The human capital function: Sectoral externalities *

Florian I. Schumacher
Mestrando PCE/UEM

Joilson Dias
Universidade Estadual de Maringá

Resumo

Apresentam-se evidências de que os efeitos de espraiamento do conhecimento têm um traço setorial importante, ao menos na economia brasileira. Utilizando a base de dados da PNAD 2008, é mostrado que o salário indivídual é influenciado positiva e significativamente pela fração de pessoas no seu setor que tem pelo menos um mestrado. Adicionalmente, são mostradas evidências de que a fração de pessoas de determinada unidade federativa empregadas no setor de serviços às empresas exerce um efeito positivo e significativo sobre o salário dos indivíduos que vivem nesse estado. Ambos os resultados suportam o modelo teórico de Romer (1990) e evidenciam a existência de externalidades do capital humano em nível setorial. A terceira contribuição é fornecer evidências adicionais sobre a existência de retornos crescentes à educação no Brasil, revelando que estes retornos do *não* surgem em qualquer setor agregado específico.

Palavras-chave: Capital humano; externalidades setoriais do conhecimento; retornos crescentes de escala; desenvolvimento econômico.

Abstract

We show evidence that so called knowledge spillovers have an important sectoral trace, at least in the Brazilian economy. Using data from the Brazilian household survey PNAD 2008 we show that the wage of an individual is influenced positively and significantly by the fraction of people in his sector that have at least a master degree. Additionally we show evidence that the fraction of people in a federal state engaged in business services has a positive and significant effect on the wage of individuals who live in that state. Both findings support the theorical model of Romer (1990) and show evidence of sectoral human capital externalities. Our third contribution is to provide additional evidence on the existence of increasing returns to schooling in Brazil, revealing that these returns do *not* arise in any specific aggregate sector.

Keywords: Human capital; sectoral knowledge externalities; increasing returns to scale; economic development.

Área ANPEC: 5 – Crescimento, Desenvolvimento Econômico e Instituições

Classificação JEL: I28; O41; O47.

*Paper submitted to the 38th *Encontro Nacional de Economia*, ANPEC, Salvador, Brazil, 2010.

1 Introduction

There is good evidence to think that the accumulation of human capital has not only direct (internal) effects on the productivity of the individual himself but also indirect (external) effects on the productivity of other people that maintain some kind of relationship (maybe quite indirect) with the individual. Less clear is how exactly these external effects arrise and the mechanism through which they spread into the economy.

We show evidence that so called knowledge spillovers have an important sectoral trace, at least in the Brazilian economy. Using Mincerian-like log wage per hour regressions we show that the Mincerian rate of return to (an additional year of) schooling of an individual is influenced positively and significantly by the fraction of people in his sector that have at least a master degree. This even holds after controlling for occupation, ethnicity, gender, metropolitan area, fixed regional effects and other factors.

Additionally our treatment of the data enables us to show evidence that the fraction of people in a federal state engaged in business services has a positive and significant effect on the Mincerian rate of return to schooling for individuals living in that state. Both findings support the theorical model of Romer (1990).

Although some empirical evidence suggest that the external effects of human capital are small [see, for instance, Acemoglu & Angrist (2001), Duflo (2004), and Ciccone & Peri (2006)], the studies done so far has addressed mainly the effects on local labor markets, such as cities or, in the case of Acemoglu & Angrist (2001), states. There are few studies that have attempted to estimate these effects on more aggregated levels, such as sectors. In Brazil, there are no such studies to our knowledge.

All results where drawn from the Brazilian household survey PNAD 2008[1]. Before turning to the methodological procedure we do a theoretical excursion for reader that is not familiar with it.

2 Theoretical background

2.1 Endogenous growth models

Endogenous growth models have their origin in the mid 1980's, although earlier work had been done. The main contribution came from Romer (1990), but also important insights were made by Romer (1986) and Lucas (1988), among others. These models are called "endogenous" because they explain how technological progress occurs. The motivation for them come from the idea that technological progress is stronlgy associated with investments in human capital and research activity, and ultimately with appropriate institutions. The neoclassical model of economic growth is irreconcilable with orthogonal technological progress and so motivated the emergence of endogenous growth models, even if, as observed by Acemoglu (2008), the former has a R^2 fit to data that reaches almost three quarters of one if augmented with human capital[2] [see Mankiw, Romer & Weil (1992)].

The endogeneization of technological progress constitute an essential step in growth modelling, and is directly related to intrinsic aspects of knowledge and its diffusion. As technology in its strict

[1]The microdata from the 2008 edition of the survey is freely available at `http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2008/microdados.shtm`.

[2]The concept of *human capital* can be understood as an input into the production function that results from all "investment" people do that enhance their productivity. Examples of investments that tend to improve productivity are education, on-the-job training, health improvement, and migration for better job opportunities.

sense is made of knowledge, a non-rival good, it is intrinsically different from capital (be it human or physical) or labor, which are rival and excludable. Additionaly, Romer (1990) introduced the idea that knowledge (or "ideas") is at least partially excludable[3], resulting largely from intentional profit-seeking behavior of economic agents.

The non-rival nature of knowledge and thus technology is intrinsically related to non-convexities of the production function, i.e. increasing returns to scale. As noted by Jones (2000), once a new idea generated, its reproduction does not have any costs beyond the physical medium, which is usually a rival good. A fixed cost of production and a low and constant marginal cost imply a decreasing average cost and increasing returns to scale. As the average cost is always greater than the marginal cost pricing marginal cost results in negative profits. As a result, no company would invest resources into a new idea if it could not get any pricing power with the invention. Thus, the economy of ideas naturally leads to imperfect competition, one of the main assumptions of Romer's model.

Increasing returns to scale violates Euler's theorem and makes it impossible for all inputs being paid their marginal contribution to output. There are at least five ways to solve this: one can assume that technology is exogenously and freely available (SOLOW, 1956); that technology is supplied by government which collects taxes (SHELL, 1966); that technology results from the allocation of resources between a research sector and a final-goods sector (UZAWA, 1965); that technology is an unintentional byproduct of the accumulation of other factors (e.g. human capital) (ARROW, 1962; LUCAS, 1988); or that technology is invented and improved by the profit-maximizing behavior of private companies that seek to obtain monopoly rents (ROMER, 1990).

Whilst all procedures make the non-rival input technology essential to economic growth, and some of them (Arrow, Uzawa, and Lucas) shed light on the importance of human capital externalities, only Romer's strategy offers a market-based explanation for investments in technological progress.

2.2 Endogenous technological change and externalities

If we accept that the production function is homogenous of degree one in respect to the upper-bounded inputs (capital and labor), the only potential source of sustained long-run growth is an unbounded technological input, which must enter the function with non-decreasing returns. The main issue in modelling endogenous technological change, therefore, is to specify where the non-decreasing returns to technology (or knowledge) come from.

There are mainly two ways through which the non-decreasing returns to technology could arise. One way we can think of is through learning-by-doing, as emphasized by Arrow (1962) and Lucas (1988). The other way is through only partially excludable benefits from research activity, as emphasized by Romer (1990). While the market incentives approach is much more sensitive, the externalities approach is nonetheless complementary because in either case spillovers arise from the non-rival caracter of knowledge. The market incentives approach only requires that knowledge is partially excludable.

In the Romer (1990) model non-decreasing returns to technology come explicitly from the knowledge accumulation equation

$$\dot{A} = \delta H_A A, \tag{2.1}$$

[3]The example of a non-rival but excludable technological input cited by Romer is a design for a new capital good. Even though the vast majority of designs results from research and development activities of profit-maximizing private firms, these designs are all non-rival because they can be easily reproduced and used by many firms and individuals at the same time.

which is a linear function of the human capital devoted to research, H_A, and, more importantly, of the already attained technological level, A. The justification is that accumalated knowledge makes reasearchers more productive today than, say, fifty years ago. This in fact means that there are knowledge spillovers in the research sector of the economy, as demonstrated by Romer when he writes that "All researchers can take advantage of A at the same time" (p. S83).

In contrast, in the Lucas (1988) model non-decreasing returns come from the human capital accumulation equation

$$\dot{h} = h(t)^{\zeta} G[1 - u(t)], \tag{2.2}$$

simplified to

$$\dot{h} = h(t)\delta[1 - u(t)], \tag{2.3}$$

where $1 - u(t)$ is the time devoted to human capital accumulation, ζ is assumed to be one and consistent with the evidence provided by Rosen (1976), and the function G is assumed to be linear for simplicity. Human capital enters the production function through an effective labor input *and* an avarage human capital term that enters the production function because individual productivity is supposed to be positively affected by average human capital. The idea behind is that very productive workers (with high human capital) affect positively the productivity of less productive works surrounding them.

The main problem with Lucas' model is that it implicitly assumes an unbounded stock of human capital, something that is unlikely to be true if we understand human capital as defined by Mankiw (1995, p. 298):

> Knowledge refers to society's understanding about how the world works. Human capital refers to the resources expended transmitting this understanding to the labor force. Put crudely, knowledge is the quality of society's textbooks; human capital is the amount of time that has been spent reading them.

Another way to formulate the accumulation of human capital is given by McDermott (2002), who makes it a function of the "degree of specialization per unit of work" (p. 8), $M/e_M n$, the level of human capital, h, and the fraction of effort devoted to research, e_L:

$$\dot{h} = \left(\frac{M}{e_M n}\right)^{\gamma} h^{1-\gamma} e_L. \tag{2.4}$$

The parameter γ is less than one and the "degree of specialization" is given by

$$M = b_1 e_{Ma} h_a n N_c, \tag{2.5}$$

where e_M is the amount of effort devoted to working in the "market" sector (as opposed to the "primitive" sector), h_a is the average stock of individual human capital, n is the number of households that populate each "city", and N_c is the number of "cities" of the economy.

In another paper, McDermott (1999) specifies the accumulation of human capital as a function of its individual stock, h, the fraction of a day's effort spent learning, e_L, and exogenously growing "world technology", Ω:

$$\dot{h} = h^{1-\gamma}(\kappa\Omega)^{\gamma} e_L. \tag{2.6}$$

This kind of specification makes clear the role that human capital plays in knowledge absorption by people. The higher the stock of human capital is and the more effort is spent learning, the more an individual can absorb globally and freely available knowledge and enhance his own human capital.

The same idea is present in the model developed by Easterly et al. (1994), where better skilled or more qualified workers can use more capital goods (and more advanced ones) than unskilled workers, and the human capital accumulation is given by

$$\dot{h} = \mu e^{\varphi u} A^{\gamma} h^{1-\gamma}, \tag{2.7}$$

where u denotes the time that a person devotes to the accumulation of skills instead of working, A represents the global technological frontier, and the parameters are supposed to attend the conditions $\mu > 0$ and $0 < \gamma \leq 1$. Again, the "absorption" of global knowledge by any country depends on its effort devoted to human capital accumulation and its actual level of skills.

While the model of Romer (1990) provides the microeconomic foundation for a model of the technological frontier and reasons for the growth of technology over time, the model presented in this section, based on Easterly et al. (1994), address the issue of how technology spreads across countries and is more suited to a specific country. (JONES, 1995).

The advantage of the *technology absorption* framework is that it allows long-run positive per capita income growth rates of specific countries without requiring constant returns to any production input. Besides it is better suited for non-advanced countries that push rather weakly the global technological frontier, but can still take advantage of knowledge created elsewhere by devoting effort to learning and skill acquisition. Furthermore it implies that developing countries can improve greatly their wealth by simply devoting more resources to improve general (undergraduate schooling) and specific skills (graduate schooling; on-the-job training), without the need to invest too much in research, which is often very expensive for these countries.

While in Romer (1990) knowledge has increasing returns, but with its growth rate limited by the diminishing returns to research activity, Lucas (1988) introduced into his model an idea that was present in the economic literature since Marshall (1890) at last: the external effect of human capital. Lucas argues that a portion of the average human capital of the population is absorbed by the individual through social interaction, and this would be the external effect of human capital, i.e. a positive externality associated with human capital accumulation. The sum of both effects of investment in human capital, internal and external, would have a positive effect on the aggregate production.

In respect to the scope of social interaction required to give rise to external effects of human capital, Lucas cited the work of Jacobs (1969) to argue in favor of cities as the focus of analysis. What Jacobs suggested is that the concentration of economic activity in cities was partly a result of human capital externalities derived from the exchange of ideas among workers and entrepreneurs. However, if this is the case, it is more likely that these spillovers arise within sectoral boundaries than geographic ones, because it is that kind of interaction that sensibly allows learning-by-interacting – at least after the communication and transportation revolution the world faced past century.

In fact, externalities in Romer (1990) can be imagined as emerging from human interaction in a specific sector: the research sector. In his simplified model, knowledge externalities must arise there because it is where learning takes place. But, making a link to reality, its possible that such kind of externalities arise wherever knowledge accquisition (read learning) takes place. If this is true there would be human capital externalities in schools, universities, research institutions, research and development departments within firms, and, also important, all human capital intensive economic activities. Although *a priori* all these externalities are equally important and interesting, it is Romer's intuition that brings into focus that only firms are able to internalize substantial parts of these externalities when they act under an monopolistic competition environment. Going somewhat further, if part of these monopolistic gains are transferred to employees, and if we assume that externalities get stronger the higher the human capital stock of any group of workers, the latter should be rewarded with increasing returns to their own human capital accumulation. This would allow us to verify the externalities by means of comparing returns to human capital accumulation between sectors with different human capital endowments (read stocks). We can even think of firms selling as their main product information and knowledge, thus profiting from knowledge spread. This last kind of firms we may want to group into one observable sector: the business services sector. We expect that a sector with such a feature should reveal particularly high returns to schooling and an increasing share of total income over time. This reasoning is exactly what leads us to test empirically in the forthcoming sections whether this kind of pattern is observable in our data or not.

3 Empirical evidences

One of the first attempts to measure the technological external effects of human capital was made by Rauch (1993), who estimated quasi-Mincerian income regressions. The main difference with the Mincerian equation was the inclusion of average human capital of workers in the local labor market on the right side of the equation. Specifically, Rauch estimated models of the form:

$$lnW_{j,m} = X^{'}_{j,m}\beta + \gamma_p S_{j,m} + \gamma_e S_m, \tag{3.1}$$

where $X_{j,m}$ is a vector of control variables, $S_{j,m}$ is the years of schooling of individual j who lives/works in the labor market m, s_m is the average years of schooling of workers in the labor market m, γ_p measures the *private return* to schooling and the coefficient γ_e measures the *external return* to schooling.

Rauch's results show high external returns to education, often in excess of private returns. However, as noted by Acemoglu (2008), the regressions of Rauch exploited average schooling differentials across cities which may reflect many factors that also directly affect wages, such as a high cost of living, for example. More recent studies such as Acemoglu & Angrist (2001), Duflo (2004), and Ciccone & Peri (2006) estimate the external effects of human capital to be relatively small (and even negligible) in local labor markets (ACEMOGLU, 2008). Moretti (2004) found higher externalities.

The potentially sectoral characteristic of human capital externalities was stressed by Wheaton & Lewis (2002), who found that higher degrees of specialization and sectoral concentration are associated with wage gains, signaling that human capital externalities may be limited by sectoral scope.

To isolate potentially different effects, Wheeler (2007) estimated wage equations containing both a measure of the level of aggregate or average human capital and a measure of spatial concentration of industries, and found that both effects are high and significant. This result made him conclude that locational economies and human capital externalities seem to be quite distinct phenomena.

This finding suggests that neither technological nor pecuniary human capital externalities arise only by spatial concentration. A logical extension of this analysis is the investigation of workers changing industry, but within the same city. The evidence provided by Jaffe, Trajtenberg & Henderson (1993) that the effect of geographical spillover of knowledge is limited also support this approach. As well as the study of Rosenthal & Strange (2008), which verified that the density of well educated workers is strongly associated with wage levels, and that this association is reduced by a factor of 2.5 to 3.0 when it exceeds a concentric circle of five miles (8 km) around the workplace of the individual. Fu (2007) reports a very rapid weakening of the effects of human capital even beyond a circle of three miles (4.8 km).

One of the few studies to estimate the external return to education in economic sectors is that of Heuermann (2009). In addition to finding higher returns in manufacturing than in services – something that he relates to pecuniary externalities arising from the physical capital intensity in manufacturing – Heuermann estimated returns of 1.8% to an 1 percentage point increase in the proportion of highly skilled workers and 0.6% for an equivalent increase in the proportion of workers that are not highly qualified. Another interesting result is that highly skilled workers tend to benefit from intersectoral spillover effects of human capital, while not highly skilled workers tend to benefit mainly from *pecuniary* externalities arising among sectors.

Charlot & Duranton (2004), in turn, draw attention to the importance of media in mediating the effects of human agglomeration and human capital concentration on wages. According to the authors, controlling for the intensity of communication in the workplace reduces the direct effects by 8 to 10 times. This evidence also suggests that the spillover effects of human capital are stronger within industries than across industries or within a specific geographic area.

Finally, it is possible that to perceive the spatial externalities of human capital it is necessary to consider the value of consumption that areas with high human capital have for people. In this case, it is recommended to analyze the behavior of rents as well, besides wages, as does Rauch (1993), for example. We will not follow this way, but turn now to some evidence that relates human capital externalities to wage returns to schooling.

Another indication of the existence of external effects associated with human capital accumulation is the results found by Card & Krueger (1992), Heckman, Layne-Fanar & Todd (1996), Jaeger & Page (1996) and Dias & McDermott (2003), among others. These authors perceived that the relationship between years of schooling and individual income is non-linear. There is an "sheepskin effect" (or diploma effect), which sharply increases the income of individuals completing a given school degree, and a "threshold effect", which changes the slope of the relationship between education and income when a certain level of schooling is reached.

The threshold effect is more important for our objective because, as indicated by Dias & McDermott (2003), it seems that positive externalities of education over the economy arise only after the accumulation of a certain minimum level of human capital stock. Applying a macro-Mincerian model to data from the World Value Survey on a set of 109 countries during the period 1960-1990, the authors estimated a threshold value of 4.5 years of schooling.

For the Brazilian state of Minas Gerais, Hoffmann & Simão (2005) using data from the Census 2000, and Salvato & Silva (2008) using data from PNAD 2005, found a threshold value of 10 years of schooling after which additional years of schooling significantly increase individual income. The non-linear relation between income and schooling in Brazil was also observed by Lam & Schoeni (1993) and Ramos & Vieira (1996).

In a different approach, Trostel (2004) modifies the Mincer (1974) equation to allow for an inverted U relation between education and income. At the same time, the specification of Trostel permits us to identify the return to scale in schooling. In implementing this specification to the states of Brazil using PNAD data from 2004 to 2008, Monteiro, Dias & Dias (2009) confirmed the presence of increasing returns to schooling after a minimum of 4 years of schooling.

There is very little evidence on returns to schooling within industries. In Brazil, Hoffmann & Simão (2005) estimated the returns to schooling by sector using data from Census 2000 for the state of Minas Gerais, and found that: (i) below 10 years of schooling, the rate of return per year of schooling is 5.4% in agriculture, 6.9% in manufacturing and 8.7% in services, (ii) for at least 10 years of schooling, rates of return are, respectively, 19.8%, 23.3% and 21.4%. It should be noted that because in agriculture the proportion of people with education above 10 years is only 4.7% (against 19.6% in industry and 37.1% in services) the income variation due to education is very restricted in this sector.

4 Methodological procedures

4.1 Empirical implementation

Instead of deriving an empirical testable model directly from any mentioned theory, we will derive it from an generic wage equation that is inspired by Romer (1990, 's) ideas about knowledge spillovers and pecuniary externalaties.

Basically we start from the wage equation

$$w_{ij} = f[H_i(S, X), A_j(R_j), Z], \tag{4.1}$$

that describes the real wage of individual i who works in market sector j as being a function of its own human capital, H_i, the knowledge level in the sector where he works, A_j, and a vector of labor-market variables Z. The individual human capital level is a function of his schooling, S, and his work experience, X. The knowledge level of sector j is a function of fraction of its people that work in research activities, R_j. First-order partial derivatives $\partial f/\partial H_i$ and $\partial f/\partial A_j$ are both positive.

Sector-specific knowledge grows in the Romer-like fashion $g_A = \delta R_j$, where R_j is the fraction of people in market sector j that works in research activities. The knowledge level of any sector at time t is given by the standard equation $A_j = A_0 e^{g_A t}$, where A_0 is the initial knowledge level and t is a time indicator. The individual human capital equals $H_i = \gamma e^{\Omega(S)+\Phi(X)}$.

Finally, it suffice to assume the functional form

$$w_{ij} = \rho {A_j}^{\alpha} H_i^{\beta} e^{\chi Z} \tag{4.2}$$

for equation (4.1) to be able to get to an estimatable log earnings equation of the form

$$\ln w_{ij} = a + bS + cX + dZ + eR_j. \tag{4.3}$$

The complete derivation can be found in the Appendix.

4.2 Data

All data are from the Brazilian household sample survey PNAD, volume 2008, which was carried out between October and December 2008.[4] Since 2004, the geographic coverage of the survey represents the whole country. The PNAD survey has a complex sample design and thus requires adequate treatment. The 2008 volume contains a total of 391,816 observations which represent a population of 190 million. From this total we ended up using only 198,467 observations aged 20-60, which represent 96.1 million people. The detailed procedure can be found in the Appendix.

From the total of 198,467 observations, 145,640 (73.4%) worked in the survey's reference week (21-27 September 2008). Table 1 show statistics about continuous variables used in the following regressions. Monthly individual wage has an unweighted average of R\$ 1,015 (about US\$ 560) and was truncated at R\$ 30k (about US\$ 16.600). Years of schooling range continuously from 0-15, where the upper limit represents completed higher education, and has an additional value of 17, which represents people above higher education. Avarage schooling is 8.12 years. Variables age and experience are shown in decades. Experience is work experience as informed

[4]Microdata can be directly downloaded via the link `http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2008/microdados.shtm`. The documentation is available only in Portuguese.

Table 1 – Continuous variables

Variable	N	Mean	SD	Min	Max
individual wage	145,640	1,015	1,467	1	30,000
schooling	198,467	8.12	4.41	0	17
age/10	198,467	3.73	1.13	2	6
experience/10	145,640	2.23	1.23	0	5.5
familiar income	198,467	2,143	3,004	1	158,500
family members	198,467	3.59	1.47	1	16
members $<$14 years	198,467	0.788	1.020	0	9
fesc17br	145,640	0.977	1.31	0.0258	3.37
fset43uf	198,467	2.39	0.885	0.777	5.52

Note: `fesc17br` is the (percentual) fraction of people in each sector that have at least a master degree. `fset43uf` is the (percentual) fraction of people in each state that work in the business services sector.
Source: PNAD 2008.

by the individual, and thus *not* potential experience as usual. Family income ranges from R$ 3k to R$ 158k. Families are composed by 3.6 members at avarage, and have between zero and nine children less than 14 years old. The fraction of people in each sector with more than higher education (`fesc17br`) ranges from 0.026% in sector 10 (domestic services, recycling, street trading, street cleaning and other activities) to 3.37% in sector 9 (post, telecommunications, financial intermediation, information technology, research and development, education, business services, leisure and associative activities). The fraction of people in each state that work in the business services sector (`fset43uf`) ranges from 0.78% in the Northeast Alagoas state to 5.52% in the Federal District.

Table 2 – Indicator variables

Variable	Sum	Prop. (%)	Variable	Sum	Prop. (%)
female	100,631	50.7	agriculture	14,278	9.80
formal employee	60,239	41.4	energy and mining	4,286	2.94
military	453	0.311	inferior manufacturing	11,991	8.23
civil servant	12,345	8.48	superior manufacturing	5,390	3.70
informal employee	33,714	23.1	utilities	6,860	4.71
self employed	31,947	21.9	building	12,191	8.37
employer	6,942	4.77	trade	33,957	23.3
white	90,241	45.5	transport	6,946	4.77
black	16,489	8.31	superior services	24,848	17.1
brown	90,055	45.4	inferior services	15,887	10.9
other color	1,682	0.848	government	9,006	6.18
rural	23,507	11.8	unionized	28,100	18.0
reference person	92,169	46.4	other occupation	137,188	94.2
Southeast	60,514	30.5	chief	2,065	1.42
North	24,207	12.2	manager	6,387	4.39
Northeast	60,661	30.6	metropolitan area	78,281	39.4
South	30,299	15.3	mother with $<$14 child	9,571	5.32
Midwest	22,786	11.5	married	149,009	82.8
works	145,640	73.4			

Note: `reference person` is the family's reference person according to the questionaire respondent. Sector variables are defined in the Appendix.
Source: PNAD 2008.

Descriptive statistics about a variety of dummy variables are given in Table 2. It shows the unweighted number of observations that have the value 1 for each variable in the 'sum' column, and the respective percentual proportion that number represent in total nonmissing values in the 'prop. (%)' column. So 50.7% of the sample is female, 0.31% is military worker, 8.31% is black, 4.71% works in the utilities sector, and 18% are unionized.

5 Results

Reconsider Equation (4.3). In order to estimate that equation we define S as beeing a cubic term of years of schooling, X as being a squared term of decades of work experience, and Z as a vector of the following indicator variables: job type (formal employee, military, civil servant, informal employee, self employed, employer), skin color (white, black, brown, other), rural area (yes/no), family's reference person (yes/no), region (Southeast, North, Northeast, South, Midwest), unionized (yes/no), job position (chief, manager, other), metropolitan area (yes/no). For R_j we use as a proxy the log fraction of people in each sector that have at least a master degree (`lfesc17br`). Furthermore we add the variable `lfset43uf`. As explained in Section 2.2, we expect that the (log) fraction of people in each state that work in the business services sector will reveal particularly high returns to schooling. So the first equation we estimate is

$$\ln w_{ij} = a + bS + cX + dZ + e_1(lfesc17br) + e_2(lfset43uf) + \epsilon_1, \tag{5.1}$$

where the last term, ϵ_1 is an error term. We estimate coefficients a, e_1, e_2 and vectors of coefficients b, c, d.

A possible non-selection bias arise because we observe only the wage of individuals who decided to work. We use Heckman (1979)'s procedure to correct this possible source of bias and assume that $\ln w$ is observed if

$$\begin{aligned} &a_0 + a_1 \ln finc + a_2 fmem + a_2 nch + a_2 educ + a_2 age + \\ &a_2 region + a_2 color + a_2 married + a_2 femch + \epsilon_2 > 0, \end{aligned} \tag{5.2}$$

where ϵ_2 is correlated with ϵ_1 by magnitude ρ. $finc$ stays for familiar income, $fmem$ for number of family members, nch for number of under 14 years old children in the family, $educ$ for years of schooling, age for the individual age in decades. Indicator variables $region$ and $color$ are defined as previously. The two binary variables $married$ and $femch$ indicate, respectively, whether the individual is married and whether she is female with at least one under 14 years old child.

Table 3 shows the estimated coefficients. The first column (`ols`) shows coefficients estimated by ordinary least squares with White's heteroscedastic-robust standard errors. Coefficients shown in the second column (`svy`) are also estimated by ordinary least squares but incorporates the complex sample design of the PNAD survey, i.e. sampling weights, cluster sampling and stratification. Finally, the third column (`svyheck`) shows coefficients estimated by full maximum likelihood using the procedure developed by Heckman (1979), also considering the complex sample design.

Although the statistics point out that both complex design structure and non-selection bias should be considered, Table 3 shows that most coefficients do not vary much between the specifications. Nonetheless standard errors tend to be smaller in the `ols` specification. The greatest discrepancies between coefficients are observed for the indicator variables (except `metropolitan`) and for the variable `lfset43uf`, whose coefficient is a lot smaller after the adjustments. Adjustments measure $R2$ is 0.463 for the `ols` specification. The selection equation coefficients are not shown, but all of them except `other color` are significant at the 5% level.

Table 3 – Log wage regression

Variable	ols	svy	svyheck	Variable	ols	svy	svyheck
educ	0.08960	0.09230	0.08560	rural	-0.16100	-0.16400	-0.10500
	(0.00332)	(0.00406)	(0.00388)		(0.00605)	(0.01190)	(0.00945)
educsq	-0.01230	-0.01230	-0.01080	reference	0.15100	0.16000	0.17600
	(0.00047)	(0.00057)	(0.00054)		(0.00368)	(0.00428)	(0.00436)
educcb	0.00079	0.00077	0.00067	North	0.00816	0.01070	-0.01700
	(0.00002)	(0.00002)	(0.00002)		(0.00703)	(0.01520)	(0.01640)
exp10	0.20600	0.20500	0.19300	Northeast	-0.25100	-0.26400	-0.27100
	(0.00545)	(0.00607)	(0.00637)		(0.00570)	(0.01120)	(0.01170)
exp10sq	-0.02840	-0.02830	-0.02250	South	0.01260	0.01160	-0.03080
	(0.00113)	(0.00125)	(0.00131)		(0.00514)	(0.00838)	(0.00970)
military	0.53000	0.51400	0.44800	Midwest	0.07830	0.07160	0.04540
	(0.02700)	(0.03200)	(0.03140)		(0.00587)	(0.01060)	(0.01200)
civil servant	0.18600	0.15000	0.14900	unionized	0.12300	0.11600	0.09440
	(0.00696)	(0.00912)	(0.00877)		(0.00476)	(0.00711)	(0.00686)
infor. employee	-0.16900	-0.17400	-0.09790	chief	0.51000	0.49700	0.46200
	(0.00423)	(0.00543)	(0.00550)		(0.02010)	(0.02400)	(0.02780)
self employed	-0.04860	-0.04550	0.03920	manager	0.28900	0.28700	0.24400
	(0.00547)	(0.00809)	(0.00737)		(0.01070)	(0.01250)	(0.01200)
employer	0.45600	0.43400	0.40000	metropolitan	0.09380	0.10200	0.09150
	(0.01230)	(0.01620)	(0.01530)		(0.00387)	(0.00698)	(0.00608)
black	-0.12800	-0.13500	-0.16300	lfesc17br	0.06630	0.06700	0.05770
	(0.00632)	(0.00765)	(0.00930)		(0.00128)	(0.00164)	(0.00166)
brown	-0.11200	-0.11600	-0.14300	lfset43uf	0.13900	0.12800	0.09180
	(0.00397)	(0.00484)	(0.00575)		(0.00713)	(0.01570)	(0.01340)
other color	-0.02880	0.00551	0.04260	constant	1.57000	1.54000	1.62000
	(0.02080)	(0.02730)	(0.03230)		(0.02920)	(0.06100)	(0.05250)
N	145,640	145,640	179,899	r2	0.463	0.462	
F	4,186	1,590	1,122				
rho	-0.86071	(0.00406)		λ	-0.66563	(0.00593)	

Source: PNAD 2008.

Table 4 – Returns to scale, %

ols			svy			svyheck		
educ	ρ	dρ	educ	ρ	dρ	educ	ρ	dρ
4	2.9	-0.57	4	3.1	-0.61	4	3.1	-0.56
8	4.4	1.31	8	4.3	1.23	8	4.1	1.05
11	10.5	2.73	11	10.1	2.62	11	9.1	2.25
15	25.1	4.61	15	24.3	4.47	15	21.3	3.85
17	35.3	5.56	17	34.1	5.39	17	29.8	4.66
5.2	2.6	0.00	5.3	2.7	0.00	5.4	2.7	0.00

Source: PNAD 2008.

The coefficients of the schooling variables (`escol, escolsq, escolcb`) can not be interpreted directly. Table 4 shows instead Mincerian returns to education for five levels of schooling for each specification, as well as years of schooling which mark the start of increasing returns to schooling. Calculations are identical with those of Trostel (2004). First, returns to education are similar in `ols` and `svy` specifications, but considerably lower in the `svyheck` – thus revealing considerable selection bias. Second, returns to schooling (column ρ) are increasing, going up from 3.1% with four years of schooling to 29.8% with 17 years of schooling in the `svyheck` specification. Third, increasing returns to education are achieved from 5.2-5.4 years of schooling. Columns called $d\rho$ show second-order derivatives of log wages in respect to schooling.

Because both fractions `fesc17br` and `fset43uf` are taken in logs their coefficients give us their elasticity of wage. Both are statistically significant, and exert an important effect on wages. A one percentage increase in the fraction of very educated people (at least master degree) in a given market sector is associated with a (approximate) 6.6% increase in wages of people who work in that sector. The spatial effect of the business services sector is even stronger: a one percentage increase in the fraction of people in given federal state that work in that sector is associated with a (approximate) 13.9% increase in wages of people who work in that state. These effects are, respectively, evidence for sectoral human capital externalities and for the knowledge-spreading role played by the business services sector. The inclusion of state-dummies does not change dramatically this result.

Table 5 – Sectoral returns to schooling (in %), pooled regression

Sector	**N**	**4**	**8**	**11**	**15**	**17**	$d\rho = 0$	**Aver. educ**
Agriculture	14,278	5.1	5.4	8.1	15.1	20.0	5.7	4.0
Energy and mining	4,286	1.5	1.8	4.6	11.5	16.5	5.7	8.7
Inferior manufact.	11,991	3.7	4.2	10.2	25.8	36.8	5.8	7.9
Superior manufact.	5,390	7.0	13.2	23.0	42.7	55.5	2.8	10.0
Utilities	6,860	-2.1	2.4	10.0	25.9	36.3	3.3	11.4
Building	12,191	2.8	5.1	12.2	28.6	39.9	4.9	6.2
Trade	33,957	5.1	5.4	8.1	15.1	20.0	5.7	8.9
Transport	6,946	0.3	-0.3	4.2	17.0	26.3	6.3	8.0
Superior services	24,848	2.0	7.6	14.4	26.7	34.3	0.1	11.6
Inferior services	15,887	2.3	2.6	5.3	12.3	17.2	5.7	6.1
Government	9,006	2.1	7.6	14.2	26.3	33.8	0.3	10.9
All sectors	145,640	3.1	4.1	9.1	21.3	29.8	2.7	8.1

Source: PNAD 2008.
Note: Sectoral returns are estimated through binary variables that interact with schooling variables. Returns to scale are calculaed by Trostel (2004)'s procedure. Coefficients not significant at the 5% level are set to zero. N is number of unweighted observations. Column $d\rho = 0$ shows minimum years of schooling required to achieve increasing returns to schooling.

When we analyze returns to schooling by large sectors[5] there are clearly differences between them. Table 5 shows that returns to schooling are much larger for superior manufacturing (e.g. paper, chemicals, plastic, machines and equipment, vehicles), but also relatively large for inferior manufacturing (e.g. food, textil, clothing, wood, furnishings), utilities, building, superior services (e.g. post, telecommunications, financial intermediation, information technology, research and development, education, business services) and government. We observe lower returns for agriculture, energy and mining, trade, transport and inferior services (e.g. domestic services, recycling, itinerant trade, street cleaning). That said, we observe increasing returns to schooling for all sectors reported. Most sectors demonstrate increasing returns from between five and six years of schooling, but some of them even earlier. There does not seem to be any relation between the sectoral average years of schooling and returns to schooling or returns to scale. Coefficient of variable `lfset43uf` remains significant at any usual level, but its point estimate is reduced to 9.1%.

[5]Specifically, to estimate sectoral returns to school we use binary variables and interact them with schooling variables. Variable `lfesc17br` is dropped because of perfect collinearity.

6 Concluding remarks

Knowledge spillovers, as well as technological and pecuniary human capital externalities are certainly important in the long-run economic growth process. Endogenous growth models offer an theoretical framework for thinking about the common sense relationship between them. We use ideas arising from these models to derive an empirically estimable wage regression and to test whether there is evidence for intra-sectoral human capital externalities (technological or pecuniary) in the Brazilian economy, in an attempt to find out a measure of the effect of Romer (1990)'s "research sector" on wages and thus labor productivity.

Furthermore, we point out that if knowledge is both non-rival and a production input that improves the productivity of other production factors (e.g. labor and capital), than firms that profit from selling productive information and thus knowledge should provide especially high returns. We identify the business services sector as having exactly this character. Assuming that higher returns are at least partially transferred to employees, we test the state-level effect of this sector on individual wages.

At the same time, taking the research sector as composed of highly skilled workers, we incorporate and important issue pointed out by Lucas (1988) when he argues for important "external effects" of human capital accumulation, although he was mainly concerned with spatially-limited personal interaction and learning-by-doing effects.

The encountered minimum value after which arise increasing returns to schooling (overall about 2.7 years, but around 5.5 years for most sectors) is compatible with the threshold effect evidence provided by Dias & McDermott (2003), who estimate a threshold value of 3 years of schooling for Brazil using data from the World Value Survey.

Using the sectoral fraction of highly qualified people (with at least a master degree), we estimate an sectoral research intensity externality. A one percent increase in the referred fraction leads to an approximate increase in sectoral wages of 6.6%. In turn, the state-wide effect of a one percent increase of the fraction of people engaged in the business services sector on individual wages is an estimated 13.9%. This last result holds even if we estimate rates of return to education by sectors or estimate wage regressions over sectors (not shown).

Our third contribution is to strenghen the evidence on increasing returns to schooling, i.e. returns to scale in the accumulation of one of the main inputs of the human capital function. Furthermore we show that these increasing returns are not a characteristic of any specific sector, but instead a feature that holds more or less for all great market sectors. Nonetheless, there are some important differences between sectors in regard to returns to schooling that still remain to be explained.

Finally, we think that there are good reasons to verify further the existence and magnitude of sectoral knowledge spillovers and human capital externalities. Not only do theoretical developments point out that these effects probably exist and are important to understand the process of economic development, but these effects are also not yet much explored in Brazil.

References

ACEMOGLU, D. *Introduction to Modern Economic Growth.* Princeton, NJ: Princeton University Press, 2008.

ACEMOGLU, D.; ANGRIST, J. How large are human-capital externalities? evidence from compulsory-schooling laws. In: BERNANKE, B. S.; ROGOFF, K. (Ed.). *NBER Macroeconomics Annual 2000, Vol. 15.* Cambridge: MIT Press, 2001, (NBER Chapters). p. 9–74.

ARROW, K. J. The economic implications of learning by doing. *The Review of Economic Studies*, v. 29, n. 3, p. 155–173, jun. 1962.

CARD, D.; KRUEGER, A. B. Does school quality matter? returns to education and the characteristics of public schools in the united states. *The Journal of Political Economy*, v. 100, n. 1, p. 1–40, fev. 1992.

CHARLOT, S.; DURANTON, G. Communication externalities in cities. *Journal of Urban Economics*, v. 56, n. 3, p. 581–613, nov. 2004.

CICCONE, A.; PERI, G. Identifying human-capital externalities: Theory with applications. *The Review of Economic Studies*, v. 73, n. 2, p. 381–412, abr. 2006.

DIAS, J.; MCDERMOTT, J. *Aggregate Threshold Effects and the Importance of Human Capital in Economic Development*. Columbia, jun. 2003.

DUFLO, E. The medium run effects of educational expansion: Evidence from a large school construction program in indonesia. *Journal of Development Economics*, v. 74, n. 1, p. 163–197, jun. 2004.

EASTERLY, W. et al. Policy, technology adoption, and growth. n. 4681, mar. 1994. Disponível em: <http://ideas.repec.org/p/nbr/nberwo/4681.html>.

FU, S. Smart café cities: Testing human capital externalities in the boston metropolitan area. *Journal of Urban Economics*, v. 61, n. 1, p. 86–111, jan. 2007.

HECKMAN, J.; LAYNE-FANAR, A.; TODD, P. Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *The Review of Economics and Statistics*, v. 78, n. 4, p. 562–610, nov. 1996.

HECKMAN, J. J. Sample selection bias as a specification error. *Econometrica*, v. 47, n. 1, p. 153–161, 1979.

HEUERMANN, D. F. *Human Capital Externalities in Western Germany*. Trier, abr. 2009. Disponível em: <http://www.iaaeg.de/documents/HCE

HOFFMANN, R.; SIMãO, R. C. S. Determinantes do rendimento das pessoas ocupadas em minas gerais em 2000: O limiar no efeito da escolaridade e as diferenças entre mesorregiões. *Nova Economia*, v. 15, n. 2, p. 35–62, mai./ago. 2005.

JACOBS, J. *The economy of cities*. New York: Random House, 1969.

JAEGER, D. A.; PAGE, M. E. Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics*, v. 78, n. 4, p. 733–740, nov. 1996.

JAFFE, A. B.; TRAJTENBERG, M.; HENDERSON, R. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, v. 108, n. 3, p. 577–598, ago. 1993.

JONES, C. I. R&d-based models of economic growth. *Journal of Political Economy*, v. 103, n. 4, p. 759–784, ago. 1995.

JONES, C. I. *Introdução à Teoria do Crescimento Econômico*. Rio de Janeiro: Campus, 2000.

LAM, D.; SCHOENI, R. F. Effects of family background on earnings and returns to schooling: Evidence from brazil. *Journal of Political Economy*, v. 101, n. 4, p. 710–740, ago. 1993.

LUCAS, R. E. On the mechanics of economic development. *Journal of Monetary Economics*, v. 22, n. 1, p. 3–42, jul. 1988.

MANKIW, G. The growth of nations. *Brookings Papers on Economic Activity*, v. 26, n. 1995-1, p. 275–326, 1995.

MANKIW, N. G.; ROMER, D.; WEIL, D. N. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, v. 107, n. 2, p. 407–437, mai. 1992.

MARSHALL, A. *Princípios de Economia*. 8. ed. London: Macmillan and Co., 1890.

MCDERMOTT, J. Mercantilism and modern growth. *Journal of Economic Growth*, v. 4, n. 1, p. 55–80, mar. 1999.

MCDERMOTT, J. Development dynamics: Economic integration and the demographic transition. *Journal of Economic Growth*, v. 7, n. 4, p. 371–409, dez. 2002.

MINCER, J. *Schooling, Experience, and Earnings*. New York: NBER; Columbia University Press, 1974.

MONTEIRO, W. de F.; DIAS, J.; DIAS, M. H. A. Taxa de retorno da escolaridade nos estados brasileiros: Crescente ou decrescente. In: *XXXVII Encontro Nacional de Economia*. Foz do Iguaçu: Rio de Janeiro: ANPEC, 2009. (Anais eletrônicos..., v. 37).

MORETTI, E. Estimation the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, v. 121, n. 1-2, p. 175–212, jul./ago. 2004.

RAMOS, L. A.; VIEIRA, M. L. A relação entre educação e salários no brasil. In: *A Economia Brasileira em Perspectiva: 1996*. Rio de Janeiro: IPEA, 1996. v. 2, cap. 21, p. 493–510.

RAUCH, J. E. Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics*, v. 34, n. 3, p. 380–400, nov. 1993.

ROMER, P. M. Increasing returns and long-run growth. *Journal of Political Economy*, v. 94, n. 5, p. 1002–1037, out. 1986.

ROMER, P. M. Endogenous technological change. *The Journal of Political Economy*, v. 98, n. 5, p. S71–S102, out. 1990.

ROSEN, S. A theory of life earnings. *The Journal of Political Economy*, v. 84, n. 4, p. S45–S67, ago. 1976.

ROSENTHAL, S. S.; STRANGE, W. C. The attenuation of human capital spillovers. *Journal of Urban Economics*, v. 64, n. 2, p. 373–389, set. 2008.

SALVATO, M. A.; SILVA, D. G. O impacto da educaçao nos rendimentos do trabalhador: Uma análise para regiao metropolitana de belo horizonte. In: *Anais do XIII Seminário sobre a Economia Mineira*. [S.l.]: Cedeplar, Universidade Federal de Minas Gerais, 2008.

SHELL, K. Toward a theory of inventive activity and capital accumulation. *The American Economic Review*, v. 56, n. 1/2, p. 62–68, mar. 1966.

SOLOW, R. M. A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, v. 70, n. 1, p. 65–94, fev. 1956.

TROSTEL, P. A. Returns to scale in producing human capital from schooling. *Oxford Economic Papers*, v. 56, n. 3, p. 461–484, jul. 2004.

UZAWA, H. Optimum technical change in an aggregative model of economic growth. *International Economic Review*, v. 6, n. 1, p. 18–31, jan. 1965.

WHEATON, W. C.; LEWIS, M. J. Urban wages and labor market agglomeration. *Journal of Urban Economics*, v. 51, n. 3, p. 542–562, mai. 2002.

WHEELER, C. H. Do localization economies derive from human capital externalities? *Annals of Regional Science*, v. 41, n. 1, p. 31–50, mar. 2007.

Appendices

A Derivation of the estimatable equation

Assume:

$$
\begin{aligned}
w_{ij} &= f[H_i(S,X), A_j(R_j), Z] \quad \text{with} \quad \frac{\partial f}{H_i} > 0 \;\; \frac{\partial f}{\partial A_j} > 0 && \text{(A.1a)} \\
&= \rho {A_j}^{\alpha} H_i^{\beta} e^{\chi Z} && \text{(A.1b)}
\end{aligned}
$$

$$A_j = A_0 e^{g_A t} \tag{A.2}$$

$$g_A = \delta R_j \tag{A.3}$$

$$H_i = \gamma e^{\Omega(S)+\Phi(X)} \tag{A.4}$$

Substituting (A.3) into (A.2), taking logs and surpressing the time indicator:

$$\ln A_j = \ln A_0 + \delta R_j \tag{A.5}$$

Taking logs of (A.4) and (A.1b):

$$\ln H_i = \ln \gamma + \Omega(S) + \Phi(X) \tag{A.6}$$

$$\ln w_{ij} = \ln \rho + \alpha \ln A_j + \beta \ln H_i + \chi Z \tag{A.7}$$

Substituting (A.5) and (A.6) into (A.7):

$$
\begin{aligned}
\ln w_{ij} &= \ln \rho + \alpha(\ln A_0 + \delta R_j) + \beta[\ln \gamma + \Omega(S) + \Phi(X)] + \chi Z \\
&= (\ln \rho + \alpha \ln A_0 + \beta \ln \gamma) + \alpha\delta R_j + \beta\Omega(S) + \beta\Phi(X) + \chi Z && \text{(A.8)} \\
&= a + bS + cX + dZ + eR_j
\end{aligned}
$$

Where

$$
\begin{aligned}
a &\equiv \ln \rho + \alpha \ln A_0 + \beta \ln \gamma \\
b &\equiv \beta\Omega \\
c &\equiv \beta\Phi \\
d &\equiv \chi \\
e &\equiv \alpha\delta
\end{aligned}
$$

B Stata routine for database manipulation

```
* UNZIP & MERGE PES & DOM TO PNAD.DTA
version 10
cd \$datadir
unzipfile DOM2008 , replace
infile using DCT-DOM2008.dct , clear // modif. version of SAS dict.
keep if V0104==1 // keep if interview realized
drop if mi(V0102,V0103) // drop if uniq. identif. var.s are miss.
isid V0102 V0103 // verify if var.s really uniq. identify obs.
sort V0102 V0103 , stable
save DOM2008 , replace
unzipfile PES2008 , replace
infile using DCT-PES2008.dct , clear
drop if mi(V0102,V0103,V0301) // drop if uniq. identif. var.s are miss.
isid V0102 V0103 V0301 // verify if var.s really uniq. identify obs.
sort V0102 V0103 V0301 , stable
save PES2008 , replace
merge V0102 V0103 using DOM2008 , uniqusing
tab _merge // verify if _merge==3 for all obs.
drop if _merge!=3
drop _merge
save PNAD2008 , replace
* EXTRACT DATA FROM PNAD.DTA
use V0101 UF V0102 V0103 V0301 V0403 V4611 UPA V4617 V4618 V0302 V8005 V0402\\
> V0404 V0601 V0602 V6003 V6007 V4803 V9001 V9906 V9907 V9058 V9087 V9892 V4706\\
> V4718 V4722 V4723 V4724 V4728 V4107 using PNAD2008 , clear
* DROP SOME OBSERVATIONS
drop if V8005<20 // who is under 20 years old
drop if V8005>60 // over 60 years old
drop if w==0 // works, but has zero wage
drop if V9058==0 // works, but zero hours
drop if w>30000 & !mi(w) // earns over R\$ 30,000 per month
* If a personal information is missing that should not
drop if mi(trab,lwfam,nmfam,idad10,esc,fem,ncri,regi,cor,rur,ref,metro)
* If a job information is missing that should not
drop if mi(exp10,pos,sind,ocup,branch) & trab==1
* WAGE AND SELECTION EQUATION VARIABLES
* Rendimento do TP
gen w = V4718 if V4718!=999999999999 & !mi(V4718)
* Ln(rendimento do TP por hora trabalhada)
gen lw = ln(w/(V9058*4)) if !mi(w,V9058)
* Anos de estudo
gen esc = V4803-1
recode esc (16=.)
recode esc (15=17) if V6003==11 | V6007==9
* Anos de estudo ao quadrado
gen escsq = esc^2
* Anos de estudo ao cubo
gen esccb = esc^3
* Anos de experiência em dezenas de anos
gen exp10 = (V8005-V9892)/10 if !mi(V9892)
replace exp10 = (V8005-esc)/10 if (V8005-V9892)<0
replace exp10=. if w==.
* Anos de experiência em dezenas de anos ao quadrado
gen exp10sq = exp10^2
* Binária de gênero [base: masculino]
recode V0302 (2=0)(4=1), gen(fem)
lab def labfem 0 "masc." 1 "fem."
lab val fem labfem
```

```
* Indicativa de posição no TP da SDR [base: empregado priv. formal]
recode V4706 (1 6=0)(2=1)(3=2)(4 7=3)(9=4)(10=5)(11/13=.),gen(pos)
lab def labpos 0 "empdo c/ CT" 1 "milit." 2 "func. púb." 3 "empdo s/ CT"
> 4 "conta próp." 5 "empdor"
lab val pos labpos
* Indicativa de cor [base: branca]
recode V0404 (2=0)(4=1)(8=2)(6 0 9=3),gen(cor)
lab def labcor 0 "branca" 1 "preta" 2 "parda" 3 "outra"
lab val cor labcor
* Binária de rural [base: região urbana]
recode V4728 (1/3=0)(4/8=1),gen(rur)
lab def labrur 0 "urbana" 1 "rural"
lab val rur labrur
* Binária de pessoa de referência [base: outra]
recode V0402 (2/8=0)(1=1),gen(ref)
lab def labref 0 "outra" 1 "pes.ref."
lab val ref labref
* Indicativa de macrorregião [base: Sudeste]
recode UF (30/39=0)(10/19=1)(20/29=2)(40/49=3)(50/59=4),gen(regi)
lab def labregi 0 "SST" 1 "NRT" 2 "NST" 3 "SUL" 4 "COE"
lab val reg labregi
* Binária de sindicato [base: não associado]
recode V9087 (3=0)(1=1),gen(sind)
lab def labsind 0 "ñ sind." 1 "sind."
lab val sind labsind
* Indicativa de ocupação [base: outra]
gen ocup=0 if V9906<.
replace ocup=1 if V9906>=1111 & V9906<=1230
replace ocup=2 if V9906>=1310 & V9906<=1320
lab def labocup 0 "outra" 1 "dirig." 2 "geren."
lab val ocup labocup
* Binária de região metropolitana [base: ñ metrop.]
recode V4107 (2/3=0)(1=1),gen(metro)
lab def labmetro 0 "ñ metrop." 1 "metrop."
lab val metro labmetro
* Binária de formal [base: ñ trab.]
gen trab = V4718>0 & V4718<999999999999
lab def labtrab 0 "ñ trab." 1 "trab."
lab val trab labtrab
* Rendimento mensal familiar
recode V4722 (999999999999=.),gen(wfam)
gen lwfam = ln(wfam)
* Número de componentes da família
gen nmfam = V4724
* Número de crianças <14 anos na família
gen cri = V8005<14 if !mi(V8005)
bysort V0102 V0103 V0403: egen ncri = total(cri)
* Binária de casado [base: ñ casado]
recode V4723 (6/8=0)(1/4=1)(10=.),gen(casad)
* Binária de mãe solteira com criança <14 anos [base: outra]
recode V4723 (6 8=1)(10=.)(nonm=0),gen(maecc)
* Idade em dezenas de anos
gen idad10 = V8005/10
* Idade em dezenas de anos ao quadrado
gen idad10sq = idad10^2
* VARIABLES TO MEASURE HUMAN CAPITAL EXTERNALITIES
* Ln da fração de pessoas com pós-graduação em cada branch [11 categ.]
bysort branch: egen nbr = count(branch)
gen esc17 = esc>=17 if !mi(esc)
by branch: egen nesc17br = sum(esc17)
```

```
by branch: gen fesc17br = nesc17br/nbr
gen lfesc17br = ln(fesc17br)
* Fração de pessoas empregadas no setor de serviços às empresas, por UF
gen set43 = setor==43 if !mi(setor)
bysort UF: egen nuf = count(UF)
by UF: egen nset43uf = sum(set43)
by UF: gen fset43uf = nset43uf/nuf
gen lfset43uf = ln(fset43uf)
```

C Sectoral variables construction

Table 6 – Sector variable definition

Variable	CNAE code	Description
Agriculture	11 12 13 14 20 50	Agriculture and related services, forestry and logging, fishing and aquaculture
Energy and mining	100 110 230 234 130 140 260 270 280	Extraction and processing of energetic materials and minerals (except machinery and equipment), and related services
Inferior manufact.	150 160 170 180 190 200 360	Food, textiles, clothing, leather and footwear, wood and furniture etc.
Superior manufact.	210 220 240 250 290 300 310 320 330 340 350	Cellulose, paper and publishing, chemicals, plastics, machinery and equipment, vehicles etc.
Utilities	400 410 850	Electricity, gas, water, health, social services etc.
Building	459	Building
Trade	500 530 550 700 450 710 930	Trade, repair, lodging, real estate, rental furniture, personal services
Transport	600 610 620 630	Transport and related activities
Superior services	640 650 660 670 720 730 740 800 910 920	Post, telecommunications, financial intermediation, information technology, research and development, education, business services, leisure and associative activities
Inferior services	370 531 900 950 990 998	Domestic services, recycling, itinerant trade, street cleaning and other activities
Government	750	Public administration, defense and social security

Note: Listed codes refer to the first three digits of the CNAE classification. The CNAE code can be retrieved from `http://www.ibge.gov.br/concla/cnaedom/cnaedom.php?sl=1`.