NBER WORKING PAPER SERIES

TENURE-DEPENDENT SEVERANCE COSTS AND LABOR MARKET DYNAMICS OVER THE LIFE CYCLE

Calebe Figueiredo
Neville Francis

Working Paper 34704
http://www.nber.org/papers/w34704

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2026

We are thankful to Ludmilla Mattos at the Brazilian Institute for Applied Economic Research (Ipea) for excellent research assistance. We are grateful to conference participants at CEF 2021 and IAAE. We thank Fei Li for helpful comments and discussions. The views expressed do not reflect the views of Bank of America or the National Bureau of Economic Research. All errors are our own.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w34704

Tenure-Dependent Severance Costs and Labor Market Dynamics over the Life Cycle
Calebe Figueiredo and Neville Francis
NBER Working Paper No. 34704
January 2026
JEL No. C61, C78, J64

ABSTRACT

We study the labor market effects of tenure-dependent severance pay systems that tie firing costs to workers' accumulated earnings histories. We develop an overlapping-generations search-and-matching model in which firms face increasing separation costs over the employment relationship. Using administrative employer–employee data from Brazil, we estimate the model and show that tenure-dependent severance costs induce labor hoarding among low-productivity, long-tenured workers who would not be hired if unemployed. These distortions are strongest late in the life cycle, when firms optimally delay separation to avoid severance obligations, while simultaneously imposing higher hiring thresholds on older unemployed workers with shorter continuation values. These forces imply that tenure-based severance policies protect employment histories rather than productivity, shaping labor market dynamics over the life cycle and generating allocative inefficiencies.

Calebe Figueiredo
calebeof@gmail.com

Neville Francis
University of North Carolina at Chapel Hill
Department of Economics
and NBER
nfrancis@unc.edu

1 Introduction

We investigate the consequences of a severance policy whose protection increases with tenure. Such a policy imposes additional costs on employers precisely at times when downsizing may be optimal. While our analysis focuses on Brazil, the policy is of broader interest because of the search and matching frictions it generates, its implications for workers' life-cycle labor outcomes, and the fact that similar tenure-dependent labor protection mechanisms exist in both developed and developing economies.

Labor protection laws in Brazil mandate that terminated workers receive severance compensation proportional to their earnings history. This law, spelled out in the *Warranty Fund for Time of Service* (FGTS), has two key components: (i) an Unemployment Insurance Savings Account (UISA)—mandatory individual savings accumulated during employment and accessed after dismissal—and (ii) an accompanying severance penalty imposed on firms when a dismissal is deemed unjust. The size of the firing cost increases with workers' tenure.[1] This paper studies the impact of the severance penalty on firms' labor decisions, wage determination, and workforce composition. Our focus is on the market inefficiencies that arise when employment decisions are driven by accumulated severance liabilities rather than by current productivity.

Brazil's labor market has attracted considerable attention due to the distinctive features of its labor protection system. Strong employment protection combined with a bimodal skill distribution generates sizable wage differentials between the formal and informal sectors (Loureiro et al., 2013), motivating a large literature on the welfare consequences of labor market regulation (Bargain and Kwenda, 2014). More recently, Arnold and Bernstein (2018) and Figueiredo and Francis (2020) study the aggregate implications of Brazilian labor regulations and their role during economic downturns.

[1]This severance penalty arises only if firms prematurely terminate the employment contract without justification.

Our contribution is to uncover the mechanisms through which tenure-dependent severance costs distort firms' hiring and firing decisions over the life cycle. We develop a search-and-matching model in the spirit of Diamond (1982) and Mortensen and Pissarides (1994), embedded in an overlapping-generations framework in which workers of different ages compete for the same vacancies, following Cheron et al. (2013). In the model, firms observe worker productivity upon meeting a candidate and decide whether to operate the match or continue searching, taking into account both future productivity risk and accumulated severance liabilities.

Our framework differs from Arnold and Bernstein (2018), who assume a fixed severance cost, by explicitly linking firing costs to the history of wages paid to the worker. As a result, it may be more costly to dismiss a short-tenured high-wage worker than a long-tenured low-wage worker. By modeling severance costs through accumulated FGTS balances (see Section 2), rather than tenure alone, the model captures these subtleties in firms' separation incentives.

We estimate the model using administrative data on Brazilian formal-sector employment. The estimation disciplines the distribution of worker productivity and the magnitude of severance costs using age–tenure profiles of wages, employment shares, and separation rates. The estimated model delivers three main results. First, reservation productivity declines sharply with tenure, generating strong labor hoarding incentives for long-tenured workers. Second, reservation productivity rises with age at short tenure, implying that firms require higher productivity from older unemployed workers, reducing their job-finding prospects. Third, for older workers with sufficiently long tenure, firms optimally retain even very low-productivity matches rather than incur severance costs, effectively waiting for retirement to dissolve the employment relationship. Together, these forces imply that the FGTS system protects employment histories rather than productivity, simultaneously discouraging the hiring of older unemployed workers and insulating long-tenured incumbents from separation.

Beyond characterizing hiring and separation distortions, we quantify the welfare and efficiency implications of tenure-dependent severance pay. We show that by inducing firms to retain low-productivity, long-tenured matches while discouraging the hiring of older unemployed workers, the severance system generates nontrivial output and welfare losses. These losses arise primarily from distorted separation incentives rather than from the overall level of severance generosity, highlighting the importance of the tenure profile of firing costs.

Related evidence for the United States is documented by Hairault et al. (2019), who show that older workers experience longer unemployment spells and longer job tenure than younger workers. They attribute these patterns to age-related fluctuations in job opportunities over the business cycle. In contrast, we show that similar life-cycle patterns in Brazil arise endogenously from tenure-dependent severance costs embedded in the FGTS system.

The remainder of the paper is organized as follows. Section 2 describes the institutional features of the FGTS policy. Section 3 documents key stylized facts of the Brazilian labor market. Section 4 presents the model. Section 5 demonstrates the workings of the model with numerical simulations. Section 6 describes the estimation strategy and discusses the quantitative results. Welfare analyses are conducted in Section 7. Section 8 concludes.

2 Motivation

Labor market regulation in Brazil bars direct negotiations of the terms of severance between employers and employees. Instead, upon unjust termination, firms must pay a penalty that is proportional to the worker's *Warranty Fund for Time of Service* - in Portuguese, Fundo de Garantia por Tempo de Serviço (hereafter, FGTS) - balance. FGTS is a mandatory savings account assigned to each worker upon entering the *formal* labor market. During the employee's tenure, the firm is legally obligated to invest an extra 8% of a month's salary in the employee's

FGTS account at the end of every month.

Even though it is a savings vehicle, the FGTS account differs from a regular savings account in key aspects. For one, the federal government manages all FGTS accounts. Workers have no say about which financial institution will manage their FGTS savings. Secondly, workers have no say about asset allocation. Instead, the government exogenously sets the rate of return earned by all accounts. Thirdly, workers can only withdraw FGTS funds under special circumstances, such as severe illness, death, real estate purchases, retirement and *unjustified* termination. Exit from the labor market via quitting or justified termination are not valid reasons for FGTS withdrawals.

Justified termination requires the firm to supply evidence of the fired worker's misconduct, which include continued absence and dishonest practices, among other reasons. Poor market conditions are not (legally) valid reasons for justified terminations. That is, firms cannot fire workers with justification because business is slow.[2] In addition, even if the firm has evidence of a worker's misconduct, it will be responsible for all litigation costs and court fees regardless of the final outcome of the termination process. In practice, therefore, unless the firm has overwhelming evidence of a worker's misconduct and the severance penalty is greater than the expected litigation costs, employers invariably choose to terminate employment without justification to avoid court costs.

If an employer decides to terminate a worker based on market conditions or for any other unjustified reason, then there will be a penalty that amounts to 50% of the worker's FGTS balance accrued during her tenure with the firm. For instance, if a worker quits her job with firm A after thirty years of employment to join firm B but is then terminated after one year, her severance package consists of 50% of her FGTS savings accumulated *only* in her last year of the labor force with firm B.

[2]Notice that FGTS imposes a cost on firms at the wrong time. In downturns when firms typically shed labor the FGTS policy makes this effort more costly.

In sum, if the FGTS balance increases at 8% and the severance penalty consists of 50% of accrued savings, then the size of the FGTS termination penalty is roughly 4% of wages accumulated up to the time of termination.[3] Comparatively, the personal savings rate in the U.S. between 2000-2018 is 4.8% (5.5% if we consider the period 1990-2018). Thus, terminating an employee with thirty years tenure will cost close to her lifetime savings, if the rate of return were the U.S. savings rate of recent decades.

Not surprisingly, the FGTS policy directly impacts worker allocation in Brazil; the nature of which we investigate in this paper. For instance, we observe that job mobility of long-tenured workers is remarkably limited in Brazil (see Figure 1). From the worker's perspective, quitting a job is only advantageous if there is a job offer whose compensation exceeds the present value of current earnings plus the expected FGTS compensation; given future dismissals are likely. Note that as the expected severance compensation grows with tenure it definitely narrows the set of incentive compatible offers as time on the job increases.

From the firm's perspective, it may become so costly to fire a long-tenured employee that, in practice, it is advantageous to just wait for the worker to retire. Workers are aware that unless they engage in extremely unprofessional behavior, firms have little incentive to pursue justified termination. In addition, *ceteris paribus*, workers also know that the incentives to fire them declines over time. In this context, if the dis-utility from labor increases with effort, long-tenured workers have less incentive to exert high effort levels.

Our theoretical framework rationalizes the labor market features in the data documented in this section. Importantly we are able to address labor market dynamics that result from decisions based on (expected) severance costs. In particular, separating from workers will depend not only on labor productivity but also on accumulated FGTS holdings. That is, cost saving may dictate firms hoard long-tenured workers at the expense of more productive shorter-tenured

[3]The rate of return chosen by the government rarely exceeds inflation, hence our decision to neglect it.

workers. Therefore, we may see instances where firms continue to operate at losses instead of restructuring employment, as firing workers would be more costly. On the other hand, the FGTS could also lead to age-discriminating hiring with older workers facing higher productivity hurdles. This arises as, with costly search and separation, firms may prefer workers who will yield more in present value. With shorter working lives older workers have to return that much more in productivity to become profitable.

There are also macroeconomic consequences of the FGTS policy. These can prove harmful in economic recessions. For instance, it will be costly for firms to downsize their workforce to match market conditions, forcing them to keep sub-optimally high levels of labor - the *labor hoarding* effect. Additionally, since wage reductions are almost impossible, firms may get stuck with wage rates negotiated under better economic conditions.[4]

In summary, the FGTS policy will have both micro- and macro-economic impacts on the Brazilian labor market. The former suggests important but separate age and tenure implications on worker movements. While the latter can speak to labor hoarding and jobless recoveries during and after recessions. In this paper we focus on the micro-economic effects and refer readers to Figueiredo and Francis (2020) where we address the macro-economic implications of FGTS policy.[5]

2.1 Other Countries with Similar Severance Policy

We stress that, while the FGTS policy is unique to Brazil, the essence of the severance policy are featured in other countries' labor protection policies as well. Some Latin American countries,

[4]Since 2017, employers and employees can negotiate wage reductions if certain requirements are met: firms must provide evidence that the current wage level threatens the financial viability of the business; employers cannot negotiate wage reductions on a one-by-one cases, i.e., wage reductions can only take place under collective bargaining; employees must accept the conditions, otherwise there can be no wage reductions and firms either have to fire workers and pay the severance packages or keep employees in the workforce at their current wages.

[5]Specifically, in that paper we examine aggregate labor movements during recent recessionary episodes in Brazil and how labor dynamics differs under FGTS policy from those under the frictionless environment and situations with other forms of labor adjustment policies.

and a few countries in Western Europe, have institutional details with the FGTS.

de Holanda Barbosa et al. (2017) analyzes the similarities between Brazil's FGTS and the UISAs adopted by other Latin American countries (see Chapter 7 of Holzmann and Vodopivec (2012) for a detailed analysis of UISAs in Latin America). For example, the Chilean setup resembles the Brazilian FGTS policy in that it comprises two types of funds: a pure individual fund (an account linked to the worker and financed by the employer); and a solidarity fund for all workers, financed partly by employers and partly by the government. In Ecuador, the individual fund is financed by the employee, and the solidarity fund by the employer.

de Holanda Barbosa et al. (2017) also analysed the withdrawal criteria for each country. While, in Brazil and Ecuador, withdrawal depends on whether the worker was unjustly dismissed, in Chile there is no such condition. In all these countries, withdrawal may be made at the time of retirement, and in majority of said countries withdrawal is allowed for the purchase of property and for reasons relating to health (specifically, terminal illnesses). In Peru, withdrawals are allowed for loan guarantees. Colombia and Venezuela allow withdrawals to cover approved educational expenditures.

Portugal and Spain are European countries with similar protection policies to firing workers. For example in Portugal if the Labour Court deems that a dismissal was unjust, the dismissed employee may choose between being reinstated in the company, or receive a compensation ranging between 15 and 45 days of base remuneration plus seniority bonus for each year, or fraction of year of seniority, with a minimum of three months of base remuneration plus seniority bonus. In Spain, a contract for an indefinite period may only be terminated, under Spanish law, according to legally defined causes. If an employer terminates a contract without just cause, the employee will be entitled to receive a severance compensation based on 45 days of salary per year of service in the company capped at 3.5 years of salary (which corresponds to more than

28 years of service). The employee is entitled to the salary which she would have received if the dismissal had not taken place and to an eventual compensation for damages suffered.

3 Empirical Considerations

This section discusses important features of Brazil's labor market dynamics that will guide the design of our theoretical model. We begin with the separation rates across different ages and tenure lengths.

To compute the separation rates we combine worker flow data from CAGED[6] and employed-worker stock from RAIS[7] for the 2002-2018 period. We obtain the number of workers terminated without cause, and are therefore entitled to termination-penalty compensation, from CAGED and divide by the number of workers employed at the beginning of the year. Both data sources report labor statistics categorized in groups according to age and tenure length.

Worker flow data are available on a monthly frequency whereas stock data are available annually. This discrepancy affects our separation rate calculations for workers with less than one year of tenure because these workers may have gained employment during the reference year and, thus, would not be counted in the stock of employed workers at the beginning of the year. Nevertheless, our decision to average separation rates across periods rests on the constancy of the size of each group (age, tenure) relative to the population of workers over time.

Figure 1 displays separation rates plotted against tenure for the following age groups: 25-29, 30-39, 40-49, and 50-64. Two features are particularly noteworthy: the high separation rates for workers aged 50 and above, and the downward sloping separation rate as tenure increases for all age groups.

Separation rates in the 50-64 age group are higher than those of other groups. In the most

[6]General Register of Employed and Unemployed Persons.
[7]Annual Social Information Report.

extreme case, separation rates for workers with time on the job between two and three years, in the 50-64 group, is approximately twenty times as large as the separation rates of similar-tenured workers in the 30-39 age group. The high separation rates for workers aged 50 and above is not exclusive to Brazil. Cheron et al. (2013) document that separation rates are U-shaped over the life cycle in the USA and France.[8]

Figure 1 also shows that separation rates vary considerably with tenure within age groups. In particular, separation rates decline unambiguously for all age groups as tenure increases. For workers with tenure greater than twenty years separation rates are (close to) zero. This feature of the data is central to our argument that tenure-dependent severance penalty induces inertia in labor flows.

In addition to the separation patterns across the age and tenure dimensions, the dynamics of labor productivity are essential to characterize the underlying forces driving labor flows. Since productivity is not observable, we turn our attention to the wage distribution conditional on age and tenure.

Note that our universe of workers is restricted to the formal sector and labor relations are thus regulated by the Consolidated Labor Laws.[9] Brazil's labor legislation severely limits employer and employees' ability to negotiate working hours. Therefore, labor fluctuations are essentially on the extensive margin and wage expenditures are due to hourly wages rather than to changes in the number of hours worked.

With these peculiar features of Brazil's labor market in mind, we estimate a wage regression with annual worker-level data. Our panel spans the years 2002-2017 for a cross section of ten

[8]In this sense, the relatively high separation rates in the 25-29 age group, to the reference 30-39 group, in Brazil are in line with the observed pattern for the above-mentioned countries.

[9]*Consolidação das Leis do Trabalho*

million workers who appear at least twice in the sample. Our benchmark specification is:

$$
\begin{aligned}
w_{it} = {} & \alpha_i + \alpha_t + \beta_1\, age_{it} + \beta_2\, age_{it}^2 + \beta_3\, tenure_{it} + \beta_4\, tenure_{it}^2 \\
& + \beta_5\, age_{it} \cdot tenure_{it} + \varepsilon_{it},
\end{aligned} \tag{1}
$$

where w_{it} is agent's i log-wage at time t.

Equation 1 is estimated via OLS and parameter estimates and t-statistics with clustered standard errors are displayed on Table 1. Both age and tenure have positive marginal effects on wage throughout most of the life cycle, though the magnitude of the effects decline as age and tenure increase. In fact, the marginal effect of age becomes negative after 61 years for a newly employed worker (having zero tenure). Turning points for age under five and ten years of tenure are 62 and 63.5 years, respectively.

These results are hardly surprising. Brazil's pension system has two regimes, urban and rural, and each regime has specific rules for retirement eligibility. Under the urban regime, workers must meet one of two criteria: "contribution time" or minimum age, whichever happens first. Males are eligible to collect pension if they make payments to social security for at least thirty-five years or is, at least, sixty-five years old. Females may retire after thirty years of contribution or if they turn sixty years old. Workers under the rural regime are not required to meet a contribution time cutoff and are eligible to retire at ages sixty and fifty-five for males and females, respectively.

Formal-sector jobs disproportionately target skilled professionals and high-earning workers typically retire under the contribution time criterion. On average, men retire at 55 under contribution time and 62 under minimum age criteria. The pattern is similar for women, who retire on average at 52 under the contribution time and 58 under minimum age respectively. Pensions granted[10] to new retirees under contribution time amounted to R\$2, 498.23 and R\$2, 132.33 for males and females. Under the minimum-age criterion, the respective figures are R\$1, 205.51 and R\$1, 089.47. Since pensions are proportional to lifetime earnings, individuals who are eligible to retire under contribution time regime earned significantly higher wages during their careers relative to those who retire under the minimum-age requirement.

Therefore, the negative marginal effect of age on wages for ages over 60 reflects the fact that individuals who remain in the workforce after 60 typically earn lower wages. Nevertheless, for the purpose of our analysis the positive marginal effect of age on wages during most of the individuals' careers provides valuable insight on the productivity allocation along the age dimension.

According to Table 1, the marginal effect of tenure on wages is also positive and decreasing in magnitude. Even though the presence of the interaction term implies that the marginal effect of tenure depends on the age level, our calculated turning points[11] of 120 and 153 years, for workers between ages 20 and 60, fall outside the range of feasible tenure. Hence, we can think of tenure as having a positive marginal effect on wages over an individual's work-life.

While the term "marginal effects" is used to refer to combinations of parameter estimates that yield $\frac{\partial E(w|x)}{\partial x}$ in Eq. 1, causal interpretation would be misguided at this stage. Estimates of Eq. 1 are the *outcome* of a reduced-form model estimated on data generated by an unknown data generating process. Instead, we will use the reduced-form results to discipline our theoretical model; essentially our conjecture about the data generating process.

[10]Data available on Brazil's social security office website.

[11]Tenure level after which the marginal effect of tenure becomes negative.

4 Search and Matching Life Cycle Model

We consider a model economy in which agents with varying ages are matched to firms in the spirit of Mortensen and Pissarides (1994) and Hairault et al. (2019).[12] Job search is random and workers of different ages compete for the same vacancies, in line with Cheron et al. (2013); which is the opposite of that assumed by Hairault et al. (2019) that assumes age-directed search. Job destruction is endogenous, but firms face a severance penalty whose magnitude is determined by the history of wages paid to the worker.

The severance penalty is a *de facto* layoff tax, whose effects on labor dynamics have been documented in Hopenhayn and Rogerson (1993), Saint-Paul (1995) and Mortensen and Pissarides (1999), to name a few. Their approach, however, assumed an exogenously determined lump-sum tax whereas in our framework the tax *rate* is *pre*determined outside the model.

4.1 Model Framework

Time is discrete. In each period, a new generation of workers is born and lives on for $T + 1$ periods. Agents are born unemployed and retire after T periods. All cohorts have the same size and are normalized to unit measure. We denote the measure of employed workers of age i and FGTS holdings by τ by $\rho_i(\tau)$. Age-i employed workers have measure $e_i = \int_0^{\bar{\tau}} \rho_i(\tau) d\tau$ and the measure of the unemployed of age i is given by $u_i = 1 - e_i$. Aggregate unemployment rate is $u = \frac{1}{T} \sum_{i=1}^{T} u_i$.

In a given period, firms post vacancies of measure v. Employers and unemployed individuals meet according to homogeneous function $M(u, v) = A\, u^{\alpha}\, v^{1-\alpha}$.[13] To make the notation less

[12] Our model is similar in spirit to Hairault et al. (2019) except that while their focus is on heterogeneity across age, we focus on age-tenure couplet with age driving hiring decisions and tenure important for the separation decision.

[13] We take care to say *meet* to highlight fact that in our setting there will be another hurdle before production takes place.

cumbersome, we denote job market tightness by $\theta = \frac{v}{u}$ and the probability that a firm finds a candidate for a posted vacancy by $q(\theta) = M\left(\frac{1}{\theta}, 1\right)$. Conversely, unemployed individuals meet potential employers with probability $\theta\, q(\theta) = M\,(1, \theta)$. We assume firms freely enter the market.

After a firm and worker meet, a productivity level ε is drawn for the worker from (cumulative) distribution $G\,(\varepsilon)$. Upon observing productivity ε, the firm decides whether to operate the job. Specifically, if the observed productivity is greater than the reservation productivity for a worker aged i then the firm operates the job.

The job destruction mechanism has similar structure. Specifically, a productive relationship between a firm and a worker aged i with FGTS holdings τ is broken if the worker's productivity level is lower than reservation productivity $R_i(\tau)$, *conditional* on a new productivity being drawn. The following sections discuss the inter-temporal choices of firms and workers. We then study the effect of FGTS-based severance penalty on wages and labor flows.

4.1.1 Firms

Value from a worker aged i, with tenure reflected in accumulated savings τ, at productivity level ε, is given by

$$J_i\,(\varepsilon, \tau) = max\left\{\varepsilon - w_i(\varepsilon) + \beta\left[\lambda\, J_{i+1}(\varepsilon, \tau') + (1-\lambda)\int_0^1 J_{i+1}(\varepsilon', \tau')\, dG(\varepsilon')\right], V - \delta_1 \tau\right\}$$

$$\text{s.t.} \qquad \tau' = \tau + \delta_0\, w_i(\varepsilon). \tag{2}$$

where V is the value from a vacancy and $\delta_1\,\tau$ is the severance penalty for a worker with age-FGTS profile (i, τ). The worker's FGTS balance evolves according to $\tau' = \tau + \delta_0\, w_i(\varepsilon)$. Under current legislation, $\delta_0 = 0.08$ and $\delta_1 = 0.5$.

The term right of the comma in the *max* operator denotes the value from terminating the worker - or not operating the job, which entails the benefit from a potential match in the future and the costs of posting a vacancy and compensating the worker for the termination. The cumulative nature of the severance penalty implies that long-tenured workers are more costly to fire than their ε-identical short-tenured counterparts.[14]

The term on the left-hand side in the *max* operator is the value from keeping the job operating. We assume workers carry the current productivity level over to the next period with probability λ. With probability $1 - \lambda$ a new productivity level is drawn from $G(\cdot)$. Having on-the-job productivity draws admits endogenous break-ups into the model. It also allows heterogeneity in wages and tenure within and across age groups.

The firm's period surplus is simply the labor productivity net of wages. The underlying assumption being that production displays constant returns to scale.

The value from a vacancy is given by

$$V = -c + \beta\, q(\theta) \left[\sum_{i=1}^{T-1} \frac{u_i}{u} \int_0^1 J_{i+1}\,(\varepsilon', 0)\, dG(\varepsilon') \right] + \beta\, \left[1 - q(\theta)\right] V \qquad (3)$$

where c is the cost of posting a vacancy. The term $\int_0^1 J_{i+1}\,(\varepsilon')\, dG(\varepsilon')$ denotes the expected value of operating a job with an i-aged worker given that her productivity draw is higher than the reservation productivity for her age $R_i(0)$. If the productivity drawn is lower than $R_i(0)$, which occurs with probability $G\,[R_i(0)]$, then the firm does not operate the job and enjoys value V. Hence, the term $\int_0^1 J_{i+1}\,(\varepsilon')\, dG(\varepsilon') = \int_{R_i(0)}^1 J_{i+1}\,(\varepsilon', 0)\, dG(\varepsilon') + G\,[R_i(0)]\, V$ denotes the expected value from meeting, *given* that the potential worker is age i.

Since search is random, workers aged $1, ..., T-1$ are potential matches. The probability of

[14]From this perspective a rise in the costs of terminating workers not only hinders the firm from reducing it's labor force but also provides no incentive to hire, as future costs of firings are factored.

an unemployed being aged i is $\frac{u_i}{u}$. In this sense, the term

$$\sum_{i=1}^{T-1} \frac{u_i}{u} \left(\int_{R_i(0)}^{1} J_{i+1}\left(\varepsilon', 0\right) dG(\varepsilon') + G\left[R_i(0)\right] V \right)$$

expresses the unconditional expected value from a new meeting. With probability $1 - q(\theta)$ a meting does not occur and the firm enjoys value V.

4.1.2 Workers

Workers derive utility from wages when employed and leisure when unemployed. The value from being employed at age i with accumulated savings τ is

$$E_i\left(\varepsilon, \tau\right) = \begin{cases} w_i(\varepsilon) + \beta\,\lambda\, E_{i+1}\left(\varepsilon, \tau'\right) + \beta\left(1-\lambda\right) \int_0^1 E_{i+1}\left(\varepsilon', \tau'\right) d\,G(\varepsilon') & \text{if } \varepsilon \geq R_i\left(\tau\right) \\ U_i + \delta_1\,\tau & \text{if } \varepsilon < R_i(\tau). \end{cases}$$

If the drawn productivity level is higher than reservation productivity $R_i(\tau)$, then the firm operates the job and the worker gets the continuation value above. Otherwise, the match breaks and the worker gets value from unemployment U_i plus severance compensation $\delta_1\,\tau$.

Value from unemployment comes from leisure z and the possibility of finding a match

$$U_i = z + \beta\,\theta q(\theta) \left[\int_0^1 E_{i+1}\left(\varepsilon', 0\right) dG(\varepsilon') \right] + \beta\,\left[1 - \theta q(\theta)\right] U_{i+1}.$$

A solution to the functional equations Eq. 2 is a set of reservation productivities $\{R_i(\tau)\}$ that make the firm indifferent between keeping or breaking the match and obey the terminal conditions

$$U_{T+1} = \tilde{z}, \text{ leisure from retirement}$$

$$E_T(\varepsilon,\tau) = w_T(\varepsilon) + \beta\, U_{T+1}, \text{ value from working in the last productive period, and finally}$$

$$J_T(\varepsilon,\tau) = max\,\{\varepsilon - w_T(\varepsilon) + \beta\, V, V - \delta_1\,\tau\}\,,$$

which denotes the firm's value from having a worker in her last period before retirement. If the firm chooses to operate the job, it enjoys value V with certainty in the next period.

4.1.3 Wage Determination

The equilibrium wage is determined in the Nash bargaining between the firms and workers. We denote the share of the surplus captured by the workers and firms by ϕ and $1-\phi$ respectively.[15] The surplus from a job operated by a worker aged i under productivity ε and FGTS holdings τ is denoted by $S_i(\varepsilon,\tau) = J_i\,(\varepsilon,\tau) + E_i\,(\varepsilon,\tau) - U_i$.

If drawn productivity is lower than threshold $R_i(\tau)$, then surplus is zero. On the other hand, if $\varepsilon \geq R_i(\tau)$ then

$$\begin{aligned} S_i(\varepsilon,\tau) &= \varepsilon + \beta\,\lambda\,[S_{i+1}(\varepsilon,\tau') + U_{i+1}] + \beta\,(1-\lambda)\int_0^1 S_{i+1}(\varepsilon',\tau')dG(\varepsilon') + \beta(1-\lambda)U_{i+1} - U_i \\ &= \varepsilon + \beta\,\lambda\,S_{i+1}(\varepsilon,\tau') + \beta\,(1-\lambda)\int_0^1 S_{i+1}(\varepsilon',\tau')dG(\varepsilon') - (U_i - \beta U_{i+1})\,. \end{aligned}$$

From the Nash bargaining, $E_i(\varepsilon,\tau) - U_i = \phi\,S_i(\varepsilon,\tau)$ and $J_i(\varepsilon,\tau) = (1-\phi)S_i(\varepsilon,\tau)$. From the

[15] Our undirected search setup allows us to abstract from age-specific bargaining weights as in Hairault et al (2019).

expression for $S_i(\varepsilon, \tau)$ we can derive the equilibrium wage $w_i(\varepsilon)$ as follows:

$$
\begin{aligned}
E_i(\varepsilon, \tau) - U_i &= \phi\, S_i(\varepsilon, \tau) \\
E_i(\varepsilon, \tau) - U_i &= \phi\, \left[E_i(\varepsilon, \tau) + J_i(\varepsilon, \tau) - U_i\right] \Rightarrow (1-\phi)U_i = E_i(\varepsilon, \tau) - \phi\, \left[E_i(\varepsilon, \tau) + U_i\right] \\
(1-\phi)U_i &= w_i(\varepsilon) + \beta\,\lambda\, E_{i+1}(\varepsilon, \tau') + \beta\,(1-\lambda) \int_0^1 \left[\phi S_{i+1}(\varepsilon', \tau') + U_{i+1}\right] dG(\varepsilon') \\
&\quad - \phi \left[\varepsilon + \beta\,\lambda \left(S_{i+1}(\varepsilon', \tau') + U_{i+1}\right) + \beta\,(1-\lambda) \int_0^1 \left(S_{i+1}(\varepsilon', \tau') + U_{i+1}\right) dG(\varepsilon')\right] \\
&= w_i(\varepsilon) - \phi\,\varepsilon + \beta\,\lambda\,(1-\phi)U_{i+1} + \beta\,\lambda\,\phi U_{i+1} = w_i(\varepsilon) - \phi\,\varepsilon + \beta\,(1-\phi)U_{i+1} \\
\Rightarrow w_i(\varepsilon) &= \phi\,\varepsilon + (1-\phi)\left(U_i - \beta\, U_{i+1}\right).
\end{aligned}
$$

Therefore, equilibrium wage is determined by the workers' share of surplus plus a compensation to the worker for leaving the job search. Because the value from unemployment consists of leisure plus the value from a potential match, leaving unemployment in the current period entails forgoing a new job search in the next period.

4.1.4 Worker Flows

To close the model we need to specify worker flows across employment and unemployment. To characterize the flows of workers in and out of unemployment, recall that we denote the measure of unemployed workers aged i by u_i and the measure of employed workers aged i with FGTS holdings τ by $\rho_i(\tau)$. Naturally, $\int_o^{\bar{\tau}} \rho_i(\tau) d\tau = 1 - u_i$, where $\bar{\tau}$ is the upper bound for FGTS holdings - which is age-specific though we abstract from this feature to keep the notation manageable.

The law of motion for $\{u_i\}_{i=1}^{T}$ is given by

$$u_i = u_{i-1}\left[1-\theta\, q(\theta)\right] + u_{i-1}\,\theta\, q(\theta) G\big[R_i(0)\big] + (1-\lambda)\left[\int_0^{\bar{\tau}} \rho_{i-1}(\tau) G\big[R_i(\tau)\big]\, d\tau\right],$$

where the measure of unemployed aged-i workers equals the measure of unemployed workers at age $i-1$ that either did not meet an employer, plus those unemployed who met with employers but whose productivity draw was not high enough to clear the reservation productivity at $R_i(\tau=0)$, plus the measure of workers aged $i-1$ that lost their jobs.

In equilibrium, the flow of workers aged $i-1$ who lose their jobs must equal the flow of workers who find a job. That is,

$$(1-\lambda)\int_0^{\bar{\tau}} \rho_{i-1}(\tau) G\big[R_i(\tau)\big]\, d\tau = u_{i-2}\,\theta\, q(\theta)\,\left[1-G\left(R_{i-1}(0)\right)\right].$$

Combining the law of motion and labor flow equations yields

$$u_i = u_{i-1}\left[1-\theta\, q(\theta)\right] + u_{i-1}\,\theta\, q(\theta)\, G\big(R_i(0)\big) + u_{i-2}\,\theta\, q(\theta)\,\left[1-G\left(R_{i-1}(0)\right)\right], \tag{4}$$

which allows us to characterize age-specific unemployment rates recursively.

We first demonstrate the workings of the model with numerical simulations before proceeding to estimation using the empirical productivity distribution. Details about the computational methods are discussed in Appendix C.

5 Numerical Simulation

We demonstrate the workings of the model with numerical simulation. We simulate the model for the worker-firm matches in the formal sector with productivity $G(\varepsilon)$, drawn from a Uniform(0,1) distribution.

Key to the simulation is a measure of labor market tightness (unemployment-vacancy ratio) - denoted θ - for Brazil. A ready made monthly measure of tightness is available from the *Institute of Economic Research Foundation* (in Portuguese, *Fundacao Instituto de Pesquisas Economicas - Fipe*) covering the period January 2004 through June 2017. This measure is labeled "vacancy per candidate" which we presume is the inverse of the unemployment-vacancy ratio we wish to calculate. However, given our uncertainty as to whether *candidates* and *the unemployed* are synonymous we verify their equivalence by estimating this ratio with available data. First, an index of vacancy per working aged is available for the entire sample period 2004:01 to 2017:06. However, unemployment for the entire population is only available beginning in 2012:03. We extend the unemployment series back to 2004 using "unemployment for the metropolitan area" data which is available from 2002:03 through 2016:02. That is, we use overlapping periods to forecast the missing values for the entire population. Figure 2 plots both series, calculated and the data. The two series are close which gives us confidence the number we use for calibration.

In practice, θ is determined endogenously in Eq. 3, hence calibrating the value of θ is equivalent to setting the cost of posting a vacancy. We estimate two values for θ. First, we use the average value for the data series which gives a value of $\theta = 0.47$. Second, we regress the unemployment-vacancy ratio on time trends (up to quadratic) and use the constant as our measure of thinness. This gives our second estimate of $\theta = 0.88$. Estimating the model with disparate values of θ will highlight the sensitivity of our results to this parameter.

5.0.1 Simulation Results

We find that the simulation results are uniquely sensitive to the value of θ. We first discuss the findings for $\theta = 0.47$ and highlight the key differences when we increase labor market tightness.

The results for $\theta = 0.88$ are available upon request.

Figure 3 depicts the firm's value from an experienced worker who had enough time to accrue sizable FGTS holdings. We consider two scenarios. In the first, the worker's FGTS holdings are zero - which is only possible if worker and firm have just met. The second scenario reflects the firm's value from an employee with FGTS holdings $\tau = 0.45$.

If the firm decides to terminate the worker with $\tau = 0.45$ then it attains the value of a vacancy plus the negative cash flow that severance compensation entails: $V - \delta_1 \tau = -0.225$, as the value from a vacancy is zero in equilibrium. Naturally, if firm and worker have just met, then the firm faces no severance cost if it chooses to not operate the job.

The productivity levels that make the firm indifferent between terminating the worker and keeping the match - or, in the case of $\tau = 0$, indifferent between operating the job and resuming the search - are depicted in Figure 3 as the points at which the value curves kink. At these productivity levels, referred to as reservation productivity throughout the text, the value from operating the job (the upward sloping portion of the firm's value curves) equals the value from terminating the worker (the flat range of value curves).

Note that if drawn productivity is greater than reservation productivity for the potential new hire, i.e., $\varepsilon > R_i(0)$, then the value attained from a worker with FGTS holdings $\tau = 0$ is identical to that of a worker with $\tau > 0$ even if the firm's value is, in a general sense, decreasing in τ. To understand why that is the case, first note that given the worker's age, potential new hires face the highest productivity threshold because reservation productivity is decreasing in FGTS holdings. Moreover, productivity-wage rates are age-specific only, so similarly aged workers with different FGTS holdings earn the same wage in our model. Since the firm chooses to operate the job if productivity is high enough given FGTS holdings, if the productivity draw clears the bar today then it is certainly high enough to clear reservation productivity in the next period because FGTS holdings will be higher and reservation productivity decreases with τ. Hence, $\varepsilon > R_i(0)$ ensures identical value from workers with heterogeneous FGTS holdings.

Figure 3 shows the firm's highest attainable value given FGTS holdings and productivity levels. In other words, the firm's optimal choice can be elicited from the value curve in the said figure. From this perspective, the mechanism through which the FGTS severance policy distorts the firm's decision becomes evident: if productivity lies in the range $R_i(\tau) \leq \varepsilon \leq R_i(0)$ then it is beneficial for the firm to operate the job even if doing so yields negative value. That is, the firm takes a loss because terminating the worker would ensue an even greater loss.

Figure 4 depicts reservation productivities across age profiles and FGTS levels. For potential new hires, $\tau = 0$, reservation productivity increases with age, which is not surprising when we consider the higher continuation value obtained from younger workers relative to those closer to retirement age. Since the search process is costly to employers, we can think of the higher reservation productivity of older workers as compensation to the firm for their short careers. In the extreme case of workers with only one period left before retirement, reservation productivity is extremely high because the firm knows it will face a costly search with certainty in the next period. This feature sheds light in the difficulty that older workers face when unemployed: even if they meet potential employers with the same probability as younger workers, $\theta\, q(\theta)$, the probability that firm will choose to operate the job is significantly lower vis-à-vis meeting younger workers of similar productivity.

On the other hand, employed workers that reach pre-retirement age with even relatively low FGTS holdings enjoy significant job security as they face reservation productivity zero for τ as low as 0.22. The fact that the worker will leave the workforce in the next period with certainty makes the wait for the worker's retirement more advantageous than terminating the worker in the current period.

While the reservation productivity for a worker aged T declines quickly to zero, workers who are only a few periods short of retirement face high reservation productivity along all FGTS levels. That is the case because when the firm chooses to operate a job with a worker only a few periods short of retirement, it delays the search for a new worker that could potentially be in the beginning of her career and would yield considerably high value - in a present value sense. By the same token, the firm is willing to operate the job at low productivity levels with young workers because even though their productivity is low in the current period, they will likely stay with the firm for a long time.

Although the probability of meeting an employer is the same across ages, the probability that the firm will operate the job is not. In particular, the probability that the firm operates the job, $1 - G\left[R_{i+1}(\tau)\right]$, declines with age - as seen in Figure 4. The decline in the probability of matching successfully with an employer is key to the decreasing value from unemployment along the life cycle.

The dynamics of reservation productivity across FGTS levels are determinant for flows and age-specific unemployment rates. Specifically, an employed worker aged i with FGTS holdings τ loses her job with probability $G\left[R_{i+1}(\tau)\right]$, conditional on a new productivity level being drawn. Hence, the unconditional separation rate is $(1-\lambda)\, \rho_i(\tau)\, G\left[R_{i+1}(\tau)\right]$ since measure $\rho_i(\tau)$ of workers aged i with FGTS level τ get a new draw with probability $1-\lambda$ that does not clear the reservation productivity with probability $G\left[R_{i+1}(\tau)\right]$. To understand worker flows to unemployment *given* that the worker draws a new productivity level, we display $G\left[R_i(\tau)\right]$ across ages i in Figure 4.

Reservation productivity provide valuable insight on the age-specific stock of unemployed workers as pointed out in Eq. 4. In Figure 5 we plot the age-specific separation rates for given levels of FGTS balances. Overall, we observe that separation rates decline in the level of FGTS balances as firms are reluctant to break relationships the more expensive the separation. For each age group there is a value beyond which firms will choose to hoard workers. For workers near retirement we see that this a cutoff FGTS balance is 0.24, about half the critical value of the next oldest group. These cutoff values indicate where firms find it cost-effective to wait until workers retire. Another feature of the model is that firms will quickly separate from older workers with very low FGTS - for the same accumulated FGTS younger workers have greater job security.

Figure 6 shows unemployment rates across ages. Since agents start out unemployed, unemployment rate is 1 for agents aged 1. Non-directed search implies that given that a firm and an unemployed worker met, the worker is more likely to come from a high-unemployment age group. In this sense, the low unemployment of workers aged 2 is driven by the number of new-born workers that entered the labor market in the previous period.

For ages greater than three, unemployment rates are increasing as result of age-increasing reservation productivities. The bottom panel of Figure 6 displays unemployment rates for this particular group - age greater than three - to highlight the upward sloping pattern of unemployment rates across age groups without the influence of early-age unemployment particularities.

Leisure from unemployment is constant across ages, so the curvature in the value from unemployment, Figure 7, must come from the age-specific probability of successfully matching with an employer and from the number of periods remaining in the workforce. To be clear, even if the probability of matching - i.e., meeting and operating the job - were constant across ages, value from unemployment would still be higher for younger workers as they have the potential to collect wages for a longer period. This feature of finite-horizon models has been discussed in the literature by Cheron et al. (2013), among others.

In addition to this finite-horizon effect, the age-increasing reservation productivity reinforces the declining value from unemployment across ages, a feature that Figure 7 makes evident. Note that value from unemployment declines at a decreasing rate, which makes the benefit from *not* spending another period in unemployment, $U_i - \beta U_{i+1}$, increasing with age. In other words, the curvature in the value from unemployment implies that the benefit from taking a job instead of waiting another period increases with age.

The curvature in the value from unemployment enables our model to deliver equilibrium wages that increase with age even under our assumption that the *unconditional* distribution of labor productivity is the same across ages. From $w_i(\varepsilon) = \phi\,\varepsilon + (1-\phi)\,(U_i - \beta U_{i+1})$, the term $(1-\phi)\,(U_i - \beta U_{i+1})$ increases with age for reasons previously established. However, because reservation productivity rises with age, the term $\phi\,\varepsilon$ also increases with age because the range of productivity levels that do not break the match shrinks as age increases.

Note, however, that wage is age and productivity specific but not FGTS-level specific. In this sense, our model does not imply that agents with large FGTS holdings earn more than their relative low-FGTS counterparts in a general sense, but that they may earn more *relative* to their productivity. To be specific, note that given the FGTS level and productivity draw, the firm faces a binary choice: either operate the job or terminate the worker. If the firm chooses to operate the job, then it must pay the equilibrium wage given the worker's age-productivity profile regardless of her FGTS holdings. The benefit from large FGTS holdings comes from the range of productivity levels[16] that would lead to a termination in the absence of the FGTS severance penalty but that the firm chooses to keep the match because the magnitude of the severance compensation is too large.

From such perspective, given a set of workers with FGTS holdings τ, the agents that extract an "FGTS-penalty" rent are those whose productivity levels would not be high enough to establish a match if they were unemployed. Thus the FGTS-based severance is detrimental to the firms'

[16]Shaded area in Figure 3.

economic performance because it leads to labor hoarding of low-productivity workers since value from workers with productivity above the zero-FGTS cutoff does not depend on FGTS holdings.

At the aggregate level, the FGTS policy may compromise economic performance to an even greater extent. If labor productivity is partially driven by macroeconomic factors then individual productivity outcomes are correlated and the share of low-productivity workers can be higher in recessions. Our model suggests that the FGTS level does not play a major role in labor decisions regarding high-productivity workers. Instead, labor hoarding occurs in the low end of the productivity spectrum, reinforcing the depressing effect of recessions on firms' performance.

5.0.2 Sensitivity to Labor Market Tightness

Our model results indicate that firms are willing to accept lower levels of productivity when the labor market is tighter. Specifically, the reservation productivity for a worker aged 5 with FGTS holdings less than 0.2 is considerably lower when $\theta = 0.88$ compared to when $\theta = 0.47$.

Such result is not surprising given the cost of breaking a match when the labor market is tighter is higher, as the expected value from searching is lower because of the reduced probability of meeting a candidate. While this feature is also present in standard search models, the tenure-dependent severance cost adds a new layer of complexity to the firms' decision set that unequivocally depresses reservation productivity.

6 Model Estimation and Results

This section describes the data, estimation strategy, and quantitative implications of the model. We estimate the model using administrative data on Brazilian formal-sector worker–firm matches and show that tenure-dependent severance costs generate strong age and tenure gradients in separation behavior.

6.1 Data

As discussed in Section 3, the empirical moments are constructed from administrative records of the Brazilian formal labor market, combining worker flows from CAGED and employment stocks from RAIS. These data provide age–tenure cell–level measures of employment shares, average wages, and separation rates, where separations are defined as terminations without cause divided by employment at the beginning of the period. We combine employment stocks and worker flows to obtain age–tenure profiles of employment, wages, and separations. The data provide near-universal coverage of formal employment relationships and are well suited for studying job duration and separations.

We use three sets of moments measured at the age–tenure cell level:

1. the share of employment across age and tenure bins;
2. average monthly wages;
3. separation rates, defined as terminations without cause divided by employment at the beginning of the period.

Age is grouped into six five-year bins, and tenure is grouped into seven bins measured in months. Separation rates are expressed at a monthly frequency. All moments are constructed using the most recent available year to approximate a stationary environment. Figure 9 displays the resulting age–tenure separation surface.

6.2 Estimation

We estimate the model by the method of simulated moments. The parameter vector is chosen to minimize the distance between empirical moments and their model-implied counterparts.

Labor productivity ε is assumed to lie in $[0, 1]$ and follows an unknown distribution $G(\varepsilon)$. We approximate G using a flexible discrete distribution over a fixed grid. Probabilities are parameterized using a softmax transformation, ensuring non-negativity and unit mass without imposing functional form restrictions.

In addition to $G(\varepsilon)$, we estimate two key structural parameters: the persistence of match-specific productivity, λ, and the scale of tenure-dependent firing costs, ξ_1. All remaining parameters, including the FGTS contribution rate, discount factor, and interest rate on accumulated balances, are calibrated to institutional features of the Brazilian system.

Let g^{data} denote the vector of empirical moments and $g^{\text{model}}(\theta)$ the corresponding model-implied moments. The moments include age–tenure employment shares, average wages, and separation rates. The estimator solves

$$\hat{\theta} = \arg\min_{\theta} \left(g^{\text{model}}(\theta) - g^{\text{data}}\right)' W \left(g^{\text{model}}(\theta) - g^{\text{data}}\right),$$

where W is a diagonal weighting matrix that places relatively higher weight on employment shares and separation rates. The model is solved by backward induction over age and tenure, and the objective function is minimized using a quasi-Newton algorithm.

Identification is intuitive and over-identified. The level and slope of separation rates at short tenure discipline λ, while the sharp decline in separations with tenure pins down ξ_1. The joint age–tenure profile of separations is crucial for separating productivity persistence from severance-cost effects. See the Appendix D for details of the estimation.

6.3 Results

Figure 10 displays model-implied separation rates, while Figure 9 reports their empirical counterparts. The model reproduces the central features of the data. Separation rates decline monotonically with tenure across all age groups and are near zero at long tenure lengths. At short tenure, separation rates are substantially higher for older workers, consistent with their higher reservation productivity. Although the model produces smoother separation profiles than those observed in the data, particularly for older workers at very short tenure, it closely matches the overall level and slope of separations across the age–tenure space.

Figure 11 reports the estimated reservation productivity surface $R(a, \tau)$. Three features stand out. First, reservation productivity declines steeply with tenure within each age group. As FGTS balances accumulate, firing costs rise, leading firms to tolerate increasingly low productivity realizations before separating. This mechanism generates substantial labor hoarding among long-tenured workers. Second, reservation productivity is increasing in age at low tenure. Older newly hired workers face substantially higher productivity thresholds than younger workers, reflecting their shorter remaining work lives and lower continuation values. This feature captures age-discriminating hiring incentives in equilibrium. Third, for workers close to retirement with sufficiently long tenure, reservation productivity approaches zero. In this region, firms optimally retain even very low-productivity workers rather than incur severance costs, effectively waiting for retirement to dissolve the match.

6.4 Economic Implications

The estimates highlight a fundamental asymmetry induced by tenure-dependent severance costs. On the hiring margin, firms require higher productivity from older unemployed workers, reducing their job-finding probabilities. On the firing margin, firms are reluctant to dismiss long-tenured workers, even when productivity is low, due to accumulated severance liabilities.

As a result, the FGTS system protects employment histories rather than productivity. It generates labor hoarding among low-productivity, long-tenured workers while simultaneously discouraging the hiring of older unemployed individuals. These forces jointly shape observed life-cycle labor dynamics in Brazil and provide a structural explanation for the strong age and tenure patterns in separation rates.

6.5 Quit Rates as a Placebo Test

We now contrast the patterns documented for involuntary separations with evidence from quit rates, which are not subject to the FGTS firing penalty and therefore provide a natural placebo for the firing-cost mechanism emphasized in the model.

Figure 12 reports monthly quit rates by age and tenure. In stark contrast to separations without cause (Figure 9), quit rates exhibit little systematic variation with tenure across age groups. In particular, there is no sharp decline in exit rates as tenure accumulates, nor is there a pronounced age–tenure interaction. This absence of a tenure gradient indicates that the strong inertia observed in involuntary separations is not a generic feature of employment relationships or job-match dynamics, but is instead specific to separations governed by firing costs.

To further assess whether the model can account for quit behavior, we re-estimate the model using quit rates as the empirical separation moments. Figure 13 shows the resulting model-implied separation rates. The model generates extremely small and smooth separation probabilities, several orders of magnitude below those observed for involuntary separations, and fails to reproduce the sharp tenure dependence present in the baseline estimates.

The corresponding reservation productivity surface, shown in Figure 14, highlights the source of this discrepancy. In the absence of firing costs, the model can match the low and weakly tenure-dependent quit rates only by imposing implausibly high reservation productivity thresholds at long tenures. This effectively assumes that firms retain long-tenure workers solely because they are extraordinarily productive, a mechanism at odds with Brazil's institutional setting, where the FGTS system stabilizes employment by discouraging separations through severance incentives rather than through extreme productivity sorting.

These results confirm that the age–tenure patterns documented above are specific to involuntary separations subject to firing costs, rather than reflecting general life–cycle effects, job-match learning, or mechanical tenure dependence. The quit-rate placebo thus provides a stringent falsification test that reinforces the interpretation of the baseline results and clarifies the domain of relevance of the model.

6.6 Partial-Equilibrium Policy Experiments with Endogenous Wages

We first evaluate severance pay reforms in partial equilibrium, holding worker outside options fixed and allowing wages to adjust endogenously within ongoing employment relationships. In this environment, policy affects outcomes exclusively through firms' continuation values and equilibrium wage renegotiation.

Environment. The analysis takes the estimated match-level productivity distribution and separation technology as given. Wages are determined through Nash bargaining at the match level, and separations occur whenever the continuation surplus becomes negative. Severance pay enters the firm's value function as a firing cost, but does not directly affect workers' outside option or participation decisions.

Counterfactual policies. We consider a reform that replaces the baseline tenure-increasing severance schedule with a flat severance payment that preserves the average level of firing costs in the economy. This experiment isolates the role of the severance profile while holding aggregate severance liabilities fixed.

Equilibrium wage responses. Figure 15 displays the change in equilibrium wages under the flat severance regime relative to the baseline. Wages increase substantially for older and long-tenure workers, while remaining flat or slightly declining for younger, short-tenure matches. The reallocation reflects firms' incentives to smooth surplus across matches when firing costs become less backloaded in tenure. By raising expected firing costs for long-tenure matches, the reform induces firms to share part of the increased continuation value with workers through higher wages.

Reservation productivity and separations. Figure 16 shows the corresponding changes in reservation productivity. Reservation thresholds rise sharply with tenure and age, reflecting firms' increased reluctance to sever high-tenure matches when firing costs are flattened. As a result, separation rates decline markedly for older and long-tenure workers, as shown in Figure 17. The reform compresses the tenure gradient in separation rates, yielding a more uniform separation profile across tenure bins (Figure 18).

Interpretation. These results highlight that even in the absence of changes in worker outside options or participation, severance pay reforms can generate sizable employment effects through endogenous wage adjustment. Flattening severance payments raises firms' continuation values for high-tenure matches, leading to higher equilibrium wages, higher reservation productivity thresholds, and lower separation rates among older and long-tenure workers.

Importantly, these effects arise entirely within ongoing matches and should be interpreted as partial-equilibrium responses. General equilibrium feedbacks and worker-side participation responses are deliberately shut down at this stage.

6.7 Flat Severance as a Standalone Policy Experiment

This subsection studies a counterfactual economy in which severance pay is set to a flat level, independent of worker tenure and age. Unlike the previous counterfactual analysis, the focus here is not on differences relative to the baseline, but on the internal structure of equilibrium outcomes under the flat severance regime itself. This allows us to highlight the mechanisms through which severance design shapes wages, separation behavior, and firms' continuation decisions.

Figure 21 displays the equilibrium separation rate across age and tenure cells under flat severance. Separations exhibit strong gradients along the age dimension, with lower separation rates for older workers and higher rates for younger cohorts. In contrast, variation along the tenure dimension is markedly weaker than in the baseline economy, reflecting the fact that severance costs no longer increase with tenure. As a result, firms' incentives to retain workers do not strengthen mechanically with accumulated tenure, leading to a flatter tenure profile of job destruction.

To interpret these patterns, Figure 22 shows the reservation productivity surface under flat severance. The reservation threshold increases monotonically with age and tenure, but its slope with respect to tenure is substantially muted relative to the baseline environment. This indicates that, absent tenure-contingent severance obligations, firms are less inclined to continue low-productivity matches solely because workers have accumulated tenure. Instead, continuation decisions depend more directly on worker age and match productivity.

Figure 23 reports the model-implied average wage by age and tenure under flat severance. Wages increase with both age and tenure, reflecting surplus sharing in ongoing matches, but the tenure gradient is smoother than in the baseline. This compression arises because severance pay no longer amplifies the firm's outside option differentially across tenure levels, weakening the pass-through of tenure into negotiated wages.

For reference, Figure 24 presents the baseline separation rate. Compared to this benchmark, the flat severance regime features a noticeably altered joint age–tenure structure of separations, with less pronounced tenure dependence and stronger reliance on age-related productivity differences. These results illustrate how flattening severance schedules reshapes equilibrium behavior even without changing aggregate severance generosity, by modifying firms' dynamic incentives to retain workers as tenure accumulates.

7 Efficiency and Welfare Implications of Severance-Induced Labor Hoarding

The previous sections document how tenure-dependent severance pay distorts firms' separation decisions across age and tenure. In this section, we quantify the efficiency and welfare implications of these distortions. In particular, we address three related questions: (i) how much output is lost due to the retention of low-productivity matches, (ii) how severance pay distorts the allocation of labor across cohorts, and (iii) who gains and who loses from these distortions in welfare terms.

Throughout this section, we report results under two alternative modeling closures. *Option A* corresponds to a partial-equilibrium environment in which wages adjust endogenously to severance-induced changes in firms' continuation decisions, while workers' outside options are held fixed. *Option B* allows accumulated FGTS balances to enter workers' outside options, introducing an additional channel through which tenure-dependent severance affects match surplus. Both specifications are disciplined by the same empirical moments and estimated productivity distribution, and differ only in how wages and outside options respond to severance pay.

All calculations in this section are based on the same empirical moments and estimated model described in Sections 3 and 6. Counterfactual economies differ only in the structure of severance pay; the underlying age–tenure composition of employment and the estimated productivity distribution are held fixed.

Counterfactual severance regimes. Let $F(a, \tau)$ denote the baseline severance obligation associated with separating a worker of age a and tenure τ. In addition to the baseline economy, we consider three classes of counterfactual severance regimes. First, we study the elimination of severance pay, $F(a, \tau) = 0$. Second, we consider a flat severance regime in which severance is constant across age and tenure and equal to the average severance payment in the baseline economy. Third, we analyze scaled versions of the baseline schedule defined by

$$F^{\kappa}(a, \tau) = \kappa\, F(a, \tau), \qquad \kappa \in (0, 1), \tag{5}$$

which uniformly reduce severance obligations while preserving their tenure profile. The parameter κ provides a parsimonious measure of the intensity of severance pay and allows us to trace how efficiency and welfare outcomes vary with the overall level of firing costs.

7.1 Output Loss from Labor Hoarding

Severance pay induces firms to retain matches with productivity below the efficient cutoff, generating labor hoarding. Let $R(a, \tau)$ denote the reservation productivity implied by the baseline economy for workers of age a and tenure τ, and let R^{eff} denote the cutoff that would arise in the absence of severance distortions. Aggregate output in the baseline economy is given by

$$Y = \sum_{a,\tau} E(a, \tau) \int_{z \geq R(a,\tau)} z \, dG(z), \tag{6}$$

where $E(a, \tau)$ denotes the employment share in each age–tenure cell and $G(z)$ is the distribution of match productivity. Efficient output is defined analogously as

$$Y^{\text{eff}} = \sum_{a,\tau} E(a, \tau) \int_{z \geq R^{\text{eff}}} z \, dG(z). \tag{7}$$

The difference $Y^{\text{eff}} - Y$ measures the output loss attributable to severance-induced labor hoarding. Because severance obligations rise with tenure, the gap between $R(a, \tau)$ and R^{eff} is largest for high-tenure workers, implying that output losses are concentrated among older cohorts.

Table 2 reports baseline output, efficient output, and the implied output loss for two alternative modeling closures. In both cases, tenure-dependent severance leads to a nontrivial reduction in aggregate output. Even when average severance generosity is held constant, eliminating tenure dependence substantially raises output, highlighting that efficiency losses operate through the shape of severance schedules rather than their overall level.

7.2 Cohort-Level Misallocation

Severance pay affects not only aggregate output but also the allocation of labor across age and tenure groups. By increasing the cost of separation for high-tenure workers, severance pay biases retention toward older cohorts independently of match productivity. As a result, younger workers face higher separation risk, while low-productivity matches among older workers are inefficiently preserved.

We quantify these distortions by comparing the distribution of employment across age–tenure cells under the baseline economy and under counterfactual regimes that flatten or eliminate tenure-dependent severance. Tables 3 and 4 summarize these differences using total variation distance and Kullback–Leibler divergence measures.

Although the aggregate magnitude of these distributional shifts is modest, the distortions are highly structured: the largest deviations arise along the tenure dimension, consistent with severance pay operating primarily through dynamic retention incentives rather than static wage adjustments.

7.3 Welfare Decomposition

We compute partial-equilibrium welfare for workers and firms by aggregating match-level surplus across age–tenure cells. Worker welfare reflects expected wage income and severance payments, while firm welfare reflects profits net of wages and severance obligations. Welfare comparisons are conducted relative to the efficient benchmark without severance pay.

Tables 5 and 6 summarize aggregate welfare outcomes under the baseline economy and key counterfactual severance regimes for the two partial-equilibrium closures considered. Under Option A, where wages adjust endogenously, tenure-dependent severance pay generates sizable efficiency losses. Aggregate welfare in the baseline economy is approximately 0.79 relative to the efficient benchmark, implying an output and surplus loss of about 21 percent. Eliminating severance pay entirely restores efficiency by construction, while flattening severance across tenure recovers only about 0.01 units of welfare, or roughly 5 percent of the total loss.[17]

In this environment, severance pay operates primarily as a distortion rather than a redistributive instrument. Worker welfare closely tracks aggregate welfare across all counterfactuals, while firm welfare remains near zero. For example, in the baseline economy, worker welfare

[17]The welfare gain from flattening severance is $0.799 - 0.788 \approx 0.011$, compared to a total loss of $1 - 0.788 \approx 0.212$.

equals aggregate welfare up to numerical precision, and firm welfare is on the order of 10^{-6}. This pattern reflects the fact that with flexible wages, firing costs are largely capitalized into wages, leaving firms approximately indifferent while distorting separation decisions.

Under Option B, where wages are held fixed but severance pay affects workers' outside options, aggregate welfare losses from tenure-dependent severance are smaller. Baseline welfare is approximately 0.87, compared to 0.79 under Option A. In this case, worker welfare remains nearly constant across counterfactuals, while firms experience meaningful welfare gains when severance obligations are reduced or flattened. As a result, aggregate welfare improvements from reform are positive but more muted than under Option A.

These results indicate that tenure-dependent severance pay distorts firms' separation decisions in a way that lowers aggregate efficiency through labor hoarding. By raising firing costs more steeply for high-tenure matches, severance pay reduces separation risk for incumbent workers relative to marginal matches, leading to the retention of low-productivity employment relationships.

The magnitude of the associated efficiency losses depends on the extent to which severance pay provides insurance through workers' outside options. When wages adjust endogenously, severance pay operates primarily as a distortion with little redistributive benefit, whereas allowing severance to affect outside options attenuates welfare losses by stabilizing worker payoffs. Importantly, tenure-dependent severance pay reduces allocative efficiency under both wage-adjustment and outside-option closures, indicating that the inefficiency arises from distorted separation incentives rather than from a particular assumption about wage setting or insurance.

8 Conclusion

We study the labor market effects of a severance policy in which firing costs increase with a worker's accumulated earnings history. Motivated by Brazil's FGTS system, the model features tenure-dependent layoff penalties that amount to approximately four percent of wages earned over the employment relationship.

Our central finding is that tenure-dependent severance costs distort firms' separation decisions by protecting employment histories rather than productivity. Workers who extract severance-penalty rents are typically low productivity and would not be hired if unemployed. Consistent with this mechanism, the estimated model implies a negative relationship between accumulated FGTS balances and the minimum productivity required to remain employed, while incentives to retain high-productivity workers are largely unaffected by the severance policy.

The model also rationalizes key features of Brazilian labor market data, including wages that rise with age, declining separation rates with tenure, and low job-finding rates among older unemployed workers. These patterns emerge endogenously from the interaction between search frictions, finite work lives, and tenure-dependent firing costs.

From a welfare perspective, tenure-dependent severance pay lowers allocative efficiency by backloading firing costs and distorting separation incentives. The policy induces labor hoarding among low-productivity, long-tenured workers while raising hiring thresholds for older unemployed individuals. Flattening severance schedules recovers part of the welfare loss even when average severance generosity is held fixed, underscoring that inefficiencies arise from the tenure profile of firing costs rather than their level.

An important avenue for future research is to study the interaction between FGTS-style severance policies, search frictions, and aggregate fluctuations. While Figueiredo and Francis (2020) analyze FGTS-based adjustment costs in a two-sector environment with aggregate and sector-specific shocks, their framework abstracts from search frictions. Incorporating both search and aggregate uncertainty would allow for a quantitative assessment of the cyclical efficiency costs associated with low-productivity labor hoarding.

A List of Figures

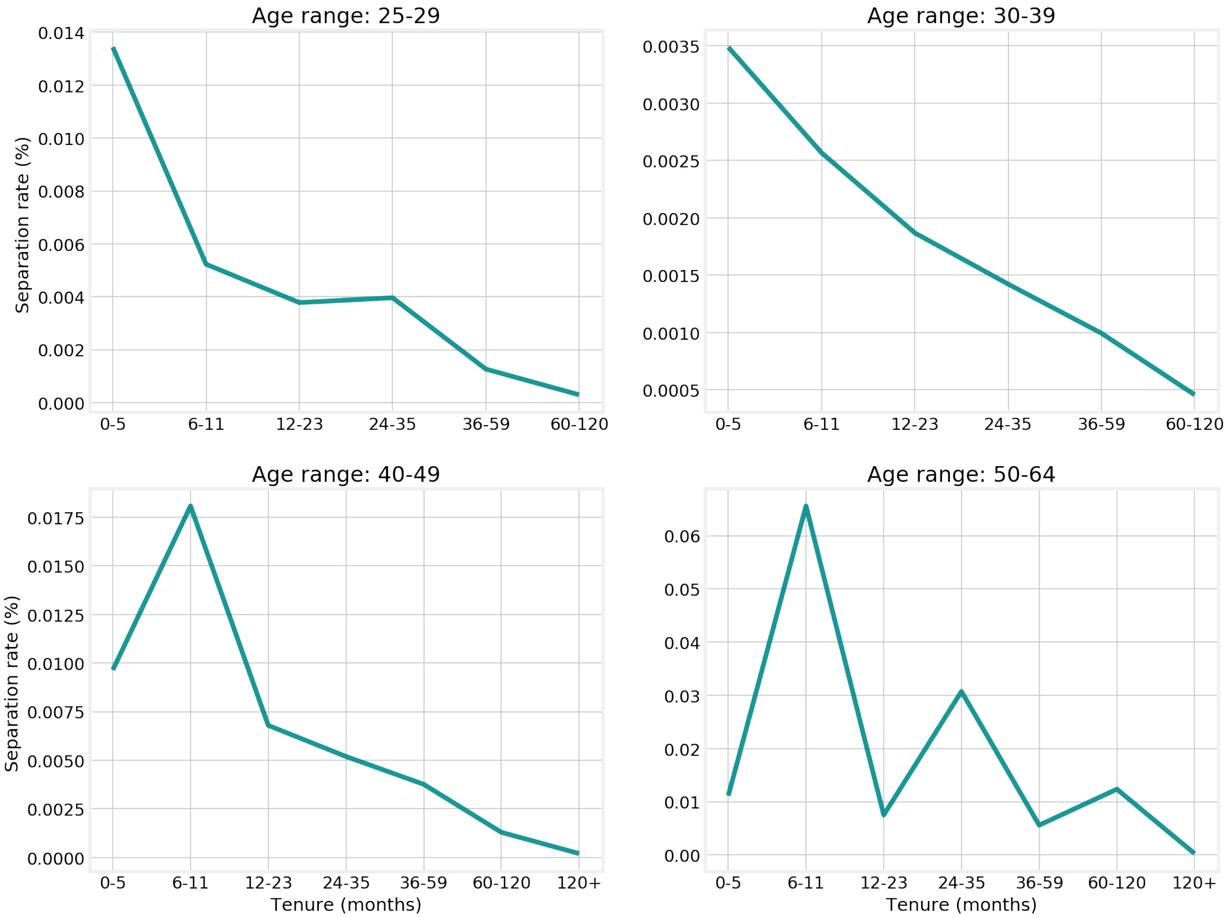

Figure 1: Separation rates by age group and tenure length. The figure plots separation rates from formal employment as a function of worker age and tenure, using Brazilian administrative data. Separation rates decline sharply with tenure and increase with age, consistent with rising severance obligations over the employment relationship.

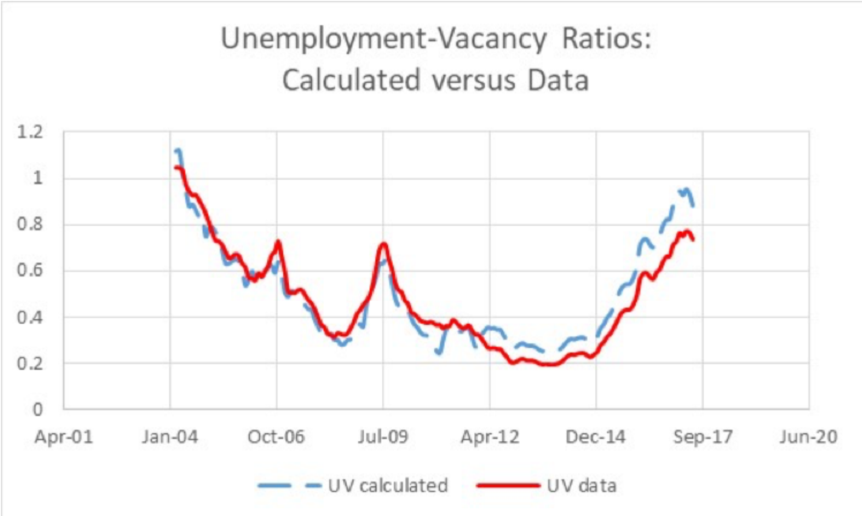

Figure 2: Labor market tightness measured by the unemployment–vacancy ratio. The figure shows the aggregate unemployment–vacancy ratio used to discipline labor market tightness in the model calibration. Higher values indicate a thinner labor market and lower job-finding probabilities.

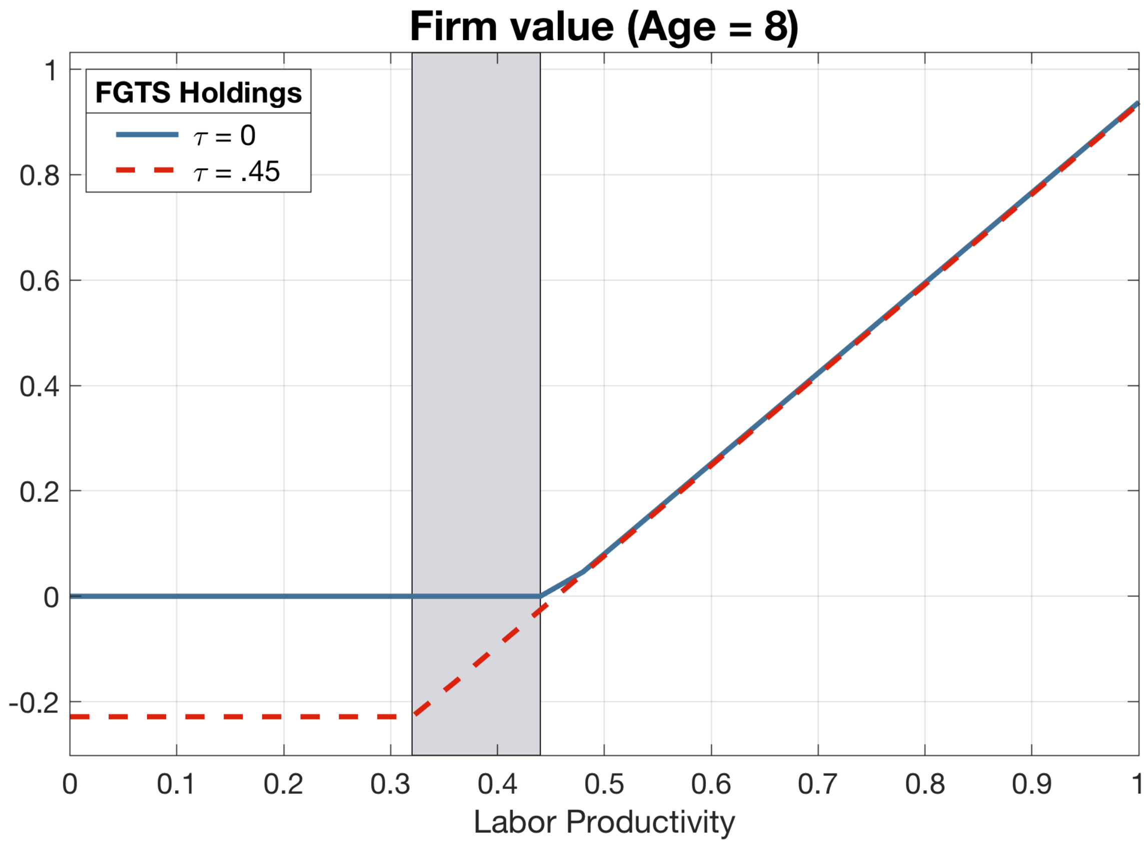

Figure 3: Firm value from employing an experienced worker. The figure plots the firm's continuation value as a function of match productivity for workers with different levels of accumulated FGTS balances. The kink identifies reservation productivity, illustrating how tenure-dependent severance costs induce firms to retain low-productivity, long-tenured workers.

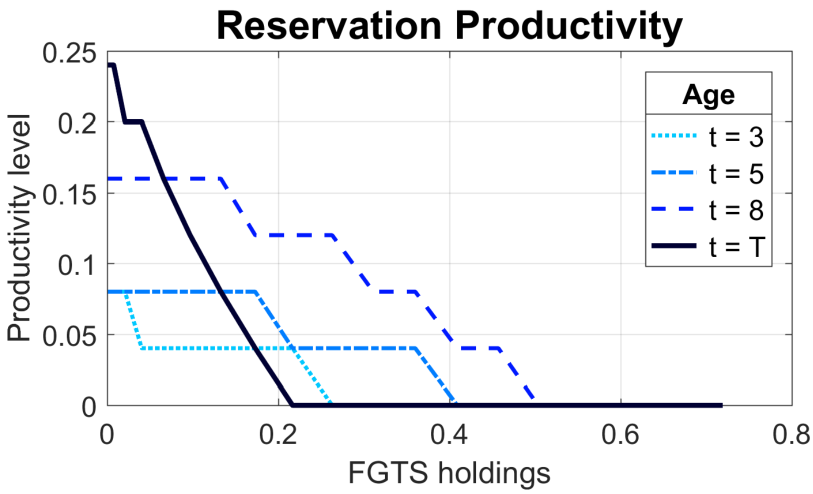

Figure 4: Reservation productivity by age and FGTS holdings. The figure reports the model-implied reservation productivity surface. Reservation productivity declines sharply with FGTS holdings (tenure), reflecting labor hoarding incentives, and increases with age at low tenure, reflecting shorter continuation values for older workers.

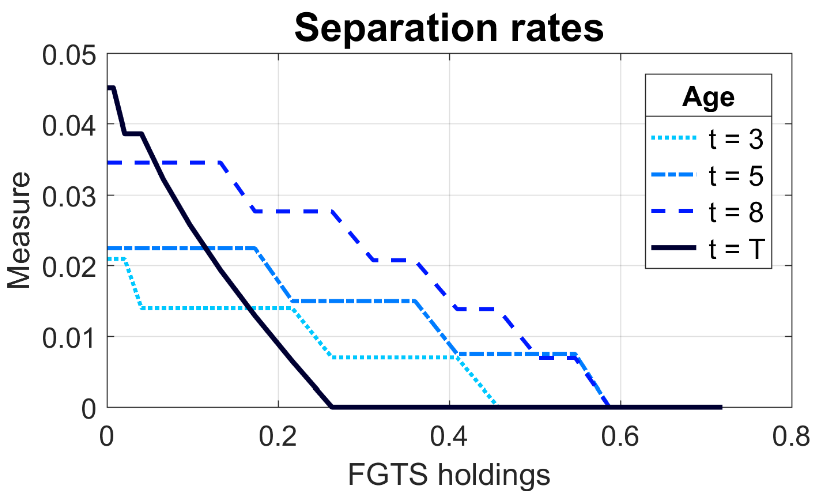

Figure 5: Model-implied separation rates by age and FGTS holdings. Separation rates decline with accumulated FGTS balances for all age groups, approaching zero at high tenure. Older workers exhibit higher separation risk at low tenure but stronger job protection once sufficient FGTS has accumulated.

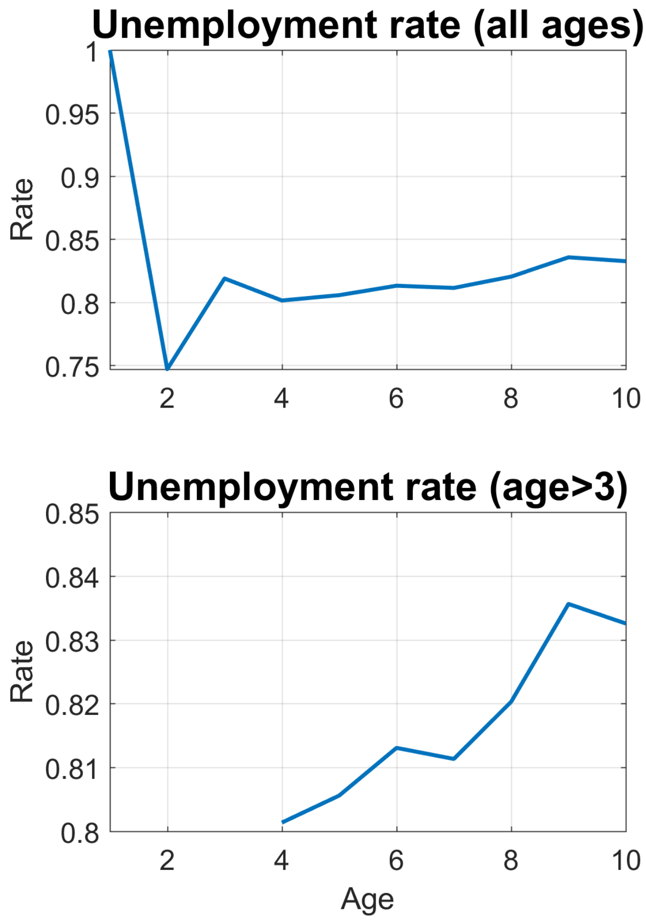

Figure 6: Unemployment rate by age. The figure reports model-implied unemployment rates across age groups, reflecting the interaction between hiring thresholds, job-finding probabilities, and separation behavior over the life cycle.

Figure 7: Age-specific value of unemployment. The figure displays the value of unemployment as a function of age. The declining and concave profile reflects both finite working lives and age-dependent job-finding prospects driven by firms' hiring thresholds.

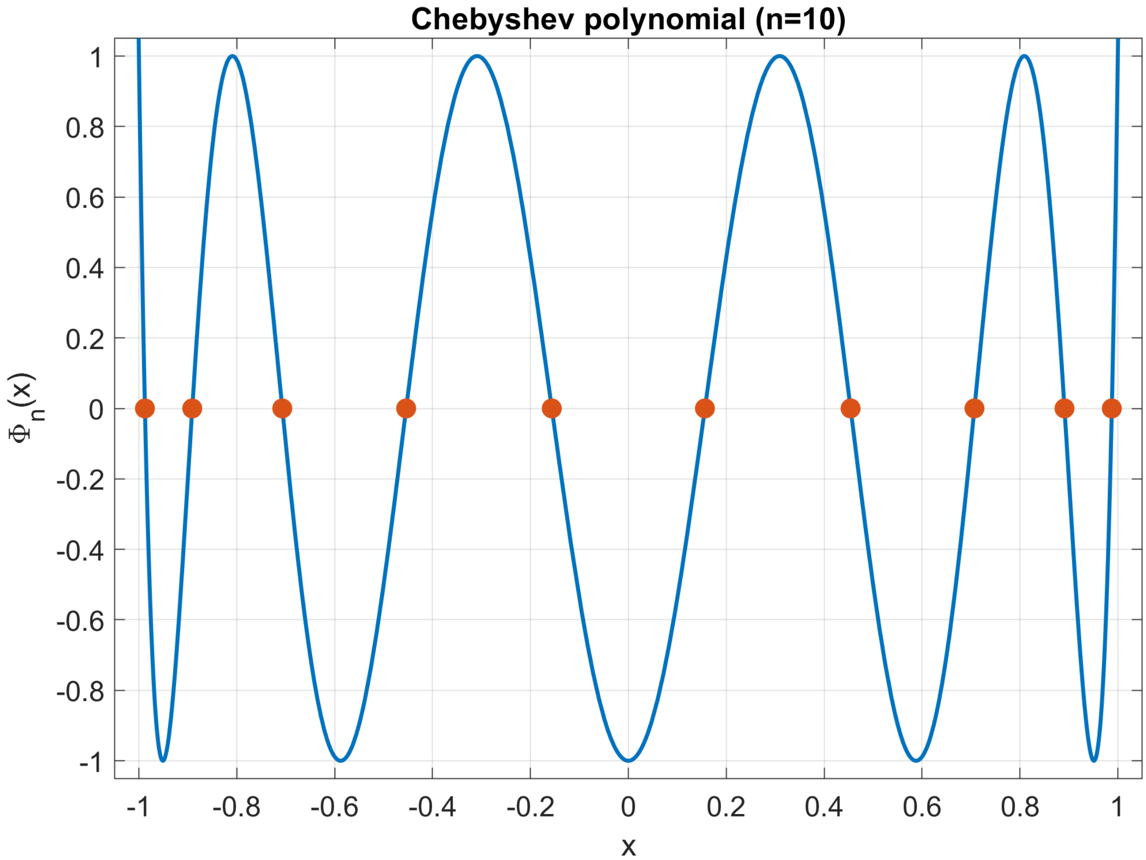

Figure 8: Chebyshev polynomial approximation of the productivity distribution. The figure illustrates the order-10 Chebyshev basis used to flexibly approximate the productivity distribution in the estimation procedure.

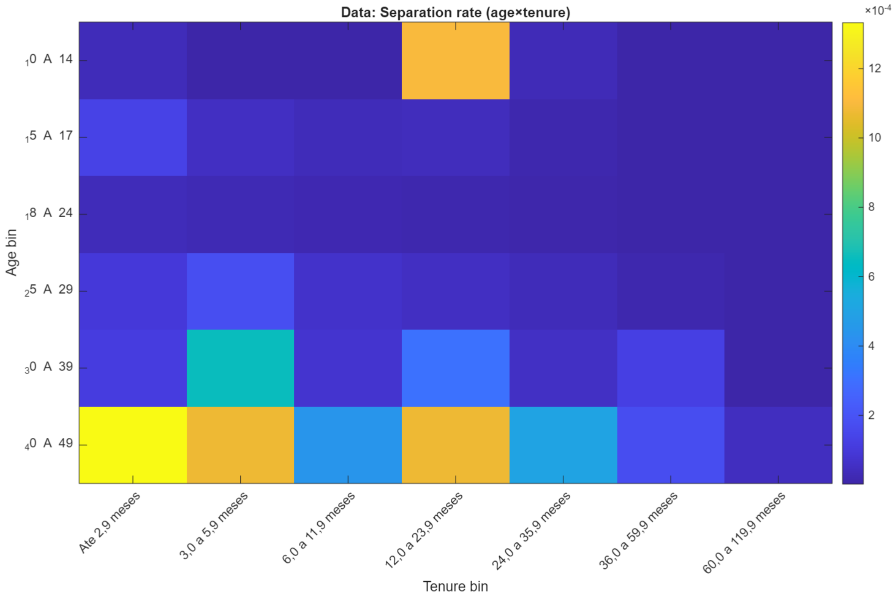

Figure 9: Separation rates by age group and tenure length. The figure plots involuntary separation rates from formal employment using Brazilian administrative data, by worker age group and tenure. Separation rates decline sharply with tenure for all age groups and are substantially higher for older workers at short tenure, consistent with rising severance obligations over the employment relationship.

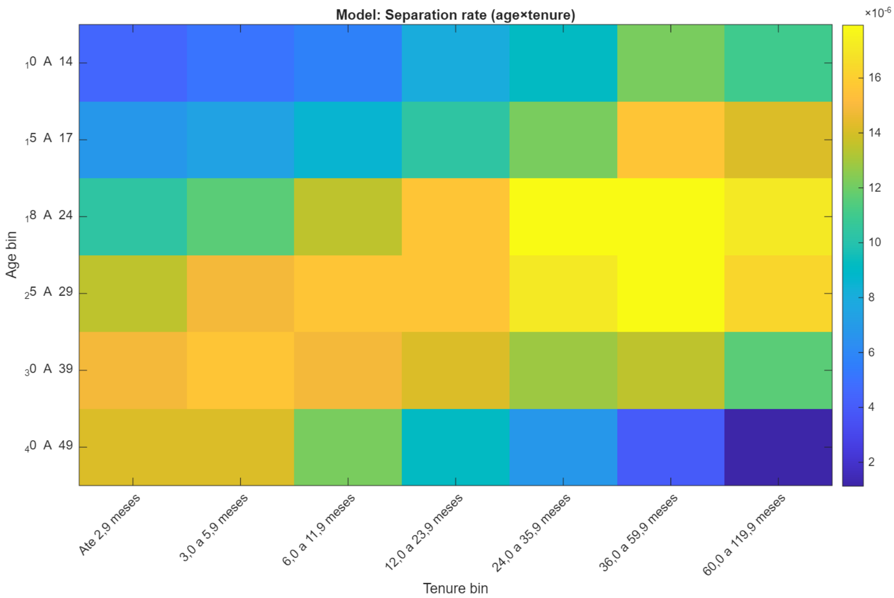

Figure 10: Separation rates by age and tenure: model. This figure shows separation rates implied by the estimated model. The model reproduces the strong decline in separations with tenure and higher separation rates for older workers early in employment, reflecting age-dependent reservation productivity and tenure-dependent firing costs.

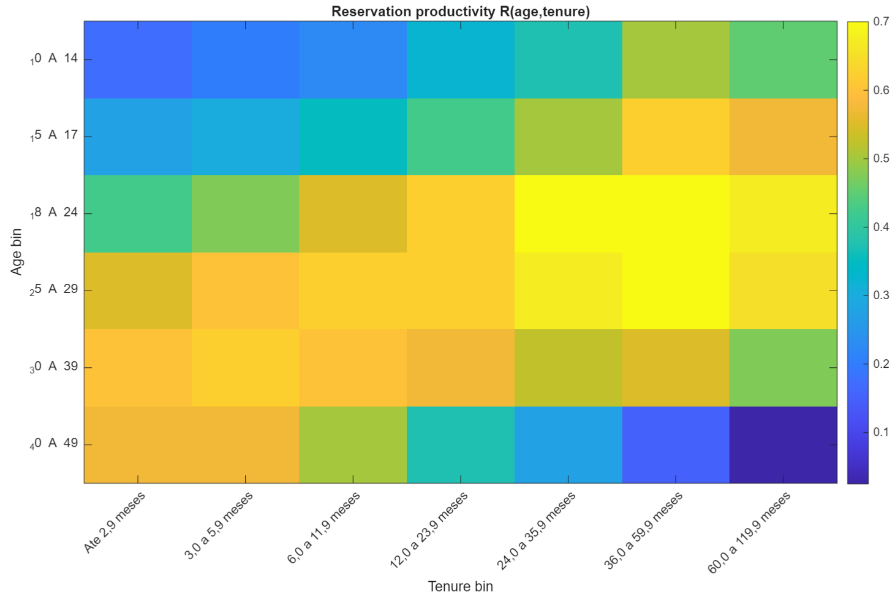

Figure 11: Reservation productivity by age and tenure. This figure reports the estimated reservation productivity surface $R(a, \tau)$. Reservation productivity declines steeply with tenure within age groups as firing costs accumulate, increases with age at short tenure due to lower continuation values, and approaches zero for older, long-tenured workers, reflecting firms' incentives to wait for retirement rather than incur severance costs.

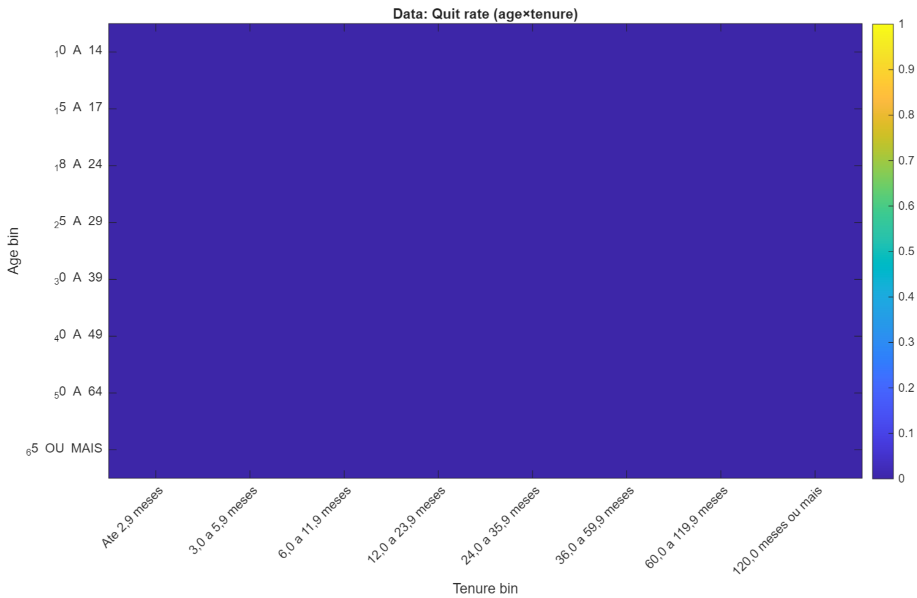

Figure 12: Quit rates by age and tenure: data (placebo). This figure reports monthly quit rates in the Brazilian formal sector by age and tenure bins. Quit rates are defined as voluntary separations divided by employment at the beginning of the period. In contrast to involuntary separations, quit rates display little systematic variation with tenure and no pronounced age–tenure gradient.

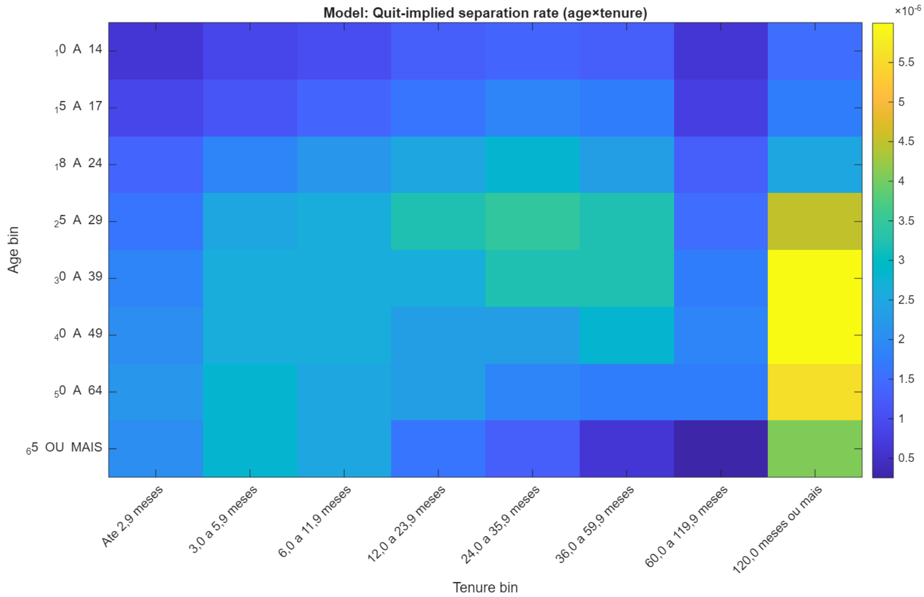

Figure 13: Quit-implied separation rates by age and tenure: model (placebo). This figure shows separation rates implied by the estimated model when quit rates are used as empirical moments. The model generates very small and smooth separation probabilities and fails to reproduce the sharp age–tenure patterns observed for involuntary separations.

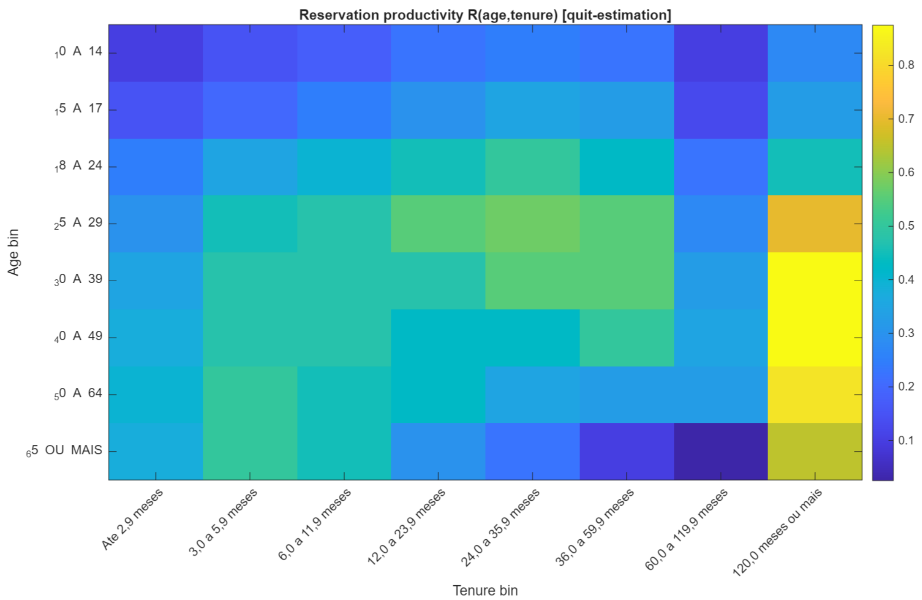

Figure 14: Reservation productivity by age and tenure under quit estimation (placebo). This figure reports the reservation productivity surface $R(a, \tau)$ implied by the model when estimated using quit rates. The model requires implausibly high reservation thresholds at long tenures to rationalize the low level and weak tenure dependence of quit rates, highlighting the limited relevance of the firing-cost mechanism for voluntary separations.

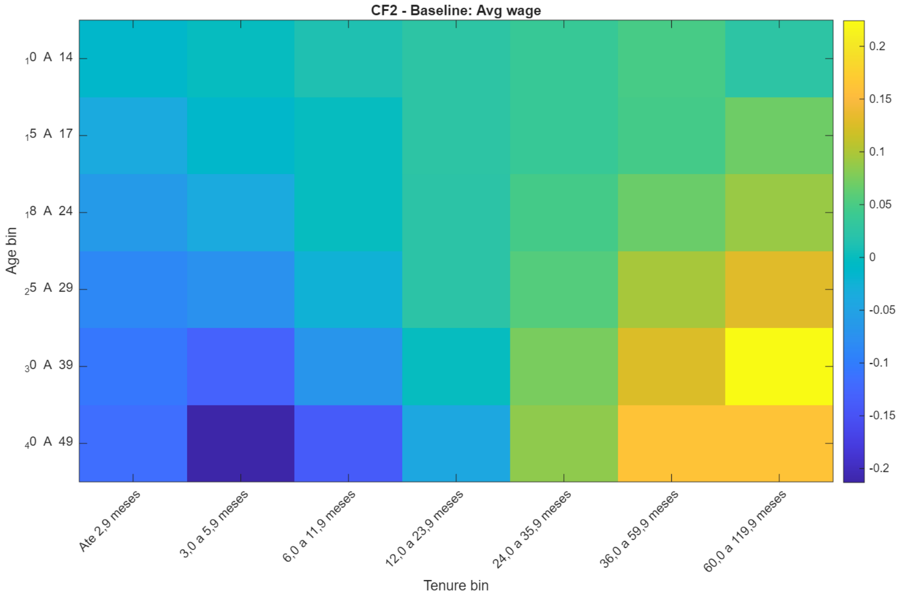

Figure 15: Change in average wages under flat severance relative to the baseline (Option A). The figure reports the difference in model-implied wages across age–tenure cells when tenure-dependent severance is replaced by a flat severance schedule that preserves average firing costs. With endogenous wage adjustment, reductions in severance gradients are capitalized into wages, with larger wage increases for high-tenure workers facing lower expected firing costs.

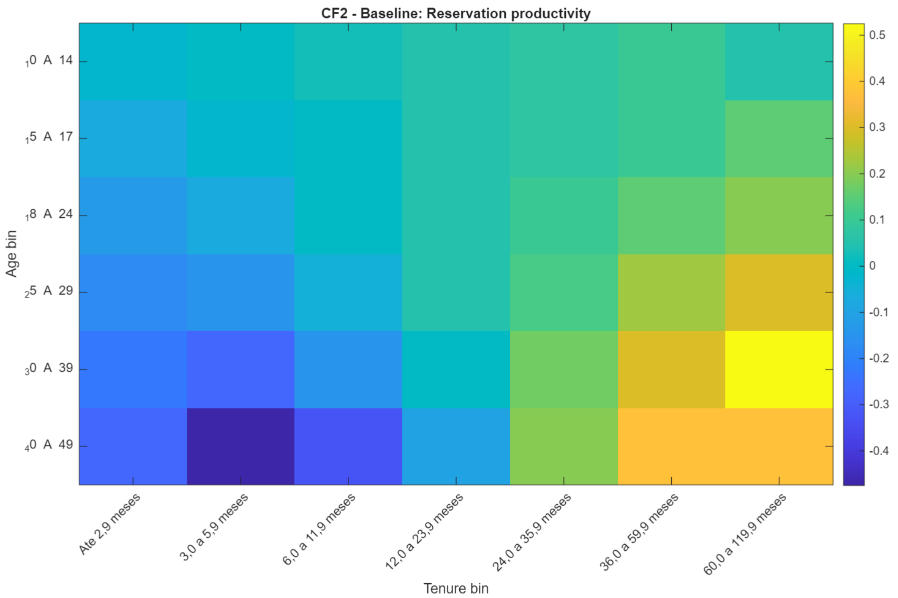

Figure 16: Change in reservation productivity under flat severance relative to baseline (Option A). Flattening severance raises reservation productivity for high-tenure matches, reducing firms' incentives to retain low-productivity workers purely due to accumulated firing costs.

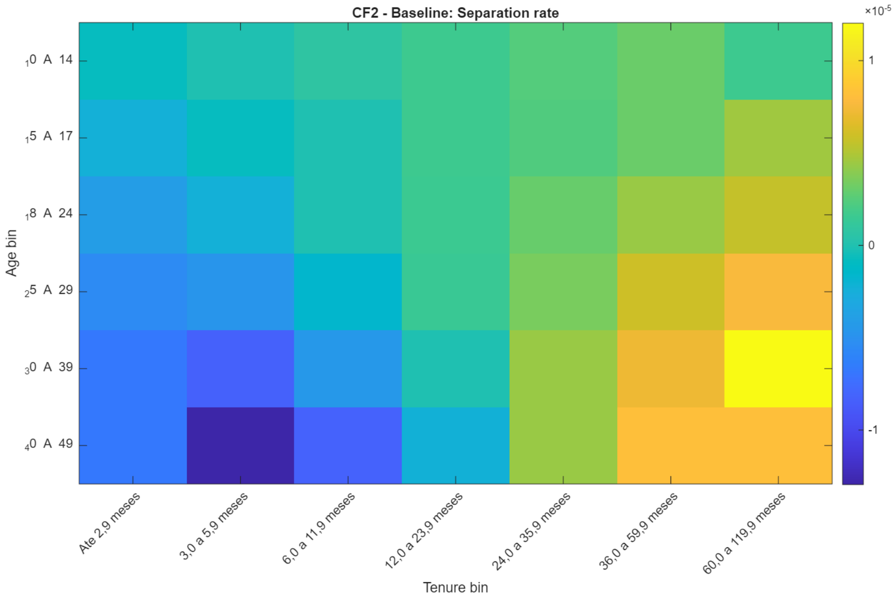

Figure 17: Change in separation rates under flat severance relative to baseline (Option A). Separation rates fall for older and long-tenure workers, compressing the tenure gradient in job destruction and reducing labor hoarding incentives.

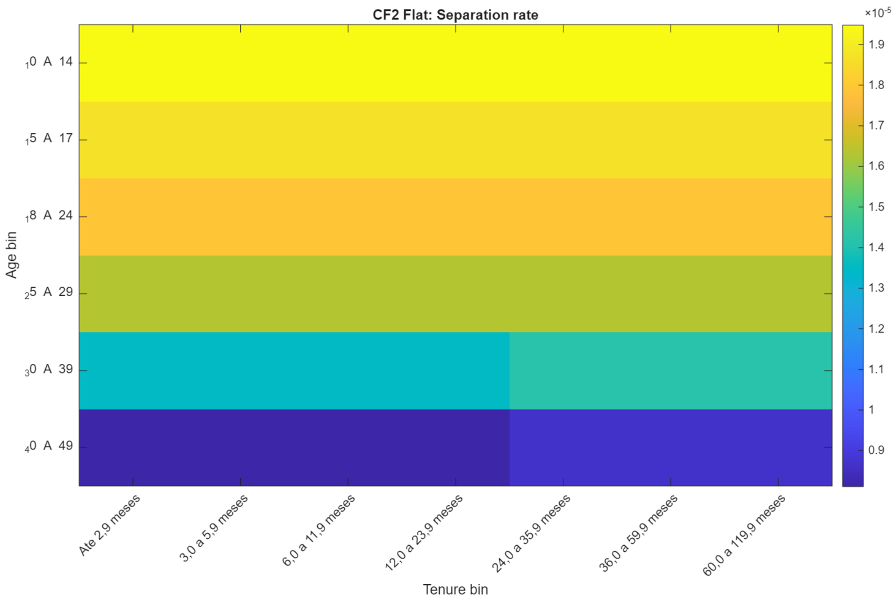

Figure 18: Separation rates under flat severance (Option A). Under a tenure-invariant severance schedule, separation rates exhibit substantially weaker dependence on tenure, indicating that firing incentives are no longer mechanically tied to employment history.

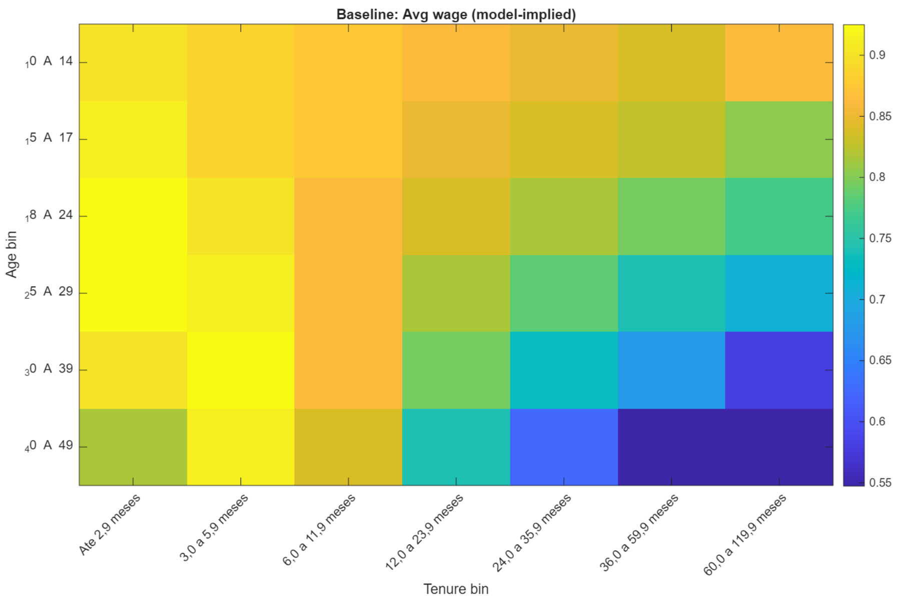

Figure 19: Baseline equilibrium wages (model-implied). The figure shows equilibrium wages across age and tenure in the baseline economy with tenure-dependent severance pay, serving as a reference point for the counterfactual wage changes in Figure 15.

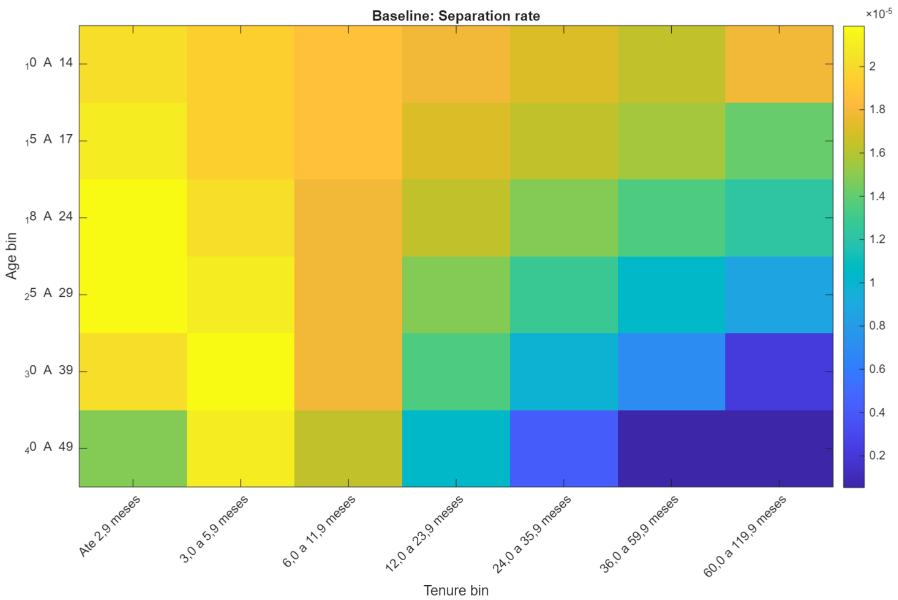

Figure 20: Baseline separation rates. The figure reports model-implied separation rates across age and tenure in the baseline economy, highlighting the strong decline in separations with tenure induced by tenure-dependent severance costs.

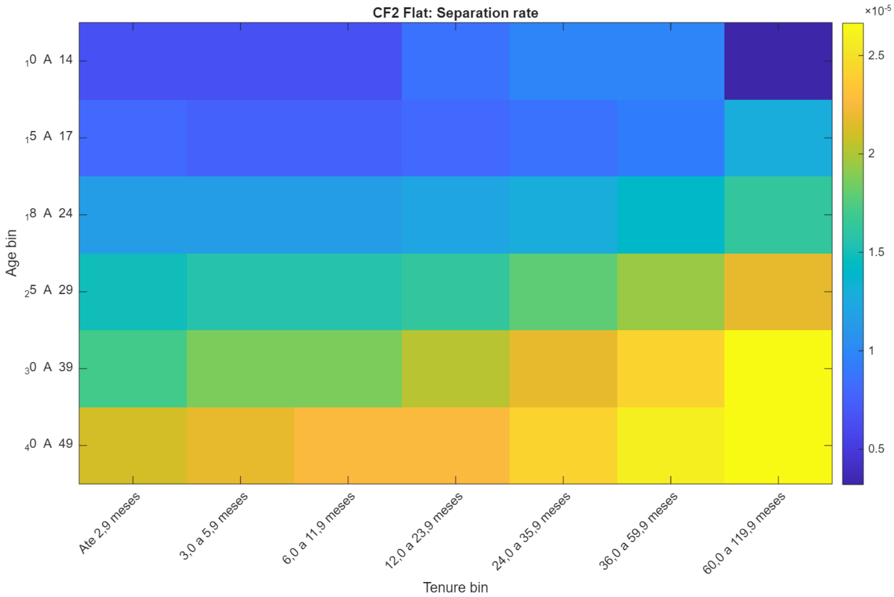

Figure 21: Equilibrium separation rate under flat severance. Separation rates vary primarily with age, while tenure gradients are substantially attenuated due to the absence of tenure-contingent severance costs.

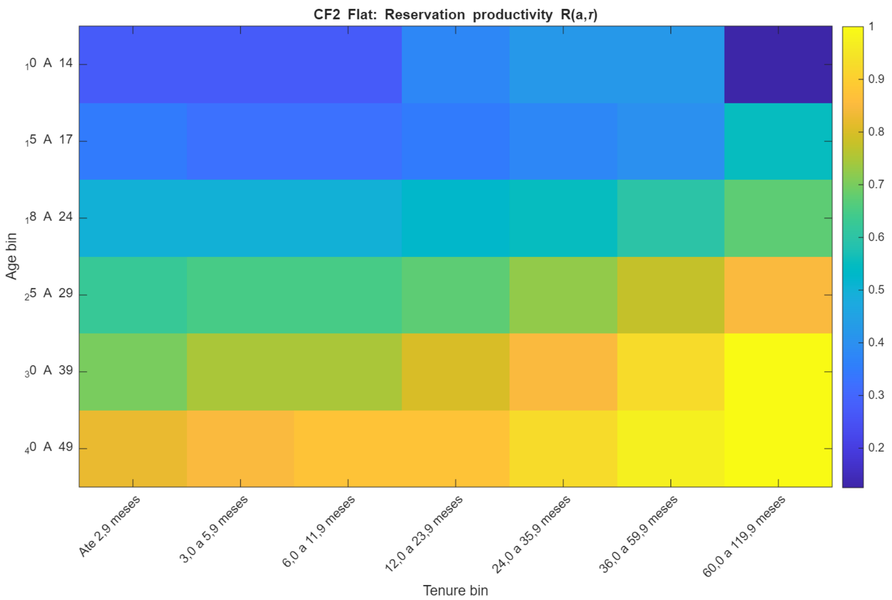

Figure 22: Reservation productivity under flat severance. The reservation threshold increases with age and tenure, but exhibits a flatter tenure profile relative to environments with tenure-dependent severance.

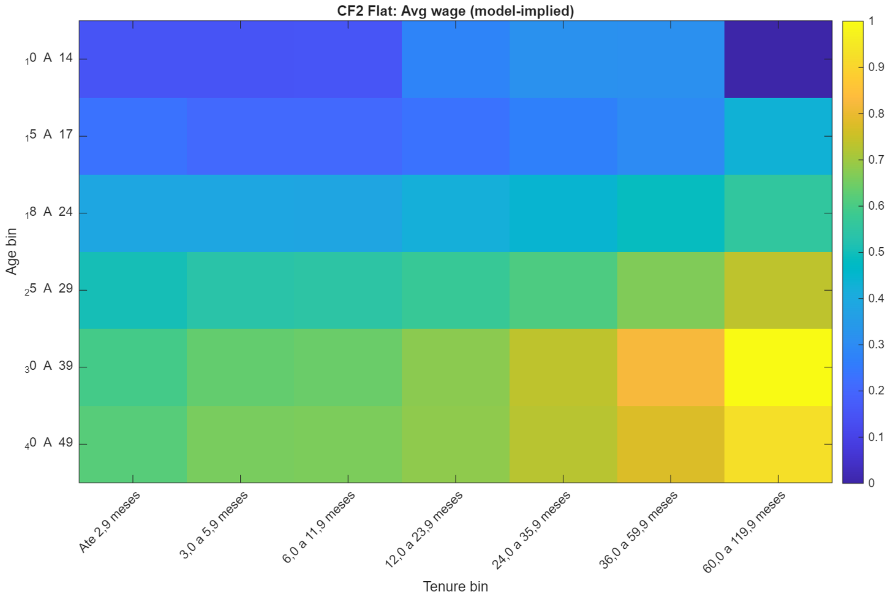

Figure 23: Model-implied average wages under flat severance. Wage profiles increase with age and tenure but display smoother tenure gradients due to weaker severance-induced bargaining effects.

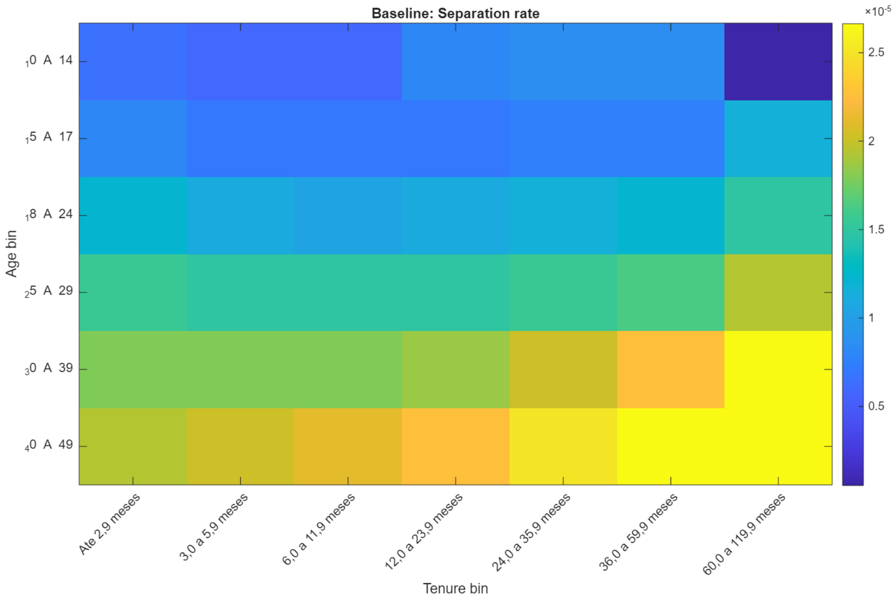

Figure 24: Baseline equilibrium separation rate. The baseline economy features stronger tenure dependence in separations, reflecting increasing severance obligations as tenure accumulates.

B List of Tables

		Benchmark model			
Parameter	Tenure	Tenure^2	Age	Age^2	Tenure · Age
Estimate	1.872	-0.009	4.223	-0.014	0.008
t-statistic	1430.87	-61.28	2679.70	-82.15	30.49
		Restricted model			
Parameter	Tenure		Age		
Estimate	1.876		4.22		
t-statistic	1446.10		2701.38		

Table 1: OLS estimates. Equation in top panel: $w_{it} = \alpha_i + \alpha_t + \beta_1\, age_{it} + \beta_2\, age_{it}^2 + \beta_3\, tenure_{it} + \beta_4\, tenure_{it}^2 + \beta_5\, age_{it} \cdot tenure_{it} + \varepsilon_{it}$, where w_{it} is log wage. Bottom panel imposes restrictions $\beta_2 = 0$, $\beta_4 = 0$, and $\beta_5 = 0$.

Table 2: Output loss from tenure-dependent severance pay

Specification	Efficient benchmark	Y_{base}	Y_{eff}	$Y_{\text{eff}} - Y_{\text{base}}$
Option A	CF1 (zero severance)	0.788459	1.000000	0.211541
Option B	CF1 (zero severance)	0.869246	0.875795	0.006549

Notes: Y_{base} denotes aggregate output in the baseline economy with tenure-dependent severance pay. Y_{eff} denotes output in the efficient counterfactual with zero severance (CF1), which eliminates severance-induced retention distortions. Option A allows wages to adjust endogenously in partial equilibrium, while Option B allows accumulated FGTS balances to enter workers' outside options. Output is measured as expected match productivity aggregated using baseline employment weights.

Table 3: Cohort-Level Employment Misallocation under Option A

Policy	Total Variation Distance ($\frac{1}{2}\|\cdot\|_1$)	Kullback–Leibler Divergence (Baseline ‖ Policy)
Baseline	0.000000	0.000000e+00
CF1: No Severance	0.000010	2.816793e−10
CF2: Flat Severance	0.000003	2.336255e−11
CF3: $\kappa = 0.25$	0.000007	1.363664e−10
CF3: $\kappa = 0.50$	0.000004	4.708322e−11
CF3: $\kappa = 0.75$	0.000002	9.153101e−12

Notes: The table reports measures of cohort-level employment misallocation relative to the baseline economy, computed from the distribution of employment across age–tenure cells. The baseline features the estimated tenure-dependent severance schedule. CF1 eliminates severance pay. CF2 replaces the tenure-dependent schedule with a flat severance level that preserves average severance. CF3 scales the baseline schedule uniformly by factor κ. Option A allows wages to adjust endogenously while holding workers' outside options fixed.

Table 4: Cohort-Level Employment Misallocation under Option B

Policy	Total Variation Distance ($\frac{1}{2}\|\cdot\|_1$)	Kullback–Leibler Divergence (Baseline ‖ Policy)
Baseline	0.000000e+00	0.000000e+00
CF1: No Severance	1.188411e−06	4.077493e−12
CF2: Flat Severance	9.317777e−07	2.427149e−12
CF3: $\kappa = 0.25$	1.008415e−06	2.913502e−12
CF3: $\kappa = 0.50$	5.863566e−07	1.013291e−12
CF3: $\kappa = 0.75$	3.103645e−07	2.752605e−13

Notes: Counterfactual definitions are identical to Table 3. Option B holds wages fixed at their baseline level and allows accumulated FGTS balances to enter workers' outside options, introducing an additional tenure-linked insurance component.

Table 5: Partial-Equilibrium Welfare Effects of Severance Pay: Option A

Policy	Total Welfare	Worker Welfare	Firm Welfare
Baseline	0.788	0.788	-3.3×10^{-6}
CF1: No Severance	1.000	1.000	1.1×10^{-9}
CF2: Flat Severance	0.799	0.799	-5.0×10^{-6}
CF3: $\kappa = 0.25$	0.933	0.933	-1.8×10^{-6}
CF3: $\kappa = 0.50$	0.871	0.871	-2.7×10^{-6}
CF3: $\kappa = 0.75$	0.824	0.824	-3.1×10^{-6}

Notes: Option A allows wages to adjust endogenously to changes in severance pay. The baseline corresponds to the estimated tenure-dependent severance schedule. CF1 eliminates severance pay entirely. CF2 flattens severance across tenure while preserving its average level. CF3 scales down severance by factor κ.

Table 6: Partial-Equilibrium Welfare Effects of Severance Pay: Option B

Policy	Total Welfare	Worker Welfare	Firm Welfare
Baseline	0.869	0.643	0.227
CF1: No Severance	0.876	0.643	0.233
CF2: Flat Severance	0.883	0.643	0.241
CF3: $\kappa = 0.25$	0.875	0.643	0.232
CF3: $\kappa = 0.50$	0.873	0.643	0.230
CF3: $\kappa = 0.75$	0.871	0.643	0.228

Notes: Option B holds wages fixed at their baseline level and allows severance pay to affect workers' outside options. Worker welfare is therefore largely insulated from severance reforms, while firms bear most of the adjustment.

C Appendix: Model Solution

A solution to our model is a set of reservation productivities $\{R_i(\tau)\}_{i=2}^{T}$ that make the firm indifferent between keeping or breaking the match and fulfill terminal conditions $J_T(\varepsilon, \tau)$. Specifically, $\{R_i(\tau)\}_{i=1}^{T}$ solves the system

$$J_i\,(\varepsilon, \tau) = max\left\{\varepsilon - w_i(\varepsilon) + \beta\left[\lambda\, J_{i+1}(\varepsilon, \tau') + (1-\lambda)\int_0^1 J_{i+1}(\varepsilon', \tau')\, dG(\varepsilon')\right], V - \delta_1 \tau\right\}$$

$$\text{s.t.} \qquad \tau' = \tau + \delta_0\, w_i(\varepsilon). \tag{8}$$

$$J_T\,(\varepsilon, \tau) = max\left\{\varepsilon - W_T(\varepsilon) + \beta\, V, V - \delta_1\, \tau\right\}.$$

In order to solve for $\{R_i(\tau)\}$, we need $\{w_i(\varepsilon)\}_{i=2}^{T}$, which, in turn, depends on value from unemployment $\{U_i\}_{i=2}^{T}$ that depends on value from employment $\{E_i(\varepsilon, 0)\}_{i=2}^{T}$.

Our solution strategy consists in approximating $\{J_i(\varepsilon, \tau)\}_{i=2}^{T}$ and $\{E_i(\varepsilon, \tau)\}_{i=2}^{T}$ numerically via Chebyshev polynomials. Chebyshev polynomials have been extensively used in Macroeconomics because of their convenient properties, such as bounded domain and orthogonal basis. In particular, the orthogonality of the columns in the Chebyshev-generated basis matrix is critical for the computational efficiency of our solution method. A more detailed treatment of Chebyshev polynomials is available in Fornberg (1996) and Boyd (2001). Fernández-Villaverde et al. (2016) compare the performance of Chebyshev polynomials to that of other solution methods in a standard RBC model.

To make our exposition clear, let $\Phi_k(x)$ denote the k-th order Chebyshev polynomial in x, wich can be computed recursively according to

$$\begin{aligned}
&\Phi_1(x) = 1 \\
&\Phi_2(x) = x \\
&\Phi_k(x) = 2\,x\,\Phi_{k-1}(x) - \Phi_{k-2}(x), \quad k \geq 2,
\end{aligned}$$

for $x \in [-1, 1]$. An equivalent approach is $\Phi_k(x) = cos\left[(k-1)\,cos^{-1}(x)\right]$.

Labor productivity is uniformly distributed in $[0, 1]$. We discretize the $[0, 1]$ interval in m points that are assigned probability $\frac{1}{m}$ each. The grid for state variable τ is generated by the roots of the Chebyshev interpolating polynomial. The roots $\Phi_n(x) = 0$ are called Chebyshev interpolation nodes and, unlike a grid with uniformly spaced nodes, Chebyshev interpolation nodes tend to populate the extremes of the $[-1, 1]$ domain (see Figure 8).

Let x_k denote the k-th Chebyshev node of a polynomial of order n. That is, $\Phi(x_k) = 0$ for $x_1, \ldots, x_n$, which will determine the grid for our model's state variable. Transformation

$\tau_k = \frac{\bar{\tau}}{2} + \frac{\bar{\tau}}{2} x_k$ maps Chebyshev node x_k into the set of feasible FGTS holdings $[0, \bar{\tau}]$. The polynomial interpolation consists of linear systems

$$J_i\left(\varepsilon_j, \tau_k\right) = a_1^{ij}\, \Phi_1(x_k) + a_2^{ij}\, \Phi_2(x_k) + \ldots + a_n^{ij}\, \Phi_n(x_k) \tag{9}$$

$$E_i\left(\varepsilon_j, \tau_k\right) = b_1^{ij}\, \Phi_1(x_k) + b_2^{ij}\, \Phi_2(x_k) + \ldots + b_n^{ij}\, \Phi_n(x_k)$$

for each age $i = 1, \ldots, T-1$ and productivity level $j = 1, \ldots, m$. In practice, fitting n-order Chebyshev polynomials implies solving an n-by-n linear system $m \times T - 1 \times 2$ times in every step of the following iterative process:

Step 1: Given starting values for polynomial coefficients $\mathbf{a}$ and $\mathbf{b}$, compute $\mathbf{J}_i$ and $\mathbf{E}_i$ as in Eq. 9.

Step 2: Use $\{\mathbf{E}_i\}$ from previous step and compute value from unemployment U_i for $i = 1, \ldots, T-1$.

Step 3: Use U_i from previous step and compute equilibrium wages $w_i(\varepsilon)$.

Step 4: Use $\mathbf{J}_i$ and $\mathbf{E}_i$ from Step 1, U_i and $w_i(\varepsilon)$ from Steps 2 and 3 and compute $\tilde{J}_i(\varepsilon_j, \tau_k)$ and $\tilde{E}_i(\varepsilon_j, \tau_k)$ as follows:

$$\tilde{J}_i(\varepsilon_j, \tau_k) = max\left\{\varepsilon - w_i(\varepsilon) + \beta\left[\lambda\, J_{i+1}\left(\varepsilon_j, \tau'_{k,ij}\right) + (1-\lambda)\int_0^1 J_{i+1}\left(\varepsilon', \tau'_{k,ij}\right)\, dG(\varepsilon')\right],\right.$$
$$\left. V - \delta_1 \tau_k \right\}, \text{ and } \tau'_{k,ij} = \tau_k + \delta_0\, w_i(\varepsilon_j)$$

$$\tilde{E}_i(\varepsilon_j, \tau_k) = max\left\{w_i(\varepsilon) + \beta\left[\lambda\, E_{i+1}\left(\varepsilon_j, \tau'_{k,ij}\right) + (1-\lambda)\int_0^1 E_{i+1}\left(\varepsilon', \tau'_{k,ij}\right)\, dG(\varepsilon')\right], U_i + \delta_1 \tau_k\right\}$$

Step 5: Coefficient updating. Stack

$$\tilde{\mathbf{J}}_{ij} = \left[\tilde{J}_i(\varepsilon_j, \tau_1), \ \dots, \ \tilde{J}_i(\varepsilon_j, \tau_n)\right]'$$

$$\tilde{\mathbf{E}}_{ij} = \left[\tilde{E}_i(\varepsilon_j, \tau_1), \ \dots, \ \tilde{E}_i(\varepsilon_j, \tau_n)\right]'$$

and update coefficients $\mathbf{a}_{new}^{ij} = \mathbf{\Psi}^{-1}\,\tilde{\mathbf{J}}_{ij}$ and $\mathbf{b}_{new}^{ij} = \mathbf{\Psi}^{-1}\,\tilde{\mathbf{E}}_{ij}$, where basis matrix $\mathbf{\Psi}$ is given by

$$\mathbf{\Psi} = \begin{bmatrix} \Phi_1(x_1) & \dots & \Phi_n(x_1) \\ \vdots & \dots & \vdots \\ \Phi_1(x_n) & \dots & \Phi_n(x_n) \end{bmatrix}.$$

Step 6: Compute updated $\hat{\mathbf{J}}_{ij} = \mathbf{\Psi}\,\mathbf{a}_{new}^{ij}$ and $\hat{\mathbf{E}}_{ij} = \mathbf{\Psi}\,\mathbf{b}_{new}^{ij}$ and distances $d_J = \underset{i,j,k}{max}\,|J_i(\varepsilon_j, \tau_k) - \hat{J}_i(\varepsilon_j, \tau_k)|$ and $d_E = \underset{i,j,k}{max}\,|E_i(\varepsilon_j, \tau_k) - \hat{E}_i(\varepsilon_j, \tau_k)|$, which measure the improvement in polynomial fit in current iteration.

Step 7: Given tolerance error Δ, if either $d_j > \Delta$ or $d_E > \Delta$, then set $\mathbf{a} = \mathbf{a}_{new}$ and $\mathbf{b} = \mathbf{b}_{new}$ and return to Step 1.

Finding good starting values for coefficients **a** and **b** is not trivial. Instead, we take advantage of the terminal conditions $J_T(\varepsilon, \tau)$ and $E_T(\varepsilon, \tau)$ which are know given $w_T(\varepsilon)$. Since $w_T(\varepsilon)$ depends only on ε, U_t and U_{T+1}, which are all known, we set $\mathbf{a}_0 = \mathbf{\Psi}^{-1}\,\mathbf{J}_T$ and $\mathbf{b}_0 = \mathbf{\Psi}^{-1}\,\mathbf{E}_T$ as initial values for **a** and **b**.

The results shown in Section 5 were obtained with Chebyshev polynomials of order $n = 23$. We avoid error propagation from starting values we start with $n = 3$, iterate until convergence, and then use the results as starting values for a polynomial of order 5 approximation - coefficients of terms of order greater than three are set to zero. We proceed with $n = 5, 7, 11, 17$ and 23. At

$n = 23$, coefficients of high-order terms are virtually zero for all ages and productivity levels. Tolerance error is set to 10^{-5}.

D Appendix: Estimation Details

This appendix provides additional econometric and computational details on the simulated method of moments (SMM) procedure used to estimate the model.

D.1 Parameterization and State Space

Labor productivity ε is assumed to lie in the compact interval $[0, 1]$ and to follow an unknown distribution $G(\varepsilon)$. We approximate G using a discrete distribution defined over a fixed grid $\{\varepsilon_j\}_{j=1}^N$, evenly spaced on $[0, 1]$. Let p_j denote the probability mass assigned to grid point ε_j.

To ensure a valid probability distribution without imposing parametric restrictions on the shape of G, probabilities are parameterized via a softmax transformation,

$$p_j = \frac{\exp(a_j)}{\sum_{k=1}^N \exp(a_k)}, \tag{10}$$

which guarantees $p_j \geq 0$ and $\sum_{j=1}^N p_j = 1$. The vector of unrestricted parameters $\{a_j\}_{j=1}^N$ is estimated jointly with the remaining structural parameters.

In addition to the productivity distribution, we estimate the persistence parameter $\lambda \in (0, 1)$ governing match-specific productivity survival and the scale parameter $\xi_1 > 0$ governing tenure-dependent firing costs. Parameter constraints are enforced through transformations,

$$\lambda = \frac{\exp(\ell)}{1 + \exp(\ell)}, \qquad \xi_1 = \exp(\kappa), \tag{11}$$

where ℓ and κ are unrestricted real-valued parameters. All remaining parameters, including the FGTS contribution rate, the discount factor, and the return on accumulated balances, are calibrated to institutional features of the Brazilian system.

The state space consists of discrete age and tenure bins that correspond exactly to those observed in the data. The model is solved by backward induction over this finite age–tenure state space.

D.2 Empirical Moments

The vector of empirical moments g^{data} stacks three sets of objects measured at the age–tenure cell level:

1. employment shares by age and tenure;
2. average monthly wages by age and tenure;
3. monthly separation rates by age and tenure.

Separation rates are defined as

$$s_{a,\tau}^{\text{data}} = \frac{\text{terminations without cause}_{a,\tau}}{\text{employment at the beginning of the period}_{a,\tau}}. \tag{12}$$

All moments are constructed from Brazilian administrative data and aligned across datasets to ensure a common age–tenure support. Cells for which moments are missing in the data are excluded symmetrically from both the empirical and model-implied moment vectors.

D.3 Weighting Matrix

The weighting matrix W used in estimation is diagonal,

$$W = \mathrm{diag}(w_1, \ldots, w_M), \tag{13}$$

where M is the total number of moments included in the objective function. Employment shares and separation rates receive relatively higher weight than wage moments to reflect their central role in identifying separation and retention behavior, which is the primary focus of the model. Results are robust to alternative reasonable choices of relative weighting.

D.4 Model Solution and Simulation

For a given parameter vector θ, the model is solved by backward induction over age and tenure. At each age–tenure state (a, τ), firms compare the continuation value of an employment relationship to the option value of termination, which depends on the accumulated firing cost implied by tenure and the FGTS contribution structure. Reservation productivity thresholds $R(a, \tau)$ are computed endogenously from the firm's indifference condition,

$$\varepsilon = R(a, \tau) \quad \Longleftrightarrow \quad \varepsilon - w_a(\varepsilon) + \beta \mathbb{E}[J_{a+1}(\varepsilon', \tau')] = V - \xi_1 \tau. \tag{14}$$

Model-implied separation rates correspond to the probability that productivity realizations fall below the reservation threshold, conditional on a productivity redraw. Specifically,

$$s_{a,\tau}^{\text{model}} = (1 - \lambda)\, G(R(a + 1, \tau)), \tag{15}$$

where $(1 - \lambda)$ is the probability of a productivity redraw and $G(\cdot)$ is the estimated productivity distribution.

Employment shares are computed by iterating implied survival probabilities forward over tenure within each age group,

$$e_{a,\tau+1} = e_{a,\tau} \left(1 - s_{a,\tau}^{\text{model}}\right), \tag{16}$$

and normalizing across all age–tenure cells to obtain model-implied employment shares.

Average wages are computed using the Nash-bargained wage schedule,

$$w_a(\varepsilon) = \phi\,\varepsilon + (1-\phi)\,(U_a - \beta U_{a+1}), \tag{17}$$

and aggregated within each age–tenure cell using the stationary distribution of productivity conditional on continued employment.

D.5 Objective Function and Optimization

The SMM estimator minimizes the quadratic distance between empirical and model-implied moments,

$$\hat{\theta} = \arg\min_{\theta} \left(g^{\text{model}}(\theta) - g^{\text{data}}\right)' W \left(g^{\text{model}}(\theta) - g^{\text{data}}\right). \tag{18}$$

The objective function is minimized using a quasi-Newton algorithm. Convergence is assessed using both changes in the value of the objective function and changes in parameter estimates. Multiple starting values are used to verify robustness to local minima.

D.6 Identification and Overidentification

Identification relies on the joint age–tenure structure of the empirical moments. The level and slope of separation rates at short tenure discipline the persistence parameter λ, which governs the frequency of productivity redraws early in employment relationships. The sharp decline in separation rates with tenure identifies the scale of firing costs ξ_1, as productivity persistence alone cannot generate near-zero separations at long tenure.

The productivity distribution $G(\varepsilon)$ is disciplined by the joint behavior of wages, employment shares, and separation rates. Wage moments restrict the mass of high-productivity matches, separation moments restrict probability mass below reservation thresholds, and employment shares discipline the implied survival dynamics across tenure.

The model is over-identified: the number of moments exceeds the number of estimated parameters, and identification is driven by cross-sectional variation across age and tenure rather than by any single moment or margin.

References

Arnold, D., Bernstein, J., 2018. Firing costs and efficiency in a frictional labor market: Evidence from brazil. Tech. rep., Princeton University.
URL https://scholar.princeton.edu/sites/default/files/joshuabernstein/files/firing_costs.pdf

Bargain, O., Kwenda, P., 2014. The informal sector wage gap: New evidence using quantile estimations on panel data. Economic Development and Cultural Change 63 (1), 117–153.

Boyd, J. P., 2001. Chebyshev and Fourier spectral methods. Courier Corporation.

Cheron, A., Hairault, J.-O., Langot, F., 2013. Life-cycle equilibrium unemployment. Journal of Labor Economics 31 (4), 843 – 882.

de Holanda Barbosa, L. N., Foguel, M. N., Bilo, C., Dec. 2017. The Brazilian Government Severance Indemnity Fund for Employees (FGTS) and unemployment insurance savings accounts in other countries: a comparative analysis of their effects on the labour market. Working Papers 164, International Policy Centre for Inclusive Growth.
URL https://ideas.repec.org/p/ipc/wpaper/164.html

Diamond, P. A., 1982. Wage determination and efficiency in search equilibrium. The Review of Economic Studies 49 (2), 217–227.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Schorfheide, F., 2016. Solution and estimation methods for dsge models. In: Handbook of Macroeconomics. Vol. 2. Elsevier, pp. 527–724.

Figueiredo, C., Francis, N., 2020. State-dependent adjustment cost and labor dynamics in recessions: Evidence using fgts policy in brazil. Tech. rep., University of North Carolina at

Chapel Hill.
URL http://dx.doi.org/10.2139/ssrn.3272449

Fornberg, B., 1996. A practical guide to pseudospectral methods. 1996. Cambridge Monographs on Applied and Computational Mathematics.

Hairault, J.-O., Langot, F., Sopraseuth, T., 2019. Unemployment fluctuations over the life-cycle. Journal of Economic Dynamics and Control 100, 334–352.

Holzmann, R., Vodopivec, M., 2012. Reforming Severance Pay : An International Perspective. Tech. rep.

Hopenhayn, H., Rogerson, R., October 1993. Job Turnover and Policy Evaluation: A General Equilibrium Analysis. Journal of Political Economy 101 (5), 915–938.

Loureiro, P. R. A., Araujo, R. A., de Souza, N. A., 2013. An evaluation of the brazilian informal labor market from 1995 to 2008. Journal of Economic Studies 40 (1), 71–87.

Mortensen, D., Pissarides, C. A., 1994. Job creation and job destruction in the theory of unemployment. Review of Economic Studies 61 (3), 397–415.

Mortensen, D. T., Pissarides, C. A., 1999. Unemployment responses to skill-biased technology shocks: the role of labour market policy. The Economic Journal 109 (455), 242–265.

Saint-Paul, G., 1995. The high unemployment trap. The Quarterly Journal of Economics 110 (2), 527–550.