Additional Material for:
The Financial Channels of Labor Rigidities
Evidence From Portugal

Edoardo Maria Acabbi* Ettore Panetti† Alessandro Sforza‡

April 2024

A Data description and cleaning

A.1 Labor market data: Quadros de Pessoal

The *Quadros de Pessoal* (henceforth QP) is a longitudinal matched employer-employee dataset, containing detailed data at the workers' and firms' level on employment composition for the firms and individual worker characteristics. The data are collected and managed by the Ministry of Labour and Social Solidarity, that draws on a compulsory annual census of all the firms employing at least one worker at the end of October each year. It does not cover the public administration and non-market services, whereas it covers partially or fully state-owned firms, provided that they offer a market service. The dataset covers approximately 350,000 firms and 3 million employees per year. In 2010 the structure of the survey was reformed and the QP was incorporated into the *Relatório Único*, an integrated reporting system to enable employers to easily provide more extensive information on workers to the Ministry. As a consequence, some very small entrepreneurial firms were exempted from filing compulsorily the questionnaire, which is why after 2009 the coverage of QP is less complete than in previous years.[1]

*Corresponding author, Universidad Carlos III de Madrid, edoardomaria.acabbi@uc3m.es.
†University of Naples Federico II, CSEF, SUERF, and UECE-ISEG.
‡University of Bologna, CEPR and CESifo.

[1] Despite this inconvenience, we use the firms' balance sheet dataset, the *Central de Balanços*, which covers all non-financial corporations in Portugal, to correctly disentangle firms' failures. This also means that in our analysis the "survivors" sample is not necessarily a balanced sample.

The dataset is available at the Bank of Portugal from 1982 to 2013, and is hierarchically made up by a firm-level dataset, an establishment-level dataset and a worker-level dataset. The firm level dataset contains information on the firm location (from regional to very narrowly defined parish level, which roughly corresponds to a neighborhood, industry of operation (CAE rev. 2.1 until 2006 and CAE rev. 3, based on NACE-Rev. 2 Statistical classification of economic activities in the European Community), total employment, total sales, ownership structure and legal incorporation. Analogous information is available on the establishment-level dataset.

The worker level dataset provides detailed information on worker characteristics and contracts. Information included comprehends workers' gender, age, detailed occupational code (the *Classificação Nacional de Profissões* (CNP94) up to 2009 and the *Classificação Portuguesa das Profissões* (CPP2010) from 2010 onward, which is based on ISCO08 International Occupational Classification Codes), detailed educational level, qualification within the firm (managerial qualification, specialized workforce or generic workers, besides trainees). At the contract level it is possible to know the precise hiring date, the kind of contract (various typologies that generally define the contract as fixed-term or open-ended), the hours arrangement (full-time versus part-time), the effective number of hours worked, and information on the compensation. More specifically, for each worker it is possible to obtain information on the base pay, any extra paid in overtimes or other extra-ordinary payments and other irregularly paid components. In contrast, there is no information on social security contributions. We winsorize the extreme 0.5 percent tails of the distribution of wages.[2]

The unique worker identifier is based on the workers' social security number, and given the extensive work on the part of the Ministry to control and certify the quality of the data in this administrative dataset, the coverage and reliability of the data is quite high (except for the discrete break in coverage for only a subset of firms in 2010 due to the new reporting requirement in the *Relatório Único*. Given that other datasets in our analysis cannot cover the same time-span, we only focus on years from 2005 to 2013 (but potentially control for observables up to 2003 in some empirical exercises).

We use the QP to extract information regarding wage policies at the firm level. This is the dataset we use in order to compute AKM firm level fixed effects, which describe what component of workers' compensation is firm specific and pertains to average firm wage policy. We compute AKM fixed effects through an AKM

[2]As regards the qualification categories, the Portuguese Decree-Law 380/80 established that firms should indicate the qualification level as in the Collective Agreement. If this is not available, firms should select the qualification level of the worker. These categories are based on the degree of complexity of tasks that the worker performs within the firm (from more basic, routine tasks to more discretionary managerial ones). The categories are defined within a 9 levels hierarchy, that we simplify into three broad categories.

regression, which is a wage regression at the worker level on firm and worker fixed effects and other workers' characteristics. The firm fixed effect captures overall generosity of payments while the individual fixed effect should capture unobservable skills. In our AKM specifications we control for sex, a third polynomial of age and educational categories dummies, all variables present in QP itself. We run the AKM regressions on worker level data from 2003 to 2008.

A.2 Firm level financial statement data: Central de Balanços

The *Central de Balanços* (henceforth CB) is a firms level balance-sheet and income statements database, managed by the Bank of Portugal. It consists of a repository of yearly economic and financial information on the universe of non-financial corporations operating in Portugal from 2005 to 2013. It includes information on sales, balance-sheet items, profit and loss statements, and cash flow statements (after 2009) for all private firms in Portugal.The CB builds on the *Informação Empresarial Semplificada*, an administrative firms' balance-sheet dataset managed by the Ministry of Finance and Public Administration. The Bank of Portugal obtains the data from the Ministry and performs extensive consistency checks to guarantee that the data are reliable and consistent over the years.
The dataset in its present form covering the universe of firms is based on information reported in the starts in 2006, even if almost the entirety of firms already existing in 2005 provided balance sheet data for that year as well together with the 2006 filing. For this reason, we actually have a very high coverage of firms' balance sheets for 2005 as well. Before 2005 the CB maintained by the Bank of Portugal was actually a survey only for the biggest firms in the country. However, given the substantially lower coverage of the population of firms before 2005, we do not rely on that data.
After 2009, in order for the data to comply with international accounting standards, there has been a major overhaul of the variables definitions in the dataset, from the *Plano Oficial de Contabilidade* (POC) to the *Sistema de Normalização Contabilística* (SNC). In all our computations, unless otherwise noted, we have personally gone through a variables' harmonization process, in collaboration with the statistics department managing the administrative datasets for researchers at the Bank of Portugal, BPLim, to guarantee comparability across periods.

The dataset contains a great amount of information on firms' balance sheets and income statements, even if the harmonization process between 2009 and 2010 makes it at time difficult if not impossible to keep consistent records for all balance sheet variables in the dataset. We use the dataset to obtain information on total assets, fixed assets, current assets total debt (not just bank debt) and interest expenditures, cash-flow and capital expenditures (after 2009), cash balances, exports and export status, trade credits, debt towards suppliers, inventories, return on equity, assets and sales, salaries, total employee related , revenues, costs and breakdowns (among which intermediate inputs, materials and services), profits. We computer value added from this dataset by adding back employee related expenditures to the firm EBITDA (which should correspond to subtracting expenditures on intermediate goods from total sales).

Given the dataset time-consistent coverage of firms operating in Portugal, we use it to identify firm exits as well. The procedure to identify a firm exit combines different criteria. Firstly, we rely on the CB on categorization of whether a firm is active, suspended activity or closed down. Secondly, we flag all the cases in which the firm will end up having 0 employees the next year but does have a positive number of employees in a given year. Thirdly, we actually check whether a firm disappears from the dataset in any given year that is not 2013 and does not re-appear at any time (and does not simply have, consequently, a gap in the data). Lastly, we label as exits the instances in which a firm disappears for more than two years, as it is likely that if the identifier reappears later it has just been reassigned to another firm (an assumption that seems to be validated by the observation that when such instance takes place the firm seems different in terms of size and sector between the two periods). In all the cases we select the criterion of exit, in case a firm matches more than one at different points in time, by looking at the case in which the firm "closed down" with the highest number of employees or, if ties are not resolved, with the lowest EBITDA.

A.3 Credit exposure level dataset: Central de Responsabilidades de Crédito

The *Central de Responsabilidades de Crédito* (henceforth CRC), is the credit registry of the Central Bank of Portugal. The dataset features available for our period of analysis (up to 2013) features bank-firm exposures above EUR50 by the universe of Portuguese credit institutions at the monthly level. The dataset does not contain credit exposure by foreign banks towards Portuguese firms, but can obviously contain credit from Portuguese banks to foreign owned firms residing and operating in Portugal.[3]

The dataset is regularly employed for supervisory purposes, and by the credit

[3]We do not believe that this fact could be a source of significant bias in any of our results, as the Portuguese economy mostly features relatively small and arguably bank-dependent firms, and for the biggest firms it is more likely for them to access directly debt markets instead of creating ties with foreign banks. Most foreign banks, moreover, operate Portugal incorporated subsidiaries in the country, the credits of which would regularly appear in the CRC.

institutions themselves to obtain information on potential debtors. It contains detailed information on the number of credit relationships, the corresponding amounts and the kind of exposure: short- and long-term, credit granted but still not materialized (potential), credit overdue, written-off or renegotiated. From 2009 onwards, but unfortunately not before, it is possible to obtain information closer to loan-level (i.e. it is possible to keep track of exposures which consist into the sum of loans with very detailed similar characteristics instead of seeing an aggregate number by kind of coarsely defined exposure) and more details about the exact maturity of each exposure and the collateral posted by each firm, if any (real collateral or guarantees, fraction of the value of the loan backed by it). Given the nature of our analysis and the period of interest, we mostly focus on obtaining a consistent representation of the information available in the dataset before 2009. For our analysis and given the time frequency in other data sources we average debt exposures at the yearly level. We use "regular" credit in our specifications as measure of credit, which corresponds to credit in good standing and in use by the firm. Credit is defined as short-term if the maturity is below 1 year or it is a credit line with undefined maturity (post-2009 data) or is categorized as commercial, discount or other funding short-term pre-2009. We group together short-term loans, credit lines with defined short-term maturity and credit lines with undefined maturity because the latter category of credit lines comprehends all those exposures that, once withdrawn by the customer, should undergo renegotiation with the bank in order to be rolled-over. This feature makes them very liquid instruments that, similarly to short-term loans, is subject to short-term credit rates volatility and rollover risk. Credit lines always constitute above 3/4 of short-term credit as we define it. Long-term credit is thus obtained as the remainder in regular credit.

A.4 Banks balance sheet dataset: Balanço das Instituições Monetárias e Financeiras

The "Balanço das Instituições Monetárias e Financeiras" (henceforth BBS) is the balance-sheet dataset for credit institutions that we employ. It is a proprietary dataset of the Bank of Portugal with the balance sheets of the universe of financial monetary institutions operating in the country. The dataset is utilized by officers of the bank in order to monitor the health of financial monetary institutions operating in the country and the overall stability of the system. In the dataset, for each balance-sheet item (liability or asset) it is possible to see which is the kind of counterparty involved (i.e. the kind of institution, government, private or non-governmental body, creditor or debtor), the maturity of the item in question if relevant (time deposits, on demand deposits, interbank long-term or short-term exposures) and the nationality of the counterparty (extra-EU or each EU country separately). The data are reported at the monthly level.

The measure of interbank funding which is the basis of our instrument is computed from this dataset as the ratio of the average (yearly) short-term foreign interbank borrowing by the bank over total assets. Foreign short-term interbank borrowing is computed as the sum of short-term deposits with maturity up to 1 year and repos where the counterparty if a foreign financial institution (obviously not a central bank).

In matching the BBS and the CRC, we also took care of harmonizing and making bank definitions consistent across datasets given the existence of many mergers and acquisitions in the Portuguese banking system during the period. Each M&A event between 2000 and 2013 (for institutions with at least 1 percent of total credit in a given month) was taken into consideration in order to make sure that credit flows across institutions were rightly accounted for, and definitions of bank codes across datasets and across time were consistent.

A.5 Banks balance sheet dataset: Sistema Integrado de Estatísticas de Títulos

The *Sistema Integrado de Estatísticas de Títulos* (henceforth SIET) is a proprietary dataset of the Bank of Portugal. It includes debt securities (i.e. banknotes, commercial papers, bonds, etc.) with maturity both short term (up until 1 year) and long term (more than 1 year), and capital (i.e. shares and other means of participation) but neither derivatives nor REPOs. For both debt securities and capital, SIET collects data about emissions and portfolio holdings. For emissions, SIET collects flows and stocks relative to national issuers, on a title-by-title and issuer-by-issuer bases. For portfolio holdings, SIET collects flows and stocks on an investor-by-investor and title-by-title basis. Through SIET we obtain holdings of sovereign debt, or more in general any government-issued debt instrument held by banks on their balance sheet.

A.6 Commuting zone definitions

Given the relevance of the concept of commuting zone, especially for the analysis of labor market reallocation, we obtained data on the definition of commuting zones for Portugal from Afonso and Venâncio (2016).

A.7 Labor market data: Occupational Information Network

Given the availability of definitions of occupations at the worker level in the QP, we were able to obtain occupation characteristics through the Occupational Information Network (O*NET) database. The O*NET database is a widely used database in labor economics and is the primary source of data in the United States for categorization of occupation characteristics. It is based on the combination of the analysis of responses to questionnaires on occupations administered to sampled employers and employees, and is updated four times a year with new data or updates to current categorizations.

We used O*NET in order to create indexes on job categorizations in terms of education, experience and training requirements. For each occupation a categorization is provided regarding the level of experience required (with possible scores ranging from 1 to 12, from less than high-school to post-graduate level), the level of previous experience (from 1 to 11, from none to more than 10 years), the level of on-site training (classes, courses, instructions sessions organized by the employer) or on-the-job training (that is, work carried out under the supervision of more experienced workers) required to being able to carry out the required tasks (from 1 to 9, from a short demonstration to years of training). Moreover, we also extracted for each occupation the categorization of the "job zone" (with a score from 1 to 4 in ascending order of "sophistication" of required vocational preparation levels), which is a further categorization created by expert O*NET analysts that combines all the previous four categories in a unique index. We obtained a separate occupational index as well for each category by averaging the scores, taking into account the frequency of each score for each response.

In order to combine the data, we first worked on making profession definitions consistent across time in our dataset, and then merged our occupational code to O*NET through a ISCO08-ONETSOC10 crosswalk. Given the change in occupational codes from the *Classificação Nacional de Profissões* (CNP94) to the new *Classificação Portuguesa das Profissões* (CPP2010) in 2010 in order to update the categorization and making in compliant to the *International Standard Classification of Occupations* (ISCO2008) categorization, we created a crosswalk based on the frequency of cross-occupational code changes from 2009 to 2010 in the QP within the same firms. We used the cross-walk in Hardy et al. (2018) to merge our ISCO08 codes to (ONET)SOC10 (Standard Occupational Codes). We then averaged all the occupational scores and indexes obtained from ONET across occupations in order to obtain a time consistent 3-digits ISCO08 occupational categorization.[4]

[4]The fact that obviously the occupational categorizations are neither bijections nor injections across sets made it difficult in some cases to reassign the occupational codes. We tried to use the official crosswalk at first, but noticed that it created very big discontinuities for the frequency of observation

We used O*NET version 23.3, and more specifically the education, training and experience files.[5]

B Production function estimation

B.1 Productivity and output elasticities estimation

For the estimation of output elasticities, markups and ultimately revenue total factor productivity (TFPR) we use different methodologies. First of all, we consider a three-factors of production gross output (y) function, where factors are labor (l), physical capital (k) and an intermediate input (m). We consider both a simple Cobb-Douglas specification where the elasticity of substitution among the factors of production is restricted to be 1 and a translog specification, which relaxes the above assumption. The (log-)production function is thus expressed as a function of log-inputs as:

$$y_{i,t} = f(l_{i,t}, k_{i,t}, m_{i,t}) + \omega_{i,t} + \varepsilon_{i,t} \tag{1}$$

where $f(l_{i,t}, k_{i,t}, m_{i,t})$ is

$$\beta_l l_{i,t} + \beta_k k_{i,t} + \beta_m m_{i,t} \tag{2}$$

in the Cobb-Douglas case, and:

$$\beta_l l_{i,t} + \beta_k k_{i,t} + \beta_m m_{i,t} + \sum_{x \in \{l,k,m\}} \beta_{xx} x_{i,t}^2 + \sum_{j \in \{l,k,m\}, j \neq x} \quad \sum_{x \in \{l,k,m\}} \beta_{jx} j_{i,t} x_{i,t} \tag{3}$$

in the translog case. ω in the equation represents the firm's level of technical efficiency (or total factor productivity, TFP).[6]

In our estimation gross output is measured as total firm sales (coming from QP when available and using CB firm revenues, which correspond to the QP definition of sales, for all other firms), deflated by 2-digit industry gross output deflators. Labor is measured as the firm wage bill (coming from QP when available or using CB total

of some professions. We noticed on the other hand that within firms changes in occupational codes seemed to be very consistent, and as such a more valid "revealed preference" categorization on the part of employers of their employees actual occupation. We then decided to limit ourselves to a 3-digits categorization in order to have a meaningful number of workers for occupation, and in order to minimize the inconsistencies in the cross-categorizations of occupational codes between CNP and CPP.

[5] https://www.onetcenter.org/dictionary/23.3/excel/education_training_experience.html

[6] The CES production function is a specific case of the general translog production function, and can be obtained by applying a second order Maclaurin approximation (which implies the parameterization of the Cobb-Douglas case as point around which the approximation is performed) to the log of $y = \left(\sum a_x x_i^\rho\right)^{\frac{\nu}{\rho}}$. The CES entails some specific parameters restrictions with respect to an unconstrained translog specification, which should thus be considered as a more general specification.

salaries for all other firms), which differently from total headcount (or full-time equivalent count) partially accounts for labor quality, and is deflated by the consumer price index. The intermediate input is the sum of the cost of intermediate goods and supplied services, deflated by 2-digit industry intermediate inputs deflators. For physical capital we use a capital series that we constructed following the perpetual inventory method (PIM) in the baseline specifications or the book value of (net) fixed assets (both tangible and intangible). In the latter case the book value of (net) fixed assets is deflated by 2-digit industry capital goods formation deflator.[7] For the PIM on the other hand we estimate the following equation:

$$K_{i,t} = (1 - \delta_{i,t})K_{i,t-1} + \frac{I_{i,t}}{def_t} \tag{4}$$

at the firm level. Instead of using the book value of yearly depreciation for fixed assets, we use a level of 7 percent for all firms.[8] From 2009 onwards we can measure directly firm level capital investment form the cash-flow statement (unavailable for earlier years) as the total yearly capital expenditure in both tangible and intangible capital formation. For the other years, or when the variable is missing, we use the variation in book fixed assets, deflated by the yearly capital goods formation deflator, as a measure of investment. We take the earliest year available level of fixed assets, deflated by the industry capital goods formation deflator, as starting value for the series. For incumbents firms in the dataset, the earliest year is 2005, and their starting value of real capital is thus just an approximation. We use the results based on PIM capital as our baseline.[9]

The estimation is carried out yearly, for all firms in the CB from 2005 to 2013, at a level of aggregation that is close to the 2-digit industry level.[10] For the estimation of output elasticities we remove from the dataset firms with a revenue labor share lower than or greater than 1 percent, firms with a revenue material labor share lower than 10 percent or greater than 1, and the firms with a sum of labor and material shares above 1.2. We also drop the lowest and highest 1-percent quantiles of labor and material

[7]All the price indexes for Portugal, apart from the CPI, are obtained from the OECD STructural ANalysis Database (STAN) (http://www.oecd.org/sti/ind/stanstructuralanalysisdatabase.htm).

[8] 7 percent is less than the maximum level of depreciation tax deduction that firm would get by deflating capital the most each year. For this reason, even if imperfect, it is a plausible measure of yearly depreciation. Given that we are unable to decompose in a time-consistent way the subcomponents of capital formation, the approximation is necessary.

[9]Our results are qualitatively insensitive to the measure of fixed capital that we used for the estimation of output elasticities and productivity, and in many cases are also almost quantitatively indistinguishable.

[10]Given that there is a change in industry definitions in QP (see Appendix A) and some subgroups are small, we aggregate some of the subgroups. The resulting industry definitions are conceived to be time-consistent across the different CAE versions of industrial definitions.

shares. We are left with 275,093 unique firms and 139,735 firm-year observations on average. Firm level (log) TFP is calculated as the residual from the estimation of the production function according to the various specifications. The estimated residual, the productivity shock, can be written as

$$\xi_{i,t} = \hat{\omega}_{i,t} + \nu_{i,t} = \omega_{i,t} + \varepsilon_{i,t} \tag{5}$$

where $\hat{\omega}$ represents the "transmitted" component of productivity (that is, the one that the firm takes into account while making input decisions) and $\nu_{i,t}$ should represent an unexpected shock. Given that the residual $\nu_{i,t}$ in the estimation might also arise because of any measurement error in output, inputs and prices, we calculate productivity either as the full residual from the production function estimation, or the residual

$$\hat{\omega}_{i,t} = \hat{y}_{i,t} - \hat{f}(l_{i,t}, k_{i,t}, m_{i,t}) \tag{6}$$

where $\hat{y}_{i,t}$ is obtained as the estimated gross output from a regression of output on a third order polynomial of all inputs of production. The latter form aims to eliminate any component of the realization of gross output that appears not to be related to the planned input choice, and remains unexplained by it, thus limiting the concern on the influence of measurement error. In the main text we show results based on this latter measure of productivity, but our results are qualitatively unchanged regardless of the measure we use. We use the full residual, as standard in the literature, for the productivity decomposition.[11]

The estimation of output elasticities and productivity generally presents problems related to the nature of input choice itself. On the one hand, input choice is likely to be very strongly correlated with (expected) productivity itself, and as such the direct estimation of the log production function by OLS would very likely be subject to biases given endogeneity determined by simultaneity.[12] On the other hand, there is generally an implicit selection bias for the firms observed in the dataset, given that more productive firm tend to be more resilient in normal times.

We address the first issue by following the literature in industrial organization on the identification by means of proxy variables (Olley and Pakes, 1996; Levinsohn and Petrin, 2003).[13] This methodology consists into substituting unobserved productivity

[11]See Petrin and Sivadasan (2013) for a similar exercise.

[12]Given Portugal's labor market institutional features, it would not look unreasonable to consider labor as a quasi-fixed input in production, with a greater degree of flexibility than capital but still less flexible. Our method of estimating labor elasticity is consistent regardless of this matter, but if labor not a fully flexible input in production it cannot be utilized to estimate firms' markups.

[13]The proxy-variable approach is the most frequently used in the industrial organization literature. Alternatives are fixed effects, first order conditions, the dynamic panel approach or the use of plausible instruments.

in the production function by a proxy variable, a choice variable assumed to have an invertible mapping with productivity itself. In our case, we use the intermediate input as the proxy variable (as in Levinsohn and Petrin (2003)).

The estimation is subdivided in two stages: in the first stage output is non-parametrically regressed on the inputs (and importantly, the proxy variable, which is an input in our case), in order to retrieve expected output and an estimate of the residual:[14]

$$y_{i,t} = \phi(l_{i,t}, k_{i,t}, m_{i,t}) + \varepsilon_{i,t} \tag{7}$$

We follow De Loecker and Warzynski (2012) and Ackerberg et al. (2015) in the estimation of all the relevant output elasticities at the second stage, which allows for consistent estimation even in presence of dynamic effects of the labor choice on the other inputs. The second stage estimation relies on the assumption that productivity at the firm level follows a Markov process:

$$\omega_{i,t} = g(\omega_{i,t-1}) + \eta_{i,t} \tag{8}$$

For a given guess of parameters **fi** one can obtain an estimate of productivity:

$$\hat{\omega}_{i,t}(\mathbf{fi}) = \hat{\phi} - (\hat{\beta}_l l_{i,t} + \hat{\beta}_k k_{i,t} + \hat{\beta}_m m_{i,t}) \tag{9}$$

for the Cobb-Douglas case of:

$$\hat{\omega}_{i,t}(\mathbf{fi}) = \hat{\phi} - \left(\hat{\beta}_l l_{i,t} + \hat{\beta}_k k_{i,t} + \hat{\beta}_m m_{i,t} + \sum_{x \in \{l,k,m\}} \hat{\beta}_{xx} x_{i,t}^2 + \sum_{j \in \{l,k,m\}, j \neq x} \sum_{x \in \{l,k,m\}} \hat{\beta}_{jx} j_{i,t} x_{i,t} \right) \tag{10}$$

in the translog case. One can thus non-parametrically regress $\hat{\omega}_{i,t}$ on its own lag and obtain the estimated innovation to productivity $\nu_{i,t}(\mathbf{fi})$). It is then possible to estimate all the output elasticities and subsequently TFP by GMM relying on moment conditions of the form:

$$\mathbb{E}[\eta_{i,t}(\mathbf{fi}) z^j] = 0 \qquad j \in \{l, k, m\} \tag{11}$$

in the Cobb Douglas case and

$$\mathbb{E}[\eta_{i,t}(\mathbf{fi}) z^j] = 0 \qquad j \in \{l, k, m\} \tag{12}$$

$$\mathbb{E}[\eta_{i,t}(\mathbf{fi}) z^j z^h] = 0 \qquad j, h \in \{l, k, m\} \tag{13}$$

in the translog case. The z variables are instruments for the various inputs. Given

[14] We use a third order polynomial of inputs in this first stage regression.

the standard assumptions on input dynamics, k can be a valid instrument of itself, whereas we use lags of labor and intermediate inputs as instruments, and according interactions for higher order terms.[15,16,17]

We address the problem of possible selection bias of firms into the dataset by trying to control for the probability of survival in the law of motion of productivity, as suggested by Olley and Pakes (1996). We actually augment the estimation of Equation (8) by adding the estimated survival probability obtained by fitting a probit model on year dummies and input levels.[18]

We compute productivity and elasticities for robustness by estimating the Cobb-Douglas and translog productions functions by straight OLS as well, adding year fixed effects to the estimation. All results in the main body of the paper are qualitatively (and quantitatively) robust to these different estimation procedures.

In the Cobb-Douglas case the estimated coefficients for each input are also output elasticities, which are consequently fixed within each industry (the Cobb-Douglas specification does not admit any variation in input revenue shares and elasticities across firms within the same estimation sample). In the translog case, on the other hand, the elasticity of substitution across any inputs is not restricted to be 1 and elasticities can vary depending on each firms' input mix utilized. For any input x, given the other two inputs j and h, the estimated output elasticity can be obtained as:

$$\hat{\theta}^{x}_{i,t} = \hat{\beta}_x + 2\hat{\beta}_{xx}x_{i,t} + \hat{\beta}_{xj}j_{i,t} + \hat{\beta}_{xh}h_{i,t} \tag{14}$$

Tables 1 and 2 show average estimates of input elasticities using all the different estimation methodologies, and with different measures of the capital input. Reassuringly, the estimated elasticities and markups are in line with the recent studies performing similar estimations (Blattner et al. (2023) for Portugal, Fonseca and Van Doornik (2022) for Brazil and Lenzu and Manaresi (2018) for Italy).

[15] In order for lags of wage bill and intermediate inputs to be valid instrument for their respective current values, one would need the prices to be correlated over time, an assumption that is quite plausible and surely confirmed in our data as regards the dynamics of wages.

[16] In the Cobb Douglas case we also add orthogonality conditions for the lag of capital and the second lag of intermediate inputs. Given the amount of parameters to estimate and the computing time required for the procedure, we do not add overidentifying restrictions in the translog case.

[17] If labor was indeed a dynamic input, the estimation of its elasticity would remain consistent anyway, as the orthogonality condition would a fortiori be valid for its lag.

[18] We carried out a the same procedure by augmenting Equation (8) with the estimated failure probability as in Antunes et al. (2016), but did not notice any material difference in final outcomes.

B.2 Markups and marginal products

The estimation of output elasticities makes it possible to also estimate firms' markups and evaluate the marginal revenue product of inputs in production.

In order to estimate firm level markups, we rely on the procedure laid out by De Loecker and Warzynski (2012), who use the first-order condition of the flexible inputs to impute the ratio of prices to costs. We use the intermediate input for this task, given that, as discussed above, labor is likely to be a dynamic input in our context, and is surely subject to some degree of adjustment costs. The markup can be obtained as

$$\hat{\mu}_{i,t} = \hat{\theta}^{m}_{i,t} \left(\frac{P_{i,t} Q_{i,t}}{P^{m}_{i,t} M_{i,t}} \right) \tag{15}$$

As in De Loecker and Warzynski (2012), we can only imperfectly measure the expenditure share of materials in gross output, given the likely presence of measurement error in the estimation of Equation (1). For this reason, we divide gross output in equation (15) by $\exp(\hat{\varepsilon}_{i,t})$, the residual from the first stage regression in the production function estimation procedure. This correction helps eliminating any variation in expenditure shares coming from variation in output not correlated with $\phi(l_{i,t}, k_{i,t}, m_{i,t})$, that is "output variation not related to variables impacting input demand".[19]

Given the estimated markups and elasticities, it is possible to obtain estimates of the distortion in labor and capital utilization, namely the differences (gaps) between their estimated marginal products and their cost. Taking into account a model in which firms compete monopolistically and choose their input demand level at each period, we can derive revenue marginal product (MRP) as

$$MRP^{X}_{i,t} \equiv \frac{\partial (P_{i,t}(Q_{i,t}) Q_{i,t})}{\partial X_{i,t}} = \underbrace{P_{i,t} \frac{\partial Q_{i,t}}{\partial X_{i,t}}}_{VMP^{X}_{i,t}} \left(\underbrace{1 + \frac{Q_{i,t}}{P_{i,t}} \frac{\partial P_{i,t}}{\partial Q_{i,t}}}_{\mu^{-1}_{i,t}} \right) = \theta^{X}_{i,t} \frac{P_{i,t} Q_{i,t}}{X_{i,t}} \frac{1}{\mu_{i,t}} \tag{16}$$

and as such MRP - cost gaps as

$$\text{MRPK-cost gap}_{i,t} = \hat{\theta}^{k}_{i,t} \frac{P_{i,t} Y_{i,t}}{K_{i,t}} \frac{1}{\hat{\mu}_{i,t}} - R_{i,t} \tag{17}$$

$$\text{MRPL-cost gap}_{i,t} = \hat{\theta}^{l}_{i,t} \frac{P_{i,t} Y_{i,t}}{L_{i,t}} \frac{1}{\hat{\mu}_{i,t}} - W_{i,t} \tag{18}$$

[19] We mainly focus on the estimates of markups and marginal products coming from the Ackerberg et al. (2015) translog specification, as in the Cobb-Douglas case elasticities do not vary within industry, and as such markups for instance are solely determined by the ranking in corrected expenditure shares, and not by possible variation in output elasticities and inputs utilization.

$R_{i,t}$ consists of the depreciation rate, which we keep at 7 percent as in the PIM exercise, and the average interest rate paid by the firm on its debt, which is the ratio of interest expenditures to total debt. When the information is missing, similarly to Fonseca and Van Doornik (2022) we impute interest rates as the average yearly interest rate at the 2-digit industry level.[20] For the average wage $W_{i,t}$, we divide the total wage bill by the number of employees (either taken from the QP when available, or as the full-time equivalent count in the CB for the remaining firms).[21]

These gaps convey information on how much a firm is constrained in the demand for an input (in case the gap is positive) or is overusing it and likely the optimal downward adjustment in its usage is hindered by adjustment costs (negative gaps).

Table 3 displays our estimates of costs, marginal revenue products and gaps. Even in this case, quite reassuringly, our estimates of gaps are in the same ballpark of magnitude of recent studies performing similar exercises.

[20]It is not possible to obtain more precise interest rates estimates for different kind of loans and credit instruments for the years of the analysis. The variation in results is minimal if using finer definitions of industry.

[21]For this estimation, one would ideally want to have more precise estimates of the marginal costs of inputs of production than the average yearly estimates of firm wage and user cost of capital. Reassuringly, studies in which data allow to gauge the distinction between average and marginal cost levels do not seem to find dramatic differences in gaps estimated according to the different costs definitions (see Lenzu and Manaresi (2018) for the difference in estimated gaps using average versus marginal wages.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451. [On pages 11, 13, 16, and 17]

Afonso, António, and Ana Venâncio. 2016. "The relevance of commuting zones for regional spending efficiency." *Applied Economics*, 48(10): 865–877. [On page 6]

Antunes, António, Homero Gonçalves, and Pedro Prego. 2016. "Firm default probabilities revisited." *Banco de Portugal Economic Studies*, 2(2): 21–45. [On page 12]

Blattner, Laura, Luisa Farinha, and Francisca Rebelo. 2023. "When Losses Turn into Loans: The Cost of Weak Banks." *American Economic Review*, 113(6): 1600–1641. [On page 12]

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102(6): 2437–2471. [On pages 11, 13, 16, and 17]

Fonseca, Julia, and Bernardus Van Doornik. 2022. "Financial Development and Labor Market Outcomes: Evidence from Brazil." *Journal of Financial Economics*, 143(1): 550–568. [On pages 12 and 14]

Hardy, Wojciech, Roma Keister, and Piotr Lewandowski. 2018. "Educational upgrading, structural change and the task composition of jobs in Europe." *Economics of Transition and Institutional Change*, 26(2): 201–231. [On page 7]

Lenzu, Simone, and Francesco Manaresi. 2018. "Do Marginal Products Differ from User Costs? Micro-Level Evidence from Italian Firms." mimeo. [On pages 12 and 14]

Levinsohn, James, and Amil Petrin. 2003. "Estimating production functions using inputs to control for unobservables." *Review of Economic Studies*, 70(2): 317–341. [On pages 10 and 11]

Olley, Steven G., and Ariel Pakes. 1996. "The dynamics of productivity in the telecommunications equipment industry." *Econometrica*, 64(6): 1263–1297. [On pages 10 and 12]

Petrin, Amil, and Jagadeesh Sivadasan. 2013. "Estimating lost output from allocative inefficiency, with an application to chile and firing costs." *Review of Economics and Statistics*, 95(1): 286–301. [On page 10]

C Tables

Table 1: Revenue elasticities and markups, PIM capital

	CD		TSLOG		ACF CD		ACF TSLOG	
	Mean	IQR	Mean	IQR	Mean	IQR	Mean	IQR
θ^L	0.20	0.13	0.21	0.18	0.22	0.11	0.21	0.19
	(0.0001)		(0.0003)		(0.0001)		(0.0003)	
θ^K	0.03	0.02	0.03	0.03	0.02	0.01	0.03	0.02
	(0.00003)		(0.00005)		(0.00001)		(0.00004)	
θ^M	0.74	0.19	0.73	0.21	0.71	0.14	0.72	0.21
	(0.0002)		(0.0003)		(0.0002)		(0.0003)	
RS	0.97	0.04	0.97	0.05	0.96	0.02	0.96	0.04
	(0.0001)		(0.0001)		(0.0002)		(0.0001)	
μ					1.33	0.34	1.26	0.16
					(0.0009)		(0.0003)	

The table displays descriptive statistics regarding firm-level production function parameters, returns to scale and markups. Mean, interquartile ranges and block bootstrapped standard errors (by firm) for the mean. We show estimates for the two specifications of the gross production function (Cobb-Douglas and translog) and two methodologies we use. The first two columns are estimated by simple OLS, whereas the second two are estimated following the method by Ackerberg et al. (2015), which accounts for endogeneity in the choice of inputs use and we correct for firm selection. See appendix B for details regarding the estimation procedure. Returns to scale are computed as $\sum_X \theta^X_{i,t}$ $X \in \{L, K, M\}$. Markups are estimated according to the method laid out by De Loecker and Warzynski (2012), see appendix B.2 for details regarding the estimation procedure. The table results are based on estimates of the production function where capital is measured according to the perpetual inventory method (PIM).

Table 2: Revenue elasticities and markups, book v. of capital

	CD		TSLOG		ACF CD		ACF TSLOG	
	Mean	IQR	Mean	IQR	Mean	IQR	Mean	IQR
θ^L	0.20	0.13	0.21	0.18	0.23	0.14	0.21	0.18
	(0.0001)		(0.0003)		(0.0002)		(0.0003)	
θ^K	0.03	0.02	0.03	0.03	0.02	0.01	0.04	0.02
	(0.00003)		(0.00005)		(0.00001)		0.00005	
θ^M	0.74	0.19	0.73	0.21	0.72	0.17	0.71	0.20
	(0.0002)		(0.0003)		(0.0002)		(0.0003)	
RS	0.97	0.04	0.97	0.05	0.97	0.02	0.95	0.04
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
μ					1.34	0.33	1.25	0.14
					(0.0008)		(0.0002)	

The table displays descriptive statistics regarding firm-level production function parameters, returns to scale and markups. Mean, interquartile ranges and block bootstrapped standard errors (by firm) for the mean. We show estimates for the two specifications of the gross production function (Cobb-Douglas and translog) and two methodologies we use. The first two columns are estimated by simple OLS, whereas the second two are estimated following the method by Ackerberg et al. (2015), which accounts for endogeneity in the choice of inputs use and we correct for firm selection. See appendix B for details regarding the estimation procedure. Returns to scale are computed as $\sum_X \theta_{i,t}^X \; X \in \{L, K, M\}$. Markups are estimated according to the method laid out by De Loecker and Warzynski (2012), see appendix B.2 for details regarding the estimation procedure.
The table results are based on estimates of the production function where capital is measured as the net book value of balance sheet.

Table 3: MRPs, user costs and gaps, full CB

	Mean	p50	p10	p90	Mean	p50	p10	p90
Panel A								
r	0.07	0.05	0.00	0.14				
w	10.59	9.27	5.21	17.35				
Panel B	PIM capital				Book v. capital			
MRP^L	13.39	10.53	3.64	25.54	13.19	10.29	3.96	24.64
	(0.0271)				(0.0262)			
MRP^K	0.38	0.21	0.06	0.83	0.33	0.18	0.05	0.68
	(0.0010)				(0.0010)			
$Lab. Gap$	2.37	1.23	-4.19	9.81	2.20	0.93	-3.68	8.80
	(0.0153)				(0.0136)			
$Cap. Gap$	0.23	0.09	-0.09	0.70	0.18	0.05	-0.10	0.54
	(0.0010)				(0.0009)			

Panel A reports descriptive statistics regarding the distribution of measured interest rates and firm level (average) wages. Interest rates are measured as the ratio of interest expenses of the firm over the total stock of debt (as reported in CB, which comprehends both bank debt and any other form of debt financing for the firm). The average wage is simply calculated as the ratio of salaries to employees, where total salaries are taken from CB and employees are either employment as measured from QP or full time equivalent employment from CB is the former data is missing. Panel B reports descriptive statistics regarding marginal products and marginal products-cost gaps. The labor marginal product and gap are measured in thousands of Euros. Block bootstrapped standard errors (by firm) displayed for the means. We report statistics both for the marginal products and gaps based on elasticities and values of variables when capital is computed according to the perpetual inventory method (PIM) or the net book value.
See appendix B.2 for details regarding the computations.