include a range of employment-related attributes. We define dummies that identify public sector workers and self-employed workers (thus the reference category is private sector employees). We also define a dummy that takes value 1 for full-time workers (value 0 for part-time workers), the number of usual weekly work hours, a dummy that takes value 1 for those who usually work overtime (0 otherwise), a dummy that takes value 1 for those who have multiple jobs (0 for those who have one job only), and a dummy that takes value 1 for those who are married and their partner is employed (0 otherwise). We also consider hourly wages, and factor variables for worker occupation, worker industry, and family income.[7]

[6]This contrasts with the time use variable measuring primary childcare. Time with children involves any activity (whether primary childcare or not) done in the presence of children, which captures a different dimension of childcare than primary childcare (Bosworth et al., 2025; Cuevas-Ruiz et al., 2025).

[7]Occupation in the ATUS represents the worker's specific technical function, while industry refers to the economic sector. Occupation is originally provided as the four-digit Census occupation codes, and then it is collapsed in the ATUS into 22 broader occupations, see https://www.atusdata.org/atus/four_digit_to_two_digit_occ.shtml. Similarly, industry is initially provided as the four-digit Census industry codes, and then it is collapsed in the ATUS into 52 categories, see https://www.atusdata.org/atus/four_digit_to_two_digit_ind.shtml. Family income captures all sources of labor and non-labor income of household members, and is defined in the following brackets: less than \$5,000; \$5,000–\$7,499; \$7,500–\$9,999; \$10,000–\$12,499; \$12,500–\$14,999; \$15,000–\$19,999; \$20,000–\$24,999; \$25,000–\$29,999; \$30,000–

Finally, we consider some urban and regional variables included in the ATUS questionnaires. We define a dummy variable that takes value 1 for homeowners (0 for renters). We also consider factor variables that capture the State of residence of respondents, the region of residence (i.e., Northeast, Midwest, South and West), and the Metropolitan Statistical Area (hereafter, MSA) of residence. We also use information on the metropolitan status, defined as a dummy that takes value 1 for those residing in a metropolitan area (either central city or balance of MSA), and value 0 for those in a non-metropolitan area; and the MSA size, which includes the following categories: 100,000–249,999; 250,000–499,999; 500,000–999,999; 1,000,000–2,499,999; 2,500,000–4,999,999; and 5,000,000 or more.

**Ancillary data**

We use some information not included in the ATUS, which we define at the State×year level, to analyze some of the hypotheses proposed in Section 2. These include state GDP, unemployment rate, unemployment insurance, and minimum wage. State×year GDP is taken from the US Bureau of Economic Analysis Regional Data, and is defined as the real GDP (millions of chained 2017 dollars). State×year unemployment insurance is defined using data from the CPS, and measures the average income from unemployment benefits reported by unemployed respondents. State×year unemployment rates are also defined using data from the CPS. Finally, State×year minimum wage is taken from the Federal Reserve Economic Data (Bank of St. Louis).

## 3.4 Descriptive statistics

Table 1 reports summary statistics for the main shirking variables, separately for female and male workers. On average, women report 24.7 minutes of shirking per day, while men report 28.3 minutes, with the gender difference being statistically significant at standard levels. The share of workers who report positive shirking time is 0.57 among women and 0.61 among men, implying a difference of 0.04 which is also highly significant. Moreover, conditional on being a shirker, average shirking time amounts to 43.5 minutes for women and 46.7 minutes for men, whereas the average duration of a shirking episode is 15.7 minutes for women and 17.2 minutes for men, while the average number of shirking episodes per day is 0.99 for women and 1.11 for men.

Figure 1 plots the density of shirking time for male and female workers who report positive shirking time. The distribution is similar for men and women, although the density for men

$34,999; $35,000–$39,999; $40,000–$49,999; $50,000–$59,999; $60,000–$74,999; $75,000–$99,999; $100,000–$149,999; and $150,000 and over.

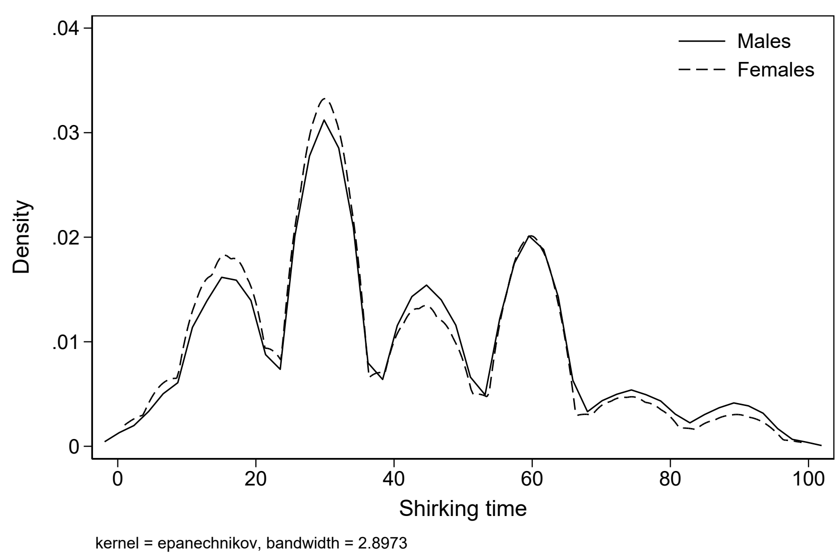
Table 1: Main summary statistics

| | Female | | Male | | |
|---|---|---|---|---|---|
| VARIABLES | Mean | St.Dev. | Mean | St.Dev. | Diff |
| Shirking time | 24.716 | 33.304 | 28.349 | 35.590 | 3.633*** |
| Being a shirker | 0.568 | 0.495 | 0.607 | 0.488 | 0.039*** |
| Shirking time (shirkers) | 43.511 | 33.687 | 46.729 | 35.057 | 3.218*** |
| Shirking episode duration | 15.682 | 20.549 | 17.163 | 21.090 | 1.481*** |
| # shirking episodes | 0.988 | 1.173 | 1.114 | 1.241 | 0.126*** |
| | 24,035 | | 23,664 | | |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Shirking time is measured in minutes per day. *** significant at the 1%, ** significant at the 5%, * significant at the 10%. Additional descriptives shown in Table A.1.

is slightly shifted to the right. Overall, the figure indicates that the distribution of shirking time is not smooth, but exhibits several local modes at specific values, for both men and women, e.g., at 15 minutes per day, at 30 minutes per day, at 45 minutes per day, and at 60 minutes per day. This suggests that shirking time tends to concentrate at a small number of commonly observed durations, indicating that many workers accumulate similar amounts of non-work time at the workplace over the day.

Figure 1: Density of shirking time




*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day and are shirkers (i.e., report positive shirking time). Shirking time is measured in minutes per day.

Figure 2 shows the evolution over the analyzed period of three outcomes: the average shirking time, the average daily shirking time among workers who report positive shirking, and the share of workers who report positive shirking. Average shirking time, including both workers who shirk and those who do not shirk, shows a decline over the analyzed period. However, once we separately analyze workers who report positive shirking, and the rate of shirkers, we observe that this decline is due to a sharp reduction from about 60% to about 50% in the rate of shirkers over the period, i.e., the reduction in shirking time corresponds to the extensive margin (i.e., to the decision to be a shirker). Conversely, once we focus on the intensive margin and analyze shirking times only for those workers who report positive shirking, we observe that shirking time shows a U-shape trend over the analyzed period, fluctuating around the averages reported in Table 1 for both males and females.

# 4 Empirical strategy

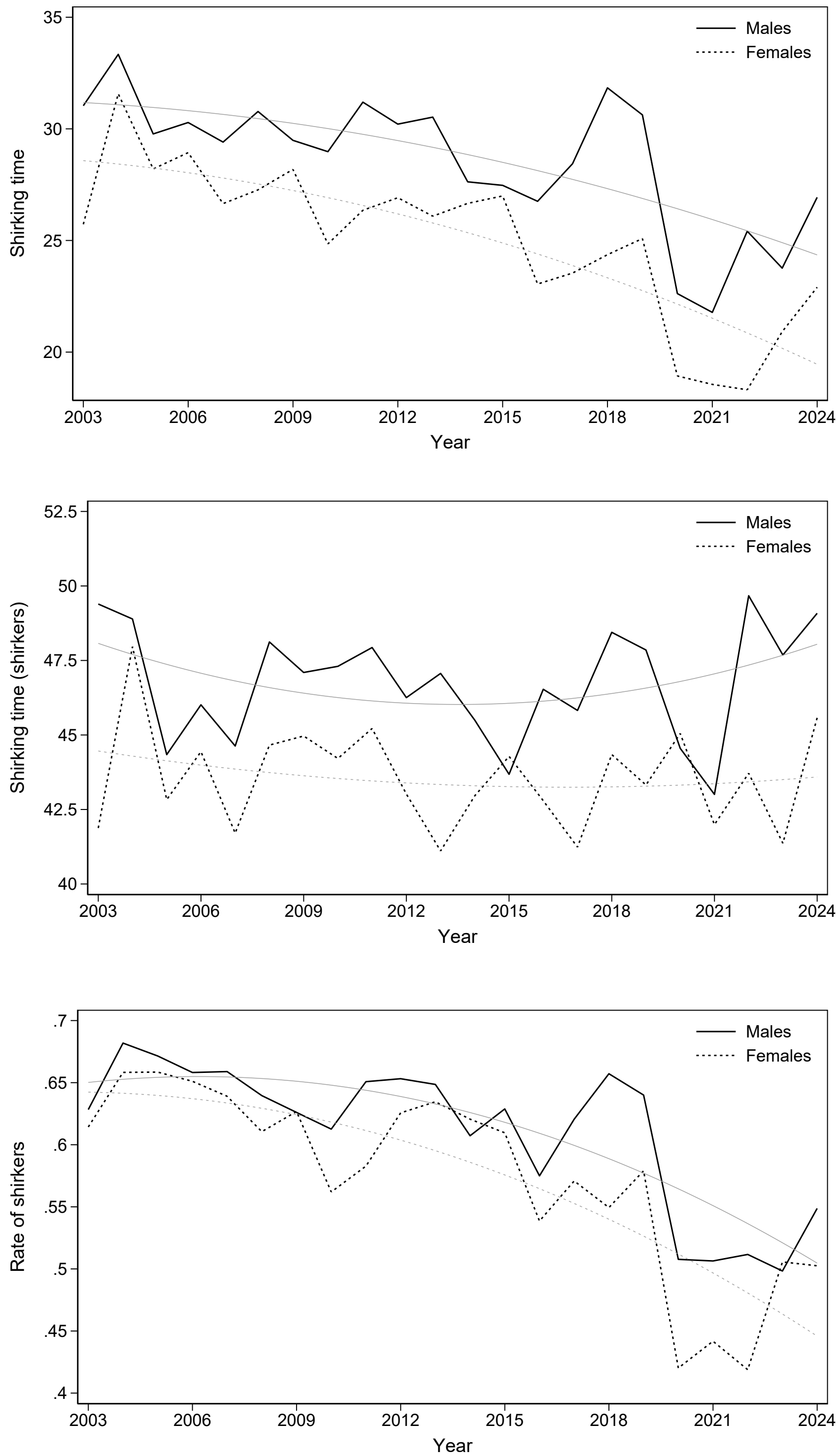
## 4.1 Variable selection based on predictive performance

Our empirical strategy aims to identify, among a large set of candidate covariates, those explanatory variables that provide the highest predictive power for three related outcomes: (i) shirking time in the full sample of workers; (ii) shirking time conditional on being a shirker (i.e., conditional on positive shirking time), which captures the intensive margin of shirking behavior; and (iii) being a shirker (i.e., the probability of being a shirker, defined as a binary indicator), which captures the extensive margin of shirking behavior. Variable selection is conducted using a data-driven strategy based on out-of-sample predictive performance, rather than on in-sample goodness of fit or statistical significance (Giménez-Nadal et al., 2019). An outline of the procedure is provided in Table 2.

The set of candidate predictors is common across the three dependent variables, and includes calendar variables (year and month), demographic characteristics (gender, age, education, race, and native status), household characteristics (couple status, family size, number of children, age of the youngest child, and time spent with children), employment characteristics (public sector worker, self-employed worker, full-time status, usual weekly working hours, overtime work, hourly wage, family income, partner's employment status, multiple job holding, occupation, and industry), housing and geographic characteristics (homeownership, state of residence, region of residence, metropolitan status, MSA of residence, and MSA size), time-use variables (paid work time, start of the working day, end of the working day, commuting time, leisure time, childcare time, unpaid work time, compulsory work breaks time, and number of work breaks), and local labor market conditions measured at the State×year level (GDP, unemployment insurance benefits, unemployment rate, and minimum wage). All the categorical variables are included as factor variables.

Variables are selected so as to minimize out-of-sample prediction error. For the contin-

Figure 2: Evolution of shirking over time




*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Shirking time is measured in minutes per day.

Table 2: Outline of the variable selection procedure

| Step | Procedure |
|---|---|
| 1 | Select the outcome variable $Y$ (shirking time, shirking time (shirkers), or being a shirker) and the full set of candidate predictors $\mathcal{X}$. |
| 2 | Initialize the selected set of predictors $\mathcal{S} = \emptyset$. |
| 3 | For each candidate variable $X_j \in \mathcal{X} \setminus \mathcal{S}$:<br>$i$) For iteration $k = 1, \ldots, 1000$:<br>- draw a bootstrap sample of individuals with replacement to form the training set;<br>- define the test set as individuals not selected in the training set;<br>- estimate a linear model of $Y$ on $\mathcal{S} \cup \{X_j\}$ using the training set;<br>- predict $\tilde{Y}$ in the corresponding test set;<br>- compute the absolute prediction error $\frac{1}{n_k} \sum_{i \in \text{test}_k} \lvert Y_i - \tilde{Y}_i \rvert$, with $n_k$ size of test set.<br>$ii$) Compute the mean absolute prediction error across the 1000 bootstrap iterations. |
| 4 | Select the variable $X_{j^*}$ that minimizes the mean absolute prediction error. |
| 5 | Update $\mathcal{S} = \mathcal{S} \cup \{X_{j^*}\}$. |
| 6 | Repeat Steps 3, 4, and 5 until all variables are ranked and $\mathcal{S} = \mathcal{X}$. |
| 7 | Store the ordered sequence of selected variables and the associated prediction errors. |

uous outcomes (i.e., shirking time, and shirking time among shirkers), this criterion is the mean absolute prediction error, while for the binary outcome it is the mean classification error, or misclassification rate. Out-of-sample performance is evaluated using a bootstrap procedure at the individual level with 1000 replications per candidate variable. For each iteration, a bootstrap sample of individuals is drawn with replacement to form the training set, and individuals not selected in that draw constitute the corresponding test set. For each candidate model, we estimate a linear regression for the continuous outcomes or a linear probability model for the binary outcome, using sampling weights and robust standard errors. Predictions are then computed for the test sample, and absolute prediction errors (or classification errors) are calculated. The mean absolute error or mean classification error associated with a given set of regressors is defined as the average prediction error across the 1000 bootstrap replications.

Variable selection follows a sequential forward selection algorithm as in Giménez-Nadal et al. (2019). The procedure starts from an empty model. At each step, for every candidate explanatory variable not yet selected, a model that includes all previously selected variables plus the candidate variable is estimated, and its mean absolute error or mean classification error is computed using the procedure described above. The variable that yields the largest reduction in the mean absolute error or mean classification error is then selected and added permanently to the model. This process is repeated until all candidate explanatory variables are ranked. The outcome of the procedure is therefore an ordered list of variables according to their contribution to predictive performance, together with the associated prediction error at each step.

## 4.2 Advantages of the prediction-based selection approach

Our variable-selection strategy differs from standard forward selection procedures commonly used in applied work, which rely on in-sample statistical significance, changes in the $R^2$, or changes in the adjusted $R^2$. These criteria are informative about goodness of fit *within* the estimation sample, but they are not designed to assess how well a model generalizes to new data. As a result, they may favor specifications that overfit sample-specific noise and deliver poor predictive performance out of sample (Hastie et al., 2009; James et al., 2013).

By contrast, our approach is inspired by machine learning techniques and explicitly targets out-of-sample accuracy by selecting variables that minimize prediction error on observations not used for estimation. The distinction between explanatory fit and predictive performance has been emphasized in the statistical literature, which shows that models optimized for in-sample fit need not be optimal for prediction (Breiman, 2001b; Shmueli, 2010). In particular, forward selection based on sequential hypothesis testing is known to suffer from instability and from biased inference due to repeated testing, while selection based on the adjusted $R^2$ remains sensitive to overfitting when the set of potential regressors is large relative to sample size (Harrell, 2015).

Using resampling methods to evaluate out-of-sample performance provides a direct assessment of a model's ability to generalize beyond the estimation sample. Cross-validation and bootstrap-based approaches have long been advocated as principled tools for model comparison when the objective is prediction rather than parameter estimation (Stone, 1974; Efron and Tibshirani, 1994, 1997; Hastie et al., 2009). In this context, minimizing the mean absolute error or the misclassification rate corresponds to choosing the specification that is expected to perform best when applied to new observations drawn from the same population.

This focus on predictive performance is particularly relevant in our context, where the objective is not to estimate the causal effect of a specific regressor, but rather to characterize which observable dimensions of workers' characteristics, job attributes, time use behaviors, and local labor-market conditions are most informative about shirking behavior. When the set of potential covariates is large and covers a range of different variables, standard approaches based on statistical significance can lead to models that are sensitive to small changes in the sample, and heavily dependent on arbitrary threshold choices. Conversely, a prediction-based approach provides a framework to assess the relative empirical relevance of different variables, which is especially useful in a setting where theory offers limited guidance.

An additional advantage of the procedure is that it produces a complete ranking of covariates according to their marginal contribution to predictive performance. This ranking is not tied to arbitrary significance thresholds and is robust to the inclusion of large sets of correlated regressors. As a result, the selected ordering summarizes the relative importance of different dimensions of individual characteristics, job attributes, time use patterns, and local labor-market conditions for predicting shirking behavior, while maintaining a clear interpretation of the underlying statistical model.

From an intuitive perspective, our approach shares some similarities with tree-based methods such as random forests (Breiman, 2001a). In both cases, model evaluation is based on repeated resampling of the data and on the assessment of out-of-sample predictive performance rather than on in-sample fit. Moreover, as in random forests, the relative importance of covariates is inferred from their contribution to reducing prediction error across multiple resamples, which allows variables to be ranked according to their predictive content. The main difference is that our procedure operates within a linear regression framework and relies on a sequential forward-selection algorithm, rather than on non-parametric trees and aggregation across a large number of randomized models. As a result, our strategy preserves the interpretability of standard econometric models while adopting a prediction-oriented criterion similar in spirit to that used in modern machine-learning methods.

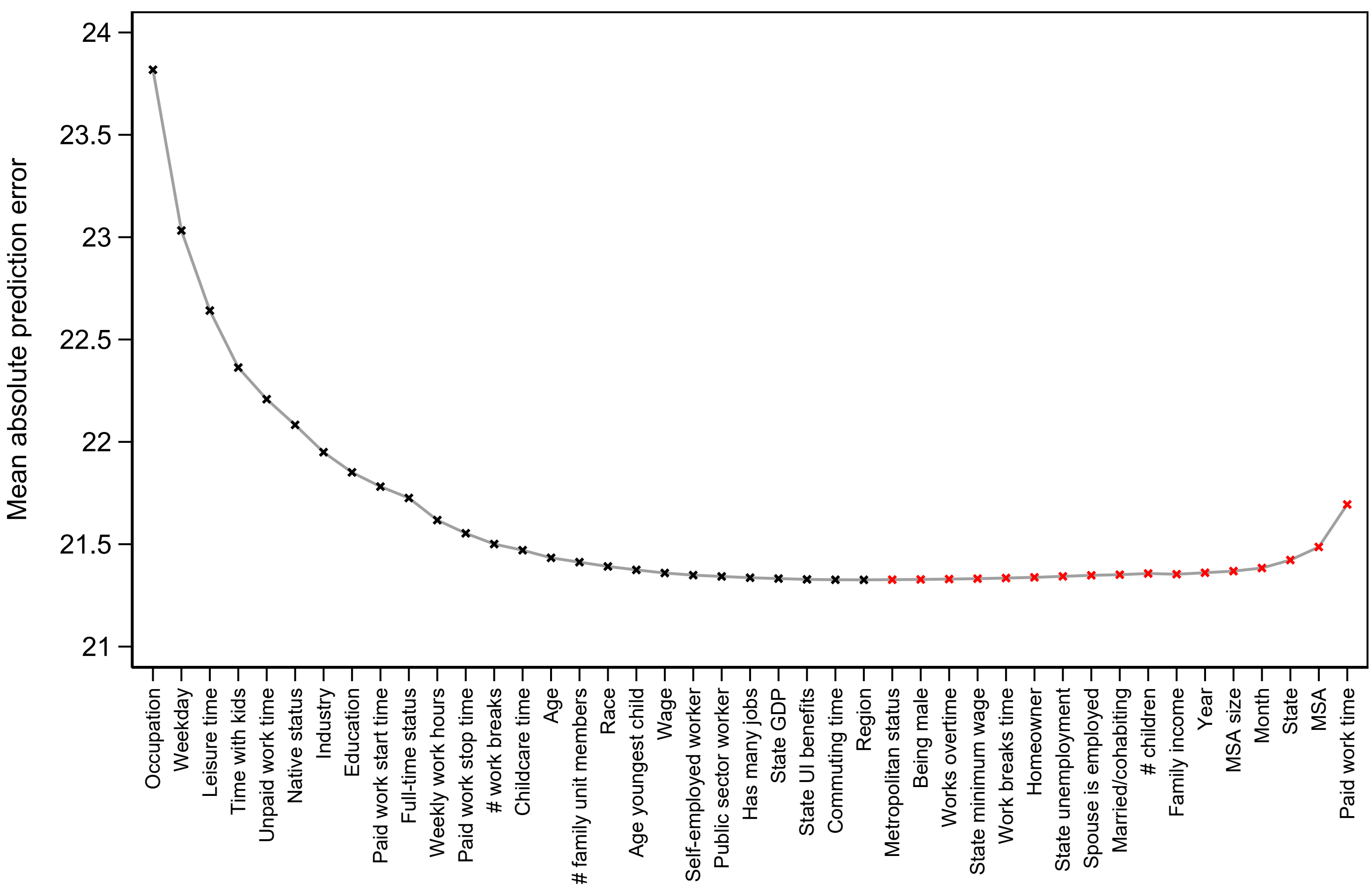
A related alternative is the Least Absolute Shrinkage and Selection Operator (Lasso), introduced by Tibshirani (1996), which performs variable selection using a penalty term that shrinks coefficients, setting some of them exactly to zero. Relative to our approach, the Lasso can be advantageous when the set of candidate predictors is very large and highly collinear. However, the Lasso also has disadvantages, as the model depends on the choice of the penalty parameter, which requires cross-validation, and estimated coefficients are not directly comparable to unpenalized estimates. Moreover, when the goal is interpretation and inference, the Lasso requires debiasing corrections to avoid misleading inference (Belloni et al., 2014; Chernozhukov et al., 2018). In contrast, our selection strategy is explicitly organized around out-of-sample predictive accuracy and yields a transparent ranking of covariates, which facilitates interpretability and keeps the subsequent estimation stage closer to standard econometric practice.

# 5 Results

## 5.1 Variable selection

Figures 3, 4 and 5 summarize our variable-selection results based on out-of-sample prediction performance. Figure 3 plots the mean absolute prediction error (in minutes) for shirking time for the overall sample, including individuals who report positive shirking as well as individuals who report zero shirking. Figure 4 plots analogously but for workers who report positive shirking time only. Figure 5 reports the analogous selection path for the binary outcome "being a shirker".[8]

Figure 3: Mean absolute prediction errors: shirking time



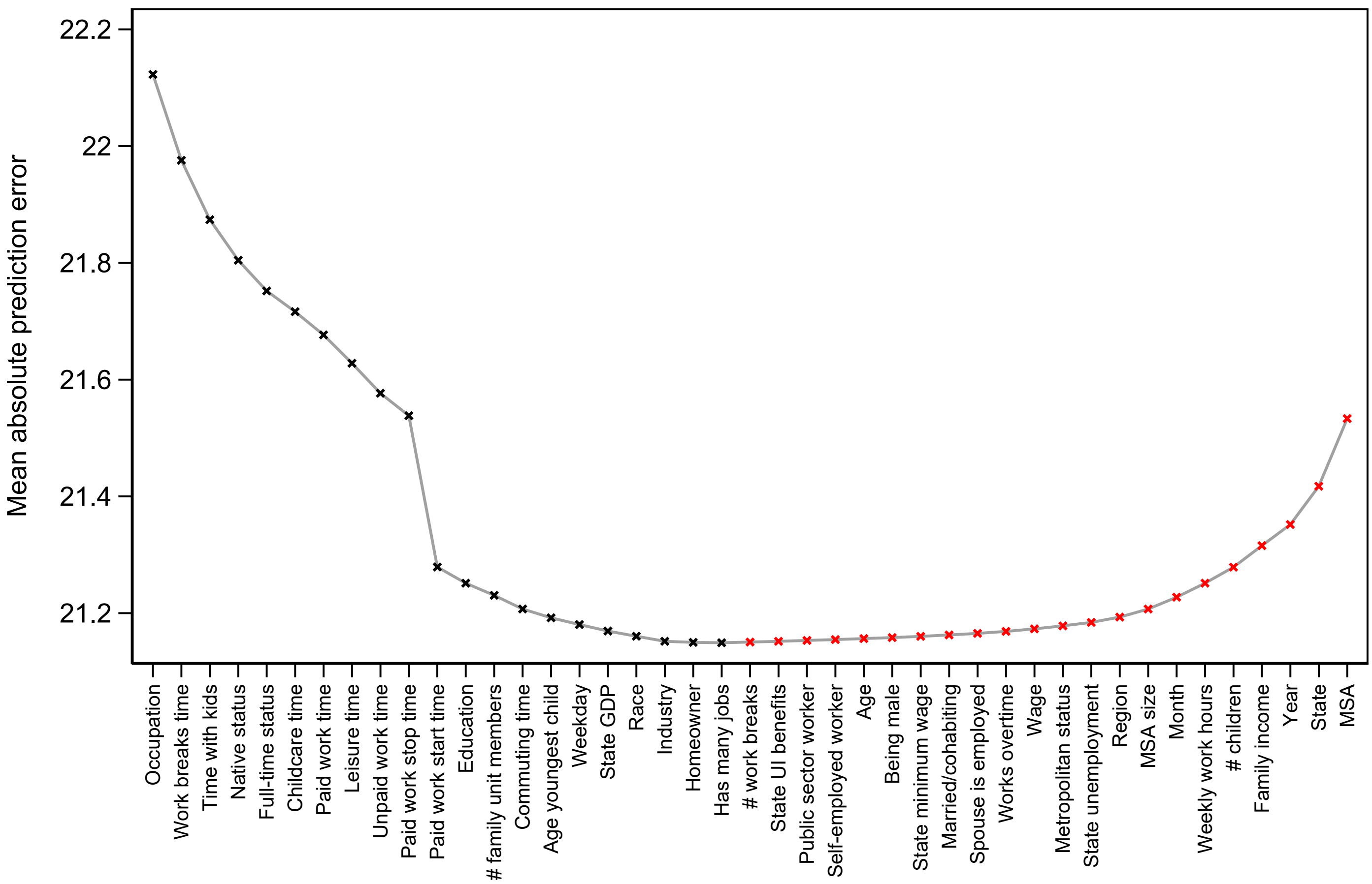

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. The dependent variable is shirking time, measured in minutes per day.

Figure 3 shows that as regressors are added sequentially, following the ordering induced

[8]Details are shown in Appendix Table A.2.

by our procedure, the mean absolute prediction error falls sharply in the first steps, indicating that a small set of variables delivers most of the predictive gains. In particular, the initial reductions are associated with adding worker occupation, weekday, leisure time, time with kids, unpaid work time, worker native status, worker industry, education, and the start time of paid work. Afterwards, improvements in predictive performance become progressively smaller. The curve reaches a flat pattern at around 20 explanatory variables, and reaches its minimum at the 26th regressor (region of residence). Then, it slightly increases as additional controls are included (red markers), suggesting diminishing returns and overfitting beyond the selected specification.

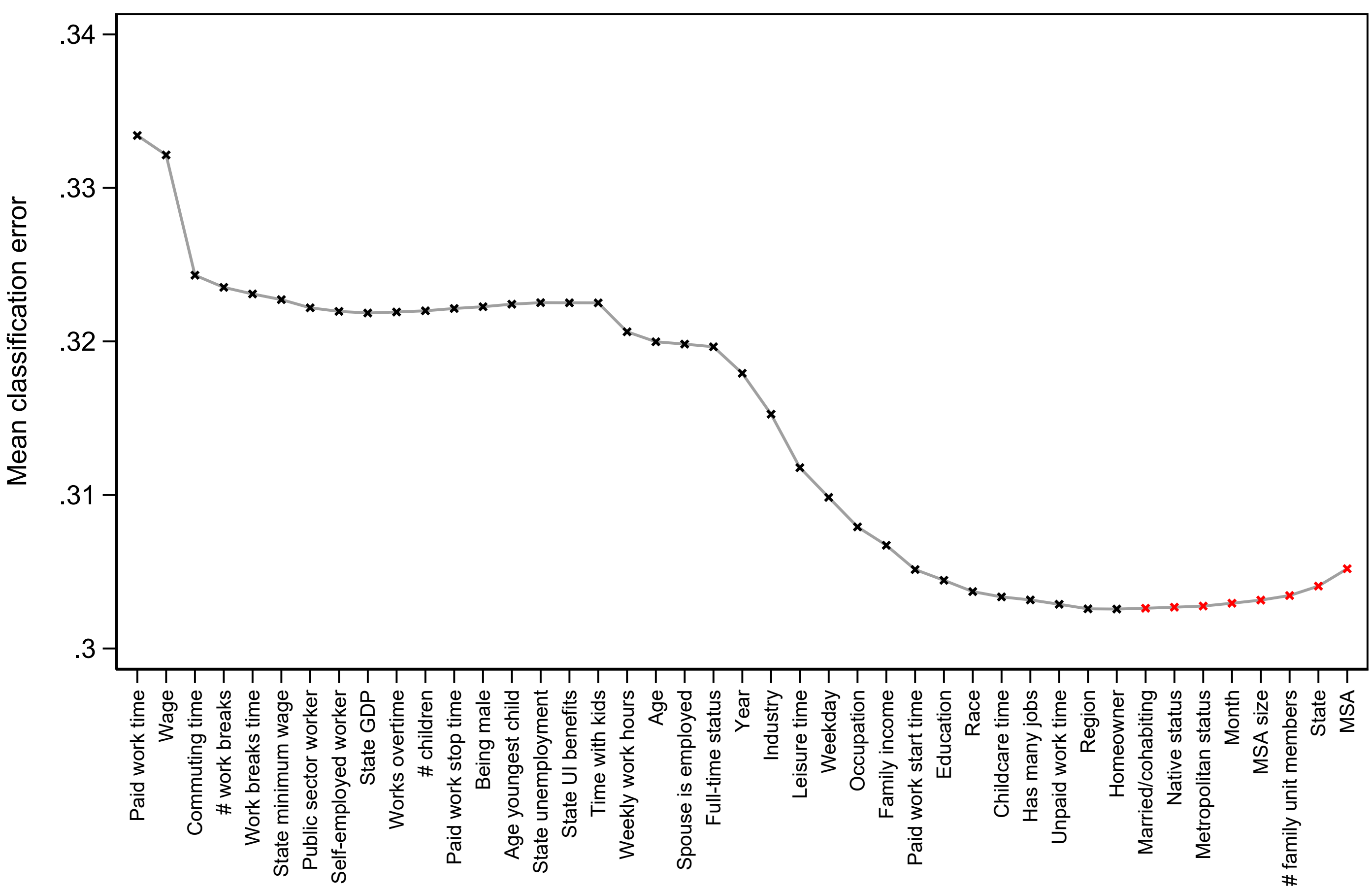
Figure 4: Mean absolute prediction errors: shirking time (shirkers)




*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day and are shirkers (i.e., report positive shirking time). The dependent variable is shirking time, measured in minutes per day.

Figure 4 shows the ranking of variables that are more relevant, in terms of predictive performance, for the *intensive* margin of shirking. The ranking seems broadly consistent with Figure 3 in the sense that work-related variables and time-allocation measures dominate early in the sequence, but the ordering shifts. For instance, worker occupation is the best predictor of the intensive margin of shirking time, followed by time spent in work breaks, time spent with children, native status, full-time status, childcare time, paid work time, leisure time, unpaid work time, and the stop and start times of paid work. Afterwards, the pattern becomes flat though still decreasing. This suggests that the variables education, family size, commuting time, age of the youngest child, weekday, State×year GDP, race, industry, homeownership, and working in multiple jobs still contribute to explain shirking time, though their predictive performance becomes marginal. Conversely, the variables ordered afterwards do not contribute to the model in terms of their predictive performance. That is to say, the mean absolute prediction error of the model worsens once these are taken into account.

Figure 5: Mean classification errors: being a shirker




*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. The dependent variable is the dummy being a shirker.

Figure 5 shows the ranking of variables for the *extensive* margin of shirking. First, the pattern shows a sharp decrease when including paid work time, wages and commuting time, followed by a flat region in the predictive performance related to the number of work breaks, work break time, State×year minimum wage, being a public sector worker, being self-employed, State×year GDP, working overtime, number of children, stop time of paid work, being male, age of the youngest child, State×year unemployment, State×year unemployment benefits, and time with kids.

This is followed by another sharp decrease of the predictive performance. This suggests that variables in the flat region may be relevant and connected with shirking, but provide redundant information until some additional variables (i.e., those in the second sharp decrease of the prediction error) are also included, namely weekly work hours, age, partner's employment status, full-time status, year, industry, leisure time, weekday, occupation, family income, and paid work start time. Afterwards, the pattern becomes again flat but still decreasing for education, race, childcare time, having multiple jobs, unpaid work time, region, and homeownership. The remaining variables worsen the predictive performance of the model.

## 5.2 Estimation results

Once we have found the variables to be taken into account in terms of their predictive performance to model shirking time, we regress the relevant dependent variables (i.e., shirking time, shirking time for shirkers only, and the binary outcome "being a shirker") against the appropriate regressors. That is to say, for each dependent variable, we only consider the regressors that contribute to the predictive performance as illustrated in Figures 3, 4 and 5.

An important remark is that the variable selection procedure is designed to maximize predictive performance, not statistical significance. It is therefore expected that, once the selected covariates are included jointly, some of them will fail to reach conventional significance levels, particularly because regressors may be correlated. In this setting, a non-significant coefficient suggests that the explanatory variable's conditional correlation with the outcome is not precisely estimated given the rest of the regressors, but that variable still contributes to lowering prediction error through its joint information content with the other regressors. Accordingly, we interpret the estimates by focusing on the direction and magnitude of the main coefficients of interest, while viewing the selected regressors whose associated coefficients are not significant at standard levels primarily as controls that improve model fit and stabilize the estimates.

Table 3 shows estimates for demographic covariates. Comparing columns (1), (2) and (3), one can analyze whether variables relate to shirking time overall in columns (1), and whether said correlation is driven by the intensive margin in column (2), and/or by the extensive margin in column (3). Being male only matters for the binary outcome "being a shirker", and its associated coefficient is not statistically significant. Worker age relates negatively to shirking time, though it seems that its primary role relates to the extensive margin, suggesting that older individuals have a lower probability of being shirkers. Education relates negatively to shirking, and those workers with university education are less likely to be shirkers, compared to their counterparts. Race also seems to relate to the extensive margin of shirking, and Whites seem to have a lower probability of being shirkers than their non-White counterparts. On the other hand, worker native status relates to the intensive margin of shirking, and being a worker of Hispanic origin relates positively to shirking time, conditional on being a shirker.

Table 3: Estimates - demographics

| | (1) Shirking time | | (2) Shirking (shirkers) | | (3) Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Being male | | | | | -0.003 | (0.007) |
| Age | -0.079*** | (0.022) | | | -0.002*** | (0.000) |
| Educ.: secondary | -2.350 | (2.541) | -4.288* | (2.474) | -0.019 | (0.022) |
| Educ.: University | -6.832*** | (2.588) | -3.017 | (2.523) | -0.078*** | (0.023) |
| Race: black | 3.365*** | (0.654) | 0.614 | (0.703) | 0.038*** | (0.009) |
| Race: other | 3.156*** | (1.036) | 0.321 | (1.015) | 0.050*** | (0.010) |
| Native status: hispan | 5.541*** | (0.692) | 3.068*** | (0.765) | | |
| Native status: asian | 1.551 | (1.509) | 2.501 | (1.566) | | |
| Native status: other | 2.489*** | (0.793) | 0.039 | (0.841) | | |
| # family members | 0.786*** | (0.181) | 0.435** | (0.199) | | |
| # children | | | | | 0.013*** | (0.003) |
| Age youngest child | -0.126*** | (0.044) | -0.047 | (0.048) | -0.001** | (0.001) |
| Time with kids | -0.013*** | (0.001) | -0.010*** | (0.002) | -0.000*** | (0.000) |
| Spouse is employed | | | | | 0.012** | (0.006) |
| Year effects | No | | No | | Yes | |
| Constant | 32.708*** | (4.643) | 49.931*** | (4.525) | 0.637*** | (0.063) |
| Observations | 47,699 | | 26,486 | | 47,699 | |
| Adjusted $R^2$ | 0.118 | | 0.341 | | 0.179 | |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. Reference for education: primary (or less than primary). Reference for race: white. Reference for native status: native. Year effects shown in Appendix Table A.3. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

Regarding household composition, Table 3 indicates that the number of family members relates positively to shirking time, and said correlation is driven by the intensive margin, while the number of children increases the probability of being a shirker, although it does not relate to the intensive margin of shirking behavior. The age of the youngest child, on the other hand, relates negatively to the probability of shirking. Time spent in the presence of children also relates negatively to shirking, both in the intensive and the extensive margin. Finally, although worker marital status seems not to be related to shirking behavior, the employment status of the partner relates positively to the probability of being a shirker.

Table 4 shows estimates for the employment-related variables. First, results suggest that these variables are primarily related to shirking time through the extensive margin, that is, through the probability of being a shirker. For instance, self-employed workers, having multiple jobs, weekly work hours and wage rates relate negatively to being a shirker. Conversely, working in the public sector and being a full-time worker relate positively to the probability of being a shirker. Moreover, being a full-time worker also relates to increased shirking time, conditional on being a shirker.

Table 4: Estimates - employment variables

| | (1) Shirking time | | (2) Shirking (shirkers) | | (3) Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Self-employed worker | -7.615*** | (1.562) | | | -0.147*** | (0.030) |
| Public sector worker | 2.727*** | (0.781) | | | 0.055*** | (0.011) |
| Full-time worker | 9.672*** | (0.896) | 5.306*** | (0.997) | 0.157*** | (0.012) |
| Has many jobs | -1.639** | (0.796) | 1.009 | (0.918) | -0.043*** | (0.010) |
| Weekly work hours | -0.290*** | (0.028) | | | -0.005*** | (0.000) |
| Wage | -0.047** | (0.020) | | | -0.001* | (0.000) |
| Works overtime | | | | | 0.008 | (0.007) |
| Family income effects | No | | No | | Yes | |
| Occupation effects | Yes | | Yes | | Yes | |
| Industry effects | Yes | | Yes | | Yes | |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. Family income effects shown in Appendix Table A.4; occupation effects shown in Appendix Table A.5; industry effects shown in Appendix Table A.6. The intercept, # observations and adjusted $R^2$ are shown in Table 3. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

Family income also relates to shirking behavior, and said correlation seems driven by the extensive margin. Detailed results are shown in Appendix Table A.4, and suggest that workers in households with very low-income families (e.g., $5,000 to $7,499) or very high-income families ($150,000 and over) have a higher probability of being shirkers. On the other hand, worker occupation and industry matter for both the intensive and the extensive margin, although Figures 4 and 5 suggest that occupation (i.e., the worker technical function) is more relevant than industry (the economic sector). Detailed results are shown in Appendix Tables A.5 and A.6, revealing large heterogeneity in terms of occupation, especially in the extensive margin of shirking, and large heterogeneity in terms of industry in the intensive margin.

This is consistent with the predictive performance of variables, as industry was among the first variables in terms of predictive power for shirking time, while it barely contributed to the predictive performance of the model for the binary outcome "being a shirker".

Table 5: Estimates - diary and time use variables

| | (1) Shirking time | | (2) Shirking (shirkers) | | (3) Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Weekday | 4.223*** | (0.538) | 2.567*** | (0.805) | 0.097*** | (0.007) |
| Leisure time | -0.047*** | (0.002) | -0.030*** | (0.004) | -0.001*** | (0.000) |
| Unpaid work time | -0.030*** | (0.003) | -0.036*** | (0.005) | -0.000*** | (0.000) |
| Childcare time | -0.028*** | (0.003) | -0.031*** | (0.005) | -0.000*** | (0.000) |
| Commuting time | 0.003 | (0.005) | -0.007 | (0.007) | 0.001*** | (0.000) |
| Paid work time | | | -0.058*** | (0.004) | 0.000*** | (0.000) |
| Paid work start time | -0.017*** | (0.002) | -0.101*** | (0.004) | -0.000*** | (0.000) |
| Paid work stop time | 0.013*** | (0.001) | 0.120*** | (0.004) | 0.000*** | (0.000) |
| Work breaks time | | | 0.252*** | (0.066) | -0.003*** | (0.000) |
| # work breaks | 12.806*** | (1.459) | | | 0.202*** | (0.018) |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. The intercept, # observations and adjusted $R^2$ are shown in Table 3. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

Table 5 shows estimates for diary variables and time use variables, which seem relevant for both the intensive and the extensive margin. For instance, working during weekdays (compared to weekends) relates to both increased shirking time and an increased probability of being a shirker. On the other hand, leisure, unpaid work and childcare times relate to decreased shirking in both margins. On the other hand, commuting time relates only to the probability of being a shirker, although it seems not to be associated with shirking time conditional on being a shirker.

Paid work time also relates to shirking time and to the probability of being a shirker, though the correlations are in the opposite direction. Longer work hours relate to increased shirking time, and at the same time relate negatively to the probability of being a shirker. In addition to this, the times when workers start and stop working also relate to shirking behavior significantly. Starting to work early relates to decreased shirking time, whereas stopping work late relates to increased shirking, both relationships working in the extensive and in the intensive margins. Finally, time spent in work breaks relates to increased shirking time (conditional on shirking), but also relates negatively to being a shirker, although the number of work breaks relates to an increased probability of being a shirker.

Finally, Table 6 shows the coefficients for regional attributes. The State GDP relates to an increased probability of being a shirker, although its coefficient is not significant for the

Table 6: Estimates - regional and housing variables

| | (1) Shirking time | | (2) Shirking (shirkers) | | (3) Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Homeowner | | | -0.072 | (0.554) | -0.007 | (0.007) |
| State×year GDP | 0.000 | (0.000) | 0.000 | (0.000) | 0.000*** | (0.000) |
| State×year UI benefits | -0.006*** | (0.001) | | | -0.000 | (0.000) |
| State×year min. wage | | | | | 0.001 | (0.001) |
| State×year unemp. | | | | | 0.132 | (0.270) |
| Region: Midwest | -0.880 | (0.609) | | | -0.019** | (0.009) |
| Region: South | -1.774*** | (0.579) | | | -0.039*** | (0.010) |
| Region: West | 1.091 | (0.675) | | | -0.011 | (0.009) |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. Reference for region: Northeast. The intercept, # observations and adjusted $R^2$ are shown in Table 3. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

intensive margin. Conversely, unemployment insurance benefits seem to relate to shirking time overall, although the results do not help understand whether said correlation is driven by the intensive or the extensive margin, as this explanatory variable was not selected for the intensive margin, and the coefficient is not significant for the extensive margin. Minimum wages and local unemployment seem not to be related to shirking behavior. Finally, results show some regional differences, as workers in the Midwest and South have a lower probability of being shirkers, compared to workers in the Northeast and in the West.

## 5.3 Discussion

Taken together, the results indicate that the determinants of shirking behavior differ depending on whether we look at shirking time for all workers, shirking time conditional on positive shirking, or the probability of being a shirker, which makes it useful to distinguish between intensive and extensive margins. Several covariates appear to operate mainly through the extensive margin. This is the case for some demographics, several employment-related variables, and regional and housing characteristics. By contrast, some other variables are more closely linked to the intensive margin, influencing shirking time among those who do shirk. Moreover, time-use variables matter for both margins, but not always in the same direction. Overall, these patterns suggest that one should carefully consider how to analyze shirking behavior, and that being a shirker and shirking times should be analyzed separately.

Despite that, the results also indicate that predictive performance remains relatively low even at the minimum of the mean absolute prediction and classification errors. For shirking time, the minimum mean absolute error is about 21.3 minutes (21.2 minutes if we analyze shirking time only among shirkers). These magnitudes are sizable, given that the average shirking time is about 25 to 28 minutes per day (44 to 47 minutes conditional on being a shirker). Thus, even the selected specifications that optimize predictive performance leave prediction errors that are of the same order of magnitude as average shirking in the full sample and roughly half of average shirking among shirkers.

Similarly, mean classification errors for the binary indicator of being a shirker also remain high even at the minimum. The mean classification error decreases modestly to about 0.30. The fact that the minimum remains close to the unconditional error implied by always predicting the modal outcome suggests that even the best mode to identify shirkers improves only to a limited extent the baseline prediction of everyone being a shirker. These patterns of limited out-of-sample accuracy suggest that shirking behavior is only partially explained by observables and is likely influenced by stochastic elements and unobserved heterogeneity such as daily task composition, monitoring, work organization, or individual effort. In other words, shirking is inherently difficult to predict in a way that makes inference feasible with standard demographic, employment, and time-allocation variables typically observed in surveys.

Returning to the hypotheses presented in Section 2, hypothesis 1 is confirmed given the descriptives in Table 1, as a significant proportion of workers report some shirking throughout their work day, and those who shirk spend about 45 minutes doing non-work while at work, excluding compulsory work breaks. Despite that, the results do not support evidence aligning with gender differences as proposed in hypothesis 2, as being male is either not considered a relevant predictor or not significant when modeling shirking behavior. Conversely, the results do confirm the existence of racial and ethnic differences, as estimates in Table 3 indicate that non-Whites have a greater likelihood of being shirkers, though shirking times do not differ racially. Besides, Hispanic origin workers spend more time shirking than their non-Hispanic counterparts, partially aligning with hypothesis 2.

Hypotheses 3 and 4 cannot be directly tested in the analysis, but some occupations and industries (e.g., those with a higher proportion of blue collar workers) have been linked to more supervision than others (Levenson and Zoghi, 2007; Ross and Zenou, 2008). For instance, occupations such as management, business and financial, computer and mathematical science, architecture and engineering, life, physical, and social science, community and social service, legal occupations, education, training, and library, arts, entertainment, sports, and media, healthcare practitioner and technical, protective service, food preparation and serving, building and cleaning and maintenance, personal care and service, and sales and related could be considered as not supervised or slightly supervised (Levenson and Zoghi, 2007; Ross and Zenou, 2008; Giménez-Nadal et al., 2018). This would leave office and administrative support, farming, fishing, and forestry, construction and extraction, installation, maintenance, and repair, production, and transportation and material moving as supervised occupations.

Similarly, industries such as finance, insurance, real estate, rental and leasing, professional and technical services, management of companies and enterprises, publishing (non-internet), motion picture and sound recording, broadcasting and internet publishing, telecommunications, data processing and other information services, education, arts, entertainment and recreation, membership organizations, and public administration could be considered not supervised or slightly supervised. On the other hand, agriculture, forestry and fishing, mining, construction, utilities, manufacturing, transportation and warehousing, wholesale and retail trade, waste management and remediation, hospitals and other health care services, social assistance, accommodation, food services and drinking places, repair and maintenance, personal and laundry services, and private households could be considered supervised industries.

Occupation and industry effects are relevant regressors for shirking behavior overall, and for the intensive and extensive margins when studied separately.[9] However, they do not provide full support for hypothesis 3. For instance, some heavily supervised occupations (e.g., office and administrative support, farming, construction and extraction, installation and repair, production, transportation and material moving) show more shirking than other occupations, and the probability of being a shirker is also greater among workers in these occupations. Despite that, among those who shirk, some heavily supervised occupations, especially production and office and administrative support, show lower shirking time.

As for industries, several supervised industries (e.g., mining, construction, transportation and warehousing) show lower shirking time relative to the reference category (agriculture). On the intensive margin, many supervised industries display negative and sometimes significant coefficients, indicating less shirking among shirkers. Despite that, the extensive margin does not show a consistent reduction in the likelihood of being a shirker in supervised industries. Hence, the monitoring hypothesis receives moderate support for total shirking in the intensive margin, but weak support for the extensive margin, in terms of worker industry.

As for the routine task hypothesis 4, one could consider blue-collar, operational occupations as a proxy for routine intensity (Giménez-Nadal and Sevilla, 2022). We find that routine-intensive occupations such as production, transport, construction, and cleaning do not show lower overall shirking. Instead, estimated coefficients for these occupations on shirking (including all workers) are generally positive. Workers in these occupations also tend to have a higher probability of being shirkers. However, some routine occupations (e.g.,

[9]Detailed coefficients for occupation and industry effects are shown in Appendix Tables A.5 and A.6, respectively.

production, office and administrative support, food preparation and serving, building, cleaning, maintenance, transportation and material moving, construction and extraction) show lower shirking among shirkers, but this pattern is not uniform. Therefore, the routine task hypothesis is not supported for overall shirking or the extensive margin, and only weakly supported for the intensive margin.

Regarding hypothesis 5, estimates indicate that the self-employed do less shirking than their employee counterparts. However, Table 4 indicates that said difference arises exclusively from the extensive margin. In other words, the self-employed have a significantly lower probability of doing any shirking at all than their private sector employee counterparts. Moreover, our analysis also compares public sector and private sector employees, and we find that the former have a larger probability of doing shirking at work than the latter, though this is only reflected in the extensive margin.

Hypotheses 6, 7 and 8 derived from the hypotheses of urban efficiency wages literature, which our results confirm, aligning with the results by Ross and Zenou (2008) and Giménez-Nadal et al. (2018). For instance, we find that leisure time relates to shirking both in the extensive and the intensive margin. That is, those workers who enjoy more leisure throughout the day are less likely to shirk, and if they shirk, they spend less time shirking than their counterparts who enjoy less time in leisure. In other words, our results confirm a substitution relationship between leisure and shirking, as postulated by Ross and Zenou (2008). In addition to this, we find that wages are also related to shirking, as workers with a higher wage are less likely to be shirkers. However, our results indicate that this relationship is relevant only in the extensive margin, but not in the intensive margin. Similarly, we find that commuting time relates positively to the probability of being a shirker. In words, workers who spend more time commuting are more likely to do some shirking at work. However, we do not find evidence that commuting time relates to shirking time (i.e., that said relationship operates in the intensive margin).

As for shirking and the timing of work, our results are not fully consistent with hypothesis 9.[10] Starting to work early in the morning relates to less shirking time throughout the workday, and also to a lower probability of being a shirker. On the other hand, stopping working late in the evening relates to more shirking time, and to a greater probability of being a shirker. Despite the associated coefficients being highly significant both in the extensive and the intensive margin, the magnitudes suggest that the impact of timing is particularly relevant in the intensive margin, and not in the extensive margin. In other words, our results

[10]However, our regressors are start and stop times of paid work, which is only a proxy for the timing of work. A different analysis exploiting the exact minutes of shirking during the day (e.g., using tempograms as in Hamermesh, 1999, 2002) could shed light and provide additional insights on the timing of shirking while at the workplace. This is beyond the scope of this paper and is left for further research.

suggest that starting to work early in the morning reduces shirking time among shirkers, and finishing the workday late in the evening increases shirking time also among shirkers, while the impact on the probability of being a shirker is relatively small.

Finally, hypotheses 10, 11 and 12 referred to the macroeconomic and labor market conditions. We do not find evidence in line with hypothesis 10, as the State×year GDP is not significant or has a negligible associated coefficient in economic terms. Regarding local unemployment, our strategy for identifying relevant variables did not include it as an appropriate predictor of shirking time. Though it is a relevant predictor for identifying shirkers, the associated coefficient is close to zero and not statistically significant at standard levels. Thus, we cannot provide evidence in line with Burda et al. (2016, 2020) regarding how local labor market conditions relate to shirking behavior. Similarly, UI benefits are not in line with hypothesis 12 either. This variable seems relevant for predicting shirking time in general terms only, but did not enter the model for the intensive margin, and the associated coefficient for the extensive margin is not statistically significant at standard levels. Besides, the sign of the coefficient for shirking time considering all workers in column (1) goes against hypothesis 12, as we find that more generous UI benefits relate to decreased shirking time. In summary, our approach does not allow us to provide evidence that a better macroeconomic context or better labor market conditions relate to shirking behaviors as postulated by Burda et al. (2016, 2020) and Belloc et al. (2025).

# 6 Conclusions

This paper studies who shirks at work and which observable characteristics are most informative about shirking behavior using time use diaries from the ATUS data over the period 2003–2024. We focus on three related outcomes: shirking time in the full sample of workers available in the ATUS, shirking time conditional on being a shirker (i.e., on doing some shirking during the diary day), and the probability of being a shirker, thus distinguishing intensive and extensive margins. We implement a prediction-oriented sequential forward-selection procedure, based on machine learning, that ranks a broad set of candidate covariates by their out-of-sample predictive performance.

Our results indicate that shirking is quantitatively non-trivial. 56.8% of women and 60.7% of men report positive shirking time. Furthermore, among those who shirk, shirking times average to about 45 minutes per day. In the estimation stage, several correlates differ across margins. Education and age are negatively related to shirking through the extensive margin, while racial and ethnic differences are also present in the likelihood of being a shirker. Employment variables matter primarily through the extensive margin too, with self-employment, multiple-job holding, weekly work hours, and wages being negatively related to being a shirker, while public-sector and full-time status are positively related. Time-use variables play a central role. Leisure, unpaid work, and childcare time are negatively related to shirking in both margins; commuting time is positively related to the probability of being a shirker. Timing of work also matters, as earlier paid-work start times relate to lower shirking, while later stop times relate to more shirking. By contrast, the macro and labor-market variables considered in the analysis provide limited support for the hypotheses linking shirking to the macroeconomic context, unemployment, or UI generosity.

The analysis has some limitations. First, even after selecting covariates to optimize out-of-sample predictive performance, predictive accuracy remains limited, suggesting that shirking is only partially captured by observables typically available in surveys, and is likely influenced by stochastic elements and unobserved heterogeneity such as daily task composition, monitoring and work organization, or individual effort. Besides, the data is cross-sectional, which prevents establishing any type of causal relationship. The analysis is therefore correlational, and all the results should be interpreted as conditional correlations only. Relatedly, our measure of shirking is based on self-reported activities and location, which may underestimate or misclassify some non-work behaviors or introduce measurement error.

Despite these limitations, the findings have some practical implications. The variables most informative for shirking are predominantly job-related and time-allocation measures, whereas local macro and labor-market indicators add little predictive content. This suggests that, in practice, interventions aimed at influencing shirking are more likely to operate through workplace-level margins, such as how workdays are scheduled, how breaks are structured, and how jobs are organized, rather than through broad macroeconomic or labor-market policies. At the same time, the limited predictive accuracy documented in the paper indicates that a large share of shirking remains unexplained by standard observables, so any policy or managerial approach based solely on observable worker or job characteristics will necessarily face significant limitations.

Moreover, our analysis opens doors for future research. First, we have documented that the determinants of shirking differ depending on whether one models total shirking time, shirking time among shirkers, or the probability of being a shirker, so separating intensive and extensive margins is empirically important. Second, our measure of shirking focuses on time spent on non-work activities while physically present at work, while some authors have focused on other measures such as workers' absence from the workplace or reduced attendance. To the best of our knowledge, these alternative dimensions of shirking have not been explored following an approach based on predictive performance, which could shed light on alternative determinants of worker performance (e.g., Antosz et al., 2020; Brunello et al., 2025; Martín-Román et al., 2026). Finally, our measures of work timing are based on paid-work start and stop times and therefore provide a proxy, rather than a direct observation, of the within-day timing of shirking episodes. Complementary analyses using within-day profiles of shirking could provide additional evidence on whether shirking concentrates at specific hours of the workday.

## References


Aguiar, M. and E. Hurst (2007). Measuring trends in leisure: The allocation of time over five decades. *Quarterly Journal of Economics 122*(3), 969–1006.

Antosz, P., T. Rembiasz, and H. Verhagen (2020). Employee shirking and overworking: modelling the unintended consequences of work organisation. *Ergonomics 63*(8), 997–1009.

Belloc, I., J. I. Gimenez-Nadal, and J. A. Molina (2025). Extreme temperatures and non-work at work. Discussion Paper 18038, IZA.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Bonke, J. (2005). Paid work and unpaid work: Diary information versus questionnaire information. *Social Indicators Research 70*(3), 349–368.

Bosworth, S. J., J. I. Gimenez-Nadal, and A. Sevilla (2025). Parental time investments and instantaneous well-being in the United States. *Scottish Journal of Political Economy 72*(1), e12402.

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

Breiman, L. (2001b). Statistical modeling: The two cultures. *Statistical Science 16*(3), 199–231.

Brodsky, A. and T. M. Amabile (2018). The downside of downtime: The prevalence and work pacing consequences of idle time at work. *Journal of Applied Psychology 103*(5), 496.

Brunello, G., M. De Paola, and L. Rocco (2025). Pension reforms, longer working horizons, and absence from work. *Journal of Population Economics 38*(1), 9.

Burda, M., K. R. Genadek, and D. S. Hamermesh (2016). Not working at work: Loafing, unemployment and labor productivity. Working Paper 21923, National Bureau of Economic Research.

Burda, M. C., K. R. Genadek, and D. S. Hamermesh (2020). Unemployment and effort at work. *Economica 87*(347), 662–681.

Chen, P. and P.-A. Edin (2002). Efficiency wages and industry wage differentials: A comparison across methods of pay. *Review of Economics and Statistics 84*(4), 617–631.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal 21*(1), C1–C68.

Cuevas-Ruiz, P., J. I. Gimenez-Nadal, S. Manfredi, and A. Sevilla (2025). A new model of parental time investments: A paradigm shift for addressing gender inequality in the labor market. *Oxford Review of Economic Policy*, forthcoming.

Efron, B. and R. Tibshirani (1997). Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association 92*(438), 548–560.

Efron, B. and R. J. Tibshirani (1994). *An introduction to the bootstrap.* Chapman and Hall/CRC.

Elciyar, K. and A. Şimşek (2021). An investigation of cyberloafing in a large-scale technology organization from the perspective of the theory of interpersonal behavior. *Online Journal of Communication and Media Technologies 11*(2).

Flood, S. M., L. C. Sayer, and D. Backman (2025). American time use survey data extract builder: Version 3.3 [dataset]. College Park, MD: University of Maryland and Minneapolis, MN: IPUMS, 2025.

Gibbons, R. and L. Katz (1992). Does unmeasured ability explain inter-industry wage differentials? *Review of Economic Studies 59*(3), 515–535.

Giménez-Nadal, J. I., J. A. Molina, and A. Sevilla (2023). Effort at work and worker well-being. In D. S. Hamermesh and S. W. Polachek (Eds.), *Time Use in Economics*, Volume 51, pp. 35–53. Emerald Publishing Limited.

Giménez-Nadal, J. I., J. A. Molina, and J. Velilla (2018). Spatial distribution of US employment in an urban efficiency wage setting. *Journal of Regional Science 58*(1), 141–158.

Giménez-Nadal, J. I., J. A. Molina, and J. Velilla (2019). Modelling commuting time in the us: Bootstrapping techniques to avoid overfitting. *Papers in Regional Science 98*(4), 1667–1685.

Giménez-Nadal, J. I. and A. Sevilla (2012). Trends in time allocation: A cross-country analysis. *European Economic Review 56*(6), 1338–1359.

Giménez-Nadal, J. I. and A. Sevilla (2022). Work effort in the uk: Trends and explanations. Discussion Paper 15329, IZA.

Guryan, J., E. Hurst, and M. Kearney (2008). Parental education and parental time with children. *Journal of Economic Perspectives 22*(3), 23–46.

Hamermesh, D. S. (1990). Shirking or productive schmoozing: Wages and the allocation of time at work. *ILR Review 43*(3), 121–S.

Hamermesh, D. S. (1999). The timing of work over time. *Economic Journal 109*(452), 37–66.

Hamermesh, D. S. (2002). Timing, togetherness and time windfalls. *Journal of Population Economics 15*(4), 601–623.

Hamermesh, D. S., K. R. Genadek, and M. C. Burda (2021). Racial/ethnic differences in non-work at work. *ILR Review 74*(2), 272–292.

Harms, T., D. Berrigan, and J. Gershuny (2019). Daily metabolic expenditures: estimates from us, uk and polish time-use data. *BMC Public Health 19*(Suppl 2), 453.

Harrell, F. E. J. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer.

Hastie, T., R. Tibshirani, J. Friedman, et al. (2009). *The elements of statistical learning*. Springer.

Horrigan, M. and D. Herz (2004). Planning, designing, and executing the BLS American Time-Use Survey. *Monthly Labor Review 127*, 3.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An introduction to statistical learning: with applications in R*. Springer.

Kan, M. Y. (2008). Measuring housework participation: The gap between "stylised" questionnaire estimates and diary-based estimates. *Social Indicators Research 86*(3), 381–400.

Krueger, A. B. and L. H. Summers (1988). Efficiency wages and the inter-industry wage structure. *Econometrica*, 259–293.

Lazear, E. P. (1999). Personnel economics: Past lessons and future directions. Working Paper 6957, National Bureau of Economic Research.

Levenson, A. R. and C. Zoghi (2007). The strength of occupation indicators as a proxy for skill. Working Paper 404, US Bureau of Labor Statistics.

Lim, V. K. (2002). The IT way of loafing on the job: Cyberloafing, neutralizing and organizational justice. *Journal of Organizational Behavior 23*(5), 675–694.

Martín-Román, Á. L., A. Moral, and S. Pinillos-Franco (2026). Gender differences in the duration of sick leave: Economics or biology? *Economics & Human Biology*, 101573.

Mazzocco, M. (2007). Household intertemporal behaviour: A collective characterization and a test of commitment. *Review of Economic Studies 74*(3), 857–895.

Neal, D. (1993). Supervision and wages across industries. *Review of Economics and Statistics*, 409–417.

Ross, S. L. and Y. Zenou (2008). Are shirking and leisure substitutable? an empirical test of efficiency wages based on urban economic theory. *Regional Science and Urban Economics 38*(5), 498–517.

Shapiro, C. and J. E. Stiglitz (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review 74*(3), 433–444.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 289–310.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological) 36*(2), 111–133.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58*(1), 267–288.

Van Ommeren, J. N. and E. Gutiérrez-i Puigarnau (2011). Are workers with a long commute less productive? an empirical analysis of absenteeism. *Regional Science and Urban Economics 41*(1), 1–8.

Vogelsang, T. (2024). Time is (not) money: Incentive effects of granting leisure time. *European Accounting Review 33*(2), 435–459.

Zenou, Y. (2009). *Urban labor economics.* Cambridge University Press.

Zhang, X., H. Guo, L. Ma, and G. Zhang (2025). Minor and serious cyberloafing in the workplace: antecedents and effects on job satisfaction. *Behaviour & Information Technology 44*(2), 387–406.

# Appendix

Table A.1: Additional summary statistics

| | Female | | Male | |
|---|---|---|---|---|
| VARIABLES | Mean | St.Dev. | Mean | St.Dev. |
| Age | 39.868 | 11.288 | 39.492 | 11.009 |
| Primary ed. | 0.010 | 0.101 | 0.021 | 0.145 |
| Secondary ed. | 0.521 | 0.500 | 0.548 | 0.498 |
| University ed. | 0.468 | 0.499 | 0.431 | 0.495 |
| Being white | 0.803 | 0.398 | 0.825 | 0.380 |
| Being black | 0.122 | 0.328 | 0.088 | 0.283 |
| Being US native | 0.749 | 0.433 | 0.702 | 0.457 |
| Being Hispan | 0.153 | 0.360 | 0.180 | 0.384 |
| Being Asian | 0.037 | 0.188 | 0.044 | 0.205 |
| Living in couple | 0.551 | 0.497 | 0.596 | 0.491 |
| Family size | 3.015 | 1.419 | 3.119 | 1.496 |
| # kids | 0.794 | 1.068 | 0.834 | 1.126 |
| Age of youngest kid | 3.388 | 5.124 | 3.039 | 4.902 |
| Time with kids | 76.905 | 154.745 | 58.131 | 129.672 |
| Public sector worker | 0.210 | 0.407 | 0.139 | 0.346 |
| Self-employed worker | 0.008 | 0.088 | 0.012 | 0.108 |
| Full-time worker | 0.823 | 0.382 | 0.934 | 0.248 |
| Weekly work hours | 39.771 | 10.346 | 44.334 | 10.495 |
| Works overtime | 0.161 | 0.367 | 0.235 | 0.424 |
| Wage | 22.843 | 19.985 | 27.357 | 28.330 |
| Family income | 12.263 | 3.410 | 12.383 | 3.361 |
| Spouse employed | 0.540 | 0.498 | 0.472 | 0.499 |
| Has many jobs | 0.107 | 0.309 | 0.096 | 0.295 |
| Paid work time | 458.859 | 140.328 | 497.484 | 141.384 |
| Starts working at | 8.683 | 2.577 | 8.283 | 2.621 |
| Stops working at | 17.242 | 2.791 | 17.450 | 2.910 |
| Work breaks time | 0.690 | 7.470 | 1.166 | 10.587 |
| # work break episodes | 0.024 | 0.168 | 0.033 | 0.197 |
| Commuting time | 33.855 | 35.365 | 42.665 | 43.476 |
| Leisure time | 77.937 | 78.153 | 87.720 | 86.563 |
| Childcare time | 26.240 | 56.502 | 16.993 | 44.824 |
| Unpaid work time | 60.437 | 69.770 | 29.400 | 48.889 |
| | 24,035 | | 23,664 | |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. The times of start and stop working are measured in hours. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

Table A.2: Mean absolute prediction errors

| Shirking time | | Shirking time (shirkers) | | Being a shirker | |
|---|---|---|---|---|---|
| Variables (ordered) | MAE | Variables (ordered) | MAE | Variables (ordered) | MCE |
| Occupation | 23.818004 | Occupation | 22.122851 | Paid work time | 0.33341689 |
| Weekday | 23.032558 | Work breaks time | 21.975716 | Wage | 0.33214616 |
| Leisure time | 22.641659 | Time with kids | 21.874053 | Commuting time | 0.32431535 |
| Time with kids | 22.363335 | Native status | 21.804565 | # work breaks | 0.32352366 |
| Unpaid work time | 22.208529 | Full-time status | 21.752056 | Work breaks time | 0.32309184 |
| Native status | 22.083130 | Childcare time | 21.716419 | State minimum wage | 0.32272422 |
| Industry | 21.949660 | Paid work time | 21.676662 | Public sector worker | 0.32218963 |
| Education | 21.851266 | Leisure time | 21.628109 | Self-employed worker | 0.32195965 |
| Paid work start time | 21.781595 | Unpaid work time | 21.576763 | State GDP | 0.32185124 |
| Full-time status | 21.725946 | Paid work stop time | 21.538311 | Works overtime | 0.32191703 |
| Weekly work hours | 21.617938 | Paid work start time | 21.279101 | # children | 0.32199646 |
| Paid work stop time | 21.553352 | Education | 21.251344 | Paid work stop time | 0.32214781 |
| # work breaks | 21.500784 | # family members | 21.230624 | Being male | 0.32226197 |
| Childcare time | 21.470918 | Commuting time | 21.207210 | Age youngest child | 0.32242687 |
| Age | 21.433673 | Age youngest child | 21.192093 | State unemployment | 0.32252634 |
| # family members | 21.412510 | Weekday | 21.180582 | State UI benefits | 0.32251881 |
| Race | 21.391494 | State GDP | 21.169375 | Time with kids | 0.32251247 |
| Age youngest child | 21.374673 | Race | 21.160503 | Weekly work hours | 0.32062903 |
| Wage | 21.359449 | Industry | 21.151864 | Age | 0.31997983 |
| Self-employed worker | 21.349040 | Homeowner | 21.150000 | Spouse is employed | 0.31982713 |
| Public sector worker | 21.342590 | Has many jobs | 21.149368 | Full-time status | 0.31965282 |
| Has many jobs | 21.336544 | # work breaks | 21.150564 | Year | 0.31792481 |
| State GDP | 21.332236 | State UI benefits | 21.15183 | Industry | 0.31526413 |
| State UI benefits | 21.328385 | Public sector worker | 21.153305 | Leisure time | 0.31177876 |
| Commuting time | 21.326488 | Self-employed worker | 21.154809 | Weekday | 0.30984082 |
| Region | 21.326092 | Age | 21.156500 | Occupation | 0.30791999 |
| Metropolitan status | 21.327031 | Being male | 21.158214 | Family income | 0.30672143 |
| Being male | 21.328114 | State minimum wage | 21.160271 | Paid work start time | 0.30514431 |
| Works overtime | 21.329885 | Married/cohabiting | 21.162663 | Education | 0.30444381 |
| State minimum wage | 21.331874 | Spouse is employed | 21.165371 | Race | 0.30370695 |
| Work breaks time | 21.335037 | Works overtime | 21.168782 | Childcare time | 0.30336220 |
| Homeowner | 21.338268 | Wage | 21.173174 | Has many jobs | 0.30316545 |
| State unemployment | 21.343148 | Metropolitan status | 21.178326 | Unpaid work time | 0.30288010 |
| Spouse is employed | 21.348243 | State unemployment | 21.184237 | Region | 0.30258550 |
| Married/cohabiting | 21.351270 | Region | 21.193373 | Homeowner | 0.30256897 |
| # children | 21.357120 | MSA size | 21.207126 | Married/cohabiting | 0.30262147 |
| Family income | 21.353556 | Month | 21.227412 | Native status | 0.30268999 |
| Year | 21.360814 | Weekly work hours | 21.251452 | Metropolitan status | 0.30276486 |
| MSA size | 21.368737 | # children | 21.278834 | Month | 0.30294997 |
| Month | 21.383863 | Family income | 21.315715 | MSA size | 0.30315165 |
| State | 21.422963 | Year | 21.351918 | # family members | 0.30344887 |
| MSA | 21.486697 | State | 21.417447 | State | 0.30406117 |
| Paid work time | 21.694417 | MSA | 21.533369 | MSA | 0.30519514 |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. The dependent variables are shirking time (measured in minutes per day), shirking time for shirkers (measured in minutes per day), and the dummy being a shirker, respectively. MAE stands for mean absolute error. MCE stands for mean classification error.

Table A.3: Year effects

| | Shirking time | | Shirking (shirkers) | | Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| 2004 | | | | | 0.051*** | (0.016) |
| 2005 | | | | | 0.044*** | (0.015) |
| 2006 | | | | | 0.042*** | (0.016) |
| 2007 | | | | | 0.045*** | (0.016) |
| 2008 | | | | | 0.013 | (0.016) |
| 2009 | | | | | 0.016 | (0.017) |
| 2010 | | | | | -0.009 | (0.018) |
| 2011 | | | | | 0.017 | (0.017) |
| 2012 | | | | | 0.030* | (0.016) |
| 2013 | | | | | 0.039** | (0.016) |
| 2014 | | | | | 0.003 | (0.016) |
| 2015 | | | | | 0.007 | (0.017) |
| 2016 | | | | | -0.033* | (0.017) |
| 2017 | | | | | -0.011 | (0.018) |
| 2018 | | | | | -0.004 | (0.018) |
| 2019 | | | | | 0.005 | (0.019) |
| 2020 | | | | | -0.099*** | (0.019) |
| 2021 | | | | | -0.067** | (0.028) |
| 2022 | | | | | -0.094*** | (0.022) |
| 2023 | | | | | -0.080*** | (0.020) |
| 2024 | | | | | -0.061*** | (0.021) |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. Reference for year: 2003. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

Table A.4: Family income effects

| | Shirking time | | Shirking (shirkers) | | Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| $5,000 to $7,499 | | | | | -0.075* | (0.040) |
| $7,500 to $9,999 | | | | | 0.029 | (0.038) |
| $10,000 to $12,499 | | | | | -0.003 | (0.033) |
| $12,500 to $14,999 | | | | | 0.007 | (0.033) |
| $15,000 to $19,999 | | | | | -0.011 | (0.030) |
| $20,000 to $24,999 | | | | | -0.006 | (0.029) |
| $25,000 to $29,999 | | | | | 0.003 | (0.028) |
| $30,000 to $34,999 | | | | | 0.026 | (0.028) |
| $35,000 to $39,999 | | | | | 0.016 | (0.028) |
| $40,000 to $49,999 | | | | | -0.018 | (0.027) |
| $50,000 to $59,999 | | | | | -0.006 | (0.027) |
| $60,000 to $74,999 | | | | | -0.014 | (0.027) |
| $75,000 to $99,999 | | | | | -0.031 | (0.027) |
| $100,000 to $149,999 | | | | | -0.048* | (0.027) |
| $150,000 and over | | | | | -0.087*** | (0.028) |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. Reference for family income: less than $5,000. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

Table A.5: Occupation effects

| | Shirking time | | Shirking (shirkers) | | Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Business, financial | 2.245*** | (0.831) | 0.187 | (1.012) | 0.029** | (0.013) |
| Computer and math science | 2.682** | (1.221) | 1.254 | (1.867) | 0.012 | (0.015) |
| Architecture and engineering | 6.968*** | (1.195) | 1.693 | (1.241) | 0.130*** | (0.018) |
| Life, physica, social science | 4.380*** | (1.458) | -2.021 | (1.632) | 0.095*** | (0.023) |
| Community, social service | 1.033 | (1.579) | -1.533 | (1.956) | -0.001 | (0.021) |
| Legal occupations | 4.343** | (1.723) | 1.437 | (2.287) | 0.059** | (0.024) |
| Education, training, and library | 4.745*** | (1.041) | 1.876 | (1.262) | 0.075*** | (0.016) |
| Arts, entertainment, sports, media | -1.671 | (1.323) | -1.931 | (1.721) | -0.018 | (0.022) |
| Healthcare practitioner, technical | 3.764*** | (1.156) | -0.728 | (1.320) | 0.084*** | (0.016) |
| Healthcare support | 5.685*** | (1.479) | -3.122** | (1.522) | 0.123*** | (0.021) |
| Protective service | 2.478 | (1.870) | -0.267 | (2.266) | 0.027 | (0.025) |
| Food preparation, serving | 4.527** | (1.803) | -3.589 | (2.308) | 0.082*** | (0.023) |
| Building, cleaning, maintenance | 11.854*** | (1.788) | -0.841 | (1.891) | 0.158*** | (0.019) |
| Personal care service | 3.276** | (1.533) | -1.708 | (1.923) | 0.027 | (0.024) |
| Sales and related | 1.332 | (0.893) | -0.762 | (1.028) | 0.020 | (0.014) |
| Office and admin. support | 6.695*** | (0.803) | -2.005** | (0.898) | 0.102*** | (0.012) |
| Farming, fishing, forestry | 5.617 | (3.428) | -7.713*** | (2.781) | 0.149*** | (0.046) |
| Construction, extraction | 15.368*** | (1.773) | 2.757 | (1.782) | 0.196*** | (0.020) |
| Installation, maintenance, repair | 9.692*** | (1.218) | 0.507 | (1.175) | 0.128*** | (0.019) |
| Production | 15.889*** | (1.173) | -4.317*** | (1.172) | 0.250*** | (0.015) |
| Transport | 11.612*** | (1.431) | -1.264 | (1.423) | 0.169*** | (0.017) |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. Reference for occupation: management. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.

Table A.6: Industry effects

| | Shirking time | | Shirking (shirkers) | | Being a shirker | |
|---|---|---|---|---|---|---|
| VARIABLES | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Forestry & fishing | -4.622 | (5.869) | -8.011 | (6.329) | 0.040 | (0.079) |
| Mining | -9.456** | (4.022) | -10.195*** | (3.690) | -0.011 | (0.060) |
| Construction | -6.463* | (3.501) | -7.883*** | (3.050) | 0.009 | (0.046) |
| Nonmetallic minerals (mfg.) | 1.748 | (4.691) | -6.051 | (3.722) | 0.067 | (0.055) |
| Primary & fabricated metals (mfg.) | 0.227 | (4.111) | -5.389 | (3.767) | 0.039 | (0.047) |
| Machinery (mfg.) | -3.401 | (3.768) | -6.033* | (3.285) | 0.040 | (0.050) |
| Computer & electronic products (mfg.) | -3.875 | (3.689) | -4.772 | (3.340) | -0.015 | (0.049) |
| Electrical equipment (mfg.) | 0.054 | (4.038) | -7.791** | (3.572) | 0.107** | (0.052) |
| Transportation equipment (mfg.) | -1.560 | (3.526) | -8.864*** | (3.088) | 0.059 | (0.047) |
| Wood products (mfg.) | -1.313 | (5.777) | -6.424 | (5.310) | 0.080 | (0.056) |
| Furniture (mfg.) | -5.248 | (4.062) | -14.318*** | (3.461) | 0.041 | (0.060) |
| Food (mfg.) | 5.156 | (3.762) | -6.860** | (3.108) | 0.090* | (0.047) |
| Beverage & tobacco (mfg.) | -3.626 | (4.847) | -8.300** | (4.185) | 0.030 | (0.069) |
| Textiles, apparel & leather (mfg.) | 8.913 | (6.402) | -0.716 | (5.230) | 0.062 | (0.052) |
| Paper & printing (mfg.) | -6.030 | (4.078) | -10.609*** | (3.306) | 0.026 | (0.054) |
| Petroleum & coal products (mfg.) | 0.202 | (6.186) | -0.858 | (5.372) | 0.031 | (0.081) |
| Chemicals (mfg.) | 1.190 | (3.758) | -3.194 | (3.185) | 0.036 | (0.049) |
| Plastics & rubber (mfg.) | 3.036 | (4.486) | -6.641* | (3.415) | 0.077 | (0.058) |
| Other / not specified mfg. | -2.075 | (4.528) | -5.623 | (4.231) | 0.058 | (0.050) |
| Wholesale trade | -2.878 | (3.397) | -6.895** | (2.897) | -0.002 | (0.045) |
| Retail trade | 0.023 | (3.386) | -7.576*** | (2.909) | 0.094** | (0.044) |
| Transportation & warehousing | -4.532 | (3.601) | -5.798* | (3.314) | -0.019 | (0.046) |
| Utilities | -3.075 | (3.639) | -9.321*** | (3.347) | 0.018 | (0.049) |
| Publishing (non-internet) | -5.762 | (3.748) | -11.550*** | (3.512) | -0.014 | (0.055) |
| Motion picture & sound | 9.414 | (7.989) | 2.796 | (8.766) | 0.105 | (0.067) |
| Broadcasting (non-internet) | -4.035 | (4.462) | -10.515** | (5.327) | 0.024 | (0.058) |
| Internet publishing/broadcasting | -3.305 | (6.636) | -18.447** | (9.295) | -0.024 | (0.102) |
| Telecommunications | -1.961 | (3.841) | -5.816* | (3.407) | -0.004 | (0.050) |
| ISP & data processing | -12.829*** | (4.772) | -5.932 | (4.321) | -0.181** | (0.079) |
| Other information services | 1.377 | (5.925) | -5.508 | (5.885) | -0.001 | (0.073) |
| Finance | -4.146 | (3.396) | -7.148** | (2.990) | 0.013 | (0.045) |
| Insurance | -4.111 | (3.472) | -5.908* | (3.049) | -0.019 | (0.046) |
| Real estate | -5.316 | (3.634) | -3.429 | (3.348) | -0.041 | (0.049) |
| Rental & leasing | -0.108 | (5.361) | 0.240 | (5.402) | -0.009 | (0.062) |
| Professional & technical services | -6.822** | (3.356) | -6.305** | (2.958) | -0.050 | (0.044) |
| Management of companies | -12.382** | (6.240) | -17.384*** | (6.364) | -0.106 | (0.115) |
| Administrative & support | -4.728 | (3.468) | -7.964*** | (3.036) | 0.029 | (0.045) |
| Waste management & remediation | -7.888* | (4.482) | -9.075** | (3.938) | -0.037 | (0.068) |
| Education | -4.767 | (3.419) | -6.351** | (2.993) | 0.005 | (0.045) |
| Hospitals | 2.783 | (3.530) | -4.207 | (3.139) | 0.119*** | (0.045) |
| Health care (excl. hospitals) | -4.824 | (3.386) | -5.374* | (2.948) | 0.011 | (0.045) |
| Social assistance | -6.303* | (3.543) | -7.582** | (3.251) | -0.026 | (0.047) |
| Arts, entertainment & recreation | -0.372 | (3.684) | -4.624 | (3.355) | 0.048 | (0.049) |
| Accommodation | -3.155 | (4.798) | -11.392** | (4.534) | 0.137*** | (0.050) |
| Food services & drinking places | -9.559*** | (3.600) | -13.661*** | (3.389) | -0.003 | (0.048) |
| Private households | -21.652*** | (3.827) | -13.454*** | (3.807) | -0.240*** | (0.058) |
| Repair & maintenance | -8.330** | (3.676) | -7.272** | (3.329) | -0.021 | (0.051) |
| Personal & laundry services | -7.494** | (3.752) | -10.276*** | (3.422) | 0.036 | (0.053) |
| Membership organizations | -3.793 | (3.946) | -1.617 | (4.440) | -0.045 | (0.048) |
| Public administration | -4.778 | (3.540) | -6.096** | (3.103) | -0.026 | (0.046) |

*Notes*: The sample (ATUS 2003-2024) is restricted to employed workers who work during the diary day. Robust standard errors in parentheses. Reference for industry: agriculture. *** significant at the 1%, ** significant at the 5%, * significant at the 10%.