# Labor Market Search, Informality, and On-The-Job Human Capital Accumulation

Matteo Bobba* Luca Flabbi† Santiago Levy‡ Mauricio Tejada§

June 10, 2019

**Web Appendix - Not For Publication**

## Contents




*Toulouse School of Economics, University of Toulouse Capitole, Toulouse France. E-mail: matteo.bobba@tse-fr.eu.

†Department of Economics, University of North Carolina, Chapel Hill USA. E-mail: luca.flabbi@unc.edu.

‡Brookings Institution, Washington DC USA. E-mail: santiagolevy4@gmail.com.

§Department of Economics (ILADES), Universidad Alberto Hurtado, Santiago Chile. E-mail: matejada@uahurtado.cl.

# A Data

## A.1 Sample Selection

Table A.1 shows descriptive statistics on relevant variables as we move from the original sample to the estimation sample. We focus on cross-sectional statistics because many relevant longitudinal statistics are not available in the overall sample. Specifically, transition probabilities are affected by attrition and employment durations are left censored; the only reliable statistics are the on-going unemployment durations, which we present in the Table. In Column 1, we report the raw data of the Mexican labor market survey (ENOE) for two stacked cohorts of male workers entering in the first quarter of the year 2013 and in the first quarter of the year 2014 who are interviewed for up to five consecutive quarters. In Column 2, the same average characteristics are displayed for the sub-set of male workers with secondary education, while in Column 3 we further restrict the sample to the remaining selection criteria detailed in Section 2.2 of the paper. Finally, in Column 4, we consider only those individuals from the sample of Column 3 that we can track longitudinally for five consecutive quarters. This is the estimation sample used throughout the analysis of the paper.

Table A.1: Descriptive Statistics Across Different Samples

| | (1) Original Sample | (2) Restricted Education | (3) Restricted All | (4) Balanced Panel |
|---|---|---|---|---|
| Proportions: | | | | |
| Formal Employees | .396 | .448 | .548 | .599 |
| Informal Employees | .300 | .302 | .287 | .265 |
| Self-Employed | .258 | .202 | .109 | .089 |
| Unemployed | .045 | .048 | .057 | .047 |
| Mean Wages: (Hourly) | | | | |
| Formal Employees | 33.1 | 26.3 | 24.2 | 23.9 |
| Informal Employees | 20.3 | 19.6 | 18.8 | 18.3 |
| Self-Employed | 30.6 | 31.1 | 24.2 | 23.0 |
| Mean Duration (months) | | | | |
| Unemployed | 1.83 | 1.65 | 1.59 | 1.56 |
| Sample Size: | | | | |
| First quarter | 184,209 | 62,071 | 23,882 | 4,936 |
| Overall | 542,378 | 183,825 | 64,732 | 24,680 |

NOTE: Wages and Incomes figures are reported in Mexican pesos (exchange rate: 10 Mex. pesos ≈ 1 US dollars in 2005). The formality status of the job is defined according to whether or not workers report having access to health care through their employers.

When compared to the nationally-representative figures of Column 1, the selected sample features a similar wage ranking across the three labor market states, albeit a higher proportion of formal workers and a lower share of self-employed individuals. In terms of education, Column 1 included workers that are both more skilled (College graduate or more) and less skilled (primary education only) than our sample. The first group dominates on average wages, in particular for formal employees, leading to higher average wages in Column 1 than in all the other Columns. There are some but not major differences in both labor market proportions and earnings between Column 3 and Column 4, indicating that the underlying determinants of sample attrition are largely idiosyncratic. We will return to this last point in some detail below (see point a). In conclusion, there are some important differences between our estimation sample and a nationally representative sample but these differences are mainly due to focusing on our specific education group. Within this education group, the main difference between a nationally representative sample and our estimation sample is the proportion of self-employed. It results from our focus on the "necessity" self-employment state as described in the paper (Section 2.1).

## A.2 Attrition Analysis

Given the definition of our estimation sample, the overall attrition rate in the sample of column 3 of Table A.1 is defined as the probability that a given individual has at least one missing survey round out five consecutive quarterly rounds. The overall attrition rate is 26%. We have checked whether this attrition rate varies systematically across the same labor market outcomes considered in Table A.1 by running a regression over the sample of column 3 and controlling for an indicator variable for whether or not the observation is eliminated due to attrition. Results reported in Table A.2 show that the OLS coefficient of the attrition indicator variable is significantly different from zero only on the labor market proportion of unemployed workers. It results from the fact that unemployed workers have a slightly higher attrition rate (31% instead of 26%). Taken together, we think the table shows attrition does not significantly alter the composition of the longitudinal sample used in the analysis.

Table A.2: Sample Attrition and Labor Market Outcomes

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Log Earnings (Hourly) | | | Labor Market Proportions | | | |
| | F | I | SE | U | F | I | SE |
| Attrition (1=yes) | -0.011 | 0.036 | 0.025 | -0.004 | -0.016 | 0.013 | 0.006 |
| | (0.018) | (0.028) | (0.081) | (0.002) | (0.011) | (0.013) | (0.009) |
| Mean Dep. Var. | | | | 0.055 | 0.575 | 0.274 | 0.096 |
| Number of Obs. | 9251 | 4418 | 1542 | 16099 | 16099 | 16099 | 16099 |
| Number of Clusters | 1861 | 1248 | 572 | 2756 | 2756 | 2756 | 2756 |

NOTE: OLS estimates. Fixed effects at the Municipality×Sector (4-digit) included but not reported. Standard errors clustered at the Municipality×Sector level are reported in parenthesis.

## A.3 Job Spells

We use two sets of information to infer status transitions within jobs. The first is the information about the formality status and the second is the information about job spells in the same job.

The first information is provided in the data set in the same way as the general information on formality status we use in the paper. As mention in Section 2.2, we identify the formal or informal status of the job depending on whether the employee reports having access to health benefits through their employers. This definition has strong foundation in the literature and it is reported for all the employees in each quarter of the ENOE sample.

The second information is provided through a question on the starting date of the job. In the last quarter of the survey, i.e. at the end of the observation window, individuals are asked when they started their current job. If the job started within the last two year, the precise starting date (month and year) is reported. If the job started more than two years before, the exact date is not asked but the fact that the job started more than two years before is recorded. When the precise starting date is recorded, we build continuous job spells by simply using the starting date and the observation of labor market status in each quarter.

When only the fact that the job started more than two years before is recorded, it is in principle equally straightforward to extract the information relevant for us. Since these agents are employed at the end of the observation period, since the observation period is one year and they declare to have started the current job more than two years before, they should have been at the same job for the entire observation period. In terms of the relevant information for status transition within job, we could then simply assign them to one job spell in the same job over the one year we observe them. However, some further investigation on the data has led to use a more conservative definition of same-job spell for this group. We have found that some individuals belonging to this group report episodes of search over the period (about 4%) and others report big differences in the economic sector they are working in (about 30% change sector at the NAICS 1-digit level). Both pieces of information are in principle consistent with working at the same job but they are not very credible to us. In the first case, we would have to assume that the worker is losing the job, has one or more episodes of search and then goes back exactly at the same job. In the second case, we would have to assume that the worker is employed in a firm operating in a given sector in one quarter and in another sector in the following quarter. While firms may operate in different sectors, we find it hard to believe that the same job in the same firm is transferred between sectors as different as the NAICS 1-digit.[1] We find it more plausible that the change of sector signals an actual change of job and that any job found after an episode of search is effectively a different job that the one held before searching. In conclusion, we have decided to assign the workers in this group as working in the same job only if, over two consecutive quarters of observation:

1. they are continuously employed, and

2. they operate in the same NAICS 1-digit sector.

Finally, for all the individuals who were either unemployed or self-employed at the time the question was asked and who were employed at some point in the previous year we consider the NAICS definition of the economic sector of the firm (at the 1-digit level). We have therefore assigned those individuals to the same job spell according to the rules [1] and [2] defined above.

[1] 1-digit sectors are very aggregated. In our sample they include the following 8 categories of economic activities, with the relative frequencies in parenthesis: mining and construction (13%), manufacturing (28%), trade, transportation, postal and warehousing services (31%), financial, insurance, real-estate, and business-related services (7%), education and health services (2%), cultural, temporary accommodation, and food and beverage preparation services (8%), other services, except government activities (7%), legislative, governmental and justice administration activities (4%).

# B Additional Material on the Model

## B.1 Wages

Wages are set by bargaining upon observing the labor market human capital $a_k$, the outside option $V_s(k,q)$, the match-specific productivity $x$ and the formality status posted by the firm $f$. We assume the axiomatic Nash-bargaining solution reported in equation (9). The resulting analytical expression for wages of employees hired formally and informally are:

$$\begin{aligned} w_1(x;k,q) = {} & \alpha(1+t)^{-1}\left[y(x,k) + \tau_{1,k}\sum_{k'=k+1}^{K}\max\{F_0(x;k',q), F_1(x;k',q), 0\}\Pr[k'|k]\right] \\ & + (1-\alpha)(1+\beta_1\tau t)^{-1}\Bigg[(\tilde{\rho}+\tau_{1,k})V_s(k,q) - \beta_1 b_1 \\ & -\tau_{1,k}\sum_{k'=k+1}^{K}\max\left\{\begin{array}{c}(1-f)E_0(x;k',q)+fE_1(x;k',q), \\ \max\{V_0(k',q),V_1(k',q)\}\end{array}\right\}\Pr[k'|k]\Bigg] \end{aligned} \tag{B.1}$$

$$\begin{aligned} w_0(x;k,q) = {} & \alpha\left[(1-c)y(x,k) + \tau_{0,k}\sum_{k'=k+1}^{K}\max\{F_0(x;k',q), F_1(x;k',q), 0\}\Pr[k'|k]\right] \\ & + (1-\alpha)\Bigg[(\tilde{\rho}+\tau_{0,k})V_s(k,q) - \beta_0 B_0 \\ & -\tau_{0,k}\sum_{k'=k+1}^{K}\max\left\{\begin{array}{c}(1-f)E_0(x;k',q)+fE_1(x;k',q), \\ \max\{V_0(k',q),V_1(k',q)\}\end{array}\right\}\Pr[k'|k]\Bigg] \end{aligned} \tag{B.2}$$

## B.2 Value Functions

Assume an individual searching in the labor market with human capital $a_k$, and potential self-employment income $q$. This agent will receive two possible shocks: meeting an employer and incurring human capital downgrading. The value of this state can be written in recursive form as follows:

$$\begin{aligned} (\tilde{\rho}+\lambda_s+\gamma_{s,k})V_s(k,q) &= (1-s)\xi + sq + \beta_0 B_0 \\ &+ \lambda_s\int_x \max\{(1-f)E_0(x;k,q)+fE_1(x;k,q), V_s(k,q)\}dG(x) \\ &+ \gamma_{s,k}\sum_{k'=1}^{k-1}\max\{V_0(k',q),V_1(k',q)\}\Pr[k'|k], \end{aligned} \tag{B.3}$$

where to simplify the notation we define $\tilde{\rho} \equiv \rho + \delta$. The first row represents the flow value, which is a function of the searching state (either unemployment or self-employment). When workers meet an employer, a match-specific productivity $x$ is drawn and they receive either a formal or informal employee offer. The worker then decides if accepting the offer or not by maximizing over the two possible value function. When the worker receives a human capital downgrading shock, he moves to the lower level $a_{k'}$ and decides if continue searching in the current state – being that unemployment or self-employment – or switch to the other state. Note that the formality status $f$ is endogenous and posted by the firm, as we show in Section 3.2.

When an agent is working as an employee, two shocks are possible: termination and human capital upgrading. The value of the employee state in recursive form is therefore:

$$
\begin{aligned}
(\tilde{\rho}+\eta_f+\tau_{f,k})E_f(x;k,q) &= w_f(x;k,q)+(1-f)\beta_0 B_0+f\beta_1 B_1[w_1(x;k,q)] \\
&+ \tau_{f,k}\sum_{k'=k+1}^{K}\max\left\{\begin{array}{c}(1-f)E_0(x;k',q)+fE_1(x;k',q),\\ \max\{V_0(k',q),V_1(k',q)\}\end{array}\right\}\Pr[k'|k] \\
&+ \eta_f\max\{V_0(k,q),V_1(k,q)\}
\end{aligned} \tag{B.4}
$$

The first row represents the flow value, which is a function of the wage and the formality–status–specific benefit (either $B_0$ or $B_1$). The second row shows that when the worker upgrades the labor market human capital, the formality status and the searching state are both updated optimally. This generates an interesting dynamic usually ignored in the literature: formality status may change *within* the same employer and job termination may occur *endogenously*. Finally, the third row shows that when the match is exogenously terminated, the agent has to go back to the searching state.

The value functions for the demand side of the market are as follows. Employers post vacancies and search for workers to fill them. The value of a filled job is consistent with the worker's side and defined as:

$$
\begin{aligned}
(\tilde{\rho}+\eta_f+\tau_{f,k})F_f(x;k,q) &= (1-f)\pi_0(x;k,q)-f\pi_1(x;k,q) \\
&+ \tau_{f,k}\sum_{k'=k+1}^{K}\max\{F_0(x;k',q),F_1(x;k',q),0\}\Pr[k'|k]
\end{aligned} \tag{B.5}
$$

The flow value is defined by the firm's profit, defined in equation (7) and (8). A filled job is subject to the same shocks we discussed for the worker's side: a termination shock $\eta_f$, which sends the firm back to a value of zero, and a human capital upgrading shock $\tau_{f,k}$. When the human capital upgrading shock hits, the employer enters a new negotiation with the worker and decides optimally the formality regime and whether or not keeping the worker.

### B.3 Numerical Solution and Simulation

We solve the model using value function iteration. We discretize the state space by using a grid of 100 equally spaced points in the interval $[0, 150]$ for both $x$ and $y$. The human capital distribution is already assumed discrete over 10 equally spaced points in the interval $[1, 5.5]$.

Since a match can change the formality status or a worker can decide to search for a new job when receiving an upgrading shock, all the value functions are dependent on each other and therefore the value function iteration is performed as a block. Specifically, we guess $V_s(k,q)$, $E_0(x;k,q)$, $E_1(x;k,q)$, $F_0(x;k,q)$ and $F_1(x;k,q)$ over the grid points in the state space and then we jointly iterate the Bellman's equations (B.3) to (B.5) (using the definitions of wages and profits) until convergence is achieved on these value functions. To approximate the integral in equation (B.3), we discretize the distribution $G(x)$ over the grid points of $x$ (using the midpoint intervals between the grid points as support) and we compute the expected value as in a discrete probability distribution.

Finally, to compute the probability of jumping to any higher level of human capital starting in a given $k$, we use a discretized truncated negative exponential distribution. In particular, let $m$ be the size of the jump in the human capital grid and assume that $m \sim Q_f(m;\nu_f)$ with $Q_f(\cdot)$ a negative exponential distribution with parameter $\nu_f$. Given that in our model $m \in [1, K]$, we define the truncated distribution as:

$$Q^T(m) = \frac{Q(m) - Q(0.5)}{Q(K+0.5) - Q(0.5)}$$

Then the discrete approximation of the probability of jumping $m$ steps can be computed as:

$$\Pr[m = j|k] = \begin{cases} Q^T(j+0.5) & j = 1 \\ Q^T(j-0.5) - Q^T(j+0.5) & j = 2, ..., K-k-1 \\ 1 - Q^T(j-0.5) & j = K-k \end{cases} \tag{B.6}$$

The maximum $K$ is adjusted to account for the number of steps that are left in the human capital support.

We simulate the model constructing labor market careers. To characterize all the optimal decisions involved in the dynamics of each career, we use direct comparisons between the solved value functions. Since we discretize the state space, we use linear interpolation to approximate the value functions and wages off the grid. We simulate 5,000 individual careers for 540 months. Each individual is assigned a potential self-employment income $q$ drawn from $R(q)$ and starts his career searching for a job with an initial human capital level equal to $a_1$. The lifetime duration is drawn from a negative exponential distribution with rate $\delta$. The optimal decision in the search state with respect to being unemployed or self-employed is made comparing $V_0(k, q)$ and $V_1(k, q)$ given $k$ and $q$. In the search state, individuals meet firms and receive downgrading shocks. The durations of these events are draws from negative exponential distributions with rates $\lambda_s$ and $\gamma_{s,k}$, respectively. If the meeting with a firm occurs first, a productivity $x$ is drawn and firms and individuals decide whether to complete the match and at what wage and formality status. If the match is realized, the individual leaves the searching state with a human capital level of $a_k$ and if not, the searching process continues. If a downgrading shock hits, a new search process starts for the same individual but with human capital $a_{k-1}$. While working as employees, individuals receive termination and upgrading shocks. As before, we simulate a competing risk model where the durations of these events are draws from negative exponential distributions. In this case, the rates are $\eta_f$ and $\tau_{f,k}$, respectively. If the termination shock arrives first, the individual starts a new search process with human capital $a_k$. On the contrary, if the upgrading shock arrives first, then the individual is upgraded to $a_{k'}$ and an optimal decision is made regarding whether to remain in the match and at what wage and formality status. This process continues until the arrival of the termination shock, which sends the agent back to search. Once the lifetime is complete, the individual dies and he is replaced by a new individual that starts his career with $q$ potential self-employment income and with $a_1$ human capital level.

As time passes in the simulation, the distributions of the labor market states and the human capital levels stabilize, which means that the model has reached the steady state invariant distributions. For estimation, we use a panel of five quarters extracted from a time window in which these distributions are in steady state.

To minimize the quadratic form (17) in the Simulated Method of Moments we use the downhill simplex (Nelder-Mead) algorithm. In each iteration of the simplex algorithm, the quadratic form is evaluated by solving and simulating the model, a procedure that is computationally very intensive. In particular, the simulation of career paths procedure is the most computationally intensive task in the process of estimation. On top of that, because the simplex method is a derivative free optimization algorithm, it requires a nontrivial number of evaluations of the quadratic form before obtaining convergence. To make the computation more efficient, the value function iteration is fully vectorized and the simulation procedure is parallelized. To have a sense of the computational burden, in a 28 core Intel(R) Xeon(R) CPU with 2.60 GHz one round of solution and simulations takes approximately 3 minutes (one quadratic form evaluation), while the complete estimation process takes roughly 500 evaluations of the quadratic form and around 26 hours.

Finally, the weighting matrix is constructed using the bootstrapped variance of the chosen moments in the quadratic form, being the bootstrap samples random samples (with replacement) of individuals of the size equal to the number of total individual in the database. Additionally, the standard errors of the estimators are also calculated using these bootstrapped samples. We use the estimated parameters in the simplex algorithm in each bootstrap iteration.

# C Additional Material on the Estimation

## C.1 Monte Carlo Experiment

To asses the reliability of our estimator we performed the following Monte Carlo procedure. In the first step, using the estimated parameters of the model, we solved and simulated the model to generate a 5 quarter balanced panel of synthetic data. In the second step, we applied our estimation procedure to the synthetic data. In this step, we keep the same values of the original estimation for all the convergence criteria of the simplex algorithm and the initial values of the minimization process. Due to time constraints we performed just one replication of this Monte Carlo procedure. Table C.1 compares the point estimates obtained by applying our estimation procedure on the original data with the point estimates obtained by applying the same estimation procedure on the synthetic data. We find them close enough to lend some credibility to our estimation method.

Table C.1: Monte Carlo Experiment

| Parameter | Real Data | Synthetic Data |
|---|---|---|
| $\lambda_{\{s=0\}}$ | 0.5051 | 0.5072 |
| $\lambda_{\{s=1\}}$ | 0.0782 | 0.0810 |
| $\eta_{\{f=0\}}$ | 0.0573 | 0.0594 |
| $\eta_{\{f=0\}}$ | 0.0317 | 0.0325 |
| $\mu_x$ | 1.6835 | 1.6902 |
| $\sigma_x$ | 1.0099 | 1.0098 |
| $\mu_q$ | 1.5733 | 1.5805 |
| $\sigma_q$ | 0.9464 | 0.9462 |
| $\gamma_{\{s=0\}}$ | 0.1617 | 0.1588 |
| $\gamma_{\{s=1\}}$ | 0.0472 | 0.0455 |
| $\tau_{\{f=0\},1}$ | 0.0460 | 0.0455 |
| $\tau_{\{f=0\},2}$ | -2.9775 | -2.9461 |
| $\tau_{\{f=1\},1}$ | 0.0576 | 0.0586 |
| $\tau_{\{f=1\},2}$ | -1.6230 | -1.6400 |
| $c$ | 0.0514 | 0.0496 |
| $\nu_{\{f=0\}}$ | 0.4958 | 0.4960 |
| $\nu_{\{f=1\}}$ | 1.3988 | 1.4022 |
| $\xi$ | -8.9533 | -9.0456 |

## C.2 Calibrated Parameters

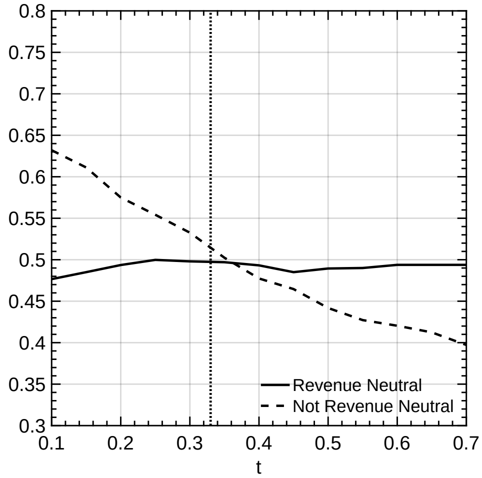
As we discuss in 4, some of the parameters describing the Mexican institutional setting are fixed to values derived from the current legislation, from aggregate data and from previous contributions. The complete list of parameters, together with values and sources, is reported in Table C.2.

Table C.2: Fixed Parameters

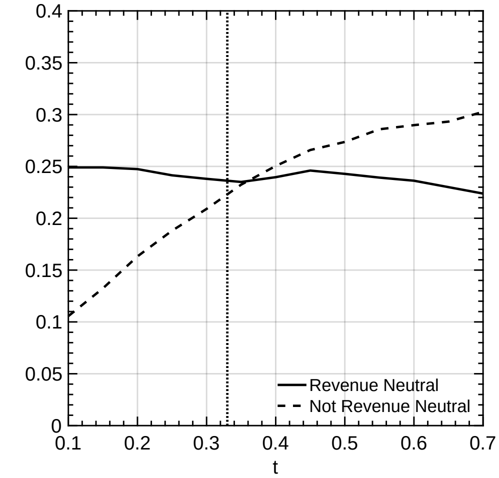
| Parameter | Value | Source |
|---|---|---|
| $\alpha$ | 0.5000 | Symmetric Bargaining case [?] |
| $\beta_0$ | 0.9082 | ? |
| $\beta_1$ | 0.6705 | ? |
| $B_0$ | 4.2700 | Updated from ? |
| $\phi$ | 0.5500 | ? |
| $t$ | 0.3300 | ? |
| $b_1$ | 4.5470 | Based on average observed wages (see equation 16) |
| $\rho$ | 0.0500 | ?? |
| $\delta$ | 0.0013 | Based on average life of 65 years |

# D Additional Material on the Policy Experiments

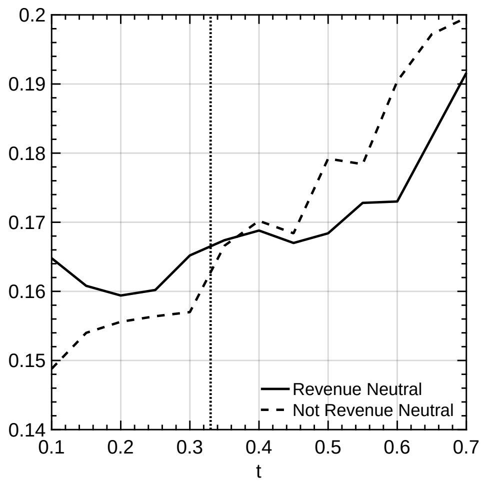
Figure D.1: Impacts of Policy 1 – Employment Status



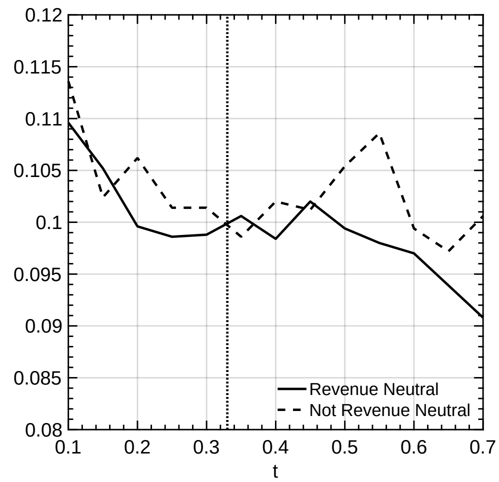

(a) Share of Formal Employees




(b) Share of Informal Employees



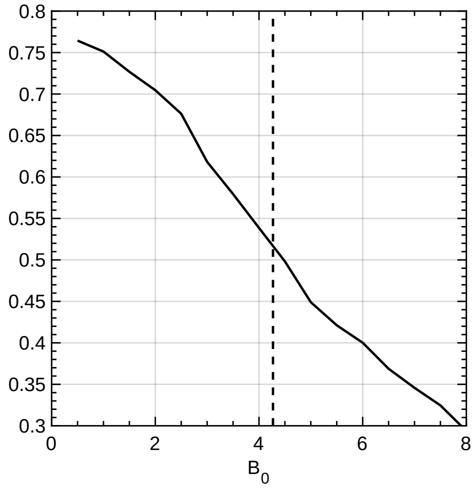

(c) Share of Self-Employed



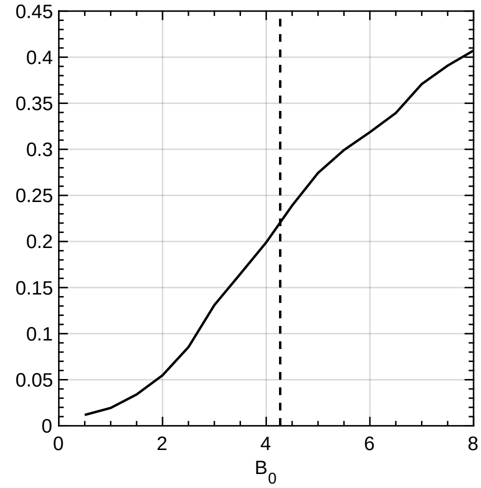

(d) Share of Unemployed

Note: Simulated samples of 5,000 worker-level observations for each quarter based on the estimates reported in Table 3. The vertical lines are set at the institutional values for the Mexican labor market in 2013-2014. See Table C.2 for details.

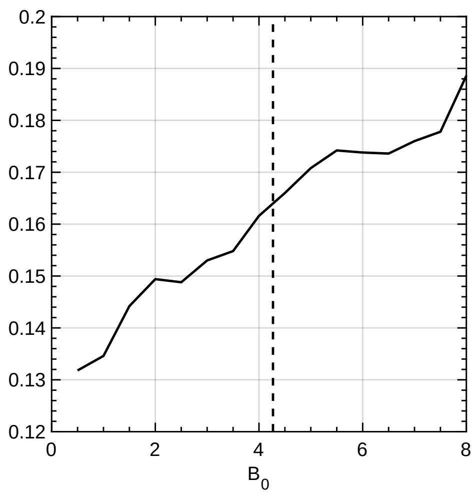
Figure D.2: Impacts of Policy 2 – Employment Status



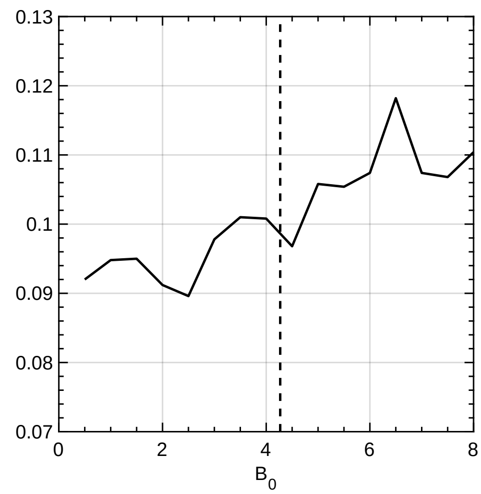

(a) Share of Formal Employees




(b) Share of Informal Employees



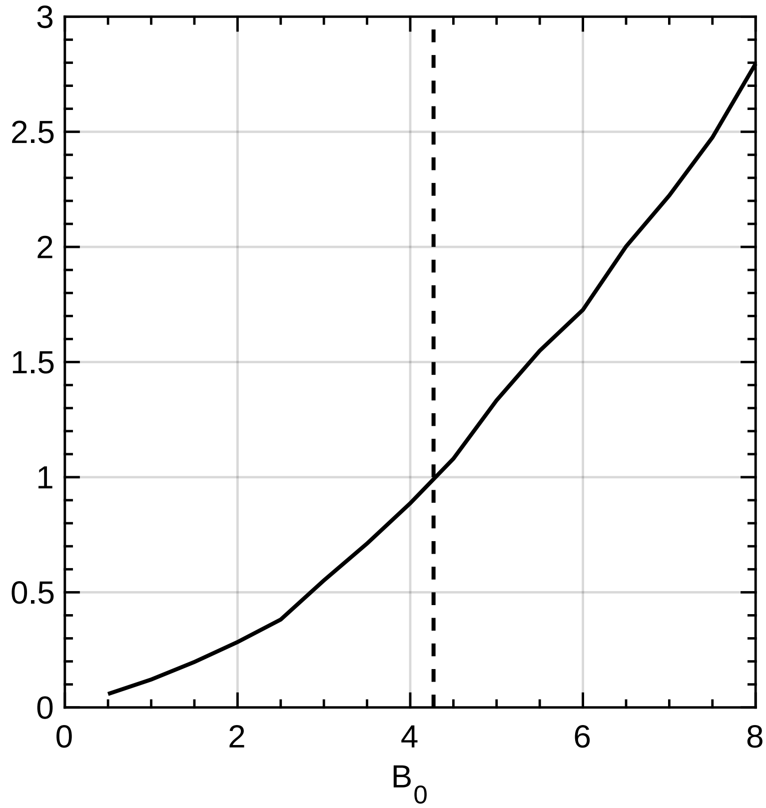

(c) Share of Self-Employed




(d) Share of Unemployed

Note: Simulated samples of 5,000 worker-level observations for each quarter based on the estimates reported in Table 3. The vertical lines are set at the institutional values for the Mexican labor market in 2013-2014. See Table C.2 for details.

Figure D.3: Impacts of Policy 2 – Fiscal Cost




NOTE: Table Reports the ratio (Total Cost for $B_0$)/(Value of Production). Benchmark = 1. Simulated samples of 5,000 worker-level observations for each quarter based on the estimates reported in Table 3. The vertical lines are set at the institutional values for the Mexican labor market in 2013-2014. See Table C.2 for details.