The Cognitive Load of Financing Constraints: Evidence from Large-Scale Wage Surveys*

Clémence Berson† Raphaël Lardeux‡ Claire Lelarge§

June 2022

Abstract

In this paper, we take advantage of the implicit cognitive exercise available in standard Labor Force Surveys in order to provide a new, population-wide quantification of the cognitive load associated with financial constraints (Mullainathan and Shafir, 2013). This quantification is based on a well-defined index of worker-level attention which filters out rounding behavior and reporting biases. We estimate it using unsupervised clustering techniques and find that workers perceive their own wages with a degree of uncertainty of around 10%, which through the lens of a simple rational signal extraction model translates into estimates of workers' attention ranging from 30% to 84% depending on their wage, education, tenure and gender. Most importantly, the attention of the lowest paid 30% of workers is cyclical and increases steadily (by 17 percentage points) in the ten days preceding payday, before immediately dropping on that day. We show theoretically that this pattern is indicative of end-of-month financing (liquidity) constraints. Furthermore, it reveals that these financing constraints induce cognitive costs arising from the not too concave (or convex) costs of achieving high levels of attention, and the convex costs of *maintaining* them over time. Our model identifies a lower bound for the annual attention cost burden incurred by financially constrained workers, which ranges between €10 and €50 depending on risk aversion.

Keywords: Behavioral Inattention, Cognitive Costs, Wage Volatility, Financing Constraints

JEL Classification: C83, D14, I32, J31

*This paper was presented at the INSEE D2E, Geneva GSEM, ENS Paris-Saclay, CREST X-ENSAE and HEC seminars as well as at the BdF/AMSE 2018, EEA 2020, RCEA 2022, IAAE 2022 and ESEM 2022 Conferences. We would like to thank (without implying) our discussants and the members of Raphaël's PhD committee (especially Pierre Boyer, Elise Coudin, Nicolas Jacquemet, Rafael Lalive and Olivier L'Haridon) for their comments. This work is indirectly supported by public grants from the French National Research Agency (ANR) as part of the *Investissements d'avenir* program (reference: ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD) and of the Idex Paris-Saclay TANED chair.

†Banque de France and Sciences Po, LIEPP. E-mail: clemence.berson@banque-france.fr

‡CRED, DREES. E-mail: lardeuxsraphael@gmail.com

§Université Paris-Saclay and CEPR. E-mail: claire.lelarge@universite-paris-saclay.fr

1 Introduction

Classic economic theory assumes that agents are fully rational and able to process the information available in their economic environment. However, this framework has since been increasingly challenged by a series of empirical and theoretical contributions which reveal that many individual decisions are flawed and suffer from psychological biases (Kahneman et al., 1991; Kahneman, 2011). Faced with a complex informational environment and endowed with limited cognitive capacities, agents tend to focus their attention on the most relevant or, less fortunately, the most salient characteristics of their economic environment, while they pay less attention to other characteristics or misunderstand them. Such inattention biases have been documented in the literature, mainly with respect to the perception of the characteristics of consumption goods. Examples include evidence about inattention to the mileage of used cars sold on the second-hand market (Lacetera et al., 2012), to goods' "shrouded attributes" (Gabaix and Laibson, 2006), to sales taxes (Chetty et al., 2009; Taubinsky and Rees-Jones, 2017), or inattention to energy pricing (Ito, 2014; Allcott, 2011). A growing body of literature in psychology, economics and cognitive sciences further suggests that the degree of (in)attention is in fact related to the overall cognitive load of agents, which determines the amount of cognitive resources available for individuals' decision making and task completion (Mullainathan and Shafir, 2013).

Our paper contributes to this literature by proposing a novel, large-scale measure of attention that we are able to estimate in the entire population of French wage earners using Labor Force Survey data. The object of attention in our setting is the amount of wages paid-out, which (obviously) is economically important and relevant since it affects a wide array of economic decisions. Most importantly in our setting, it constitutes the main component of the budget of French workers, particularly for those at the bottom of the wage distribution. Interestingly, our data furthermore allow us to control for the exposure of these low-wage workers to potential financing constraints by varying distance to payday, as in Carvalho et al. (2016) and Mani et al. (2020), but in a much larger population.[1] This feature of our setting allows us to test whether the amount of attention allocated to the monitoring of the budget constraint depends on the magnitude of financing constraints and to quantify the associated costs, i.e. the corresponding cognitive burden they generate.

Critical to this research strategy is our original measure of the attention that workers allocate to their own wages. We leverage large-scale data from the French Labor Force Survey (LFS) by reinterpreting the item requesting workers to report their own wage as a cognition exercise, which enables us to propose a well-defined and *direct* measure of workers' attention (Gabaix, 2019),[2] which is based on a comparison of wages that are self-reported by workers with their fiscal, arguably "true" counterparts. This comparison

[1] A similar strategy is also implemented in the real-life quasi-experiment of Mani et al. (2013): there, harvest time plays the same role as payday in our setting. Our paper thus also provides evidence about the external validity of the results in Mani et al. (2013), but in a much wider setting (their sample only contains approximately 500 Indian farmers), and for the case of a developed country. Even more recent contributions vary financing constraint by implementing quasi-experimental methods. For example, in Kaur et al. (2021), the experiment staggers the timing of wage payments: some workers are paid earlier and receive a cash infusion while others remain liquidity constrained.

[2] In this respect, our empirical strategy to measure attention has the same flavor of the "Lab-in-the-Field" methods described in Gneezy and Imas (2017): first, it is conducted in the field (in the general population of French workers), thus ensuring a large degree of external validity. Second, it takes advantage of an implicit but well-defined cognitive exercise available from the LFS, thus preserving a reasonable level of control in our measures of attention.

requires building a statistical mixture model addressing the issues raised by perception or reporting biases and rounding, which is estimated with unsupervised clustering techniques.

We find that over our period of analysis (2005 to 2015) workers tended to perceive their own wages with a degree of uncertainty of around 10%. Through the lens of a simple rational signal extraction model, this amounts to estimates of worker attention ranging between 30% and 84% depending on wage, education, tenure and gender. Secondly, we rely on a feature of the sampling scheme of the French LFS which makes the date of interview and critically, its distance to payday orthogonal to worker characteristics and therefore exogenous to all our dimensions of interest. This enables us to document whether attention evolves over time in different populations of workers. This point is highly debated in the literature (Carvalho et al., 2016; Mani et al., 2020). We find that in our data, the 30% lowest-paid workers, who are most likely to struggle to make ends meet each month, exhibit suggestive patterns of cyclicality: their attention is minimal in the middle of the month and increases steadily until payday, suggesting that their budget constraint becomes increasingly tight during this period of the month and requires closer monitoring. Finally, their attention drops immediately once payday is reached. This feature of the monthly cycle of attention is not compatible with a pure informational story, whereby workers would be measured as more *attentive* simply because they are more informed.[3] Conversely, we show that this feature of the data is well rationalized by a mechanism of credit constraints with costly budget constraint monitoring. In such a framework, the cyclical pattern of attention of the data identifies the shape of the attention cost function. It reveals that the costs of achieving high levels of attention are not too concave (or even convex), and the costs of *maintaining* attention over time are convex. These features of the attention cost function explain why workers find it excessively costly to maintain high levels of attention over the entire month even when it would improve consumption smoothing. Overall, we estimate that lower bounds for the annual attention cost premium incurred by financially constrained workers range between €10 and €50 depending on workers' risk aversion.

As workers are not incentivized to report accurate amounts in the LFS, we expect that they only give information involving the least effort and, most often, simply provide approximate answers off the top of their heads. In this respect, our use of survey data is similar to Ferrario and Stantcheva (2022): their research shows that the answers of respondents who have not previously thought carefully about the survey items are informative in that they reflect "*what a respondent thinks and will keep thinking, absent more learning or targeted reflection*". Similarly, in our setting, the survey data offer a large-scale depiction and quantification of the real-life variations in attention and cognitive load in the general population of French workers. Furthermore, our parsimonious "structural" model enables us to apply a revealed preference strategy, and use the previous information in order to identify the magnitude of the time-varying external incentives determining these patterns, which we interpret in terms of financing constraints. Overall, our results are indicative of the fact that the bottom 30% of workers in the wage distribution have to keep a precise record of their wage in

[3]Indeed, such a story would predict that attention should increase discretely (jump) to reach a peak at payday, and then decrease steadily until the next payday. These predictions are however rejected in the data, since workers' attention rather increases steadily during the last 15 days of the month.

their mind and are thus subject to a higher mental burden, especially in the last days before payday (Mani et al., 2013; Shah et al., 2018; Schilbach et al., 2016). This proportion constitutes an estimate of the share of the overall population suffering from liquidity financing constraints which (to our knowledge) is new in the literature. Estimating the size of this population of agents who are at a kink[4] of their intertemporal budget sets is important, as they typically feature high marginal propensities to consume and play a critical role in heterogeneous agent macro models (Kaplan and Violante, 2018). Furthermore, our results confirm that scarcity (in the form of tighter end-of-month budget constraints) implies a higher cognitive load, which opens the door to further implications in terms of individual performance and decision making for the affected people (Haushofer and Fehr, 2014; Gneezy et al., 2020; Mani et al., 2013; Mullainathan and Shafir, 2013; Kaur et al., 2021).

Beyond our population-wide quantification of the cognitive load arising from financing constraints, we make an original methodological contribution to the literature by providing a novel and *direct* measure of attention. In contrast, most of the papers which rely on non-experimental data *indirectly* elicit measures of attention from choice or action data, and their departure from "fully rational" benchmarks. A notable recent exception is Handel and Kolstad (2015), who combine choice data from a health plan with a survey specifically designed to measure different dimensions of agents' information sets and risk preferences, and are able to identify the separate contributions of these factors to agents' decisions. While their empirical proxies are optimally designed for their research question, the relatively small sample size (ca. 1,700 non-missing observations) and the absence of sample recurrence limit their dataset. In contrast, our empirical proxies are less optimal but, as we show in the paper, still contain a lot of information about workers' attention efforts. Moreover, the statistical power of our setting is magnified by the large-scale, recurrent and scientific sampling procedure of our data, which prevents any selection issue on workers' unobservable characteristics for our empirical strategy to measure financing constraints and the associated attention cost.[5]

Lastly, our paper relates to several strands of literature beyond those mentioned above. First, the empirical literature in labor economics has documented that the distributions of workers' self-reported wages are close to discrete because of rounding, while those constructed from administrative sources are more continuous. This feature of the data was documented in studies of wage rigidities (e.g. Biscourp et al., 2005 for the case of France or Dickens et al., 2007 for a wider set of 16 countries) and wage dynamics (Pischke, 1995) as worker's rounding behavior significantly complicated the analyses in these contexts by introducing a noise in the measurement of wage variations. This tilted researchers to rely on administrative rather than survey data as the latter became increasingly available. Pischke (1995) also proposed a statistical methodology to

[4]Typically generated by a borrowing limit of the type we consider in Section 2.

[5]Further recent contributions investigate the determinants of differentiated and sometimes inconsistent responses of agents that are collected via surveys and in experiments, thus suggesting that combining the two sources of information requires special care and might not be generalizable to many settings, especially when attempting to measure parameters related to risk aversion. For example, Belzil et al. (2021) show that the the welfare gains that can be elicited from the survey data on Canadian high school students' perceptions of the costs and benefits to higher education and from (incentivized) experiments are inconsistent. The authors argue that this is due to the fact that in the survey, the stakes are null and reporting false intentions and expectations is costless. This critique however does not apply to our setting, as we only rely on survey data to describe features of the *information set* of agents, and not their *intentions*, as explained above (Ferrario and Stantcheva, 2022).

try and correct for this feature of survey data in analyses of wage dynamics. Our work indirectly contributes to this literature, as our enhanced understanding of workers' reporting behavior delivers alternative, more structural methods of smoothing and de-biasing survey data. More importantly, we propose turning this "old" problem on its head and consider that the noise introduced by workers is an object of study in itself rather than purely a nuisance in the data,[6] since it contains invaluable information about workers' level of attention. Lastly, our paper points to the debate in macroeconomics which discusses how (alternative) mechanisms of imperfect information extraction and processing have aggregate consequences, particularly via their impact on households' marginal propensities to consume (Sims, 2003; Luo, 2008; Reis, 2006).

The remainder of the paper is organized as follows. Section 2 presents our theoretical set-up and clarifies the links between attention and financial constraints. Section 3 details the data and our two empirical proxies of wages while Section 4 contains our empirical model and estimation strategy for our index of attention. Section 5 reports the results obtained from our refined exercise of variance decomposition, with a particular focus on the monthly cycle of attention exhibited by low-wage workers. Finally, Section 6 concludes.

2 Theoretical Framework

In this section, we propose a parsimonious model of costly attention designed to guide our empirical investigations and the interpretation of our results.

Baseline set-up. We consider a worker i who is paid on a monthly basis and has to make daily decisions about consumption between two subsequent paydays. This time interval is normalized to 1 without loss of generality, as well as the price of the consumption good. The objective function of this worker and for the month under consideration is written as:

$$U^{(0)}(C_t, F(.)) = \int_0^1 u(C_t)\mathrm{d}t - R_A \int_{\mathbb{R}+} \left(\bar{C} - W\right) F(W)\mathrm{d}W - R_B \int_0^{\bar{C}} \left(\bar{C} - W\right) F(W)\mathrm{d}W \qquad (1)$$

where C_t denotes instantaneous consumption and W is the wage payment, which from the perspective of the worker is imperfectly known and therefore uncertain. In this objective function, the baseline utility function u is assumed to be strictly concave, which implies that it is optimal to smooth the total monthly consumption $\bar{C} = \int_0^T C_t \mathrm{d}t$ over time, i.e., $C_t = \bar{C}\ \forall t \in [0;1]$.[7]

We assume that it is possible to transfer revenue (or rather, liquidities) across time. We model this di-

[6] See also Binder (2017) for a similar strategy of "making the most" of survey data, which we discuss below in Section 5.

[7] Equation 1 adopts a minimalist specification and neglects in particular within-month time discounting. An alternative would have been to introduce subjective as in Laibson (1997) or O'Donoghue and Rabin (1999) time-varying discount factors, but at the cost of significantly higher complexity since hyperbolic discounting is usually associated with the introduction of multiple "selves" which complicate the definition of the relevant welfare criterion (as different selves have different utilities). In addition, we lack the data about the consumption profiles of each worker which are required to quantify these parameters. Introducing discount factors in Equation 1 would mainly affect the consumption profile C_t (thus potentially allowing for decreasing patterns over time) and would weaken the incentives to pay attention and to exert this effort early, but the section's other insights would still hold qualitatively.

rectly in the objective function via the last two integral terms, which are simply to be interpreted as the reduction in the continuation value of the next period from starting it with debt rather than being debt-free, or alternatively as the increase in the continuation value arising from starting it with savings (Allcott et al., forthcoming).[8] Specifically, the first term captures the potential ability of the worker to transfer income across time symmetrically, at a potentially idiosyncratic interest rate R_A. This parameter is typically driven down to 0 in the case of workers who are unable to transfer revenue in the next period, e.g. because they lack access to the necessary financial products. The second term, parameterized by R_B, captures the potential asymmetry between the cost of borrowing and the gains from saving. It is typically negligible for financially (or liquidity) unconstrained workers, but large for those facing large interest rates premia or psychological costs upon borrowing. For example, Allcott et al. (forthcoming) document that borrowers are typically willing to pay a significant premium for an experimental incentive to avoid (future) payday borrowing.[9] In what follows, we will label these costs (either financial or psychological) as “financing constraints”.

A critical object of Equation 1 is the probability density function F of imperfectly observed wages. We consider at this stage that workers are only able to formulate “default” guesses W^d about their true (“fiscal”) wage, W^f. A natural benchmark to consider is that they face Gaussian relative uncertainty in terms of their wage, i.e., that the relative errors they make when guessing their wage are normally distributed:

$$w^d = w^f + \eta \Longleftrightarrow w^f = w^d - \eta \text{ with } \eta \sim \mathcal{N}\left(0, \sigma^2\right) \tag{2}$$

where w^d and w^f denote the logarithms of W^d and W^f, respectively. We specify the error term η in the logarithmic space in order to abstract from scale effects. From the perspective of the worker, the guessed wage w^d is known but the true value w^f is not, such that F has a lognormal probability density function with mean w^d and variance σ^2. One benefit of this Gaussian benchmark is its simplicity.[10] It furthermore corresponds to the class of distribution with maximal (Shannon) entropy within the class of absolutely continuous density functions with a given variance σ^2: in other words, these distributions are the least informative for workers and therefore correspond to a well-defined “worst case” scenario.[11]

[8]The benefit of this modeling strategy (which resembles a standard Bellman equation of an - unspecified - dynamic model) is that we do not need to observe or explicitly model decisions in subsequent periods, as their continuation value is captured in the two integral terms.

[9]Allcott et al. (forthcoming) actually equate the concavity of the overall continuation value term as their main measure of risk aversion (beyond the concavity of u) as there is no heterogeneity in financial fees across agents in their setting. In our Equation 1, it captures both financing constraints and loss (risk) aversion.

[10]Our set-up focuses on the analysis of the uncertainty parameters σ and σ_m, which appear to be most relevant and important in the empirical section of the paper. However, the model could easily be adapted to allow that workers' perceptions w^d of their own wage could be affected by biases, in which case the expectation of w^d is no longer w^f. Increasing the level of attention could allow them to mitigate such biases, thus allowing them to decrease the costs that these “mistakes” might generate (either in terms of financing costs if $\mathbb{E}(w^d) > w^f$, or in terms of utility derived from consumption if $w^f > \mathbb{E}(w^d)$). There are two reasons why the version of the main text does not incorporate this extension. First, as explained in Section 4, our data do not allow us to separately identify such cognitive biases from pure (under- or over-) reporting strategies. Second, such an extension does not ultimately appear to be empirically relevant: our results show that worker level biases do not fluctuate during the month, such that they do not seem to be correlated with attention.

[11]Probability density functions of maximal entropy are usually introduced as starting points in Bayesian update processes of the type considered in this section. Our empirical application also focuses on the Gaussian case, as the empirical literature considers that the lognormal hypothesis is a good approximation of the distribution of wages and income in general (Heckman and Sattinger, 2015). However, our estimation procedure could easily be adapted to alternative parametric assumptions via inverse transform sampling.

Incorporating all these elements into Equation 1 yields:

$$U^{(0)}(\bar{C}, w^d, \sigma) = u(\bar{C}) - \bar{C}\left(R_A + R_B \Phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right)\right) + e^{w^d + \frac{\sigma^2}{2}}\left(R_A + R_B \Phi\left(\frac{\ln \bar{C} - w^d}{\sigma} - \sigma\right)\right) \quad (3)$$

The first order condition of optimality with respect to $\bar{C}$ can be written as:[12]

$$u'\left(\bar{C}\right) = R_A + R_B \Phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right) \geq R_A, \quad (4)$$

where Φ denotes the cumulative density function of a normalized Gaussian random variable (with mean 0 and variance 1). Equation 4 shows that risk-neutral workers facing no financial constraints ($R_A \geq 0$, $R_B = 0$) equate their marginal utility of consumption with the marginal utility gain of transferring revenue across time, R_A. In contrast, if workers are loss averse or face additional financing constraints in the form of an interest rate premium R_B, they will lower total consumption $\bar{C}$ in order to create a buffer preventing them from consuming beyond W^f with a too high likelihood.[13]

Equation 4 also shows that financially unconstrained and risk-neutral workers (with $R_B = 0$) are unaffected by the uncertainty surrounding W. In contrast, financially constrained workers have a lower consumption when they face a higher level of uncertainty. Indeed, differencing Equation 4 in the neighborhood of the optimal consumption level $\bar{C}$ gives:

$$\left.\frac{\partial C}{\partial \sigma}\right|_{\bar{C}} = -\frac{\frac{R_B}{\sigma}\frac{w^d - \ln \bar{C}}{\sigma}\phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right)}{\frac{R_B}{\sigma \bar{C}}\phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right) - u''(\bar{C})} \quad (5)$$

This quantity is negative as long as $w^d \geq \ln \bar{C}$, which has to hold at least in expectation to ensure that the consumption path is sustainable.[14] However, this result does not imply that it is necessarily welfare improving to benefit from lower perceived wage uncertainty in our setting, despite the concavity of u, as Equation 5 does not take the opportunity to transfer revenue across time (via R_A) into account. Using the

[12] This results from the following derivation:

$$u'\left(\bar{C}\right) = R_A + R_B \Phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right) + \frac{R_B}{\sigma}\frac{\bar{C}}{\bar{C}}\phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right) - \frac{R_B}{\sigma}\frac{1}{\bar{C}}\underbrace{e^{w^d + \frac{\sigma^2}{2}}\phi\left(\frac{\ln \bar{C} - w^d}{\sigma} - \sigma\right)}_{\bar{C}\phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right)}$$

The last two terms cancel each other.

[13] This results from the concavity of u, which also determines the magnitude of this buffer. Whenever u is steeper, i.e. whenever u' is higher all else equal, the level of consumption is set at a higher level and the buffer is lower.

[14] To be precise, we assume that the following condition holds for all workers: $u'(e^{w^f}) < R_A$. For unconstrained workers ($R_B = 0$) whose optimal consumption level $\bar{C}^{NC}$ is such that $u'(e^{\ln \bar{C}^{NC}}) = R_A$, we have $\ln \bar{C} < w^f = \mathbb{E}[w^d]$ given the concavity of u. If workers are financially constrained, given Equation 4, their optimal consumption level is lower than $\bar{C}^{NC}$. This implies that $\mathbb{E}[\ln \bar{C} - w^d] \leq 0$ for all workers, which ensures that the consumption path is sustainable across months.

Envelope theorem, we get:[15]

$$\frac{\partial U^{(0)}}{\partial \sigma} = -\bar{C}\phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right)\left(R_B - \sigma.\frac{R_A + R_B\Phi\left(\frac{\ln \bar{C}-w^d}{\sigma} - \sigma\right)}{\phi\left(\frac{\ln \bar{C}-w^d}{\sigma} - \sigma\right)}\right) \tag{6}$$

Therefore, workers benefit from lower uncertainty whenever financing constraints are high, i.e., whenever R_A is small relative to R_B:

$$\frac{\Phi\left(\frac{\ln \bar{C}-w^d}{\sigma} - \sigma\right)}{\sigma}\left(\frac{\phi\left(\frac{\ln \bar{C}-w^d}{\sigma} - \sigma\right)}{\Phi\left(\frac{\ln \bar{C}-w^d}{\sigma} - \sigma\right)} - \sigma\right) \geq \frac{R_A}{R_B} \tag{7}$$

Equation 7 typically does not hold in the case of the financially unconstrained workers ($R_B = 0$) as the term on the right-hand-side converges to infinity in their case, while the term on the left remains finite. Workers who are financially constrained only have an incentive to decrease σ by increasing their attention effort (relative to the default) in order to better monitor their monthly budget constraint. Thus, attention is strongly correlated with financing constraints and, in this respect, can be interpreted as an indirect indicator of their magnitude.

Introducing attention. We now introduce the possibility that the workers under consideration make efforts to improve their knowledge of the realized value of w at a date $\tau \in [0; 1]$. They begin with the default optimal consumption policy computed above from optimizing $U^{(0)}$, and decide at date τ whether to improve their knowledge of w and adjust consumption if it is optimal to do so. The new objective function is written as:

$$\begin{aligned} U^{(m)}(\bar{\bar{C}}, \bar{C}) \quad &= \quad \tau u(\bar{C}) + (1-\tau)u\left(\bar{\bar{C}} + \frac{\tau(\bar{\bar{C}} - \bar{C})}{1-\tau}\right) - K(m)h(1-\tau) \\ &\quad -\bar{\bar{C}}\left(R_A + R_B.\Phi\left(\frac{\ln \bar{\bar{C}} - w^r}{\sigma_m}\right)\right) + e^{w^r + \frac{\sigma_m^2}{2}}\left(R_A + R_B\Phi\left(\frac{\ln \bar{\bar{C}} - w^r}{\sigma_m} - \sigma_m\right)\right) \end{aligned} \tag{8}$$

where $\bar{C}$ maximizes the initial objective function in Equation 4 while $\bar{\bar{C}}$ is the new optimal monthly consumption to be computed.

More importantly, m denotes the amount of "attention". To fix ideas, we model it as the amount of effort that is allocated to searching for a complementary signal s, which is assumed to be orthogonal to w^d without loss of generality,[16] but also Gaussian with mean w^f and variance σ_s^2. This signal allows the workers to decrease their mean squared error and improve their estimate of the "correct" wage via Bayesian updating

[15] See Appendix A.1 for further exercises of comparative statics.

[16] This only corresponds to a normalization assumption.

as:

$$w^r = m\ s + (1-m)\ w^d \tag{9}$$

$$= w^d + \underbrace{m\ (s - w^d)}_{\text{Residual adjustment}} \tag{10}$$

where w^r is the updated guess (which is ultimately "reported" in our data). In this equation, parameter $m = \frac{\sigma^2}{\sigma^2+\sigma_s^2}$ controls the Bayesian update process and can be interpreted as an "attention" parameter ranging between 0 and 1. If both w^d and s are distributed as Gaussian random variables, then the Bayesian update w^r is also distributed as a Gaussian random variable with probability density φ_m, mean w^f and variance σ_m^2:

$$\sigma_m^2 = (1-m)\ \sigma^2 \Longleftrightarrow m = 1 - \frac{\sigma_m^2}{\sigma^2} \tag{11}$$

Ultimately, the updated distribution of the wage remains log-normal, which rationalizes the specification of Equation 8.

Lastly, the cost of collecting this signal depends on the date τ at which the signal is obtained via the $h(1-\tau)$ term (the earlier, the more expensive) and on the overall informativeness of the signal via the $K(m)$ term. The latter can be understood as a pure cost of effort, or as the opportunity cost of reallocating mental resources to the monitoring of the budget constraint (Shah et al., 2018).[17]

For ease of understanding, consider first that the level of attention m and the date of information collection τ are exogenous. The first order condition for $\bar{\bar{C}}$ becomes:

$$u'\left(\bar{\bar{C}} + \frac{\tau\ (\bar{\bar{C}} - \bar{C})}{1-\tau}\right) = R_A + R_B \Phi\left(\frac{\ln \bar{\bar{C}} - w^r}{\sigma_m}\right) \tag{12}$$

Equation 12 shows that consumption is set to optimize the allocation of income across time, as in Equation 4 but with a more accurate depiction of the (expected) financing constraints. To simplify, we assume first that $w^r \approx w^d$, i.e. that the Bayesian updating process does not effectively alter the level of workers' guesses too much. In this case, whenever $m \in [0;1]$, the uncertainty σ_m surrounding the budget constraint is lower than previously, which enables workers to lower their precautionary buffer[18] ($\bar{\bar{C}} \geq \bar{C}$). As a consequence, the instantaneous consumption flow rises to a level which takes account of both this lower buffer and of the catch-up term $\frac{\tau.(\bar{\bar{C}}-\bar{C})}{1-\tau}$, which is due to the fact that consumption until date τ was too low. In some cases, however, the Bayesian update process could lead workers to revise the level their guess significantly, ie. w^r could significantly differ from w^d. In the latter case, this revision could counter the buffer effect and lead to

[17] Our assumption that the two dimensions of attentional effort (time consistency $1-\tau$ and depth m) are multiplicatively separable is obviously a simplification, as complementarities could be at play. However, estimating such complementarities would be excessively demanding in terms of data, and we think Equation 8 is already a useful first benchmark to test against the data, despite its restricted form.

[18] This follows from the concavity of u in Equation 12: it is easy to check that $\bar{\bar{C}} = \bar{C}$ does not qualify (it is too low) as $\Phi\left(\frac{\ln \bar{C} - w^d}{\sigma_m}\right) > \Phi\left(\frac{\ln \bar{C} - w^d}{\sigma}\right)$ (keeping the assumption that $w^r \approx w^d$).

a decrease in consumption at the end of the month.[19]

It is also informative to inspect how utility varies as either m or τ increases. Relying on the Envelope theorem, we get:

$$\frac{\partial U^{(m)}}{\partial m} = \frac{\sigma_m}{1-m}\frac{\bar{\bar{C}}}{2}\phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}\right)\left(R_B-\sigma_m\frac{R_A+R_B\ \Phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}-\sigma_m\right)}{\phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}-\sigma_m\right)}\right)-K'(m)h(1-\tau) \tag{13}$$

$$\frac{\partial U^{(m)}}{\partial \tau} = K(m)h'(1-\tau)-\left(u\left(\bar{\bar{C}}+\frac{\tau\ (\bar{\bar{C}}-\bar{C})}{1-\tau}\right)-u(\bar{C})\right)+\frac{\bar{\bar{C}}-\bar{C}}{1-\tau}u'\left(\bar{\bar{C}}+\frac{\tau\ (\bar{\bar{C}}-\bar{C})}{1-\tau}\right) \tag{14}$$

Equation 13 is analogous to Equation 5 and can only be positive if the workers under consideration face high values of R_B (relative to R_A), and if the ultimate gain from being better informed (in terms of consumption, net of financial costs) is higher than the marginal cost of paying attention, $K'(m).h(1-\tau)$. In other words, financially constrained workers only have an incentive to increase the attention level they allocate to monitoring their budget constraint, and they will only do so if the attention cost is not too high.

Lastly, Equation 14 clarifies the trade-offs involved in the determination of τ. Delaying the information collection effort lowers the cost associated with maintaining attention via the term h'. However, it comes at a cost: if the initial guess w^d was too low, then consumption is maintained for too long at a sub-optimal low level. Conversely, if the initial guess w^d was too high, then consumption has to be adjusted downwards by more than if the signal was collected earlier, thus leading to disproportionately low levels of utility at the end of the period (given the concavity of u). Ultimately, the optimal date for collecting additional information s about wages depends on the relative magnitude of the utility gain of increasing consumption smoothing on one hand, and the costs in terms of having to maintain the attention effort for a longer period of time on the other hand.

Endogeneizing attention. We now consider the more credible situation where both m and τ are endogenously selected by workers. The previous developments imply that financially constrained workers only have an incentive to set a strictly positive level of attention m. Among this population of workers, the first order optimality conditions are such that $\frac{\partial U^{(m)}}{\partial m}=0$ and $\frac{\partial U^{(m)}}{\partial \tau}=0$ in Equations 13 and 14, respectively (in the cases of interior solutions).

Taking the full differential of all three first order conditions enables an analysis of how the optimal attention level m, or equivalently, the standard deviation $\sigma_m=(1-m)^{\frac{1}{2}}.\sigma$, co-evolve with τ. Working with

[19] To simplify, we focus in this section on settings where both the prior w^d and the signal s are unbiased estimators of w^f. In reality, however, as previously discussed in footnote 10, they may be affected by behavioral biases, respectively b and B. In the latter case, w^r would also be affected by a bias $m.B+(1-m).b$. If B is very different from b, this could drive a large adjustment between w^d and w^r. Again, our data do allow us to test this assumption, but provide little supporting evidence, as shown in Section 5.3 (Figure 5, Panel C).

Equations 12 and 14, and for $\bar{\bar{C}}$ close to $\bar{C}$ we get:[20]

$$\frac{\mathrm{d}m}{\mathrm{d}\tau} \approx \frac{h''(1-\tau)\ K(m)}{h'(1-\tau)\ K'(m)} \tag{15}$$

As K, K', and h' are all likely to be strictly positive, the sign of $\frac{\mathrm{d}m}{\mathrm{d}\tau}$ is informative of the sign of h'', that is to say, of the shape of the cost function associated with maintaining a given level of attention over time. If $\frac{\mathrm{d}m}{\mathrm{d}\tau}$ is measured as negative in the data, then h is concave. In contrast, if it is positive, then this reveals that the cost h of maintaining attention across time is convex.

We also show in Appendix A that Equations 12 and 13 deliver bounds for K''. The derivation shows that the sign of $\frac{\mathrm{d}m}{\mathrm{d}\tau}$ is informative of the sign of K'': if $\frac{\mathrm{d}m}{\mathrm{d}\tau} < 0$, then K'' is necessarily negative and large in absolute value, i.e. K is necessarily concave. In the opposite situation where $\frac{\mathrm{d}m}{\mathrm{d}\tau} > 0$, then K'' is either positive or slightly negative but small in absolute value. This implies that K is either convex or not too concave. This result is intuitive, as in the opposite case, it would be cheap to pay attention during the entire month.

Synthesis of empirical predictions. To summarize, our simple theoretical framework delivers the following theoretical predictions:

- Financially constrained (or risk-averse) workers are likely to pay more attention overall (i.e. increase m, lower σ_m) in order to monitor their budget constraints.
- Financially constrained (or risk-averse) workers only have an incentive to vary their attention level over the month, and pay more attention m as the budget constraints tighten (or equivalently, to decrease their level of uncertainty σ_m).
- For financially constrained (or risk-averse) workers, the cyclicality of the attention effort is informative of the features of the cost function associated with paying attention. Most importantly, an increase in attention over the month (as suggested by our empirical results reported below) reveals that the cost of *maintaining* attention is convex while the cost of attaining a high level of attention is either convex or (at least) not too concave.

To our knowledge, our setting offers one of the few opportunities to approach the overall attention cost function and document its qualitative features. In the remainder of the paper, we aim to test these predictions against the data and we will contrast the results obtained in different populations of workers who are likely to face different magnitudes (and types) of financing constraints:

- First, *low-wage workers* are more likely to hit their financing constraints (embodied in the second integral term of Equation 1) and are therefore more likely to be exposed to the cost term R_B. We therefore expect that those workers will exhibit lower overall uncertainty σ_m, higher attention m and

[20] See Appendix A for the full derivation and discussion of these results.

higher cyclicality of these quantities during the month, from payday to payday. Our setting will allow us to estimate a (rough) bound in terms of wage, separating workers facing such financial constraints which they have to address by varying their (costly) attention level, and those who do not need to do so.

- Second, we investigate whether *women*, who are likely to be more risk-averse than men according to the literature (Borghans et al., 2009; Croson and Gneezy, 2009), exhibit different behaviors in terms of attention, both overall (in level) and in evolution across time.

- Similarly, we also investigate whether *workers with a shorter tenure* exhibit specific patterns. We hypothesize that, on average, these workers have accumulated less wealth as an additional safety buffer, such that they may exhibit higher values of R_B in Equation 1, all else equal. Furthermore, these workers are likely to be less informed about bonuses, payment of overtime, and other firm-specific dimensions of their employer's wage policy. This implies that the volatility σ of their default prior w^d is likely to be high, thus amplifying the expected costs associated with losses in the second integral term of Equation 1, and increasing their incentives to pay attention to mitigate this volatility.

- Lastly, we will also be able to incorporate proxies of education as a coarse measure of financial literacy in our empirical analysis. Financial literacy is likely to affect the relative gains and costs of paying attention, as collecting signals s and implementing Bayesian updates may not be straightforward for less-educated workers (Lusardi and Mitchell, 2014). This would be reflected in the attention cost functions K and h and would lead to a higher relative cost of paying attention in their case.

3 Data and Descriptive Statistics

3.1 Data Sources

Our main dataset is the French "Enquête sur les Revenus Fiscaux et Sociaux" (ERFS) survey. This dataset is constructed from the French implementation of the EU-wide Labor Force Survey (LFS) and is matched with income tax files. The resulting dataset provides us with values of wages that are self-reported by workers in the LFS, which can be compared with information about taxable wage income that is directly provided by the fiscal administration. These files have been used in several papers, mostly in the field of public economics (Aghion et al. (2017); Garbinti et al. (2018) among others).[21]

Table 1 describes the structure of the ERFS files in greater detail. The LFS component of the dataset consists of 4 rotating panels of individuals who are interviewed in 6 consecutive quarters, either face to face

[21] More closely related to our own work (but in French) is Biscourp et al. (2005), who show that self-reported, rounded data, in the absence of adequate econometric treatment of the type in Section 4, do not allow us to accurately quantify nominal wage rigidity. Prati (2017) also proposes a measurement of hedonic recall biases using a dataset which also contains self-reported values of wages (from SalSa, "Enquête sur les Salaires Auprès des Salariés) and administrative information about annual wages at the worker level sourced from the DADS. We introduce this statistical source in Section 4 in order to construct complementary indicators of wage volatility. The SalSa dataset has however no panel dimension and the sample is six times smaller than our final estimation sample (ca. 3,000 observations, against more than 19,000 in our final sample).

or by phone. They are asked about their wage in the first and sixth waves of the interviews. Each panel of the LFS is then matched with the fiscal data for the fourth calendar quarter. This implies that Panels 3 and 4 (in Table 1) are matched twice with the fiscal data. This feature provides a short but very useful panel dimension. Our estimation sample is ultimately restricted to workers surveyed in the Panels 3 and 4 of the LFS, between 2005/2006 and 2015/2016.[22]

The sampling scheme of this dataset is critical for us. It is based on pre-defined clusters of 17 to 23 accommodations which are randomly sampled and then exhaustively surveyed within two-week windows. The small size of these clusters and their randomization ensure that the precise dates of the interviews are a little clustered but still close to randomly sampled. As a result, they appear to be uncorrelated with observed (and unobserved) characteristics of workers, as documented in the methodological documentation of the Statistical Institute[23] as well as in Table 9 in Appendix C. This orthogonality between the dates of interviews and worker-level characteristics is a critical feature of the survey which allows for temporal analyses (presented in Section 5.3) that are unaffected by self-selection issues (Carvalho et al., 2016; Mani et al., 2020).

Table 1: Description of the Panel Structure of ERFS: the Matched Fiscal (POTE) Data and French Labor Force Surveys

		Year $t-1$				Year t				Year $t+1$			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Panel 1	Wave of Labor Force Survey	1	2	3	4	5	6						
	Wage reported in LFS	×					×						
	Fiscal wage				×								
Panel 2	Wave of Labor Force Survey		1	2	3	4	5	6					
	Wage reported in LFS		×					×					
	Fiscal wage				×								
Panel 3	Wave of Labor Force Survey			1	2	3	4	5	**6**				
	Wage reported in LFS			×					×				
	Fiscal wage				×				×				
Panel 4	Wave of Labor Force Survey				**1**	2	3	4	5	6			
	Wage reported in LFS				×					×			
	Fiscal wage				×				×				

Notes: In this table, Q1, Q2, Q3 and Q4 denote quarters 1 to 4 in a given year. Each "Panel" of the LFS is surveyed six times, during six consecutive quarters. The rolling panel structure of the survey implies that there are four types of panel, each surveyed over a different set of quarters. Self-reported information about wages is only collected in the first and sixth interrogations, while fiscal information is only introduced once a year in interrogations corresponding to the fourth quarter.

3.2 Harmonization of Wage Concepts

One difficulty for our analysis is that the baseline concepts of wage differ in the LFS and in the fiscal data. For a large fraction of surveyed workers, it is, however, possible to adjust fiscal wages and make them

[22] One important variable (job title) is missing from the source files in 2013 and 2017, which means that our estimation sample is restricted to the following years: 2005/06, 2006/07, 2007/08, 2008/09, 2009/10, 2010/11, 2011/12, 2014/15, 2015/16.

[23] The two-stage sampling scheme results from the trade-off between pure randomization (clusters of size one at the limit) maximizing statistical power - i.e. minimizing the correlation across characterisics of workers within clusters - and the mitigation of data collection costs. The Statistical Institute computed the optimal cluster size allowing for a negligible impact in terms of clustering while allowing for significantly lower data collection costs.

comparable to the wage concept of the LFS.[24]

LFS concept of wage. The LFS concept of wage corresponds to the monthly wage earned by workers for their main job in the LFS survey, including monthly bonuses. The following question is asked in face-to-face interviews: "What is the total monthly compensation that you receive from your main occupation?" The answer given by workers can refer to wages that are either gross or net of social contributions.[25] In principle, workers could retrieve this information from their bank account (net wages) or their payslip (both gross or net wages). However, they are not required to do so, first, in order to limit interviewing times and, second, to mitigate feelings of intrusion and thus increase response rates. Furthermore, payslips for a given month are typically edited and given to workers with a delay of several weeks, so that they are not always in possession of them at the time of the LFS interviews. As a consequence, answering this question boils down to a wage nowcasting exercise. In France, the main component of a wage is typically highly stable, as it corresponds to workers' core labor contract. Nowcasting a wage therefore means that workers have to remember this component and guess the variable elements (monthly bonuses, overtime), if any. The strategy used by workers to complete this task depends on their level of financial literacy. In addition, as discussed in Section 4, their survey reporting behavior may not be entirely truthful. One of the important challenges of this paper is address these potential biases and devise a robust estimation strategy for their actual level of attention.

Concept of wage in the fiscal files. Income tax returns contain information about taxable wage income. Since 2006, the corresponding item in the fiscal form is pre-filled with information provided by employers. Taxpayers are allowed to correct the reported amount if required, but most wage earners do not have to alter anything. Fiscal wage earnings encompass wages (including overtime), bonuses and the following additional components, when applicable: paid leave, payments on termination of contract, complementary financial support from the employer or a staff committee ("comité d'entreprise") when the corresponding amount is higher than €1,830. Furthermore, social contributions and the deductible fraction of the "generalized social contribution" are excluded.[26]

Harmonization and resulting sample selection. The concepts of wages in the LFS and in the fiscal sources are broadly consistent, apart from a small number of difficulties. First, as previously stated, workers are allowed to report wages that are either gross or net of social contributions.[27] As documented in Appendix Table 8, only 60% of full-time, employed workers actually accept to report their wage in the LFS. Most of the

[24]As explained below in Section 4 in greater detail, our analytical framework is robust to time-invariant differences in wage concepts. The main purpose of the treatments described in this section and of the further selections described in Section 3.3 is to remove (as much as possible) the differential components that are likely to evolve over time.

[25]This is the first question in the "Labor Income" block of the survey. The exact question (SALMEE) is: "Quelle rémunération totale mensuelle retirez-vous de votre profession principale ? (salaire du dernier mois, y compris primes et compléments mensuels)". A technical document intended for the interviewers of the National Statistical Institute specifies that payments to complementary health insurance made directly by employers on behalf of employees should be excluded, and that the question refers to the concept of the wage slip. The variable TYPSAL indicates whether reported wages are to be understood as gross or net of social contributions.

[26]2.4% of the generalized social contribution (CSG) and 0.5% of the contribution to the reduction of the social debt (CRDS) are not deductible.

[27]Net wage corresponds to gross wage net of social contributions but not of income taxes.

respondents (ca. 90%) report net wages. To ensure consistency in our comparisons between self-reported and fiscal wages, we simply select net wages as our baseline concept and discard the remaining 10% of workers who report gross wages.[28] We then adjust the fiscal variables are then adjusted in accordance with the official tax schemes and compute net wages. Second, a small number of occupations benefit from very specific tax treatments, which means that the fiscal wage item is a poor proxy of their actual wages. We therefore also discard the corresponding workers:[29] journalists, artists, apprentices, childminders, local officials, scholarships (i.e. students), and any workers hired by individual employers under the CESU[30] (pre-financed vouchers) scheme.

The greatest difficulty for us stems from the fact that wages in the LFS correspond to monthly wages, while the fiscal information is an annual aggregate, which we normalize throughout by 12 to obtain a monthly average. For workers with smooth career paths, the difference between the two concepts is negligible. Where a worker has experienced unemployment spells or job/employer changes, however, the annual fiscal average is likely to be a poor proxy of the monthly wage surveyed in the LFS. We therefore restrict our sample to workers with *tenure* in the same firm longer than 15 months. Obviously, this is a little disappointing, as shorter-tenure workers would have been a very interesting population to study, but only very few of them actually report their wages, such that by discarding them, we only loose fewer than 1.4% of observations. In contrast, this selection has two main benefits for our purposes: first, it selects relatively stable workers whose annual averages are a good approximation of monthly wages. Second, the 15-month threshold prevents fiscal proxies from being polluted by any severance payments as a result of job changes. To mitigate this risk even further, we also discard workers who report having multiple employers and those working part-time, as the latter are more likely to work irregular overtime than full-time workers.

Finally, one last minor difficulty lies in the treatment of bonuses and other elements of incentive pay, as in the LFS most workers report wages that are net of bonuses, particularly annual bonuses,[31] whereas they are included in the fiscal proxy. This difficulty is relatively easy to solve using the rich set of complementary indicators which describe the various components of incentive pay that are available from the LFS. Our overall strategy is to leave the self-reported wages untouched, whether gross or net of bonuses. Instead, we rely on the information about bonuses provided by the LFS in order to net these components out of the fiscal wage when necessary (i.e. when self-reported wages are declared to be net of such bonuses). Ultimately, we focus on workers whose resulting fiscal wage ranges between €1,000 and €4,000 in order to avoid cases that are too atypical, both at the bottom and at the top of the wage distribution.

Elicitation of the behavioral parameters in the LFS (survey) data. As will be explained in detail in Section 4, the econometric comparison between fiscal and self-reported wages will enable us to measure

[28] This also corresponds to the concept of wage that is best understood by workers (and most salient to them), as it appears clearly at the bottom of all payslips (both gross and net wages are compulsory payslip items under French law) and it corresponds in almost all cases to the actual payments made by employers to their workers' bank accounts until 2019.

[29] Unfortunately, information about the precise occupation of workers is missing from the 2013 and 2017 files. This explains the temporal "hole" in our dataset in 2012/2013 and 2013/2014 mentioned in footnote 22.

[30] The acronym stands for "Chèques-Emploi Services Universels".

[31] Monthly bonuses are either incorporated into wages or reported separately (PRIM).

the worker-level uncertainty parameter σ_m and the corresponding level of attention m, both overall and at different dates in the calendar month. At this stage, it is important to note that workers who are requested to report their wage in the LFS are *not* incentivized to do so accurately[32] in any way. In addition, the interviews during which this survey item is requested are typically short: the INSEE methodological document states that the first interview, which is performed face to face, typically takes less than 17 minutes. The sixth interview also occurs face to face, and only takes 7 minutes on average. In these circumstances, we should not expect workers to make any particular effort in responding to the survey. The response strategy which is least costly to them is probably to report amounts spontaneously, off the top of their heads, thus providing a direct measure of uncertainty at the date of the interview and which is not affected by any artefact or experimental incentive. However, as explained in Section 2, we do expect that the external incentives to pay attention to their wage may be differentiated across populations of workers with different magnitudes of liquidity constraints, and that it might fluctuate over time within the population of the most financially constrained workers. The fact that the LFS itself is neutral in terms of incentives will enable us to more accurately elicit these external heterogeneity and fluctuations.

3.3 Description of the Estimation Sample

As explained above, fiscal wages can be converted to the relevant wage concept, i.e. the wage concept of the LFS for only a sub-sample of workers. In future years, the French legal framework may enable wider matching of the LFS data with a greater variety of administrative sources, which would make this selection process useless.[33] At this stage, however, our estimation sample has to be restricted to the most "stable", full-time workers, in the sense that their work history should not create large divergences between the wages reported in the LFS and in the fiscal forms. Our resulting sample contains 19,045 observations spanning the period from 2005 to 2015. The underlying population of workers is described in Table 2. The sampling weights of the ERFS show that our sample is representative of around 2.5 million workers, which represent roughly 30% of the full population of French workers aged 15 to 64 who are continuously employed. Appendix Table 7 checks that these rates are stable across survey years. Roughly three-quarters of our sample observations correspond to male workers, and more than two-thirds are between 35 and 54 years old. Half of the sample have a high school degree or higher diploma. A similar proportion are in occupations which correspond to intermediate or higher managers, while 34% are blue-collar workers and 20% are low-skilled white-collar workers. Finally, the average tenure is 179 months (15 years), thus exceeding our selection threshold of 15 months by a large amount. Our overall selection procedure thus implies that the workers in our sample are, on average, somewhat older than the population of continuously employed workers. They are more educated and have longer tenure. Importantly, the ratio of female to male workers is significantly lower in the population of workers which we focus on than in the general population of employed workers, mainly because of our selection of full-time workers.[34] Lastly, Panel (B) in Table 2 shows that the distribution of wages is slightly shifted

[32] Note that they have also no incentive to report truthfully - but our econometric strategy will allow us to measure the corresponding biases and purge their impact on the estimates of our main parameter of interest, σ_m.

[33] The almost ideal data exist for French workers (and have been constructed within the French Institute of Statistics), but our administrative efforts so far to gain access to the matched files remain unsuccessful.

[34] In France, part-time workers are disproportionately female workers.

to the right, which is unsurprising given that the workers in our sample are selected according to tenure. These differences are, however, limited. Our filters do not select specific profiles of workers, but instead, specific and stable periods of their career paths. They mainly discard workers with recent important "breaks."

Appendix Table 8 describes in detail which step of our selection procedure is the most stringent, and what is the impact in terms of the composition of the sample. It shows that our sample size is mainly limited by non-responses to the wage items rather than by the different filters we apply to harmonize self-reported and fiscal wages. Furthermore and unsurprisingly, it is the criterion related to tenure which mainly selects more experienced workers. This also means a selection of slightly older workers, together with our criterion selecting wages ranging between €1,000 and €4,000. The criterion related to full-time workers has the main (downward) impact on the ratio of female workers. Appendix Table 9 further verifies that worker characteristics are not correlated with the specific day of their interview (within the month), a feature which is ensured by the sampling design of our data set.

As expected, our sample selection procedure removes a significant share of the divergence between self-reported and fiscal wages: the portion that is generated by career breaks. Denoting by w_{it}^r and w_{it}^f the logarithm of the self-reported and fiscal wages respectively, we obtain that relative gaps $\left|w_{it}^r - w_{it}^f\right|$ decrease on average from 15% to 10% when we eliminate the most unstable workers. In the empirical analyses which follow, we consider that the remaining 10% divergence between self-reported and fiscal wages is mainly driven by workers' recall and response behavior, of which roughly half (5.2%) is stable across time and half (4.8%) is time varying. Our estimates of attention will precisely be based on this statistical information.

3.4 Wage Distributions

We now turn to the description of our two main variables of interest and start with their distributions. Chart (A) in Figure 1 depicts the distribution of wages that are self-reported by workers in the LFS for the years 2005 to 2015, while Chart (B) depicts the distribution of fiscal wages. As a reminder, both distributions relate to the same (harmonized) concept and to the exact same population of workers.

As in Biscourp et al. (2005), we find that the distribution of wages reported by workers differs significantly from the distribution of their fiscal counterpart. The support of the two distributions is broadly similar but, while the fiscal data exhibit a rather continuous density function,[35] the distribution of self-reported wages instead resembles a mixture of several overlaid discrete distributions: the first with discrete mass points by steps of €500, the second with discrete mass points by steps of €100 and the third with discrete mass points by steps of €50. This contrast between the two distributions is clear evidence of the fact that a significant share of self-reported wages are rounded with a mixture of different levels of "coarsening" of the underlying continuous wage information. This cerates "bunching", or more precisely, mass points at each salient value

[35] Note that, in this respect, the French administrative wage data are different from the US administrative wage data and do not feature bunching at multiples of €10 even when delineating in terms of net or gross wage per hour rather than by month (Dube et al., 2018).

Table 2: Estimation Sample: Representativeness

(A) Overall Statistics

		Sample		Labor market aged 15 to 64	
		Un-weighted	Weighted (mil. workers)	All (mil. workers)	Emp. $t/t+1$ (mil. workers)
Age:	15 to 24	0.024	0.025	0.065	0.052
	25 to 34	0.208	0.237	0.214	0.242
	35 to 44	0.314	0.318	0.254	0.303
	45 to 54	0.344	0.322	0.256	0.293
	55 to 64	0.110	0.098	0.211	0.110
Gender:	Female	0.239	0.240	0.319	0.316
Education:	No diploma (low)	0.287	0.293	0.285	0.334
	Lower than high school (low)	0.180	0.186	0.175	0.183
	High school degree (high)	0.314	0.306	0.270	0.267
	Higher than high school (high)	0.219	0.215	0.270	0.215
Occupation:	Managers/professionals	0.144	0.147	0.142	0.198
	Intermediate occupations	0.310	0.313	0.195	0.274
	Low-skilled white-collars	0.205	0.208	0.190	0.253
	Blue-collars	0.342	0.332	0.218	0.274
Tenure:	Average (months)	179	172	142	145
	Std dev.	(122)	(120)	(125)	(123)
Observations:	Total	19,045	22.896	116.028	68.481
	Per annual wave	2,116	2.544	12.892	7.609

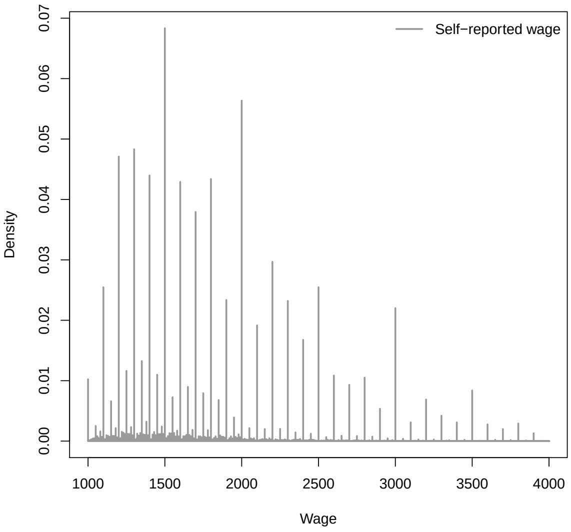
(B) Wages and Wage Errors (Ln-pts)

		Sample (weighted)	Labor market: Emp. t/t+1
Fiscal wage Distribution:	wage $\leq$ €1,200	0.100	0.200
	€1,200 $<$ wage $\leq$ €1,300	0.081	0.075
	€1,300 $<$ wage $\leq$ €1,400	0.093	0.081
	€1,400 $<$ wage $\leq$ €1,500	0.086	0.074
	€1,500 $<$ wage $\leq$ €1,600	0.083	0.071
	€1,600 $<$ wage $\leq$ €1,700	0.080	0.064
	€1,700 $<$ wage $\leq$ €1,800	0.072	0.058
	€1,800 $<$ wage $\leq$ €1,900	0.058	0.047
	€1,900 $<$ wage $\leq$ €2,000	0.050	0.041
	€2,000 $<$ wage	0.296	0.290
Reporting errors: (Average, ln-pt)	Overall	0.100	0.147
	Time varying only	0.048	0.079

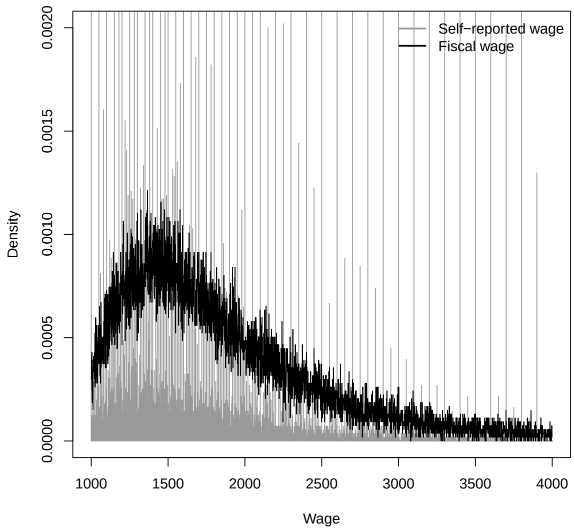
Source: ERFS survey, 2005 to 2015. In Panel (A), Columns 2 to 4, observations are weighted by their ERFS sampling weights to provide an estimate of the considered population, in million workers. As a reminder, the fiscal wages of our sample observations are restricted to the range €1,000 to €4,000 while those of the general population of workers who are employed both at t and $t+1$ (i.e. at the 2 waves of the LFS) are unrestricted. See Section 3.2 and Appendix C for a complete description of our sample selection.

Figure 1: Wage Distributions:
Self-Reported *vs.* Fiscal Data

(A) Wages reported in the LFS



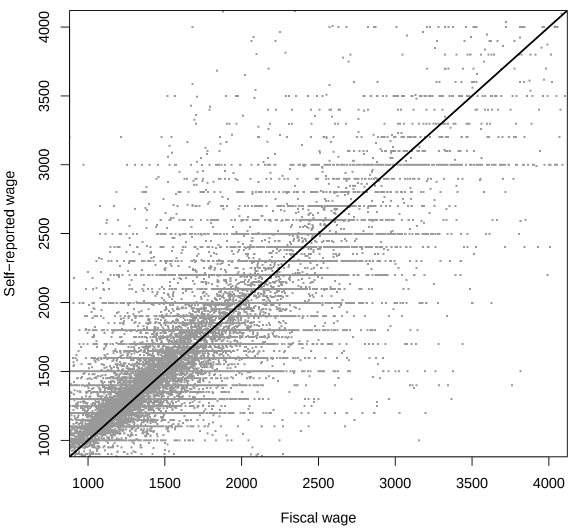
(B) Wages in the fiscal data



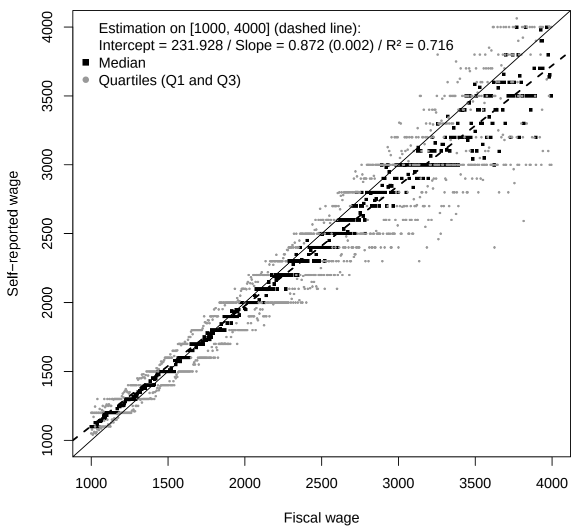
Notes: Chart (A) shows the histogram distribution of wages as self-reported by workers in the LFS, by bins of €1. Chart (B) plots in black the histogram distribution of fiscal wages, also by bins of €1, overlaid with the distribution of LFS wages in light gray for comparison. Note that the scales of y-axes differ between Charts (A) and (B). Source: ERFS (2005-2015).

Figure 2: Correspondence between Self-Reported and Fiscal Wages

(A) Raw Data



(B) Quartiles of Self-Reported Wages by Bins of €5 of Fiscal Wages



Notes: This figure plots the wage reported in the Labor Force Survey against the fiscal wage available from the income tax files over the interval €1,000 to €4,000. Source: ERFS (2005-2015).

of the corresponding discrete scales (Kleven, 2016).

The discrepancies between self-reported and fiscal wages can be further investigated by looking at the direct correspondence between the two variables. The results are reported in Figure 2 and are particularly insightful. Chart (A) shows that the relationship between fiscal and self-reported wages is not a simple 45 degree line. First, an unstructured "cloud" of data emerges and appears to be much thicker than a line. Second, a rather complex pattern of horizontal lines overlays this first unstructured "cloud". The resulting grid indicates who the rounders are. Figure 2 shows, for example, that workers who reports a wage of €2,000 in the LFS have a fiscal wage ranging between €1,000 and €3,500. Chart (B) in Figure 2 further provides the main quantiles of the distribution of self-reported wages that are associated with a given value of the fiscal wage. These quantiles almost always coincide with round numbers, due to the prevalence of rounders in our population of workers. The interquantile range increases significantly as the fiscal wage increases, and the distribution of self-reported wages shifts significantly below the 45 degree line. This pattern is confirmed by the estimated correlation between self-reported and fiscal wages, which is significantly lower than 1 (0.87) and thus indicative of some form of under-reporting. This feature has to be taken into account in our econometric strategy.

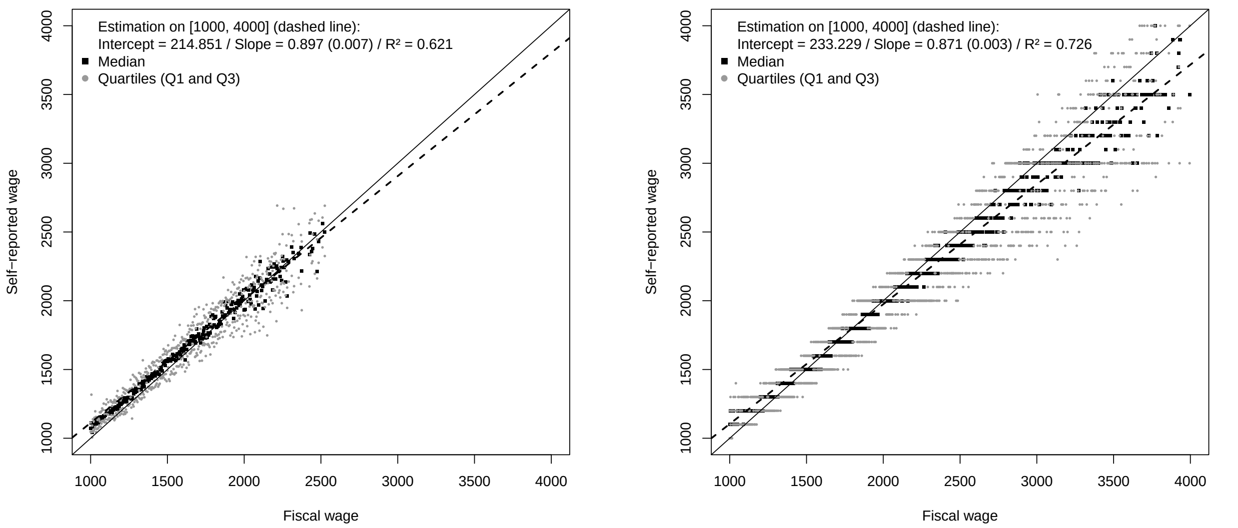
Prevalence of rounding. So far, the data clearly indicate that our sample is a mixture of different populations of rounders and non-rounders,[36] with potentially different reporting behaviors. Figure 3 and Table 3 offer a first descriptive investigation of this point. Specifically, Figure 3 breaks down Figure 2 across non-rounders (in Chart A) and workers reporting values that are multiples of €100 (in Chart B). This exercise provides first evidence that non-rounders tend to be characterized by a low fiscal wage, while workers reporting values that are multiples of €100 are distributed over the entire support of the wage distribution. Second, "rounders" tend to under-report a little more, but not by much, as the correlation between self-reported and fiscal wages is 0.87 across rounders but only slightly higher, at 0.90, across non-rounders. Interestingly, we also find that the R^2 of the regressions of the self-reported wage against the fiscal wage is not higher, and even 10 percentage points lower across non-rounders, which implies that their guesses are not necessarily more accurate than those of rounders. Our econometric analyses below will confirm this finding and suggest that, in our setting, rounding seems to be more correlated with workers' wage levels and education (most likely reflecting that being more financially literate enables workers to better evaluate the uncertainty around their own guesses) than with a lower accuracy.

Table 3 complements these descriptive statistics by providing a comprehensive description of the frequency and stability of workers' rounding behavior. It shows that in each wave of the LFS, 80% to 88% of workers report values that are multiples of €10, and the vast majority of them (73%) consistently report multiples of €10 in both of the survey interviews. In contrast, fewer than 1% of our sample observations are associated with fiscal wage values that are multiples of €10, as is also clear from the fact that the distribution of the

[36]As explained below, a worker who reports a rounded value is not necessarily a rounder. Our econometric approach addresses this issue rigorously via a mixture model.

Figure 3: Correspondence between Self-Reported and Fiscal Wages

(A) Self-reported wages that are not multiples of €5 (B) Self-reported wages that are multiples of €100



Notes: This figure plots for each bin of €5 in terms fiscal wage the quartiles of the self-reported wages. Sources: ERFS (2005-2015).

Table 3: Rounded Values in Self-Reported *vs.* Fiscal Wages

Shares of wages	at t		at $t+1$		at t and $t+1$	
Multiples of:	LFS	Fiscal files	LFS	Fiscal files	LFS	Fiscal files
€1	1	1	1	1	1	1
€10	0.794	0.008	0.877	0.008	0.727	0
€50	0.707	0.002	0.783	0.001	0.606	0
€100	0.619	0.001	0.677	0	0.478	0
€500	0.177	0	0.196	0	0.074	0
€1,000	0.077	0	0.084	0	0.031	0

Notes: This table describes the frequency of wage values that are multiples of, respectively, 10, 50, 100, 500 and 1,000, either in the LFS (self-reported wages) or in the fiscal data. Sources: ERFS (2005-2015).

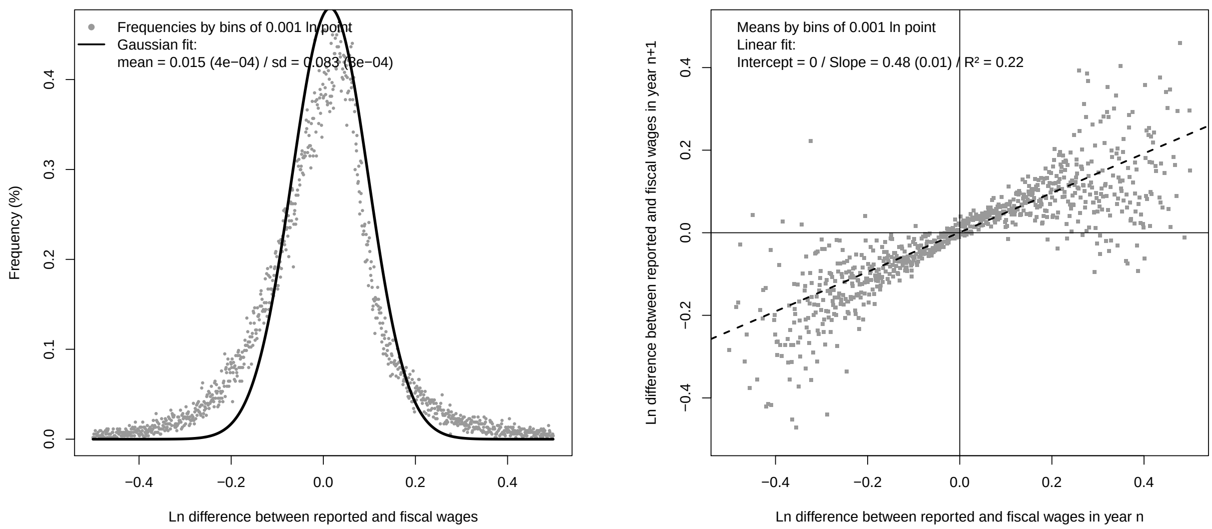
fiscal wage in Figure 1 does not feature bunching. Again, more than 60% of workers report values that are multiples of €100 in either wave of the LFS, and 48% of them consistently report multiples of €100 in both waves. These numbers confirm that rounding is actually pervasive and that it appears to be rather stable at the worker level across waves of the LFS.

Distribution of "errors". Lastly, Figure 4 provides a detailed description of the "errors", defined as the difference between the logarithm of self-reported wages and the logarithm of their fiscal counterparts. These errors are our empirical object of main interest because they capture the attention effort of workers (the higher the effort, the lower these "errors"). At this stage, we leave aside the measurement problems posed by rounding behavior and simply show in Chart (A) that their distribution is very close to Gaussian. It is almost centred on zero but features non-negligible dispersion, as the estimated standard deviation is 0.082. Chart (B) also documents that the "mistakes" that workers make are highly positively correlated across waves of the LFS, whether they are positive or negative, thus suggesting that the magnitude and direction of these mistakes are rather stable over time. In addition, the data cloud is conical, which signals some form of heteroscedasticity in the time correlations of these "errors." This feature of the data may be driven by the fact that our population of workers is composed of a mixture of rounders and non-rounders, with potentially different reporting behaviors. The empirical setting laid out in Section 4 addresses all these potential sources of concern for our strategy to measure workers' attention.

Figure 4: (ln) Differences between Self-Reported and Fiscal Wages

(A) Cross-sectional distribution

(B) Correlation across Years (Waves 1 and 2 of the LFS)



Notes: Chart (A) describes the distribution of $w^r_{it} - w^f_{it}$, where w^r_{it} denotes the logarithm of self-reported wages and w^f_{it} denotes the logarithm of the fiscal wage. The Gaussian fit of the obtained distribution has mean parameter 0.023 and standard deviation 0.082. Chart (B) plots the temporal correlations across waves of the LFS, i.e. $w^r_{it} - w^f_{it}$ against $w^r_{i,t+1} - w^f_{i,t+1}$. The obtained correlation is 0.43. Source: ERFS (2005-2015).

4 Empirical Framework

4.1 Baseline Variance Analysis Setting

Analyzing the amount of attention that workers allocate to their wage involves estimating the magnitude of their errors when they guess their wages, or more precisely, the variance of these errors. As documented in Section 3, our dataset allows precisely such a comparison of two different wage proxies: the value that is self-reported by workers in the LFS and the "correct" value indicated in the fiscal files. The difference between both provides an estimate of the variance parameter σ_m in Section 2. Overall, the empirical exercise described in this section thus boils down to a variance analysis exercise, which in our data is complicated by the prevalence of rounding and of potential additional reporting biases.[37] To that end, we take advantage of the fact that the wage of each worker is surveyed twice and that our data therefore feature a short panel dimension. This allows us to estimate the following random effect model:

$$w_{it}^{r} = w_{it}^{f} + a_i + \upsilon_{it}, \tag{16}$$

where w_{it}^{r} denotes the logarithm of the latent variable corresponding to the reported wage, w_{it}^{f} is the logarithm of the fiscal wage, a_i denotes a random effect component[38] at the worker level, and υ_{it} denotes the residual term. The a_i term typically captures (and filters out) the stable determinants of workers' reporting behavior which may be driven by the interviewing conditions of the LFS or by any environmental or worker-level characteristics. On the other hand, the residual term υ_{it} captures the time-varying discrepancy between w_{it}^{r} and w_{it}^{f}, i.e. the ignorance or mistakes of workers. The variance of υ can be interpreted as the overall measure of uncertainty at the worker level, coinciding with the variance parameter σ_m^2 introduced in Section 2.

Our central exercise of variance analysis is performed under the assumptions that a_i and υ_{it} are orthogonal and normally distributed, which is consistent with Section 2 and the suggestive descriptive statistics in Section 3:

$$\begin{aligned} a_i &\sim \mathcal{N}\left(\mu_a, \sigma_a^2\right) && (17)\\ \upsilon_{it} &\sim \mathcal{N}\left(0, \sigma_m^2\right) && (18)\\ \mathrm{Corr}\left(a_i, \upsilon_{it}\right) &= 0 && (19) \end{aligned}$$

A difficulty of our setting is that Equation 16 cannot be directly estimated in the data as a significant proportion of workers report rounded values (see Section 3). There are several ways of introducing workers' rounding behavior in our framework. Our baseline specification simply assumes that some workers implement standard, symmetric rounding, thus changing the Equation 16 model to a latent class model with a limited

[37]The psychology literature indeed documents that survey questions about earnings are often perceived as particularly intrusive and tend to generate reporting biases (Tourangeau and Yan, 2007), which are of no interest to our research question and should be filtered out.

[38]The specification with random effects rather than fixed effects is constrained by the short panel dimension of our data, and by the fact that a significant share of workers report rounded values, thus transforming our estimation problem into a limited dependent variable problem (see below) where fixed-effect estimators are biased at finite distance.

dependent variable:[39]

$$\begin{aligned}\tilde{W}_{it}^{r} &= N_i \left\lfloor \frac{e^{w_{it}^{r}}}{N_i} + 0.5 \right\rfloor \\ &= N_i \left\lfloor \frac{e^{w_{it}^{f}} e^{a_i} e^{\upsilon_{it}}}{N_i} + 0.5 \right\rfloor, \ t \in \{1, 2\} \end{aligned} \tag{20}$$

In this equation, $\tilde{W}_{it}^{r}$ corresponds to the actual wage value reported by workers in the LFS form (while w_{it}^{r} is the corresponding log-latent variable) and $N_i \in \mathcal{N} \equiv \{1, 10, 50, 100, 500, 1000\}$ denotes the class of rounding. Workers with $N_i = 1$ correspond to non-rounders, while $N_i \in \{10, 50, 100, 500, 1000\}$ corresponds to the five main classes of rounded values that are documented in Section 3.[40] In robustness checks, we also estimate a variant of Equation 20 which allows for left-digit bias, as in Busse et al. (2013) or Lacetera et al. (2012). This simply involves removing the normalizing constant 0.5 in the squared brackets in Equation 20.[41] It is also important to keep in mind that the precise level of rounding of each worker is not fully observed, as a worker reporting a multiple of €100 might actually be a non-rounder if the quantity $e^{w_{it}^{f}} e^{a_i} e^{\upsilon_{it}}$ appears to be a multiple of €100 or a rounder of lower class (e.g. €10 or €50). The estimation procedure described in Section 4.3 takes account of this feature of our estimation problem. Furthermore, we estimate one set of parameters $\left\{\mu_a^N, (\sigma_a^N)^2, (\sigma_\upsilon^N)^2\right\}$ by class of rounding N, thus allowing for any correlation between rounding and worker-level random effects a_i or uncertainty υ_{it}.[42]

4.2 Discussion

The empirical framework set out in Section 4.1 is fully flexible in the sense that it does not take a stand about the precise cognitive or reporting processes which lead to the values that are self-reported by workers in the LFS. For example, it is possible that employees first make a continuous guess, then potentially introduce reporting biases a_i, and finally decide to only report a rounded value of the resulting index to blur the information that is ultimately made public.[43] Alternatively, the worker-level biases a_i and/or rounding could be behavioral, i.e. could reflect distortions of workers' information sets rather than simply being features of their reporting strategy in the interaction with the LFS interviewer. While it is impossible to credibly identify from our data which of these two alternative stories is most relevant for the interpretation of a_i or rounding, the volatility premium term σ_m^2 can more credibly be interpreted as an actual behavioral characteristic. Indeed, it seems unlikely that this time-varying component could be manipulated in the interaction between survey respondents and interviewers or manipulated by workers. In addition, σ_m is

[39] The notation $\lfloor y \rfloor$ corresponds to the largest integer which does not exceed y.

[40] Our methodology allows, in principle, an arbitrarily large number of classes, provided sample size is sufficiently large.

[41] Our setting allows, in principle, for an even more flexible specification where the normalizing constant could be estimated. However, our results show that the baseline specification which assumes "correct" rounding actually fits the data better. Therefore, we only report the specifications with left-digit bias for comparison with the literature.

[42] This flexibility of our specification also eases estimation computationally, as it splits the different models associated with each class of rounding into independent sub-problems, which can be optimized separately at each iteration of the EM algorithm (see Section 4.3 and Appendix B). We also systematically report averages $\left\{\mu_a, (\sigma_a)^2, (\sigma_\upsilon)^2\right\}$ across classes and evaluate the standard deviations of the corresponding estimators via delta-method.

[43] Workers are also given the opportunity to report binned values. However, the bins are predefined in the survey and do not contain any information about worker-level uncertainty, σ_m. This is why we have to discard the corresponding observations reported as "missing wage information" in Appendix Table 8.

of particular interest as it corresponds to a volatility premium whose magnitude may result from workers' attention effort, as formalized in Section 2.

4.3 Estimation Strategy

Estimating Equation 20 is not entirely straightforward. To better describe the structure of the estimation problem, we introduce more compact notations. Let Ω_i denote the logarithm of the wages $\{w^r_{1i}, w^r_{2i}\}$ reported by worker i in the first and second waves of the LFS, X_i denote the logarithm of the fiscal wages $\{w^f_{1i}, w^f_{2i}\}$, $(\theta^n) = \{(\mu^n_a), (\sigma^n_a), (\sigma^n_m)\}$ denote the sub-set of parameters to be estimated which determine worker-level reporting biases and uncertainty[44] and $(\pi_n)_{n\in\mathcal{N}}$ the set of parameters controlling the probabilities of rounding. Under the assumptions described in Section 4.1, the contribution of an observation i to the ln-likelihood can be written as:

$$l(\Omega_i, N_i | X_i, \theta, (\pi_n)) = \ln\left(\sum_{n\in\mathcal{N}} \pi_n \mathbb{P}(\Omega_i | N_i = n, X_i, \theta^n)\right) \tag{21}$$

The structure of the latent class mixture model embodied in Equation 21 is made up of two different sets of parameters:

1. The six probabilities π_n for $n \in \mathcal{N} \equiv \{1, 10, 50, 100, 500, 1000\}$ that worker i rounds their wage by bins of width n. These probabilities add up to one, such that there are only 5 parameters to be estimated.
2. The six conditional probabilities $\mathbb{P}(\Omega_i | N_i, X_i, \theta^n)$ to observe Ω_i conditionally on N_i, X_i, and θ^n. These conditional probabilities are simply 0 whenever $e^{w^r_{i1}}$ or $e^{w^r_{i2}}$ are not multiples of n. In the alternative case where both quantities are actually multiples of n (i.e. $n | e^{w^r_{i1,2}}$), estimating them involves a standard model of limited dependent variable with Gaussian random effects.

Maximizing Equation 21 is difficult because of the summation within the logarithm. It is performed using an EM algorithm (Train, 2003) The general principle of this algorithm is to alternate between performing an expectation ("E") step, which creates a function for the expectation of the log-likelihood evaluated using the current estimates of the parameters, and a maximization ("M") step, which computes parameters maximizing the expected log-likelihood found in the E step. In our setting, the E step involves estimating the set of probabilities $(\pi_n)_n$ as the posterior probabilities that a worker i rounds at level n, conditionally on the observables (X_i, Ω_i). The M step involves estimating the conditional probabilities $\mathbb{P}(\Omega_i | N_i = n, X_i, \theta^n)$ which maximize the log-likelihood for given $(\pi_n)_n$. We rely on Gaussian quadrature: Appendix B provides the full details.

4.4 Identification

The identification of the various parameters underlying Equation 20 is relatively straightforward. First, the distribution of a_i is identified from the empirical distribution of workers' logarithmic differences between the (unobserved) index of guessed wages and fiscal wages $(w^r_{it} - w^f_{it})$, as in Figure 4 (neglecting rounding) but averaged over the two waves of the LFS. Second, the distribution of υ_i is identified from the subsequent

[44] We estimate one set of parameters $\theta^n = \{\mu^n_a, \sigma^n_a, \sigma^n_m\}$ per class of rounding n.

empirical distribution of the residuals that are obtained in the log space once w^r_{it} has been purged from a_i. To be precise, the worker-level measure of uncertainty embodied in the residual term v_{it} is identified from the *time variations* of discrepancies between values of w^r and values of w^f, once controlled for a_i. They reflect variations in the information sets of workers and can therefore be safely related to a measure of attention at the worker level. Lastly, the probabilities of rounding at different levels $N_i \in \mathcal{N} \equiv \{1, 10, 50, 100, 500, 1000\}$ are broadly identified from the empirical frequencies of rounded values previously reported in Table 3, but adjusted (i.e. dampened) for the probability of making large mistakes, either persistently or idiosyncratically.[45]

We apply our estimation procedure firstly, on the pooled sample of the ERFS and, secondly, by sub-populations defined in terms of broad socio-demographic characteristics. This allows us to document heterogeneity in terms of rounding, reporting biases a_i and, most importantly, uncertainty σ_m across different populations of workers defined in terms of wage, gender, tenure, and education. Importantly, we also estimate our structural parameters of interest by sub-populations defined in terms of day of LFS interview. Unfortunately, workers are almost never interviewed on the same day in their first and second LFS interviews. To overcome this difficulty, we widen the window and isolate samples of workers who are interviewed within an identical time window of ten days, both in their first and second LFS interviews. This enables us to retrieve estimates of σ_m, μ_a and σ by rolling windows of ten consecutive days. Using a simple matrix inversion procedure described in full details in Appendix B.3, we ultimately retrieve the daily estimates of all parameters of interest, most importantly σ_m (which is required to test the empirical predictions of Section 2).

4.5 Estimates of w^d, σ and Attention m

Our empirical framework so far amounts to an analysis of the variance of LFS self-reported wages. To our knowledge, it is one of the first large-scale analyses of workers' response behavior and of the potential behavioral biases that can be identified from this source. As discussed in Section 4.2, the volatility premium associated with workers' perceptions σ^2_m is of particular interest as it measures an actual premium in terms of wage volatility whose magnitude could result from conscious efforts of memorization ("attention"), as formalized in Section 2.

A first challenge is to assess whether our estimates of this volatility premium σ^2_m are large or small. Ideally, it would be useful to know what was the volatility σ^2 of the prior w^d introduced in Section 2 and to compare it to the ex-post estimated volatility σ^2_m and infer attention m as in Equation 11 of Section 2. In order to take a stand on this question, we simply propose to equate w^d and σ^2 with familiar benchmarks: respectively, the wage prediction and volatility that would be estimated by an econometrician using standard fiscal ("true") wage data. Our estimating equation takes the following form (Delaney and Devereux, 2019):

[45] Appendix B shows that large values of a_i and v_{it} are both associated with lower values of the likelihood function of Equation 21. In Section 4.1, we also introduced specifications with left-digit bias as opposed to "correct" rounding. The prevalence of left-digit bias is identified from the frequency of discontinuities in the reporting of rounded values which occur "too late" (as in Lacetera et al. (2012)) while allowing for a lower value of a_i (and a higher value of the likelihood function).

$$w_{it}^{f} = \underbrace{X_{it}\beta + \delta_j + \delta_g + \delta_i + \delta_t}_{\widehat{w_{it}^{d}}} + \eta_{it} \tag{22}$$

where, as previously, w_{it}^{f} denotes the logarithm of fiscal wages and $\widehat{w_{it}^{d}}$ is the estimated "default" (worker's prior introduced in Section 2), in logarithm. The specification in Equation 22 contains aggregate time dummies (δ_t), controls (X_{it}) for gender, age, detailed occupation,[46] detailed information about tenure (in years), firm industry (δ_j) at the one-digit level, dummies for regions (δ_g), as well as worker-level fixed effects (δ_i). Econometrician guesses, i.e. out-of-sample log-wage predictions constructed from these estimated models will make errors that will be distributed as Gaussian random variables, centered on the correct value with standard deviation $\widehat{\sigma}$. Equation 22 is estimated on eight different samples, each with a different end year (ranging from 2005 to 2010 and 2014 to 2015, as in our main ERFS sample) and up to 10 years of retrospective data, depending on each worker's labor market trajectory.[47] We compute estimates of σ^2 by retrieving the series of estimated residuals η_{it} and constructing indices of volatility at the worker level from their respective wage data history. Of course, our baseline ERFS data lack a sufficient panel dimension which would enable us to properly estimate Equation 22. We instead estimate this equation using a complementary, much wider dataset (the DADS panel)[48] which contains administrative information about wages that is very similar to fiscal wages, for one-fourth of the French population of workers. Finally, we adopt a pseudo-panel approach and simply match the obtained estimates to our main ERFS sample of workers according to year of interview, all the selection criteria in Section 3.2 (and Appendix Table 8) and the following detailed worker characteristics: gender, age, detailed occupation, education level (high/low) and tenure (short/long).[49] This procedure also yields estimates of indices of attention m via Equation 11. The associated standard deviations are computed by bootstrap (50 replications).

5 Results

5.1 Pooled Variance Analysis

Table 4 first reports the results from our pooled dataset. Each specification is estimated on the entire estimation sample, but incorporates different sets of classes of rounding. The first specification is the most parsimonious and introduces only one class - this is equivalent to neglecting rounding. The second speci-

[46] Information about education is only available on too small a sub-sample. We therefore proxy this information with detailed fixed effects for occupations at the 2-digit level.

[47] As explained below, we estimate this wage volatility using a companion dataset, the DADS panel, which allows us to track the wage history of one-fourth of the labor market population over time, whatever the nature of their employment or unemployment spells.

[48] The DADS provide us with detailed wage information and a rich set of worker-level information for just under 8% (1/12) of the French labor force since 2002, and 1/25 since 1976. See e.g. Cahuc et al. (2018) and Schmutz and Sidibé (2018) (among many others) for recent papers based on the DADS panel data. The DADS wage information is an annual aggregate that is directly reported by employers to compute social security contributions, as in the pre-filled fiscal files, which makes this source very similar to fiscal wages.

[49] It is not possible to include the same detailed set of controls as in Equation 22 because, despite the larger scale of the DADS panel dataset, some of the rare combinations of characteristics in our ERFS sample are also rare in the DADS and do not match a sufficient number of DADS workers.

Table 4: Baseline Estimation Results

Classes	€1	€10	€50	€100	€500	€1,000	Average	LnLik
Specifications	(A) Probabilities of rounding, π						Coarsening	
1 class	1.000						1.000	-262,023
	-						-	
2 classes	0.523			0.477			48.272	-191,222
	(0.005)			(0.005)			(0.501)	
3 classes	0.395		0.178	0.427			52.013	-176,802
	(0.005)		(0.004)	(0.005)			(0.507)	
4 classes	0.280	0.119	0.174	0.426			52.799	-170,421
	(0.004)	(0.003)	(0.004)	(0.005)			(0.507)	
5 classes	0.281	0.119	0.176	0.369	0.055		74.833	-169,485
	(0.004)	(0.003)	(0.004)	(0.005)	(0.002)		(1.051)	
6 classes	0.280	0.119	0.175	0.369	0.049	0.007	78.716	-169,437
	(0.004)	(0.003)	(0.004)	(0.005)	(0.002)	(0.001)	(1.311)	
Left-digit bias	0.281	0.119	0.176	0.370	0.048	0.005	76.625	-169,902
	(0.004)	(0.003)	(0.004)	(0.006)	(0.002)	(0.001)	(1.273)	
	(B) Uncertainty parameter, σ_m							AIC
1 class	0.105						0.105	524,053
	(0.000)						(0.001)	
2 classes	0.104			0.105			0.104	382,457
	(0.000)			(0.000)			(0.001)	
3 classes	0.108		0.082	0.109			0.104	353,623
	(0.000)		(0.001)	(0.000)			(0.001)	
4 classes	0.121	0.068	0.081	0.109			0.103	340,867
	(0.001)	(0.001)	(0.001)	(0.000)			(0.001)	
5 classes	0.122	0.068	0.080	0.114	0.042		0.101	339,000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)		(0.001)	
6 classes	0.122	0.068	0.080	0.113	0.046	0.077	0.101	338,910
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	0.013	(0.001)	
Left-digit bias	0.122	0.068	0.080	0.112	0.037	0.086	0.100	339,840
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	-0.016	(0.001)	
	(C) Mean of bias, μ_a							BIC
1 class	0.008						0.008	524,066
	(0.001)						(0.001)	
2 classes	0.020			-0.004			0.008	382,484
	(0.001)			(0.001)			(0.001)	
3 classes	0.025		0.003	-0.005			0.008	353,664
	(0.001)		(0.002)	(0.002)			(0.001)	
4 classes	0.030	0.012	0.003	-0.005			0.008	340,921
	(0.002)	(0.002)	(0.002)	(0.002)			(0.001)	
5 classes	0.030	0.012	0.003	-0.004	-0.018		0.008	339,068
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)		(0.001)	
6 classes	0.030	0.012	0.003	-0.004	-0.021	-0.007	0.008	338,992
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.023)	(0.001)	
Left-digit bias	0.031	0.015	0.019	0.023	0.110	0.202	0.028	339,922
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.027)	(0.001)	
	(D) Standard deviation of bias, σ_a							
1 class	0.095						0.095	
	(0.001)						(0.001)	
2 classes	0.083			0.106			0.094	
	(0.001)			(0.001)			(0.001)	
3 classes	0.076		0.094	0.109			0.093	
	(0.001)		(0.001)	(0.001)			(0.001)	
4 classes	0.072	0.087	0.093	0.109			0.093	
	(0.001)	(0.001)	(0.001)	(0.001)			(0.001)	
5 classes	0.072	0.087	0.092	0.107	0.120		0.093	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)		(0.001)	
6 classes	0.072	0.087	0.092	0.107	0.113	0.160	0.093	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.019)	(0.001)	
Left-digit bias	0.072	0.087	0.091	0.107	0.117	0.178	0.093	
	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)	(0.026)	(0.001)	
Observations	19,045							

Notes: This table reports the estimation results of our main specification. Each specification (containing 1 to 6 classes of rounding or implementing "left-digit bias") is estimated on the same sample of 19,045 workers spanning the 2005 to 2015 period (see Section 3). All estimation details are reported in Appendix B.

fication reports the best 2-class model in terms of minimization of the Akaike (AIC) and Bayesian (BIC) information criteria: this involves allowing some workers to round at €100. The third specification is the best three-class model and introduces the €50 class. The fourth and fifth specifications correspond to the best four- and five-class models and introduce the €10 and €500 classes, respectively, while the most comprehensive specification introduces the €1,000 class. This sequential exercise shows that the specification incorporating six classes of rounding best fits the data, as all likelihood ratio tests, AIC or BIC criteria reject the more parsimonious models. Moreover, Table 4 is informative of the cost of neglecting rounding in terms of bias for the estimation of our main parameters of interest. These costs are negligible in terms of average biases (μ_a), but are statistically and economically significant in terms of our two parameters of variance. To be precise, comparing the results from our first and sixth specifications shows that neglecting rounding in the estimation procedure overstates the standard deviation of the uncertainty parameter σ_m by around 5%, which means that the corresponding "volatility premium" σ_m^2 is overstated by 8%. Similarly, specifications which neglect rounding overstate the standard deviation σ_a of time-invariant biases by 2.5%, which in terms of variance amounts to 5%. In total, neglecting rounding implies overstating the overall variance by 13%.

The last row of Table 4 proposes a specification incorporating 6 classes of rounding, but with the additional assumption that workers feature left-digit biases, as in Busse et al. (2013) and Lacetera et al. (2012). Interestingly, the various selection criteria (AIC, BIC, Log-likelihood ratio) lead us to reject this assumption.[50] This result is probably unsurprising as our setting is very different from the environment described in Busse et al. (2013) and Lacetera et al. (2012). In particular, in their setting, prospective used car buyers are under time pressure when collecting and processing information about car characteristics since they have to supervise multiple simultaneous auctions. In contrast, and as explained in details in Section 3, our workers are not really subject to time pressure: although the LFS questionnaire is designed to be parsimonious, interviewers from the statistical institute are trained in order to avoid exerting any time pressure. Workers simply have to provide the best estimate of their current wage, usually without being able (or willing) to access accurate sources of information about it. There is no particular reason for them to keep truncated rather than rounded values in mind, or for their estimates to suffer from a left-digit bias, as the sequence of relevant digits to be retained is much more limited than for mileages of second-hand cars. In the remainder of the paper, we therefore safely focus on the specification with six classes of rounding associated with standard, symmetric rounding.

Our estimates in Table 4 show that around 72% of workers report rounded values of their wage: 37% report values rounded to the nearest hundred, while 18% and 12% report values rounded to fifty or ten, respectively. The remaining 5% of workers round to €500 or €1,000. These figures are broadly consistent with the gross observation counts in Table 3, but the more rigorous estimation of the mixture model tends to inflate the €50 category and deflate the €100 and €1,000 categories.

[50] Our specification contains a worker-level random effect a_i which may capture most of the biases that are stable across time, although these terms do not enter exactly the same way as the "left digit" shifter in Equation 20 and are therefore, in principle, separately identified in the data.

More importantly for our analyses, Panel (B) in Table 4 shows that the parameter governing the uncertainty premium perceived by workers, σ_m, is estimated to range between 4.6% and 12.2% of the fiscal wage, depending on the class of rounding. The estimated average across all classes of rounding is 10.1%, a value that is just below the number attained in the most populated classes of rounding (€1 and €100). Strikingly, this uncertainty parameter is highest in the class of non-rounders, and there is no monotonic correlation between this uncertainty measure and the class of rounding. This implies that in our setting, and in contrast to the findings of Ruud et al. (2014) and Binder (2017), the prevalence of rounding does *not* signal higher levels of uncertainty at the worker level. We return to this puzzle in Section 5.2, where in particular we refine the analysis by controlling for education as a proxy for financial literacy.

The distributions of the random effects a_i also appear to be somewhat differentiated across classes of rounding, in a more monotonic fashion. First, we find that the average of reporting biases μ_a tend to be positive and significant for non-rounders or for workers reporting values that are rounded at €10. The sign switches and becomes significantly negative for coarser classes of rounding: €100 and €500.[51] As a result, the estimated average of parameter μ_a over all classes of rounding is estimated to be only 1%. Second, we find that the dispersion of the random effects a_i increases as the class of rounding becomes coarser. While the estimated standard deviation σ_a is only around 7% for non-rounders, it reaches 11% among workers reporting values rounded at €100 or €500. Ultimately, the total variance of the discrepancy between self-reported and fiscal wages $(\sigma_a^2 + \sigma_m^2)$ is higher in coarser classes of rounding. But this result is mainly driven by the population variance σ_a^2 of the time-invariant parameters a_i, rather than by the variance σ_m^2 of the time-varying uncertainty parameters, u_{it}. In other words, workers' rounding behavior is more correlated with their time-invariant errors or biases than with the time-varying volatility term σ_m^2.

In Appendix D, we check the robustness of our results to even more flexible specifications. There, we investigate whether the implicit assumption in Table 4 that all our parameters of interest are stable across time at the worker level actually hold. The main take-aways of these experiments are that the rounding profile captured by the probabilities π is stable over time. Likewise, the parameter measuring mean biases μ_a is also stable over time, with none of the estimated values being statistically different in the first and second interviews of the LFS. The only parameters that increase slightly over time are the volatility parameters σ_m and σ_a. This implies in particular that surveyed workers do not appear to be better informed in the second interview than in the first: if anything, the converse holds. The pattern is, furthermore, qualitatively and quantitatively very limited. [52]

[51] In Section 5.2, we will document that this pattern is driven by a composition effect: coarser classes of rounding are populated by workers who simultaneously earn higher wages and tend to under-report their wages.

[52] These experiments show that overall, allowing our structural parameters of interest to vary in the first and second interviews of the LFS provides limited insight, at the cost of a significant loss of statistical power. Consequently, the 6-class specification in Table 4 remains our central baseline specification in the remainder of the paper.

5.2 Heterogeneity Across Different Populations of Workers

Table 5 summarizes the results of a second set of estimations which documents parameter heterogeneity in different sub-populations of workers. We first investigate whether our data feature heterogeneity between high- and low-wage workers, and simply split our sample in half at the median (€1,665). As explained in detail in Section 2, the reason for this is that low-wage workers are likely to face tighter financing constraints and pay more attention overall to their labor income. Furthermore, the rounding behavior of workers may be different at low vs. high levels of the wage distribution, as the relative amounts implied by rounding approximations are different. We also investigate whether tenure impacts workers' assessment of their own wage and further split the samples at the median (13 years): we hypothesize that workers with shorter tenure had less time to accumulate wealth as an additional safety buffer, such that they might exhibit higher financing constraints, all else equal. Furthermore, these workers are likely to be less informed about bonuses, payment of overtime, and other dimensions of their employer's wage policy. This means that the volatility of their default prior σ is likely to be higher. Lastly, we introduce two characteristics that previous literature has shown to be associated with different degrees of risk aversion (Borghans et al., 2009; Croson and Gneezy, 2009) or processes of anticipation formation (e.g. D'Haultfoeuille et al., 2018) and thus are also good potential sources of heterogeneity in our setting: education, as a proxy for numerical literacy, and gender.

To assess the impact of these four different dimensions of heterogeneity on our parameters of interest, and in the absence of precise guidance on how they will affect them, we adopt a fully non-parametric approach: we simply estimate one set of parameters per sub-sample of workers defined in terms of gender (male vs. female), education (lower than high school vs. higher than high school), tenure (higher vs. lower than the sample median) and wage (lower vs. higher than the sample median). All but one of the resulting 16 sub-samples are sufficiently populated to allow for such a separate estimation of the parameters.

Our estimates in Table 5 first highlight that the probabilities of rounding $(1-\pi_1)$ feature some heterogeneity and range between 62% and 79% depending on the sub-population of workers considered. Similarly, the average coarsening, i.e. the average rounding class in each sub-sample, ranges between €38 and €119. The measure of worker-level uncertainty σ_m varies between 8.2% and 12.4% while σ_a ranges between 6.3% and 10.1%. Lastly, the average bias parameter μ_a ranges between -4.3% and 5.0%. All these differences are statistically significant. To better gauge the orders of magnitude and relative contributions of all these parameters to overall sample heterogeneity, it is useful to apply the König-Huyghens formula to the sub-populations in Table 5. Under the identifying assumptions of Section 4, this results in the following variance decomposition equation:

$$\mathbb{V}\left(a_i+v_{it}\right)=\sum_p sh_p\sigma_{ap}^2+\sum_p sh_p\sigma_{vp}^2+\sum_p sh_p\left(\mu_{ap}-\mu_a\right)^2 \tag{23}$$

where p indexes the different sub-populations, sh_p represents the share of our sample observations that are allocated to each of them, $(\sigma_{ap},\sigma_{vp},\mu_{ap})$ denote the associated parameters and μ_a denotes the full sample

Table 5: Estimates by Detailed Sub-Samples

Sample definition				Nb obs.			Parameters (averages across classes)					
Wage (≥ med.)	Gender (Fem.)	Education (≥ H. Sch.)	Tenure (short)	ERFS	DADS	σ	$1-\pi_1$	Coarsening	σ_m	μ_a	σ_a	m
0/1	0/1	0/1	0/1	19,045	139,320	0.160	0.720	78.716	0.101	0.008	0.093	0.633
						(0.001)	(0.004)	(1.311)	(0.001)	(0.001)	(0.001)	(0.009)
0	0	0	0	2,322	1,652	0.129	0.675	65.278	0.096	0.024	0.078	0.443
						(0.007)	(0.012)	(2.873)	(0.002)	(0.002)	(0.003)	(0.090)
0	0	0	1	2,729	21,203	0.154	0.675	50.569	0.105	0.049	0.090	0.533
						(0.002)	(0.011)	(2.048)	(0.002)	(0.003)	(0.003)	(0.028)
0	0	1	0	326	7,084	0.127	0.674	75.425	0.107	0.048	0.065	0.288
						(0.003)	(0.032)	(8.669)	(0.007)	(0.008)	(0.010)	(0.120)
0	0	1	1	1,566	34,847	0.163	0.685	55.693	0.107	0.050	0.083	0.568
						(0.003)	(0.014)	(3.037)	(0.003)	(0.004)	(0.003)	(0.032)
0	1	0	0	678	2,109	0.148	0.633	51.844	0.086	0.011	0.064	0.660
						(0.004)	(0.023)	(4.463)	(0.004)	(0.004)	(0.005)	(0.057)
0	1	0	1	563	18,228	0.180	0.619	38.076	0.086	0.027	0.070	0.771
						(0.003)	(0.026)	(2.429)	(0.004)	(0.005)	(0.005)	(0.033)
0	1	1	0	298	3,966	0.141	0.645	52.569	0.088	0.031	0.085	0.610
						(0.003)	(0.035)	(8.197)	(0.007)	(0.008)	(0.008)	(0.063)
0	1	1	1	1,039	12,864	0.204	0.623	43.098	0.090	0.026	0.063	0.807
						(0.004)	(0.019)	(3.063)	(0.003)	(0.003)	(0.004)	(0.020)
1	0	0	0	2,547	3,902	0.139	0.777	117.940	0.101	-0.032	0.096	0.466
						(0.004)	(0.009)	(5.311)	(0.003)	(0.003)	(0.003)	(0.055)
1	0	0	1	945	15,805	0.168	0.745	110.238	0.124	-0.017	0.095	0.458
						(0.003)	(0.017)	(8.174)	(0.005)	(0.005)	(0.008)	(0.051)
1	0	1	0	2,132	3,747	0.154	0.792	118.683	0.094	-0.014	0.085	0.629
						(0.003)	(0.010)	(6.318)	(0.003)	(0.003)	(0.003)	(0.034)
1	0	1	1	1,921	7,751	0.193	0.794	110.309	0.107	-0.012	0.101	0.694
						(0.008)	(0.010)	(5.705)	(0.003)	(0.003)	(0.003)	(0.033)
1	1	0	0	306	1,760	0.145	0.776	91.392	0.082	-0.043	0.082	0.683
						(0.004)	(0.027)	(15.236)	(0.011)	(0.013)	(0.009)	(0.061)
1	1	0	1	63	1,937	0.185	-	-	-	-	-	-
						(0.010)						
1	1	1	0	959	1,035	0.135	0.738	109.965	0.084	-0.016	0.067	0.616
						(0.004)	(0.017)	(8.377)	(0.003)	(0.003)	(0.004)	(0.051)
1	1	1	1	651	1,430	0.237	0.765	97.990	0.095	-0.011	0.084	0.840
						(0.018)	(0.019)	(9.204)	(0.008)	(0.005)	(0.079)	(0.034)

Notes: This table reports estimates obtained for each of the sub-populations indicated in the "sample definition" columns. The standard deviations of all structural parameters are computed via the BHHH estimator described in Appendix B, except for σ and m, for which we rely on bootstrap (50 replications). Sources: ERFS (2005-2015) for parameters σ_m, μ_a, σ_a, rounding $(1-\pi_1)$ and coarsening; and DADS (1995-2015) for σ and m.

mean of residuals, i.e. the weighted average of the set of (μ_{ap}) parameters. Our estimates in Table 5 imply that the first term contributes 40% of the total variance, while the second term contributes 55% and the last term only 5%. In other words, the between-population heterogeneity in average biases μ_{ap} is small relative to the overall order of magnitude of the within-population variance parameters σ^2_{ap} and σ^2_{vp}, which contribute roughly equally to the overall variance. Interestingly, a closer look at the contribution of each sub-population to the overall variance reveals that all sub-populations of female workers contribute less than their observation weights, while 5 out of the 8 sub-populations of male workers contribute significantly more.

Table 5 also proposes an estimate of the implied attention parameter m along the lines of Section 4.5. As a reminder, the computation of this parameter implicitly relies on a benchmark, σ, which we estimate from a Mincerian equation on a companion dataset (DADS), which features a longer panel dimension than our main estimation sample. The comparison of the estimated σ_m with this benchmark implies attention parameters ranging between 0.288 and 0.840 depending on the considered sub-population, with the mass of estimates lying in the 0.4 to 0.8 range. Interestingly, these orders of magnitude and the amount of heterogeneity in the attention parameter m that we find match almost perfectly with the results from the previous literature surveyed in Gabaix (2019). He notes that in studies where the importance of the opaque attribute ranges between 0.07 and 0.24 (which has to be compared with σ ranging between 0.13 and 0.24 in our setting), measures of attention typically range between 0.25 and 0.69 (Taubinsky and Rees-Jones, 2017; Chetty et al., 2009; Lacetera et al., 2012; Hossain and Morgan, 2006).[53]

Finally, Table 6 adopts a more synthetic approach and correlates our estimated parameters with the four characteristics we introduced: wage level, tenure, gender and education. Panel (A) first investigates the gross correlations. The first column checks that the "prior" wage uncertainty estimated from a Mincerian equation, σ, is higher for higher wages (by 1.6 percentage points), female workers (by 1.8 ppt), more-educated workers (by 1.2 ppt) and those with shorter tenure (4.0 ppt). The following two columns describe workers' rounding behavior via the share of rounders on one hand and average coarsening on the other. Both appear to be positively correlated with the wage level: workers earning more than the sample median round 10 ppt more often and report coarser values (by €53) than workers earning less than the sample median. Similarly, we find that women round significantly less often overall (by 4.1 ppt) and, all else equal, rely on finer classes of rounding such that their average coarsening is lower (by €13). This is also the case of workers with shorter tenure, by approximately the same amount (€13). Moreover and all else equal, the probability of rounding is only weakly correlated with education, but we find that when rounding occurs, more educated workers tend to use coarser scales.

The fourth column collects the correlations with our main parameter of interest, i.e. the measure of uncertainty σ_m. We find that women exhibit a lower actual uncertainty parameter σ_m (by 1.5 ppt), which

[53]For example, Chetty et al. (2009) find that the attention allocated to the sales taxes of grocery store items is equivalent to 35%, while the magnitude of this tax is typically 7% of the total price. Similarly, in Lacetera et al. (2012), the attention index to the mileage of used cars sold at auction attains 69%, while the magnitude of the error implied by left-digit bias is 10% of the total mileage.

Table 6: Wage Volatility σ and Main Structural Parameters

	σ	$1-\pi_1$	Coarsening	σ_m	μ_a	σ_a	m
			(A) Gross correlations with Worker-Level Characteristics				
Wage: high	0.012*	0.105***	53.261***	0.004	-0.054***	0.015***	0.042
	(0.007)	(0.008)	(1.519)	(0.003)	(0.005)	(0.003)	(0.040)
Women	0.018**	-0.041***	-13.243***	-0.015***	-0.012*	-0.017***	0.161***
	(0.006)	(0.009)	(0.604)	(0.002)	(0.006)	(0.003)	(0.049)
Education: high	0.009	0.015*	4.757***	-0.002	0.011**	-0.006**	0.078
	(0.007)	(0.008)	(0.846)	(0.003)	(0.005)	(0.003)	(0.048)
Tenure: short	0.039***	-0.004	-13.327***	0.010***	0.011*	0.010***	0.104**
	(0.006)	(0.008)	(0.990)	(0.003)	(0.005)	(0.003)	(0.042)
			(B) Net Correlations with Worker-Level Characteristics				
σ		0.290	42.915	-0.066	-0.354***	-0.001	3.265***
		(0.174)	(61.814)	(0.057)	(0.092)	(0.069)	(0.627)
Wage: high		0.100***	52.587***	0.004	-0.050***	0.015***	-0.002
		(0.008)	(3.272)	(0.002)	(0.002)	(0.003)	(0.024)
Women		-0.046***	-14.365***	-0.014***	-0.006***	-0.017***	0.108***
		(0.009)	(2.523)	(0.002)	(0.002)	(0.003)	(0.020)
Education: high		0.011	4.193***	-0.002	0.015***	-0.006*	0.040
		(0.009)	(2.646)	(0.003)	(0.003)	(0.003)	(0.026)
Tenure: short		-0.016*	-14.814***	0.013***	0.026***	0.010***	-0.031
		(0.007)	(3.289)	(0.002)	(0.003)	(0.003)	(0.022)
Observations	15	15	15	15	15	15	15

Notes: σ_m, μ_a and σ_a are estimated in Section 5. Sources: ERFS (2005-2015) for parameters σ_m, μ_a, σ_a, rounding and coarsening and DADS (1995-2015) for σ and m. All observations are weighted by the inverse of the variance estimated for the dependent variable (with observations with higher variance being given a lower weight). In Panel (B), all standard deviations are computed by bootstrap (50 replications) to take account of the fact that σ is a generated regressor.

is associated with a significantly higher attention parameter m in the last column (by 16 ppt). In contrast, workers with shorter tenure feature a significantly higher uncertainty parameter σ_m (by 1.0 ppt). This appears to be associated with an extremely high "prior" σ, so that their underlying attention m is not lower but is actually higher than the attention parameter of workers with longer tenure by 10 ppt. Interestingly, we do not find any correlation between education and the uncertainty parameter σ_m. Given that the baseline wage volatility σ of highly educated workers is high, this ends up in a slightly positive, though statistically insignificant, correlation between education and attention m. Overall, more-educated workers tend to exhibit an attention index that is higher by 7.8 ppt than attention index of less educated workers.

Lastly, Columns 5 and 6 report the correlations between worker characteristics and the mean and standard deviation of the distribution of time-invariant biases a_i. High-wage workers tend to under-report by 5.4% and feature a higher dispersion of time-invariant biases. All else equal, women also tend to under-report on average by 1.2% but constitute a population which features a low dispersion of time-invariant biases. In contrast, both educated and shorter-tenured workers tend to over-report, all else equal, by around 1.1 ppt. However, while educated workers feature a low dispersion of time-invariant biases, the opposite is the case for workers with a shorter tenure.

One limitation of the correlations in Panel (A) is that they do not account for the fact that the various populations of workers actually face different levels of *ex ante* (prior) uncertainty, σ. Panel (B) therefore replicates the previous analysis while introducing σ as an additional control. As this regressor is generated (i.e. estimated), inference now requires a comprehensive bootstrap procedure.

The results are as follows. We find that workers featuring higher σ tend to under-report their wage, a result that is consistent with some form of risk aversion. This negative correlation between μ_a and σ attenuates the residual correlation that is obtained between μ_a and gender or between μ_a and the wage level, while it amplifies the positive correlations with education and tenure. Secondly, we find that workers featuring higher σ exhibit higher attention indices m. This results in no significant difference in terms of actual ("ex-post") uncertainty σ_m. The net correlation between attention and gender remains positive and highly significant, while the correlation with tenure disappears. For the other indicators (rounding, σ_m and σ_a), the correlations with wage level, gender, education and tenure are basically unaltered by the introduction of the additional control. In particular, as a previous literature (Ruud et al., 2014; Binder, 2017) suggested that rounding may be associated with higher uncertainty in a variety of settings, we clarify that this is not always the case. Especially, the LFS data feature no correlation between rounding and any of our indicators of uncertainty, either σ or the dummy indicating a shorter tenure.[54]

5.3 Payday and the Monthly Cycle of Attention

We now focus on documenting fluctuations in attention over time. Through the perspective of Section 2, this exercise is particularly useful as varying the "distance" to the payday is equivalent to varying the perceived financial constraint (Mani et al., 2013; Carvalho et al., 2016), thus allowing us to document correlations between the magnitude of this constraint and the level of attention that workers pay to monitoring it.[55]

Using our data, it is unfortunately impossible to directly investigate how our structural parameters of interest vary as the distance to the payday increases by estimating separate models for each day between the 1st and 31st: this would require us to restrict our sample to workers who appear to be interviewed on the same day both in their first and second LFS interviews, and there are too few such cases. A simple way to overcome this difficulty is to perform separate estimations by successive rolling windows, which we set to 10 days. This threshold is convenient, as the series of time windows indexed by $(d/d+9, d+10/d+19, d+20/d+29)$ also constitute rolling partitions of the month and can therefore be interpreted as simple estimations performed over successive, non-overlapping periods of 10 days. The resulting sub-samples of workers are described in Table 9 in Appendix C. As documented in Section 1, the dates of the interview are orthogonal to workers'

[54] Moreover, we find that the main characteristics of workers who round more (mostly higher-wage workers) are different from the main characteristics of workers facing higher uncertainty in terms of σ or σ_m (workers with shorter tenure).

[55] The payday is not strictly regulated in France. Labor law (Article L. 3242-1 of the Labor Code) only sets the maximum interval between two successive paydays: it is set at one month for workers with a monthly contracts while for employees who are not on a monthly contract (seasonal or temporary workers, etc.), payment must be made at least twice a month, at a maximum interval of 16 days. There are two main reasons why the vast majority of French employers actually synchronizes these cycles and pays wages on the last or first days of the month. First, this is the convention that prevails among French civil servants, who represent roughly 20% of the total workforce (and of our sample): this component of the workforce is consistently paid on the third working day before the end of the month, according to a precise calendar published each year by the Ministry of the Interior. Second, payroll taxes to the French social administration for a given month are due on the fifth day of the following month, which means that employers must insure that all computations are ready in the last days of the working period or in the first few days of the next. Payments to workers and to the fiscal administration often occur quasi-simultaneously.

characteristics, such that these sub-samples exhibit the same structure of characteristics as the full estimation sample. The only difference is that the workers in our initial estimation sample who were interviewed more than 10 days apart in their first and second LFS interviews have to be discarded, thus slightly reducing the statistical power of our experiment.

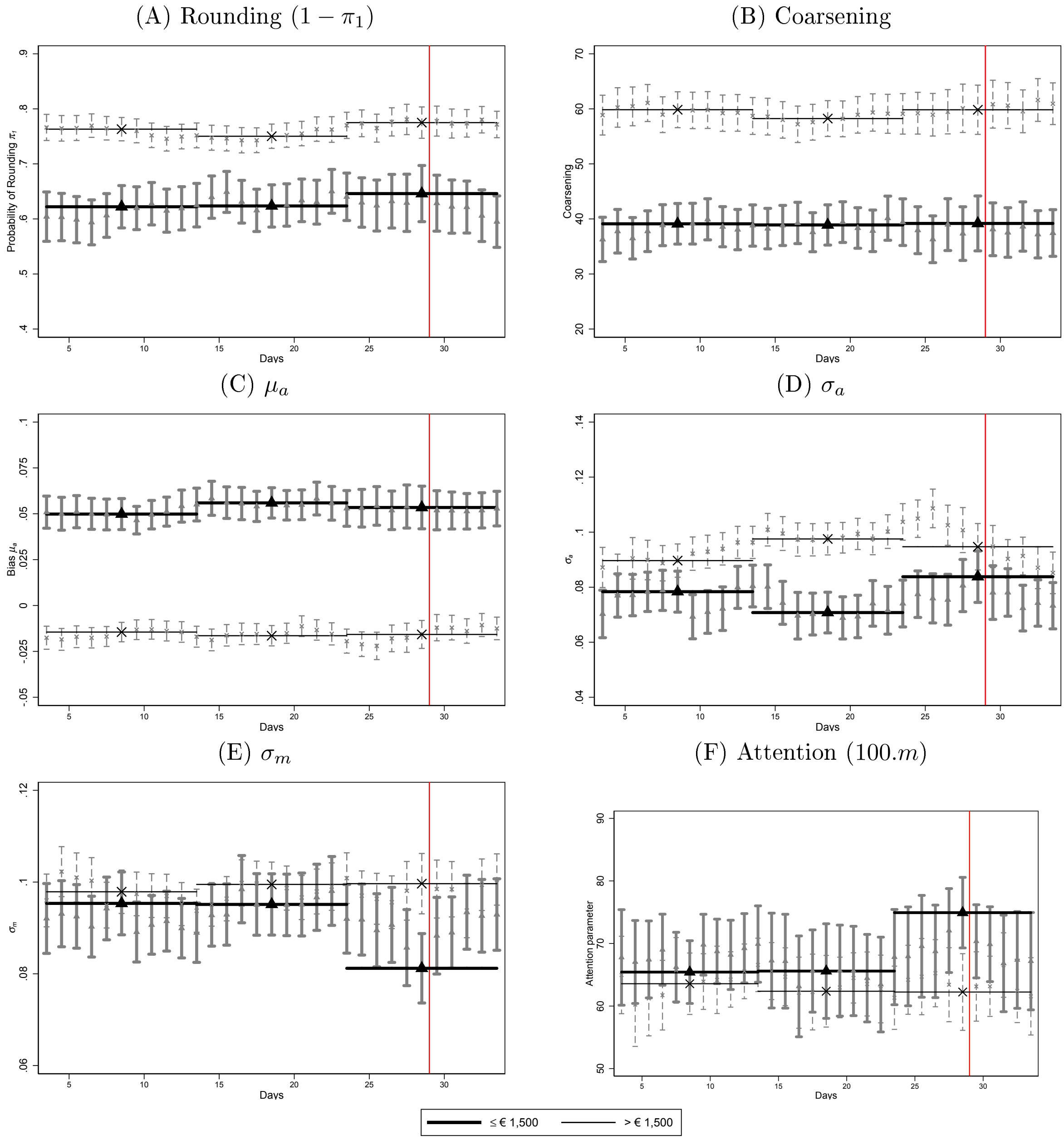
Main results. The results are reported in Figure 5, where we plot the estimates obtained for each of the 30 successive time windows of 10 days. We also outline a specific set of three non-overlapping estimates covering the first, middle and last 10 days of a given month, respectively. Figure 5 documents the results for the full population of workers and simply splits the estimates between those earning less than €1,500 and strictly more than €1,500, respectively. The first take-away from Panels (A) to (D) in Figure 5 is that our estimates of the parameters describing workers' rounding behavior (the probability to round, $1 - \pi_1$, and average coarsening) and their time-invariant biases (μ_a and σ_a) appear to be stable over time, both in the population of high- and low-wage workers. For example, the share of rounders fluctuates between 74% and 78% among high-wage workers, while it fluctuates between 59% and 65% among low-wage workers. This only represents roughly 5% and 10% of the overall average, respectively. None of these fluctuations are statistically significant. Similarly, workers earning more than €1,500 coarsen their wage by €58 to €62, which represents a very limited scope, while workers earning less than €1,500 round on average by bins of €36 to €40. Again, these fluctuations are limited in magnitude and statistically insignificant. In terms of the average bias μ_a, Panel (C) of Figure 5 confirms previous results from Table 6 that better-off workers tend to under-report their wage relative to less well-off workers. It adds to the evidence that this difference is strikingly stable over the month. Overall, high-wage workers tend to under-report their wage by around 1% to 2%, while workers earning lower wages tend to over-report by 5% to 6%. Results are similar, although a bit more volatile, in terms of σ_a: the dispersion of the time-invariant biases of better-off workers is consistently higher than the corresponding parameter estimated for less well-off workers. However, their monthly fluctuations are statistically insignificant.

Ultimately, the only parameters that exhibit a statistically significant fluctuation over the month are those capturing uncertainty, σ_m and attention, m. This pattern only emerges across low-wage workers, i.e. those who are most likely to experience financing constraints. This is striking, as it first implies that attention is uncorrelated with the parameters of bias (μ_a and σ_a), whether they are to be interpreted as a feature of workers' information sets or simply of their reporting behavior. In other words, our data reveal that increases in attention does not seem to generate mechanisms leading to bias reduction in terms of μ_a.[56].

Second, through the lens of the model in Section 2, the fluctuations of σ_m and m are meaningful: in the case of low-wage workers, it features a clear cyclical pattern whereby uncertainty drops and attention rises in the last 10 days before the payday. The magnitude of this month-end pattern is significant: -1.7 ppt in terms of uncertainty σ_m and +6.3 ppt in terms of attention index m. This corresponds to a statistically

[56] See also footnote 19 in Section 2

Figure 5: Evolution of the Structural Parameters During the Month
Low *vs.* High-Wage Workers

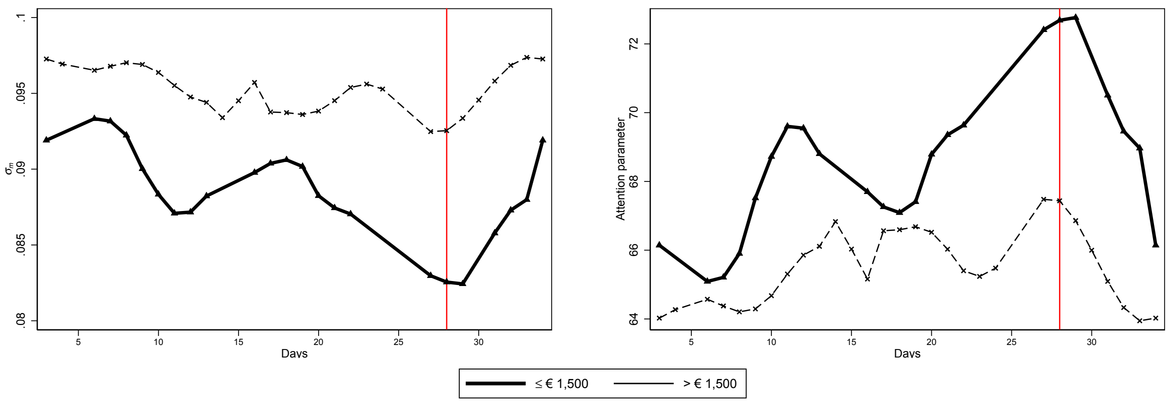


Notes: These figures show the estimates of each structural parameter of interest that are obtained over (rolling) windows of ten consecutive days. The center of the 10-day window is reported on the x-axis, while the value of the considered estimate is reported on the y-axis. These specifications are estimated separately on the sub-populations of workers earning a fiscal wage that is either lower than €1,500 (point estimates as triangle, thick confidence intervals) or strictly higher than €1,500 (point estimates as crosses, thin confidence intervals). In each panel, we also outline a specific set of three estimates covering, respectively, the first, middle and last 10 days of a given month. These time periods are materialized by horizontal bars and form a (non-overlapping) partition of the month into three sub-periods. Sources: ERFS (2005-2015) for parameters σ_m, μ_a, σ_a, rounding and coarsening and ERFS (2005-2015)/DADS (1995-2015) for m. σ is estimated to reach 0.158 across workers earning less than €1,500, and 0.162 across workers earning more than €1,500. The reported standard deviations rely on the BHHH estimator except for the indicator of attention, which requires bootstrap (50 replications).

significant drop in uncertainty σ_m of 20%.[57] The timing of this pattern suggests that it is driven by the tightening of the budget constraint in the last 10 days of the month, when the previous payday is already some distance in the past and the next is still some time away. In contrast, workers earning more than €1,500 do not exhibit any cyclical pattern, and their level of attention is on average lower - approximately at the monthly minimum of low-wage workers. This suggests that in everyday life,[58] better-off workers do not need to remember a precise record of their wages, because their budget constraints are not tight. Overall, the results shown in Panels (E) and (F) in Figure 5 are consistent with the first two predictions in Section 2.

Figure 6: Daily Estimates

(A) Uncertainty Index (σ_m) (B) Attention Index ($100.m$)



Notes: These figures complement Panels (E) and (F) in Figure 5 by providing daily estimates of the uncertainty (σ_m) and attention ($100.m$) parameters, by sub-populations of workers earning, respectively, (in terms of their fiscal wage): less than €1,500 and more than €1,500. The precise methodology underlying these estimates is described in Appendix B.3. The confidence intervals are computed by bootstrap (35 iterations) but are too wide to be meaningfully reported (see text). Sources: ERFS (2005-2015)/DADS (1995-2015).

Robustness and extensions. One limitation of Figure 5 is that reported estimates only correspond to moving averages of our main parameters of interest over 10 days. The width of this time window has the advantage of ensuring a sufficient number of observations for estimation, but it prevents us from accurately tracking the daily variation in uncertainty and attention. In Appendix B.3, we detail a methodology that allows us to retrieve daily estimates of our parameters of interest, up to a simple matrix inversion. The results of this procedure are reported in Figure 6. Unfortunately, our estimation sample is not large enough to obtain accurate estimates as we are left with too few observations per day (and even sometimes no observations at all,[59] as explained in Appendix B.3). As a consequence, the confidence intervals are wide and

[57]Notice that these values are probably only lower bounds on the actual variations in cognitive loads induced by financing constraints. The reason is that when the financing constraint tightens, workers most probably check their entire bank account rather than just their previous payslip. Our data reveal that by doing so, they gather a precise signal about their wage (the net value that is registered as a receipt on their bank account), but this component is probably only just one of the many that workers keep in mind upon checking the balance of their accounts: overall balance, upcoming expenses and bills, upcoming receipts (e.g. wages, social benefits), etc.

[58]i.e. with the likely exception of when they have to take discretionary and large investment decisions (only very few of our sample observations are likely to be in this situation).

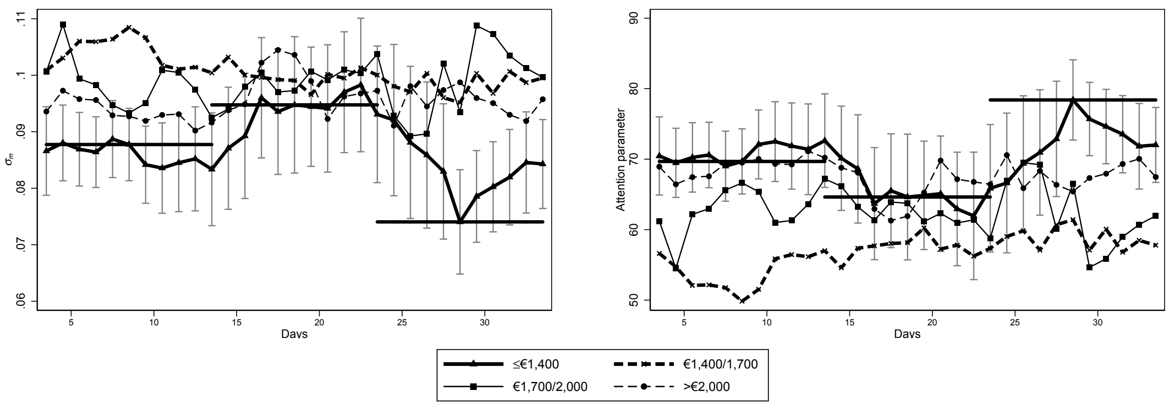
[59]We are unable to identify 9 out of the 31 daily estimates in the case of low-wage workers because of the lack of observations.

none of the monthly evolutions are statistically significant.

Despite its limited statistical power, this exercise allows us to refine the timing of the end-of-month drop in uncertainty σ_m and of the associated rise in attention m. Figure 5 clearly shows that for low-wage workers, attention starts increasing on the 17^{th} day of the month and rises until the 28^{th} day approximately, which corresponds to the payday for most workers. Unsurprisingly, the pattern for the uncertainty index σ_m is symmetric. The magnitude of this rise is of the same order as the previous estimates of Figure 5, i.e. about 15% of the baseline estimate. Attention appears to drop as soon as the payday is reached and, symmetrically, uncertainty immediately begins to rise. The pattern exhibited by the data is well rationalized by the liquidity constraint story in Section 2. Conversely, it is not consistent with alternative models of attention fluctuations such as purely passive information exposure stories whereby workers would be measured as more attentive simply because they are more informed. Indeed, such a story predicts that attention should peak on payday, but then steadily decrease until the next payday and discretely jump to a maximum instead of steadily increasing during the last 10 days of the month. In this respect, the precise timing of fluctuations in attention is critical for the identification of the cognitive load induced by financing constraints.

Figure 7: Estimates by "Narrow" Bins of Wages

(A) Uncertainty Index (σ_m) (B) Attention Index ($100.m$)



Notes: These figures complement Panels (E) and (F) in Figure 5 by providing estimates of the uncertainty (σ_m) and attention ($100.m$) parameters by sub-populations of workers earning, respectively, (in terms of their fiscal wage): less than €1,400, between €1,400 and €1,700, between €1,700 and €2,000, and more than €2,000. The precise methodology underlying these estimates is describes in Appendix B.3. The confidence intervals are computed via the BHHH estimator in the case of σ_m, and by bootstrap (50 iterations) in the case of the attention index m. They are only reported for workers earning less than €1,400 only. Sources: ERFS (2005-2015)/DADS (1995-2015).

In Figure 7, we propose a complementary exercise where we rely on the same methodology to compute estimates of our parameters of interest over narrow bins of wage rather than narrow time periods. Once again, the size of our estimation sample limits the statistical power of this experiment. However, we are able to separately describe the behavior of workers earning, respectively: less than €1,400, between €1,400 and €1,700, between €1,700 and €2,000, and more than €2,000. This exercise identifies more accurately which population of workers is likely to be financially constrained by checking whether they exhibit a cyclical attention pattern. As Figure 5 showed that workers earning less than €1,500 were on average likely to face financing constraints, we unsurprisingly find that this is also the case for workers earning less than €1,400. Yet, there is still some uncertainty about whether the threshold of €1,500 is a good proxy of the upper limit beyond which workers no longer face financial constraints. In the more detailed estimates in Figure 7, we actually find that workers in the €1,400 to €1,700 bin still show a slight pattern of increasing end-of-month attention, suggesting financial constraints, but workers earning more than €1,700 do not. We conclude that the relevant threshold that divides the population of workers into those who are most likely financially constrained and those who are not is strictly above €1,400, and strictly below €1,700, so that €1,400 or €1,500 are reasonable orders of magnitude. Based on the percentiles of the wage distribution reported in Table 2, this implies that 27% to 34% of our population of employed workers are financially constrained in the form described in Section 2, which is sizeable.

Quantitative implications. Within the theoretical framework of Section 2, our estimation results have further quantitative implications beyond the estimated share of workers facing financing constraints.

It is first possible to provide a quantification of the financial costs generating financing constraints across low-wage workers since Equation 7 delivers a lower bound on the ratio of R_B to R_A:

$$\frac{R_B}{R_A} \geq \frac{\sigma}{\phi\left(\frac{\ln \bar{C} - w^d}{\sigma} - \sigma\right) - \sigma\Phi\left(\frac{\ln \bar{C} - w^d}{\sigma} - \sigma\right)} \tag{24}$$

This ratio provides a measure of the asymmetry between workers' ability to transfer revenue from the current to the next period (with a gain R_A) and their ability to borrow, ie. to transfer revenue from the next period to the current one, at total cost $R_A + R_B$.

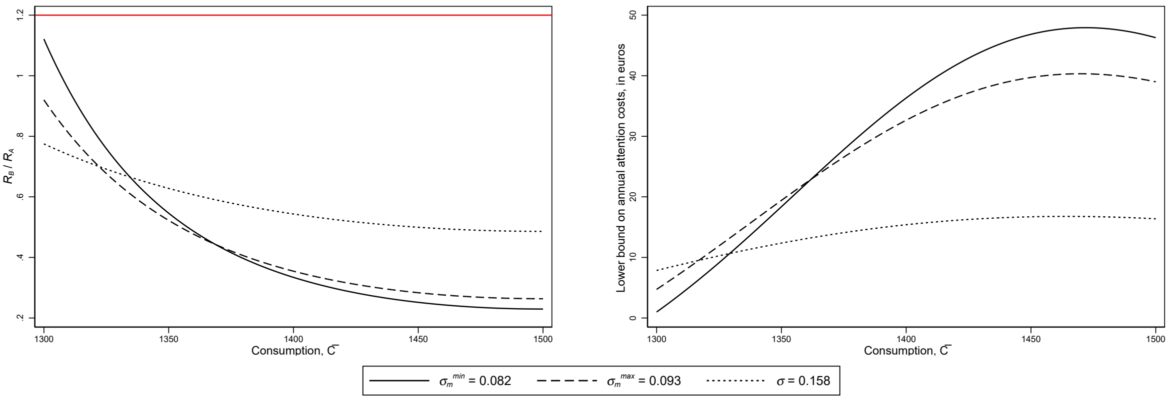
Confronting Equation 24 to the data is relatively straightforward. First, w^r can be simply set to €1,500, the threshold we estimate for financing constraints. Second, σ can be evaluated at the value estimated in the data according to the methodology laid out in Section 4.5. The maximum (as well as the minimum) values estimated for σ_m in the population of workers earning less than €1,500 are alternative admissible values for calibration. The only quantity which is unobserved in our data is $\bar{C}$: we therefore simply remain agnostic about its value and compute our bound of interest in a range of admissible values, between €1,300 and €1,500.[60]

The result of this exercise is reported in Panel (A) of Figure 8. Our results suggest that, among financially constrained workers, the cost of borrowing (be they charged by banks or rather behavioral) are on average 20% to 40% larger than the gains associated with savings among those featuring moderate risk aversion (implying $\bar{C} \geq$ €1,400). They are even larger, i.e. around 80% to 100%, across workers featuring higher

[60] Let us remind that these values map to the size of the security buffer ($W - \bar{C}$) decided by workers and thus capture their degree of risk aversion as shown in Equations 4 (or 12) of Section 2.

Figure 8: Estimated Lower Bounds on Costs

(A) Relative Financing Costs, $\frac{R_B}{R_A}$ (B) Attention Costs (€)



Notes: Panel (A) plots the lower bound on R_B/R_A faced by financially constrained workers of different levels of risk aversion, as embodied in their different choices of consumption, $\bar{C}$ (and the implicit safety buffer $\bar{C}$ reveals). The bound is computed in Equation 24. The red line materializes the actual ratio of interest rates faced by workers for short term borrowing (annualized value of 1.20) relative to the interest rate earned of safe savings (annualized value of 1.01). Panel (B) plots the lower bound on the decrease in expected wage, W^r, which would generate the same desutility cost as maintaining attention to its monthly maximum for ten days or more (see main text for full details). Again, this quantity is computed for financially constrained workers of different levels of risk aversion, as embodied in their different choices of consumption, $\bar{\bar{C}}$. The bound is computed in Equation 25. R_A is set to 1.01 and R_B to 1.20.

risk aversion (implying $\bar{C} \leq$ €1,400, ie. a larger safety buffer relative to their expected wage). The red line materializes the ratio of the average fee (20%) charged by banks to households for lines of credits as of 2016,[61] divided by the rate of return (just 1%) over the same period of the most popular, risk-free "Livret A" savings account. This is a natural benchmark to consider, although the ratio $\frac{R_B}{R_A}$ in Equation 24 is in principle defined in utility terms, such that its value might depart from the actually observed values if workers are risk averse, for example. However, we easily check that this benchmark is of the correct order of magnitude and actually exceeds our estimates of the lower bound for $\frac{R_B}{R_A}$. This confirms that workers earning less than €1,500 are indeed likely to be financially constrained given the credit conditions they face.

A second aspect is that the increase in attention until the payday, i.e., until the liquidity constraint is relaxed, is informative of the shape of the underlying costs associated with increased attention. As detailed in Section 2, it first reveals that the cost of *maintaining* attention over time is high and convex in duration (i.e. the h function in Section 2 is convex). The data show (Figures 5 to 7) that the increase is steep as it is large in magnitude and concentrated over a very limited period of time. This suggests that the cost of maintaining attention is actually high relative to the utility gain it would generate in terms of consumption smoothing. Similarly, the pattern exhibited by the data reveals that the cost of achieving high attention levels is necessarily non-trivial (i.e., the K function is either convex or not too concave). The intuition is that otherwise, it would be optimal to discretely jump to the highest possible attention levels early in the

[61]This corresponds to the earliest date publicly available:
https://www.banque-france.fr/statistiques/taux-dusure-2016t3

month rather than gradually reaching those levels a few days before payday. Once again, this shows that the cost of reaching high attention levels continuously for longer periods of time is larger than the benefit it would bring in terms of smoothing consumption over that same period. Workers rather find it optimal to start low and only increase gradually their effort of attention. It also shows that the cost of ultimately achieving high levels of attention is smaller than the financial costs that are avoided.

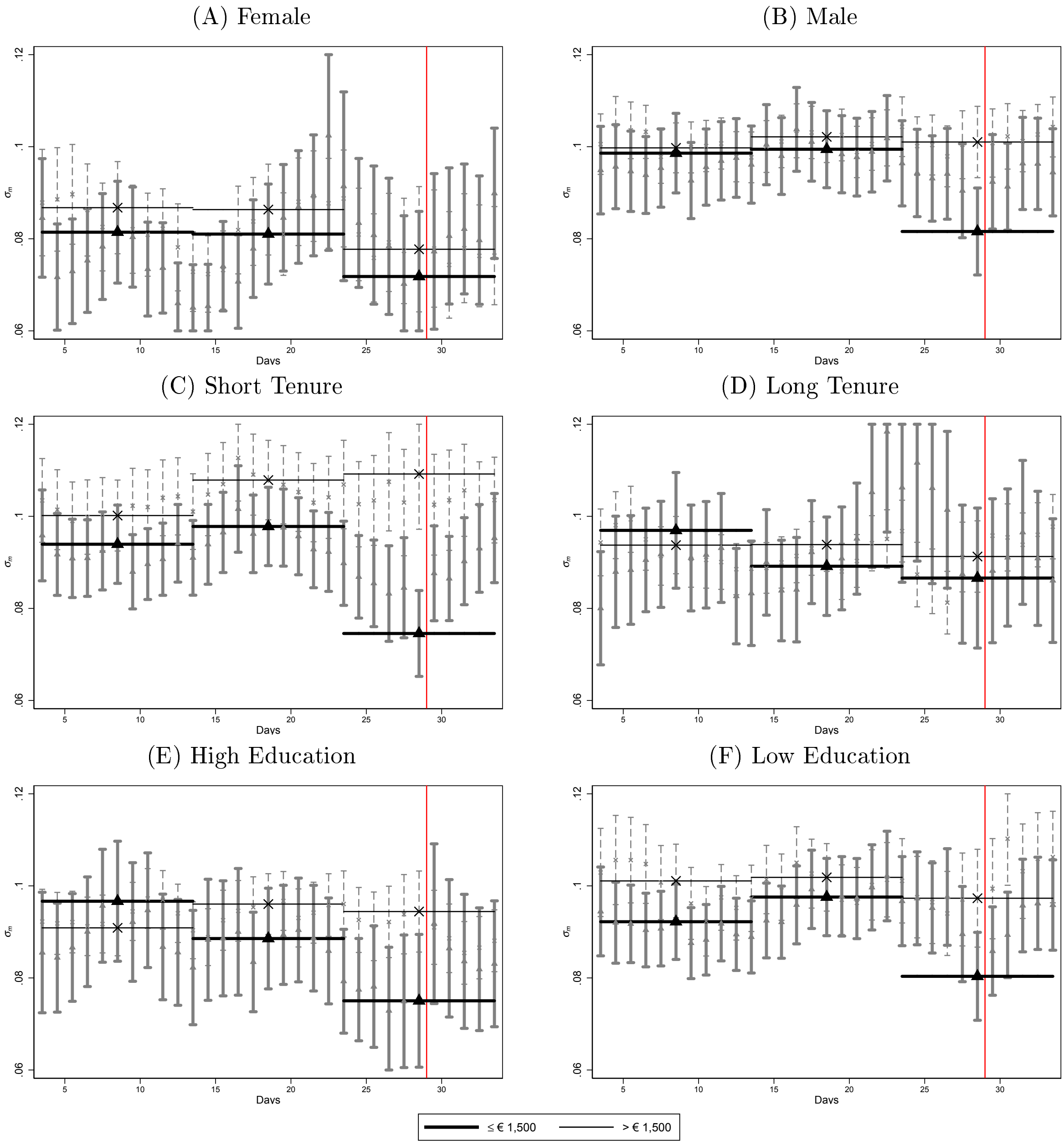
Lastly, our set-up also allows us to go one step further and provide a lower bound on global attention costs. The intuition behind this result is the following: in Figure 6, the monthly patterns of attention reveal that the cost in terms of attention of maintaining σ_m to its minimum for more than ten days (which corresponds to $\tau \leq 0.67$ with the notations of Section 2) is too high to be optimal. This implies that the attention cost associated with such an effort exceeds the utility gains which could be expected from it. Applying a linearization of the utility function of our model in Section 2 in the neighborhood of the optimal choices $(\bar{\bar{C}}, m)$, this implies that:

$$h(1-\tau).K'(m)\ |\mathrm{d}m| \geq \frac{R_A.W^r}{2}.\frac{\sigma^2}{\sigma_m}.e^{\frac{\sigma_m^2}{2}}.\left[\frac{R_B}{R_A}.\left(\phi\left(\frac{\ln\bar{\bar{C}} - w^r}{\sigma_m} - \sigma_m\right) - \sigma_m.\Phi\left(\frac{\ln\bar{\bar{C}} - w^r}{\sigma_m} - \sigma_m\right)\right) - \sigma_m\right]|\mathrm{d}m|, \quad \tau \leq 0.67 \tag{25}$$

Equation 25 provides a lower bound for the cost associated with reaching the level of $m = 0.73$ rather than $m = 0.65$ over periods of time ranging between 10 days (the actual period of increasing attention level) and one full month. However, this bound is expressed is expressed in terms of utility and is thus not directly interpretable. We therefore rather compute the increase in expected wage $\mathrm{d}W^r$ which would cause the same variation in utility as the attention cost $h(1-\tau).K'(m)\ |\mathrm{d}m|$. The derivations in Appendix 2 show that this simply amounts to normalize the right-hand side term of Equation 25 by $R_A.e^{\frac{\sigma^2}{2}}.\left(1 + \frac{R_B}{R_A}.\Phi\left(\frac{\ln\bar{C} - w^d}{\sigma} - \sigma\right)\right)$. We use the same benchmark as previously and calibrate R_B and R_A to their observed counterparts, ie. $R_B \approx 1.2$ and $R_A \approx 1.01$.[62] The result is plotted in Panel (B) of Figure 8. The obtained estimates of the lower bound for the annual attention cost range between €10 and €50. These amounts are typically in line with the associated risks, ie. the financial fees ($R_B \approx 20\%$) that would be charged to a worker who would exceed her budget constraint (€1,500) by σ_m. Such fees would typically range annually between €25 and €28. Furthermore, the lower bound in Equation 25 is tight for $\tau \approx 0.67$, ie. for the cost of maintaining high attention during periods of time that are close to those observed in the data (roughly 10 days per month). Assuming that the cost of maintaining high levels of attention across time is proportional to duration (ie. h is linear), our estimate imply that the cost of maintaining a high level of attention throughout the month would range between €30 and €150.

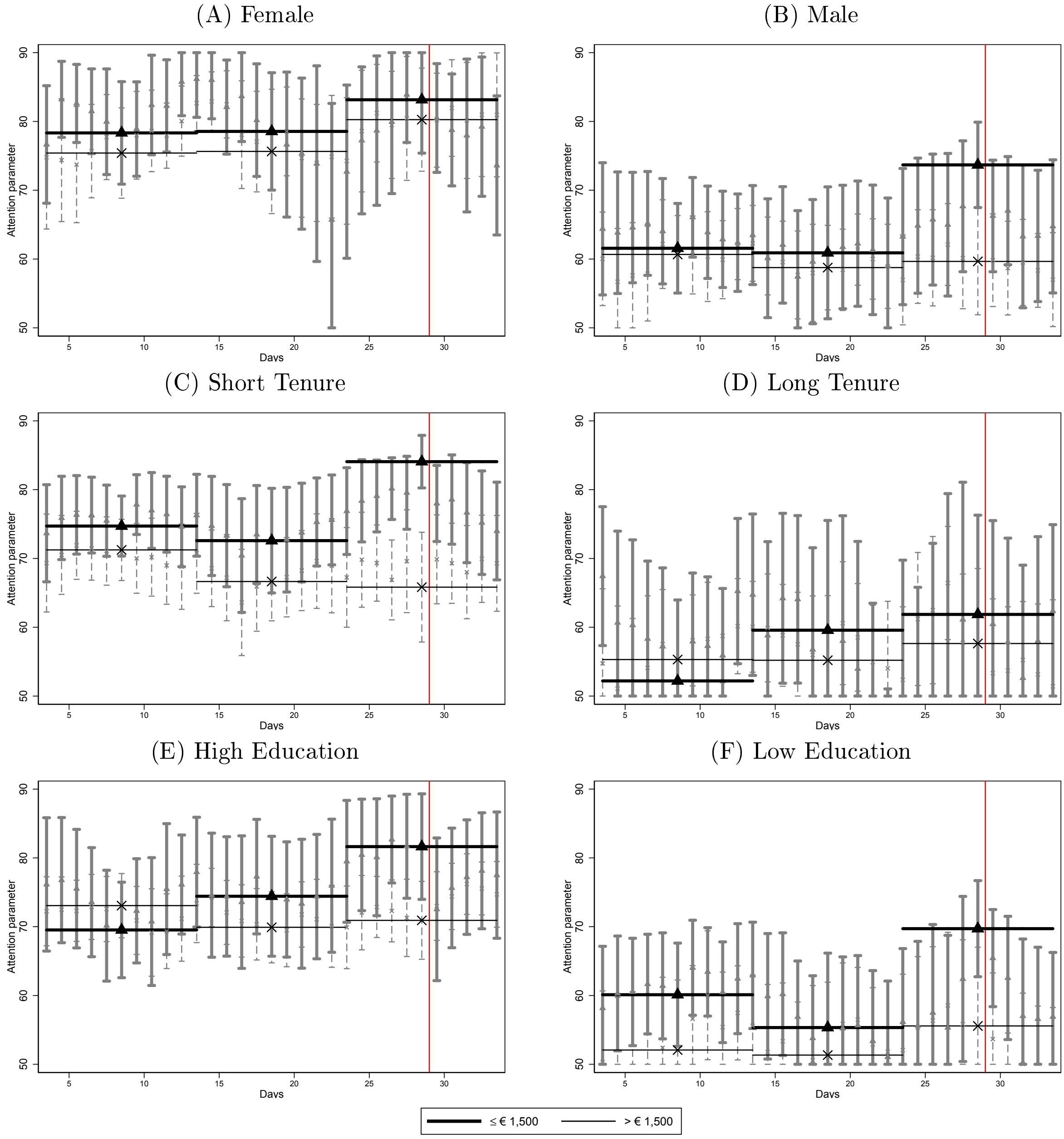
[62]Notice that by construction, the bound in Equation 25 is driven to 0 when $\frac{R_B}{R_A}$ is itself set to its lower bound (Equation 24). This is because in this situation, the worker is no longer financially constrained and do not varies its level of attention during the month, such that the exercise of comparative statics underlying Equation 25 is no longer possible.

Figure 9: Evolution of the Variance Parameter σ_m During the Month Across Different Sub-Populations of Workers



Notes: These figures complement Panel (E) in Figure 5 by providing estimates of the uncertainty parameter (σ_m) obtained across the month for different sub-populations of workers: women in Panel (A), men in Panel (B), workers with shorter or longer than median tenure in Panels (C) and (D) respectively, more educated workers in Panel (E) and less educated workers in Panel (F). Sources: ERFS (2005-2015). All standard errors are computed via the BHHH methodology.

Figure 10: Evolution of the Attention Parameter 100.m During the Month
Across Different Sub-Populations of Workers



Notes: These figures complement Panel (F) in Figure 5 by providing the estimates of attention (100.m) obtained across the month for different sub-populations of workers: women in Panel (A), men in Panel (B), workers with shorter or longer than median tenure in Panels (C) and (D) respectively, more educated workers in Panel (E) and less educated workers in Panel (F). Sources: ERFS (2005-2015)/DADS (1995-2015). All standard deviations are computed by bootstrap (50 replications).

Heterogeneity across different populations of workers. Finally, Figures 9 and 10 investigate whether the monthly pattern of attention is associated with specific worker-level characteristics. To that end, we further split our sub-period estimation samples according to the dimensions introduced above in Section 5.2: gender, tenure and education.

Considering first the comparison between women and men, Panels (A) and (B) in Figures 9 and 10 show that the end-of-month decrease in uncertainty σ_m and the corresponding increase in attention m is mainly driven by male workers. Overall, this pattern is highly attenuated across women, who appear to continuously maintain levels of uncertainty that are lower than the minimum attained in the sample of male workers and, symmetrically, levels of attention that are higher than those of men. Through the lens of our model, this behavior may be driven by the fact that women face such high financing constraints that they need to maintain a high level of attention throughout the month.

Panels (C) and (D) in Figures 9 and 10 contrast workers with shorter or longer tenure. Our estimations show that the cyclical pattern of attention is disproportionately driven by workers with shorter tenure and that, in addition, these workers continuously exhibit higher levels of attention than those with longer tenure. Again, these results are consistent with the empirical predictions in Section 2. They can be rationalized by the fact that workers with shorter tenure are likely to have accumulated a lower buffer of wealth and are thus more exposed to financing constraints. Furthermore, in France, they are more at risk of income losses as, in the case of an economic downturn, labor laws stipulate a type of seniority-based rules regarding dismissals (last in, first out). This feature of the French labor market is likely to create higher financing constraints, thus rationalizing both the higher overall level of attention and the additional increase in days to payday among shorter-tenure workers.

Lastly, Panels (E) and (F) in Figures 9 and 10 contrast the results for workers with different education levels. Interestingly, we find that the pattern of an end-of-month increase in attention is somewhat more pronounced for low-wage workers with lower education levels than for low-wage workers with higher education levels, but not by much: the increase is 16.3 ppt (27% of the base level) for the former and 13.1 ppt (17% of the base level) for the second. This suggests that neither education nor financial literacy seem to play a large role in shaping the monthly patterns of attention.

6 Conclusion

In this paper, we take advantage of a unique dataset (ERFS) which combines the values of wages that are self-reported by workers in the French Labor Force Survey together with corresponding fiscal items. This is a unique setting where both the reported answer and an accurate proxy of the correct response are known at the worker level and with a (unfortunately short but highly useful) panel dimension. This information allows us to provide a detailed description of workers' reporting behavior, which we reinterpret as a cognition test providing rich insights about the prevalence of "real-life" (behavioral) biases and most importantly, about workers' attention to their own wage and budget constraint. We propose a structural mixture model which allows for worker-level heterogeneity and which addresses the issues raised by perception or reporting biases and rounding. We estimate this model using unsupervised clustering techniques, which enables us to quantify attention at the worker level.

We find that over our period of study (2005 to 2015) workers tended to perceive their own wages with a degree of uncertainty of around 10%. Through the lens of a simple signal extraction model, this amounts to estimates of workers' attention ranging between 30% and 84% depending on their wage level, education, tenure or gender. Secondly, we use a feature of the sampling scheme of the French Labor Force Surveys which makes the date of interview, and critically, its distance to payday, orthogonal to workers' characteristics and prevents any selection issues. This enables us to show that low-wage workers who are most likely to experience a difficult "end of month", actually exhibit suggestive patterns of cyclicality. Their attention is minimal in the middle of the month and increases steadily until payday, suggesting that their respective budget constraints become increasingly tight over this final period of the month and require more monitoring. Equally interestingly, attention drops immediately once payday is reached. This feature of the cyclical evolution of attention is not compatible with a pure informational story, whereby workers are measured as being more "attentive" simply because they are more informed. Conversely, this feature is well rationalized by a mechanism of credit constraints with costly budget constraint monitoring.

As a last comparison, attention exhibits no cyclical pattern across high-wage workers and is, on average, lower - at approximately the minimum monthly level of attention reached by low-wage workers. This suggests that in everyday life (i.e. with the likely exception of when they have to take discretionary investment decisions) better-off workers do not need to remember a precise record of their wage, because their budget constraints are not tight. Overall, these results are indicative of the fact that less well-off, financially constrained workers have a greater mental burden (Mani et al., 2013; Shah et al., 2018; Schilbach et al., 2016). According to our estimates, the bottom 30% of French workers in the wage distribution are subject to these cycles and can therefore be considered as being in that situation. Our model identifies a lower bound for the annual attention cost burden incurred by financially constrained workers, which ranges between €10 and €50.

Overall, our contribution shows through real-world data that the fluctuation of mental burden is pervasive and easily detectable across less well-off workers. Our methodology could be adapted and implemented in other settings where correct information can be compared with data that are self-reported by agents, thus providing measures of their attention to different aspects of their economic environment (prices, characteristics of goods, etc.). One of our other results requires in particular further scrutiny using richer data. We show that worker-level biases are stable across time (Stango and Zinman, 2020) and therefore generally unrelated to their level of attention. This result, if confirmed, is interesting, as it suggests that nudges which increase attention are not likely to have an impact on behavioral biases.

References

Aghion, Philippe, Ufuk Akcigit, Matthieu Lequien, and Stefanie Stantcheva (2017), "Tax Simplicity and Heterogeneous Learning." Working Paper 24049, National Bureau of Economic Research.

Allcott, Hunt (2011), "Consumers' Perceptions and Misperceptions of Energy Costs." *American Economic Review*, 101, 98–104.

Allcott, Hunt, Joshua J. Kim, Dmitry Taubinsky, and Jonathan Zinman (forthcoming), "Are High-Interest Loans Predatory? Theory and Evidence from Payday Lending." *Review of Economic Studies*.

Belzil, Christian, Julie Pernaudet, and François Poinas (2021), "Estimating Coherency between Survey Data and Incentivized Experimental Data." IZA Discussion Papers 14594, Institute of Labor Economics (IZA).

Berndt, Ernst R., Bronwyn H. Hall, Robert E. Hall, and Jerry A. Hausman (1974), *Estimation and Inference in Nonlinear Structural Models*, 653–665. NBER.

Binder, Carola C. (2017), "Measuring Uncertainty Based on Rounding: New Method and Application to Inflation Expectations." *Journal of Monetary Economics*, 90, 1 – 12.

Biscourp, Pierre, Orietta Dessy, and Nathalie Fourcade (2005), "Les salaires sont-ils rigides ? Le cas de la France à la fin des années 1990." *Economie et statistique*, 386, 59–89.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and Huub Meijers (2009), "Gender Differences in Risk Aversion and Ambiguity Aversion." *Journal of the European Economic Association*, 7, 649–658.

Busse, Meghan R., Nicola Lacetera, Devin G. Pope, Jorge Silva-Risso, and Justin R. Sydnor (2013), "Estimating the Effect of Salience in Wholesale and Retail Car Markets." *American Economic Review*, 103, 575–79.

Cahuc, Pierre, Stéphane Carcillo, and Thomas Le Barbanchon (2018), "The Effectiveness of Hiring Credits." *Review of Economic Studies*, 86, 593–626.

Carvalho, Leandro S., Stephan Meier, and Stephanie W. Wang (2016), "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday." *American Economic Review*, 106, 260–84.

Chamberlain, Gary (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions." *Journal of Econometrics*, 34, 305–334.

Chetty, Raj, Adam Looney, and Kory Kroft (2009), "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99, 1145–77.

Croson, Rachel and Uri Gneezy (2009), "Gender Differences in Preferences." *Journal of Economic Literature*, 47, 448–74.

Delaney, Judith M. and Paul J. Devereux (2019), "More Education, Less Volatility? The Effect of Education on Earnings Volatility over the Life Cycle." *Journal of Labor Economics*, 37, 101–137.

D'Haultfoeuille, Xavier, Christophe Gaillac, and Arnaud Maurel (2018), "Rationalizing Rational Expectations? Tests and Deviations." Working Paper 25274, National Bureau of Economic Research.

Dickens, William T., Lorenz Goette, Erica L. Groshen, Steinar Holden, Julian Messina, Mark E. Schweitzer, Jarkko Turunen, and Melanie E. Ward (2007), "How Wages Change: Micro Evidence from the International Wage Flexibility Project." *Journal of Economic Perspectives*, 21, 195–214.

Dube, Arindrajit, Alan Manning, and Suresh Naidu (2018), "Monopsony and Employer Mis-optimization Explain Why Wages Bunch at Round Numbers." Working Papers 24991, National Bureau of Economic Research.

Ferrario, Beatrice and Stefanie Stantcheva (2022), "Eliciting People's First-Order Concerns: Text Analysis of Open-Ended Survey Questions." *AEA Papers and Proceedings*, forthcoming.

Gabaix, Xavier (2019), "Behavioral Inattention." In *Handbook of Behavioral Economics - Foundations and Applications* (B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds.), volume 2, 261–343, North-Holland.

Gabaix, Xavier and David Laibson (2006), "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics*, 121, 505–540.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2018), "Income Inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA)." *Journal of Public Economics*, 162, 63–77.

Gneezy, Uri and Alex Imas (2017), "Chapter 10 - Lab in the Field: Measuring Preferences in the Wild." In *Handbook of Economic Field Experiments* (Abhijit Vinayak Banerjee and Esther Duflo, eds.), volume 1 of *Handbook of Economic Field Experiments*, 439–464, Elsevier, North-Holland.

Gneezy, Uri, Alex Imas, and Ania Jaroszewicz (2020), "The Impact of Agency on Time and Risk Preferences." *Nature Communications*, 11.

Gourieroux, Christian and Alain Monfort (1995), *Statistics and Econometric Models*. Themes in Modern Econometrics, Cambridge University Press.

Gourieroux, Christian, Alain Monfort, and Alain Trognon (1985), "Moindres carrés asymptotiques." *Annales de l'INSEE*, 91–122.

Handel, Benjamin R. and Jonathan T. Kolstad (2015), "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review*, 105, 2449–2500.

Haushofer, Johannes and Ernst Fehr (2014), "On the Psychology of Poverty." *Science*, 344, 862–867.

Heckman, James J. and Michael Sattinger (2015), "Introduction to The Distribution of Earnings and of Individual Output, by A.D. Roy." *Economic Journal*, 125, 378–402.

Hossain, Tanjim and John Morgan (2006), "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay." *The B.E. Journal of Economic Analysis and Policy*, 5, 1–30.

Ito, Koichiro (2014), "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review*, 104, 537–63.

Kahneman, Daniel (2011), *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1991), "The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 5, 193–206.

Kaplan, Greg and Giovanni L. Violante (2018), "Microeconomic Heterogeneity and Macroeconomic Shocks." *Journal of Economic Perspectives*, 32, 167–94.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach (2021), "Do Financial Concerns Make Workers Less Productive?" Working Paper 28338, National Bureau of Economic Research.

Kleven, Henrik J. (2016), "Bunching." *Annual Review of Economics*, 8.

Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor (2012), "Heuristic Thinking and Limited Attention in the Car Market." *American Economic Review*, 102, 2206–36.

Laibson, David (1997), "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112, 443–478.

Luo, Yulei (2008), "Consumption Dynamics under Information Processing Constraints." *Review of Economic Dynamics*, 11, 366 – 385.

Lusardi, Annamaria and Olivia S. Mitchell (2014), "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature*, 52, 5–44.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao (2013), "Poverty Impedes Cognitive Function." *Science*, 341, 976–80.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao (2020), "Scarcity and Cognitive Function around Payday: A Conceptual and Empirical Analysis." *Journal of the Association for Consumer Research*, 5, 365–376.

Mullainathan, Sendhil and Eldar Shafir (2013), *Scarcity: Why Having Too Little Means So Much*. NY: Times Books.

O'Donoghue, Ted and Matthew Rabin (1999), "Doing It Now or Later." *American Economic Review*, 89, 103–124.

Pischke, Jörn-Steffen (1995), "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study." *Journal of Business and Economic Statistics*, 13, 305–314.

Prati, Alberto (2017), "Hedonic Recall Bias. Why You Should Not Ask People How Much They Earn." *Journal of Economic Behavior and Organization*, 143, 78 – 97.

Reis, Ricardo (2006), "Inattentive Consumers." *Journal of Monetary Economics*, 53, 1761 – 1800.

Ruud, Paul, Daniel Schunk, and Joachim Winter (2014), "Uncertainty Causes Rounding: an Experimental Study." *Experimental Economics*, 17, 391–413.

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan (2016), "The Psychological Lives of the Poor." *American Economic Review*, 106, 435–40.

Schmutz, Benoît and Modibo Sidibé (2018), "Frictional Labour Mobility." *Review of Economic Studies*, 86, 1779–1826.

Shah, Anuj, Jiaying Zhao, Sendhil Mullainathan, and Eldar Shafir (2018), "Money in the Mental Lives of the Poor." *Social Cognition*, 36, 4–19.

Sims, Christopher A. (2003), "Implications of Rational Inattention." *Journal of Monetary Economics*, 50, 665 – 690.

Stango, Victor and Jonathan Zinman (2020), "Behavioral biases are temporally stable." Working Paper 27860, National Bureau of Economic Research.

Taubinsky, Dmitry and Alex Rees-Jones (2017), "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Review of Economic Studies*, 85, 2462–2496.

Tourangeau, Roger and Ting Yan (2007), "Sensitive Questions in Surveys." *Psychological Bulletin*, 133, 859.

Train, Kenneth (2003), *Discrete Choice Methods with Simulation*. SUNY-Oswego, Department of Economics.

Appendix

A Proofs, Derivations and Additional Results for Section 2

A.1 Complementary Comparative Statics

It is useful to complement the analyses presented in the main text with results describing how utility varies when the expected wage w^d varies. Relying again on the Envelope theorem (but applied to Equations 3 or 3), we obtain:

$$\frac{\partial U^{(0)}}{\partial w^d} = \underbrace{e^{w^d+\frac{\sigma^2}{2}}}_{W^d.e^{\frac{\sigma^2}{2}}}.\left(R_A + R_B.\Phi\left(\frac{\ln \bar{C} - w^d}{\sigma} - \sigma\right)\right) \tag{26}$$

Equation 26 allows quantifying the marginal utility gain associated with an additional €1 of expected wage. In terms of the log-linearized variable, this shift can be expressed as:

$$\begin{aligned} \mathrm{d}w^d &= \ln(W^d + 1) - \ln(W^d) \\ &= \ln\left(1 + \frac{1}{W^d}\right) \approx \frac{1}{W^d} \end{aligned}$$

such that the marginal utility gains writes:

$$\begin{aligned} \frac{\partial U^{(0)}}{\partial w^d}\mathrm{d}w^d &\approx e^{\frac{\sigma^2}{2}}.\left(R_A + R_B.\Phi\left(\frac{\ln \bar{C} - w^d}{\sigma} - \sigma\right)\right) \\ &\approx \underbrace{\left(1 + \frac{\sigma^2}{2}\right)}_{\approx 1}.\left(R_A + \underbrace{R_B.\Phi\left(\frac{\ln \bar{C} - w^d}{\sigma} - \sigma\right)}_{\geq 0}\right) \end{aligned}$$

Therefore, €1 of additional expected wage translates into an increase in utility by R_A among workers who are not financially constrained (featuring $R_A \geq 0, R_B = 0$): this simply corresponds to the interest rate they would mange to earn for the additional revenue. In the case of financially constrained workers ($R_A \geq 0, R_B > 0$), the increase is even larger since the additional €1 allows them to bypass the further costs associated with R_B.

A.2 Derivation of Equation 15

To simplify the derivations, let us rewrite Equations 12 and 14 in terms of $\bar{\bar{C}}$, $\sigma_m = \sigma.(1-m)^{1/2}$ and τ rather than $\bar{\bar{C}}$, m and τ. We get:

$$u'\left(\bar{\bar{C}} + \frac{\tau.(\bar{\bar{C}} - \bar{C})}{1-\tau}\right) = R_A + R_B.\Phi\left(\frac{\ln \bar{\bar{C}} - w^r}{\sigma_m}\right) \tag{27}$$

$$\mathcal{K}(\sigma_m).h'(1-\tau) = u\left(\bar{\bar{C}} + \frac{\tau.(\bar{\bar{C}} - \bar{C})}{1-\tau}\right) - u(\bar{C}) - \frac{\bar{\bar{C}} - \bar{C}}{1-\tau}.u'\left(\bar{\bar{C}} + \frac{\tau.(\bar{\bar{C}} - \bar{C})}{1-\tau}\right) \tag{28}$$

where $\mathcal{K}$ is defined as $\mathcal{K}(\sigma_m) = K\left(1 - \left(\frac{\sigma_m}{\sigma}\right)^2\right)$. Therefore $\mathcal{K}'(\sigma_m) = -2\frac{\sigma_m}{\sigma^2}.K'\left(1 - \left(\frac{\sigma_m}{\sigma}\right)^2\right) < 0$ and

$\mathcal{K}''(\sigma_m) = \frac{2}{\sigma^2}.\left[2.\left(\frac{\sigma_m}{\sigma}\right)^2.K''(m) - K'(m)\right]$.

Differentiating the two equations, we get:

$$\mathrm{d}\bar{\bar{C}} = -\frac{\frac{\bar{\bar{C}}-\bar{C}}{(1-\tau)^2}.u''\left(\bar{\bar{C}}+\frac{\tau.(\bar{\bar{C}}-\bar{C})}{1-\tau}\right)}{\frac{u''\left(\bar{\bar{C}}+\frac{\tau.(\bar{\bar{C}}-\bar{C})}{1-\tau}\right)}{1-\tau}-\frac{R_B.\phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}\right)}{C.\sigma_m}}\mathrm{d}\tau - \frac{\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}.\frac{R_B}{\sigma_m}.\phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}\right)}{\frac{u''\left(\bar{\bar{C}}+\frac{\tau.(\bar{\bar{C}}-\bar{C})}{1-\tau}\right)}{1-\tau}-\frac{R_B.\phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}\right)}{C.\sigma_m}}\mathrm{d}\sigma_m \quad (29)$$

$$\mathrm{d}\bar{\bar{C}} = -\frac{1}{1-\tau}\mathrm{d}\tau + \underbrace{\frac{h''(1-\tau).\mathcal{K}(\sigma_m)}{\frac{(\bar{\bar{C}}-\bar{C})}{(1-\tau)^2}.u''\left(\bar{\bar{C}}+\frac{\tau.(\bar{\bar{C}}-\bar{C})}{1-\tau}\right)}}_{*}\mathrm{d}\tau - \underbrace{\frac{h'(1-\tau).\mathcal{K}'(\sigma_m)}{\frac{(\bar{\bar{C}}-\bar{C})}{(1-\tau)^2}.u''\left(\bar{\bar{C}}+\frac{\tau.(\bar{\bar{C}}-\bar{C})}{1-\tau}\right)}}_{*}\mathrm{d}\sigma_m \quad (30)$$

We focus now on workers who are marginally financially constrained, such that the magnitude of their adjustment in consumption is limited: $\bar{\bar{C}} \approx \bar{C}$. For such workers, the terms denoted by a star are large relative to the others, which implies:

$$\frac{\mathrm{d}\sigma_m}{\mathrm{d}\tau} \approx \frac{h''(1-\tau).\mathcal{K}(\sigma_m)}{h'(1-\tau).\mathcal{K}'(\sigma_m)} \Longleftrightarrow \frac{\mathrm{d}m}{\mathrm{d}\tau} \approx \frac{h''(1-\tau).K(m)}{h'(1-\tau).K'(m)} \quad (31)$$

which coincides with Equation 15 in the main text.

A.3 Derivation of the Bounds for K''

To simplify the derivations, let us rewrite Equations 12 and 13 in terms of $x = \frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}$, $\sigma_m = \sigma.(1-m)^{1/2}$ and τ rather than $\bar{\bar{C}}$, m and τ. We get:[63]

$$u'\left(\frac{e^{w^r}.e^{\sigma_m.x}}{1-\tau} - \frac{\tau.\bar{C}}{1-\tau}\right) = R_A + R_B.\Phi(x) \quad (32)$$

$$\mathcal{K}'(\sigma_m).h(1-\tau) = e^{w^r+\frac{\sigma_m^2}{2}}.(\sigma_m.R_A + \sigma_m.R_B.\Phi(x-\sigma_m) - R_B.\phi(x-\sigma_m)) \quad (33)$$

Differentiating the two equations, we get:

$$\mathrm{d}x = \underbrace{\frac{\frac{\bar{\bar{C}}-\bar{C}}{1-\tau}.u''(.)}{(1-\tau).R_B.\phi(x) - \sigma_m.\bar{\bar{C}}.u''(.)}}_{*}\mathrm{d}\tau + \underbrace{\frac{\bar{\bar{C}}.x.u''(.)}{(1-\tau).R_B.\phi(x) - \sigma_m.\bar{\bar{C}}.u''(.)}}_{A}\mathrm{d}\sigma_m \quad (34)$$

$$\mathrm{d}x = -\underbrace{\frac{\mathcal{K}'(\sigma_m).h'(1-\tau)}{\bar{\bar{C}}.x.\phi(x).R_B}}_{B}\mathrm{d}\tau + \frac{(\mathcal{K}''(\sigma_m) - \sigma_m.\mathcal{K}'(\sigma_m))h(1-\tau)}{\bar{\bar{C}}.x.\phi(x).R_B}\mathrm{d}\sigma_m + \underbrace{\frac{\bar{\bar{C}}.\frac{\phi(x)}{\phi(x-\sigma_m)}.(R_A + R_B.(\Phi(x-\sigma_m) - x.\phi(x-\sigma_m)))}{\bar{\bar{C}}.x.\phi(x).R_B}}_{D}\mathrm{d}\sigma_m \quad (35)$$

[63]In the derivations, we make intensive use of the following result: $\frac{e^{w^r+\frac{\sigma^2}{2}}}{\bar{\bar{C}}} = \frac{\phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}\right)}{\phi\left(\frac{\ln\bar{\bar{C}}-w^r}{\sigma_m}-\sigma_m\right)}$.

For $\bar{\bar{C}} \approx \bar{C}$, the term denoted by a star is small relative to the others. This implies:

$$\frac{\mathrm{d}\sigma_m}{\mathrm{d}\tau} \approx \frac{B}{D - A + \underbrace{\frac{(\mathcal{K}''(\sigma_m) - \sigma_m.\mathcal{K}'(\sigma_m))h(1-\tau)}{\bar{\bar{C}}.x.\phi(x).R_B}}_{E}} \tag{36}$$

Whenever $x < 0$, i.e. whenever consumption is sustainable (see footnote 14 for this to hold in expectation), term A is positive (as $u'' < 0$), term B is positive (as $\mathcal{K}' < 0$), term D is negative and term E is positive. The data show that $\frac{\mathrm{d}\sigma_m}{\mathrm{d}\tau} \leq 0$. This implies that $\mathcal{K}''(\sigma_m) \geq 0$ or that it is negative but $(\mathcal{K}''(\sigma_m) - \sigma_m.\mathcal{K}'(\sigma_m))$ remains sufficiently large (above $E.(A - D)$). Given the relations between $\mathcal{K}$ and K indicated in Section A.2, the same qualitative statements hold true for K.

B Details of the Estimation Strategy

This Appendix provides full details of the estimation strategy set out in Section 4.3. Maximizing Equation 21 is performed using an EM algorithm. Within each iteration, we approximate the conditional probabilities using Gaussian quadrature.

B.1 Gaussian Quadrature (Section 4.3)

We begin with the estimation method for the conditional probabilities $\mathbb{P}(\Omega_i | N_i, X_i, \theta)$. A standard way to proceed is to introduce a conditioning on the random effect a_i:

$$\mathbb{P}(\Omega_i | N_i = n, X_i, \theta^n) = \begin{cases} \int_a \prod_{t=1,2} \left\{ \Phi\left(\frac{\ln\left(e^{w_{it}^r} + \frac{n}{2}\right) - w_{it}^f - a_i^n}{\sigma_m^n} \right) - \Phi\left(\frac{\ln\left(e^{w_{it}^r} - \frac{n}{2}\right) - w_{it}^f - a_i^n}{\sigma_m^n} \right) \right\} \varphi(a) \mathrm{d}a \\ \qquad \text{if both } e^{w_{i1}^r} \text{ and } e^{w_{i2}^r} \text{ are multiples of } N_i = n \ \ (\text{i.e. } n | e^{w_{i1,2}^r}) \\ 0 \qquad \text{if not.} \end{cases} \tag{37}$$

Taking advantage of the fact that a is assumed to be drawn from a Gaussian distribution, these conditional probabilities can be approximated by Gauss-Hermite quadrature methods. Formally, for K quadrature points, we get:

$$\mathbb{P}(\Omega_i | N_i = n, X_i, \theta^n) \approx \begin{cases} \frac{1}{\sqrt{\pi}} \sum_{k=1}^{K} \psi_k \prod_{t=1,2} \left\{ \Phi\left(\frac{\ln\left(e^{w_{it}^r} + \frac{n}{2}\right) - w_{it}^f - \sqrt{2}\sigma_a^n z_k - \mu_a^n}{\sigma_m^n} \right) - \Phi\left(\frac{\ln\left(e^{w_{it}^r} - \frac{n}{2}\right) - w_{it}^f - \sqrt{2}\sigma_a^n z_k - \mu_a^n}{\sigma_m^n} \right) \right\} \\ \qquad \text{if } n | e^{w_{i1,2}^r} \\ 0 \qquad \text{if not.} \end{cases} \tag{38}$$

where the z_k's are the roots of the Hermite polynomial of order K and the ψ_k are the associated weights. These quantities can be maximized separately in each class of rounding N and then plugged into Equation 21. Our different sets of parameters of interest are estimated using an EM algorithm to achieve convergence.

B.2 Implementation of the EM Algorithm (Section 4.3)

The full log-likelihood of our model is obtained by summing Equation 21 over all sample observations:

$$l(\Omega, N | X, (\theta^n), (\pi_n)) = \sum_i \ln\left(\sum_{n \in \mathcal{N}} \pi_n \mathbb{P}(\Omega_i | N_i = n, X_i, \theta^n) \right) \tag{39}$$

Maximizing this formula is not straightforward. The main difficulty arises from the summation across classes of rounding within the log function in Equation 39, which makes optimization numerically difficult. A robust way to address this problem is to implement an EM algorithm.[64] The principle of this algorithm is intuitive. Considering any set of distributions Q_i defined over $\mathcal{N}$, we can write:[65]

$$\begin{aligned}
l\left(\Omega, N|X,\theta,(\pi_n)\right) &= \sum_i \ln\left(\sum_{n\in\mathcal{N}} \pi_n \mathbb{P}\left(\Omega_i|N_i=n,X_i,\theta^n\right)\right) && (40)\\
&= \sum_i \ln\left(\sum_{n\in\mathcal{N}} Q_i(N_i)\frac{\pi_n \mathbb{P}\left(\Omega_i|N_i=n,X_i,\theta^n\right)}{Q_i(N_i)}\right) && (41)\\
&\geq \sum_i \sum_{\substack{n\in\mathcal{N}\\ n|e^{w^r_{i1,2}}}} Q_i(N_i)\ln\left(\frac{\pi_n \mathbb{P}\left(\Omega_i|N_i=n,X_i,\theta^n\right)}{Q_i(N_i)}\right) && (42)
\end{aligned}$$

where the inequality in Equation 42 follows from Jensen's inequality. The principle of the EM algorithm is to construct a convenient Q_i allowing us to approximate the log-likelihood $l\left(\Omega, N|X,\theta,(\pi_n)\right)$ closely and to maximize the right-hand side of Equation 42. In particular, for:

$$Q_i(N_i=n) = \frac{\pi_n \mathbb{P}\left(\Omega_i|N_i=n,X_i,\theta^n\right)}{\sum_k \pi_k \mathbb{P}\left(\Omega_i|N_i=k,X_i,\theta^n\right)} = Q(N_i=n|\Omega_i,X_i,\theta,(\pi_n)) \qquad (43)$$

we get that the ratio $\frac{\pi_n \mathbb{P}(\Omega_i|N_i=n,X_i,\theta^n)}{Q_i(N_i)}$ on the right-hand side of Equation 42 is constant, such that Equation 42 holds with equality. For such choice of Q_i, maximizing the right-hand side of Equation 42 is therefore equivalent to maximizing the log-likelihood $l\left(\Omega, N|X,\theta,(\pi_n)\right)$.[66]

The expectation that is actually maximized at each step of the algorithm is directly derived from this insight. It is defined as:

$$\begin{aligned}
\mathcal{E}\left(\boldsymbol{\theta},(\boldsymbol{\pi_n})\,|\theta^{n,t},(\pi^t_n)\right) &= \sum_i\sum_n Q(N_i|\Omega_i,X_i,\theta^{n,t},(\pi^t_n))\ln\left(\frac{\boldsymbol{\pi_n}\mathbb{P}\left(\Omega_i|N_i=n,X_i,\boldsymbol{\theta^n}\right)}{Q(N_i|\Omega_i,X_i,\theta^t,(\pi^t_n))}\right) && (44)\\
&= \sum_n\sum_i Q(N_i|\Omega_i,X_i,\theta^{n,t},(\pi^t_n))\ln\left(\boldsymbol{\pi_n}\right)\\
&\quad +\sum_n\sum_i Q(N_i|\Omega_i,X_i,\theta^{n,t},(\pi^t_n))\ln\left(\mathbb{P}\left(\Omega_i|N_i=n,X_i,\boldsymbol{\theta}\right)\right)\\
&\quad -\sum_n\sum_i Q(N_i|\Omega_i,X_i,\theta^{n,t},(\pi^t_n))\ln\left(Q(N_i|\Omega_i,X_i,\theta^t,(\pi^t_n))\right) && (45)\\
\left\{\left(\theta^{n,t+1}\right),\left(\pi^{t+1}_n\right)\right\} &= \arg\max_{(\boldsymbol{\theta^n}),(\boldsymbol{\pi_n})} \mathcal{E}\left(\left(\boldsymbol{\theta^n}\right),(\boldsymbol{\pi_n})\,|\left(\theta^{n,t}\right),(\pi^t_n)\right) && (46)
\end{aligned}$$

Equation 45 shows that the maximization problem described in Equation 46 can be split into two inde-

[64]Our main references here are Train (2003) and lecture notes by Andrew Ng which are available online from his website.

[65]These distributions are potentially worker specific.

[66]Another important property of Q_i as defined in Equation 43 is that it coincides with the posterior probability that worker i rounds at level n, conditionally on all observables (X_i,Ω_i):

$$\begin{aligned}
Q_i(N_i=n) &= \frac{\pi_n \mathbb{P}\left(\Omega_i|N_i=n,X_i,\theta\right)}{\sum_k \pi_k \mathbb{P}\left(\Omega_i|N_i=k,X_i,\theta\right)} = \frac{\mathbb{P}\left(\Omega_i,N_i=n|X_i,\theta\right)}{\sum_k \mathbb{P}\left(\Omega_i,N_i=k|X_i,\Omega_i,\theta,(\pi_n)\right)}\\
&= \mathbb{P}\left(N_i=n|X_i,\Omega_i,\theta\right)
\end{aligned}$$

pendent sub-problems:[67]

1. The first sub-problem, embodied in the first term of Equation 45, involves (π_n) only and is easily solved as: $\pi_n^{t+1} = \frac{\sum_i Q_i(N_i|X_i,(\theta^{n,t}),(\pi_n^t))}{I}$, where I is the number of workers in the estimation sample.

2. The second sub-problem is embodied in the second term of Equation 45 and involves θ only. As the problem is also additively separable, it is easily solved class of rounding by class of rounding using the methodology described in Section B.1.

Finally, the algorithm is initiated for all classes of rounding at the following starting values: $\mu_a^0 = 0$, $\sigma_a^0 = 0.1$, $\sigma_v^0 = 0.1$ and $\pi_n = \frac{1}{Card(\mathcal{N})}$ for all $n \in \mathcal{N}$. Once convergence is reached, all standard errors are computed using the Berndt–Hall–Hall–Hausman (BHHH) estimator (Berndt et al., 1974). To that end, we take advantage of the fact that the score is easily evaluated in our setting as (Train, 2003):

$$\left.\frac{\mathrm{d}l\left(\Omega, N|X,\left(\theta^n\right),\left(\pi_n\right)\right)}{\mathrm{d}\theta}\right|_{(\theta^{n,t}),(\pi_n^t)} = \left.\frac{\mathrm{d}\mathcal{E}\left(\left(\theta^n\right),\left(\pi_n\right)|\left(\theta^{n,t}\right),\left(\pi_n^t\right)\right)}{\mathrm{d}\theta}\right|_{(\theta^{n,t}),(\pi_n^t)} \tag{47}$$

$$\left.\frac{\mathrm{d}l\left(\Omega, N|X,\left(\theta^n\right),\left(\pi_n\right)\right)}{\mathrm{d}\pi_n}\right|_{(\theta^{n,t}),(\pi_n^t)} = \left.\frac{\mathrm{d}\mathcal{E}\left(\left(\theta^n\right),\left(\pi_n\right)|\left(\theta^{n,t}\right),\left(\pi_n^t\right)\right)}{\mathrm{d}\pi_n}\right|_{(\theta^{n,t}),(\pi_n^t)} \tag{48}$$

Derivation of the score. In order to improve the accuracy of our estimates (and most importantly the estimated standard errors associated with the estimators of the main parameters of interest), we inserted the analytical expression of the score into our algorithm. To simplify notations, let us define the following parameters:

$$\overline{\alpha}_{ikt}^n \equiv \frac{\ln\left(e^{w_{it}^r} + \frac{n}{2}\right) - w_{it}^f - \sqrt{2}\sigma_a^n z_k - \mu_a^n}{\sigma_m^n}$$

$$\underline{\alpha}_{ikt}^n \equiv \frac{\ln\left(e^{w_{it}^r} - \frac{n}{2}\right) - w_{it}^f - \sqrt{2}\sigma_a^n z_k - \mu_a^n}{\sigma_m^n}$$

This allows us to re-write Equation 38 in a more compact form as:

$$\mathbb{P}\left(\Omega_i|N_i = n, X_i, \theta^n\right) \approx \begin{cases} \frac{1}{\sqrt{\pi}} \sum_{k=1}^K \psi_k \prod_{t=1,2} \left\{\Phi\left(\overline{\alpha}_{ikt}^n\right) - \Phi\left(\underline{\alpha}_{ikt}^n\right)\right\} & \text{if } n|e^{w_{i1,2}^r} \\ 0 \quad \text{if not.} \end{cases}$$

For each class of rounding n,[68] the vector of scores (at the worker level) can be computed from the derivatives of the second term of Equation 45 with respect to μ_a^n, σ_a^n and σ_m^n:

[67] Note that the third term in Equation 45 is simply a constant.

[68] Recall that we estimate one set of parameters $(\mu_a, \sigma_a, \sigma_m)$ by class of rounding, although this is not reflected in the notations in order to ease readibility.

$$\left.\frac{\mathrm{d}l}{\mathrm{d}\mu_a^n}\right|_{(\theta^{nt}),(\pi_n^t)} = -\frac{Q_i(N_i)}{\sigma_m}\sum_{k=1}^{K}\psi_k\frac{\left(\phi(\overline{\alpha}_{ik1})-\phi(\underline{\alpha}_{ik1})\right)\left(\Phi(\overline{\alpha}_{ik2})-\Phi(\underline{\alpha}_{ik2})\right)+\left(\phi(\overline{\alpha}_{ik2})-\phi(\underline{\alpha}_{ik2})\right)\left(\Phi(\overline{\alpha}_{ik1})-\Phi(\underline{\alpha}_{ik1})\right)}{\sum_{k=1}^{K}\psi_k\left(\Phi\left(\overline{\alpha}_{ik1}\right)-\Phi\left(\underline{\alpha}_{ik1}\right)\right)\left(\Phi\left(\overline{\alpha}_{ik2}\right)-\Phi\left(\underline{\alpha}_{ik2}\right)\right)}$$

$$\left.\frac{\mathrm{d}l}{\mathrm{d}\sigma_a^n}\right|_{(\theta^{nt}),(\pi_n^t)} = -\frac{Q_i(N_i)\sqrt{2}}{\sigma_m}\sum_{k=1}^{K}\psi_k z_k\frac{\left(\phi(\overline{\alpha}_{ik1})-\phi(\underline{\alpha}_{ik1})\right)\left(\Phi(\overline{\alpha}_{ik2})-\Phi(\underline{\alpha}_{ik2})\right)+\left(\phi(\overline{\alpha}_{ik2})-\phi(\underline{\alpha}_{ik2})\right)\left(\Phi(\overline{\alpha}_{ik1})-\Phi(\underline{\alpha}_{ik1})\right)}{\sum_{k=1}^{K}\psi_k\left(\Phi\left(\overline{\alpha}_{ik1}\right)-\Phi\left(\underline{\alpha}_{ik1}\right)\right)\left(\Phi\left(\overline{\alpha}_{ik2}\right)-\Phi\left(\underline{\alpha}_{ik2}\right)\right)}$$

$$\left.\frac{\mathrm{d}l}{\mathrm{d}\sigma_m^n}\right|_{(\theta^{nt}),(\pi_n^t)} = -\frac{Q_i(N_i)}{\sigma_m}\sum_{k=1}^{K}\psi_k\frac{\left(\overline{\alpha}_1\phi(\overline{\alpha}_1)-\underline{\alpha}_{ik1}\phi(\underline{\alpha}_{ik1})\right)\left(\Phi(\overline{\alpha}_{ik2})-\Phi(\underline{\alpha}_{ik2})\right)+\left(\overline{\alpha}_{ik2}\phi(\overline{\alpha}_{ik2})-\underline{\alpha}_{ik2}\phi(\underline{\alpha}_{ik2})\right)\left(\Phi(\overline{\alpha}_{ik1})-\Phi(\underline{\alpha}_{ik1})\right)}{\sum_{k=1}^{K}\psi_k\left(\Phi\left(\overline{\alpha}_{ik1}\right)-\Phi\left(\underline{\alpha}_{ik1}\right)\right)\left(\Phi\left(\overline{\alpha}_{ik2}\right)-\Phi\left(\underline{\alpha}_{ik2}\right)\right)}$$

Lastly, the gradient with respect to the π_n's can be computed from the first term of Equation 45 as:

$$\left.\frac{\mathrm{d}l\left(\Omega, N|X,\left(\theta^n\right),\left(\pi_n\right)\right)}{\mathrm{d}\pi_n}\right|_{(\theta^{nt}),(\pi_n^t)} = \frac{Q_i(N_i)}{\pi_n}$$

Inference for the parameter of attention, m. As described in Section 4.5, the index of attention m is estimated by a combination of two structural parameters, σ_m and σ, which are estimated from two different datasets. The standard errors of our estimators of σ and m are both evaluated by bootstrap (50 replications).[69]

B.3 Daily Estimates of Attention; Estimates by Narrow Bins of Wages (Section 4.5)

This Appendix describes the method which enables us to retrieve the daily estimates of our various structural parameters presented in Section 5.3. These estimates are based on a series of estimations of our baseline empirical model on a sequence of samples defined by date of interview. As mentioned in Section 4.4, workers are almost never interviewed on the same day (within a month) in their first and second LFS interviews, such that it is not possible to strictly divide our sample according to day of interview. To overcome this difficulty, we widen the window and isolate samples of workers interviewed within the same 10-day time window, both in their first and second LFS interviews. However, this does not correspond to a partition in the mathematical sense, as workers whose interview days are more than 10 days apart (e.g. on the 5$^{\text{th}}$ of the month for the first interview and on the 20$^{\text{ieth}}$ of the month during the second interview) are excluded, while workers who are interviewed less than 10 days apart are inserted into up to 10 different such samples.

This structure is complex to handle, particularly because of the limited number of observations in our dataset compared to the large number of parameters that a full mixture model would require to estimate. In Section 5.3, we rely on the following approximation: we first compute a counterfactual daily partition which would have the same 10-day rolling window observation counts as the observed data. These approximate counts of daily observations $(n_d)_d$ and the actual observed sample size series $\left(n_{d/d+9}\right)_d$ thus verify the

[69] Note that the boostrapped standard errors which we obtain for all other parameters are reassuringly numerically close (up to the third digit) to the baseline BHHH estimates that are reported in the main text. These robustness checks are available upon request.

following relationships:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & & & & & & & & & & & & \ddots & \ddots & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} . \begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_{30} \\ n_{31} \end{pmatrix} = \begin{pmatrix} n_{1/10} \\ n_{2/11} \\ n_{3/12} \\ \vdots \\ n_{30/8} \\ n_{31/9} \end{pmatrix} \tag{49}$$

As the matrix on the left is full rank, we can recover the $(n_d)_d$ via a simple matrix inversion. The difficulty for us is that this matrix inversion ends up in low and even negative counterfactual observation counts. When this arises, we simply consider that the daily estimates of our structural parameters of interest are not identified in the data and remove these days from the subsequent computations.

The principle of the remainder of our estimation strategy is to implement a two-step minimum distance estimator (Chamberlain, 1987 or "asymptotic least squares" in the terminology of Gourieroux et al., 1985 and Gourieroux and Monfort, 1995) which allows us to retrieve our daily estimates of interest from our baseline sequence of 10-day rolling window estimations. Interpreting our sequence of samples of 10-day rolling windows as approximate "mixtures" of different sets of daily estimates, we get the following relationships between our "pooled" estimates and the daily contributions which we aim to recover:[70]

$$\widehat{\pi}_{1,d/d+9} = \sum_{\tau=d}^{d+9} s_{\tau}^{d/d+9}.\pi_{1,\tau} \tag{50}$$

$$\widehat{\sigma}^2_{m,d/d+9} = \sum_{\tau=d}^{d+9} s_{\tau}^{d/d+9}.\sigma^2_{m,\tau} \tag{51}$$

$$\widehat{m}^2_{d/d+9} = \sum_{\tau=d}^{d+9} s_{\tau}^{d/d+9}.m^2_{\tau} \tag{52}$$

$$\widehat{\mu}_{a,d/d+9} = \sum_{\tau=d}^{d+9} s_{\tau}^{d/d+9}.\mu_{a,\tau} \tag{53}$$

$$\widehat{\sigma}^2_{a,d/d+9} + \widehat{\mu}^2_{a,d/d+9} = \sum_{\tau=d}^{d+9} s_{\tau}^{d/d+9}.\left(\sigma^2_{a,\tau} + \mu^2_{a,\tau}\right), \tag{54}$$

where $s_{\tau}^{d/d+9}$ denotes the share (in terms of observation counts) accounted for day τ in the sample pooling days d to $d+9$.

Equations 50 to 53 can be treated independently to retrieve the daily contributions $\pi_{1,\tau}$, $\sigma_{m,\tau}$, m_{τ} and $\mu_{a,\tau}$, respectively. Introducing more compact matrix notations, we get:

[70] As explained above, our rolling window strategy does not explicitly take account of the mixture model nature of our problem, as this problem would remain too complex (i.e. high-dimensional) in comparison with the number of our sample observations. As a consequence, Equations 50 to 54 only hold approximately, and $\widehat{\pi}_{1,d/d+9}$, $\widehat{\sigma}^2_{m,d/d+9}$, $\widehat{\mu}_{a,d/d+9}$ and $\widehat{\sigma}^2_{a,d/d+9}$ are likely to be biased estimators of the true parameters in the pooled sample. This prevents us from applying the standard estimates of the variance matrix that would have been available if our estimates had been unbiased (Gourieroux and Monfort, 1995). As explained below, we instead evaluate this matrix by bootstrap.

$$M.\begin{pmatrix} \vdots \\ \pi_{1,\tau} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \widehat{\pi}_{1,d/d+9} \\ \vdots \end{pmatrix} \Longleftrightarrow \begin{pmatrix} \vdots \\ \pi_{1,\tau} \\ \vdots \end{pmatrix} = \left(M'M\right)^{-1} M'.\begin{pmatrix} \vdots \\ \widehat{\pi}_{1,d/d+9} \\ \vdots \end{pmatrix}, \tag{55}$$

and similarly for $\sigma^2_{m,\tau}$ and $\mu_{a,\tau}$. In Equation 55, M is simply the following matrix:

$$M = \begin{pmatrix} s_1^{1,10} & s_2^{1,10} & s_3^{1,10} & s_4^{1,10} & s_5^{1,10} & s_6^{1,10} & s_7^{1,10} & s_8^{1,10} & s_9^{1,10} & s_{10}^{1,10} & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & s_2^{2,11} & s_3^{2,11} & s_4^{2,11} & s_5^{2,11} & s_6^{2,11} & s_7^{2,11} & s_8^{2,11} & s_9^{2,11} & s_{10}^{2,11} & s_{11}^{2,11} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & s_3^{3,12} & s_4^{3,12} & s_5^{3,12} & s_6^{3,12} & s_7^{3,12} & s_8^{3,12} & s_9^{3,12} & s_{10}^{3,12} & s_{11}^{3,12} & s_{12}^{3,12} & 0 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & & & & & & & & & & & & \ddots & \ddots & \\ s_1^{30,8} & s_2^{30,8} & s_3^{30,8} & s_4^{30,8} & s_5^{30,8} & s_6^{30,8} & s_7^{30,8} & s_8^{30,8} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & s_{30}^{30,8} & s_{31}^{30,8} \\ s_1^{31,9} & s_2^{31,9} & s_3^{31,9} & s_4^{31,9} & s_5^{31,9} & s_6^{31,9} & s_7^{31,9} & s_8^{31,9} & s_9^{31,9} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & s_{31}^{31,9} \end{pmatrix}$$

Note that this matrix M is no longer square (nor full rank!) once the unidentified daily parameters have been filtered out (see above). However, $M'M$ remains full rank, which is all what Equation 55 requires.

The estimation of the $\sigma^2_{a,\tau}$ parameters is slightly less direct but remains simple. Once the $\mu_{a,\tau}$ have been computed from Equation 53, they can be plugged into Equation 54 to retrieve the $\sigma^2_{a,\tau}$ parameters in the same way as previously.

The variance of the resulting daily estimates are all evaluated by bootstrap.

Lastly, this methodology can be straightforwardly adapted to alternative partitions of our estimation sample. Specifically, we also estimate sets of estimators $(\widehat{\pi}^{(b)}_{1,d/d+9}, \widehat{\sigma}^{(b)}_{m,d/d+9}, \widehat{\mu}^{(b)}_{a,d/d+9}, \widehat{\sigma}^{(b)}_{a,d/d+9})$ over sequential 10-day windows in populations (indexed by b) of workers earning respectively: less than €1,400, strictly more than €1,400, less than €1,700, strictly more than €1,700, less than €2,000 and strictly more than €2,000. This opens the door to adapting the matrices in Equations 49 and 55 in order to retrieve estimates of our parameters of interest for workers earning respectively: less than €1,400, between €1,400 and €1,700, between €1,700 and €2,000, and strictly more than €2,000 (by 10-day windows). This simply involves working on wage ranges b rather than time ranges $d/d+9$.

C Additional Descriptive Statistics

This Appendix complements the descriptive statistics in Section 3.3:

- Table 7 documents the representativity of our estimation sample year by year. It shows that the numbers in Table 2 that are computed on the pooled sample also hold in each cross-section.

- Table 8 describes the separate impact of each filter that is required to harmonize the two concepts of wage in the LFS and in the fiscal files. These filters are described and motivated in Section 3.2 and induce a selection of the most stable workers. Our sample workers are, on average, somewhat older than the population of continuously employed workers, more educated and have longer tenure. This is mainly driven by the selection on tenure (longer than 15 months). Importantly, the ratio of female to male workers is also significantly lower in the population of workers which we focus on than in the general population of employed workers, mostly because of our selection of full-time workers. Lastly (and unfortunately), item non-response (i.e. missing information about self-reported wages) induces a large drop in the number of sample observations.

- Finally, Table 9 shows that the composition of our sample is stable across samples with different days of interviews (within months). This orthogonality between survey days and worker-level characteristics simplifies our analysis of the monthly evolutions of σ_m and attention, as explained in Section 1.

Table 7: Sample Representativeness Across Time

		Sample		Labor market aged 15 to 64	
		Un-weighted	Weighted (mil. workers)	All (mil. workers)	Emp. t/t+1 (mil. workers)
Nb workers:	2005/2006	1,567	2.373	12.711	7.515
	2006/2007	1,773	2.566	12.792	7.626
	2007/2008	1,791	2.666	13.109	8.024
	2008/2009	1,816	2.609	13.241	7.966
	2009/2010	2,427	2.560	13.112	7.689
	2010/2011	2,555	2.525	13.097	7.725
	2011/2012	2,583	2.487	13.057	7.696
	2014/2015	2,250	2.516	12.350	7.071
	2015/2016	2,283	2.555	12.555	7.163

Source: ERFS survey, 2005-2015.

Table 8: Analysis of The Selection Criteria in the Pooled Estimation Sample

		Employed at t and $t+1$, aged 15 to 64, and...					
		Full time	Tenure $\geq$ 15 m.	Single employer	Non-missing wages	Non-missing net wages	Wage range (1-4 k€)
Age:	15 to 24	0.048	0.031	0.053	0.046	0.047	0.036
	25 to 34	0.244	0.220	0.244	0.248	0.240	0.250
	35 to 44	0.305	0.312	0.303	0.315	0.303	0.318
	45 to 54	0.297	0.315	0.292	0.299	0.300	0.303
	55 to 64	0.105	0.121	0.109	0.093	0.110	0.094
Gender:	Female	0.266	0.307	0.310	0.300	0.321	0.278
Education:	No diploma (low)	0.342	0.328	0.336	0.322	0.305	0.329
	Lower than high school (low)	0.181	0.180	0.184	0.182	0.182	0.187
	High school degree (high)	0.271	0.273	0.268	0.282	0.288	0.281
	Higher than high school (high)	0.206	0.219	0.213	0.213	0.225	0.202
Occupation:	Managers/professionals	0.210	0.202	0.198	0.181	0.164	0.180
	Intermediate occupations	0.284	0.279	0.277	0.279	0.268	0.312
	Low-skilled white-collars	0.218	0.246	0.248	0.248	0.270	0.222
	Blue-collars	0.289	0.273	0.277	0.292	0.298	0.287
Tenure:	Average (months)	150	164	146	153	152	158
	Std dev.	(124)	(120)	(124)	(121)	(124)	(122)
Observations:	Total	60.507	59.679	66.672	37.674	37.962	50.643
(weighted)	Per annual wave	6.723	6.631	7.408	4.186	4.218	5.627

Source: ERFS survey, 2005-2015.

Table 9: Description of (Some of the) Sub-samples with Restricted Interviewing Days

		Full sample	Days 1 to 10	Days 5 to 14	Days 10 to 19	Days 15 to 24	Days 20 to 29	Days 25 to 3
Age:	15 to 24	0.024	0.027	0.023	0.024	0.023	0.030	0.026
	25 to 34	0.208	0.205	0.204	0.200	0.209	0.211	0.204
	35 to 44	0.314	0.319	0.320	0.302	0.323	0.309	0.316
	45 to 54	0.344	0.355	0.343	0.352	0.331	0.337	0.342
	55 to 64	0.110	0.094	0.110	0.122	0.113	0.113	0.112
Gender:	Female	0.239	0.229	0.243	0.224	0.242	0.239	0.223
Education:	No diploma (low)	0.287	0.299	0.296	0.268	0.280	0.278	0.275
	Lower than high school (low)	0.180	0.177	0.170	0.163	0.181	0.186	0.188
	High school degree (high)	0.314	0.318	0.312	0.328	0.314	0.327	0.328
	Higher than high school (high)	0.219	0.206	0.222	0.241	0.224	0.209	0.210
Occupation:	Managers/professionals	0.144	0.159	0.155	0.132	0.137	0.133	0.139
	Intermediate occupations	0.310	0.302	0.303	0.303	0.318	0.324	0.303
	Low-skilled white-collars	0.205	0.203	0.204	0.194	0.202	0.207	0.207
	Blue-collars	0.342	0.337	0.338	0.370	0.343	0.336	0.351
Tenure:	Average (months)	179	179	180	183	178	179	181
	Std dev.	(122)	(121)	(122)	(123)	(123)	(124)	(121)
Observations:	Total	19,045	3,193	3,193	3,194	3,142	2,329	2,076

Source: ERFS survey, 2005-2015.

D Robustness Checks: More Flexible Specifications

In a first series of complementary regressions, we check the robustness of our results in Section 5.1 to more flexible specifications. In particular, it is possible to allow for an even greater amount of heterogeneity in the parameters, although at the cost of a loss of statistical power.

In Table 10, we allow the time-varying uncertainty parameter σ_m to be different at the two dates workers are interviewed for the LFS. This leads to Equation 18 being split into two parts:

$$\upsilon_{i1} \sim \mathcal{N}\left(0, \sigma_{m1}^2\right) \tag{20a}$$

$$\upsilon_{i2} \sim \mathcal{N}\left(0, \sigma_{m2}^2\right) \tag{20b}$$

This augmented specification therefore allows for potential learning effects between the first and second interrogations.

The results show that the estimates of σ_m obtained in the two different waves are close, but statistically different. If anything, the index of uncertainty is higher in the second interrogation than in the first one, especially for the most populated classes of rounding: non-rounders and rounders at €50, €100 or €1,000. The difference is limited in magnitude and attains 0.6 percentage points, which corresponds to 5% of the baseline estimated standard deviation σ_m. In contrast, the difference is not statistically significant in the classes of rounders at €10 and €500. This finding goes against the assumption that there could be learning effects during the interrogation process of the LFS, as this would lead to the opposite result (i.e. to a lower volatility premium at the second interrogation). We also note that all other parameters are left broadly unchanged in this augmented specification.

Table 10: Allowing for Heterogeneous σ_m Across Waves of the LFS

Classes	€1	€10	€50	€100	€500	€1,000	Average
π	0.281	0.119	0.175	0.368	0.049	0.007	78.972
	(0.004)	(0.003)	(0.004)	(0.005)	(0.002)	(0.001)	(1.315)
$\sigma_{m,1}$	0.119	0.068	0.076	0.110	0.044	0.014	0.098
	(0.001)	(0.002)	(0.001)	(0.001)	(0.008)	(0.014)	(0.001)
$\sigma_{m,2}$	0.124	0.068	0.084	0.116	0.064	0.098	0.104
	(0.001)	(0.002)	(0.001)	(0.001)	(0.006)	(0.017)	(0.001)
μ_a	0.030	0.012	0.004	-0.004	-0.018	0.019	0.008
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.010)	(0.001)
σ_a	0.073	0.086	0.091	0.107	0.116	0.151	0.093
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.011)	(0.001)
Obs.							19,045
Ln-Lik							-169,435

Notes: This table reports the results obtained when constraining the rounding behavior to be stable across time, but allowing σ_m to vary across waves of the LFS to allow for potential learning effects.

Table 11 proposes an even more drastic exercise, in which all parameters are allowed to vary by wave of

Table 11: Estimations in the Cross-Section

Classes	€1	€10	€50	€100	Average
	First LFS Interrogation				
π	0.305	0.120	0.246	0.329	46.731
	(0.004)	(0.003)	(0.006)	(0.006)	(0.500)
$\sigma_m + \sigma_a$	0.197	0.160	0.160	0.230	0.194
	(0.001)	(0.004)	(0.002)	(0.001)	(0.001)
μ_a	0.032	0.010	0.001	-0.003	0.010
	(0.002)	(0.004)	(0.003)	(0.002)	(0.001)
Obs.					19,045
Ln-Lik					-95,092
	Second LFS Interrogation				
π	0.305	0.120	0.250	0.325	46.484
	(0.004)	(0.003)	(0.006)	(0.006)	(0.500)
$\sigma_m + \sigma_a$	0.206	0.156	0.157	0.239	0.198
	(0.001)	(0.005)	(0.002)	(0.001)	(0.001)
μ_a	0.028	0.011	(0.001)	-0.010	0.007
	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)
Obs.					19,045
Ln-Lik					-96,124

Notes: This table reports the results obtained when removing the panel dimension of our dataset.

the LFS. In particular, we do not impose that workers consistently round at the same level, and we also allow worker-level biases μ_a to vary across interrogations. The specification becomes fully cross-sectional and no longer relies on the panel dimension of the data. As a result, parameters σ_a and σ_m are no longer separately identified (such that we only report the sum of σ_a and σ_m). A second difficulty is that the number of parameters to be estimated increases, and we do not manage to reach convergence of our algorithm when we insert the least populated classes of rounding, €500 and €1,000. Table 11 therefore only contains the bottom four classes of rounding.

The main take-away of the estimates in Table 11 is that the profile of rounding captured by the probabilities π appears to be very stable across time. Likewise, the parameter measuring mean biases μ_a is also stable across time, with none of the estimated values being statistically different in the first and second interviews of the LFS. The only parameter which evolves slightly across time is the overall volatility parameter, which increases by 0.4 percentage points, i.e. 2% of the baseline estimate of $\sigma_m + \sigma_a$. This difference is statistically significant and confirms the picture in Table 10 both qualitatively and quantitatively, but is limited in terms of economic magnitude. Given this result, in all analyses in Sections 5.2 and 5.3 (where we investigate the heretogeneiy of our structural estimates across sub-samples of limited size) we neglect the potential evolution of the σ_m across interviews in order to limit the number of parameters to be estimated and preserve more statistical power.[71]

[71] When allowing for heterogeneity in σ_m in these sub-sample analyses, the difference between the two estimated parameters quickly becomes insignificant as sample size (and therefore, statistical power) drops by a lot.