Turnover Taxes and Productivity
Evidence from a Brazilian Tax Reform

Bruno Caprettini Antonio Ciccone*

November 29, 2015

Abstract

We exploit a Brazilian tax reform to study the productivity losses caused by taxes on turnover, a type of tax that distorts transactions between firms and that is common in developing countries. We build a model in which many sectors exchange inputs sequentially before delivering final goods, and show that distortions to trade between firms are amplified by the "number of stages of production". We calibrate the model to the Brazilian economy and compute the productivity gains of the reform. We find that considering production processes that have 11 or more stages of production leads to estimated productivity gains of removing the turnover tax more than 4 times larger than considering production processes where firms exchange inputs only once. When production requires an infinite number of stages the gains of the reform are maximized and they are equal to around 0.05% of national income.

*Caprettini: University of Zurich. E-mail: bruno.caprettini@gmail.com. Ciccone: University of Mannheim and Universitat Pompeu Fabra. E-mail: antonio.ciccone@gmail.com. Many thanks to Paula Bustos, Fernando Broner, Stephen Redding, Esteban Rossi-Hansberg, Alessandro Tarozzi, Joachim Voth, Chris Woodruff and seminar participants at CREI and UPF, Princeton, UAB, World Bank, Inter-American Development Bank, CEPR PEDL Workshop and the 2014 Summer School in Development Economics. Part of this research was undertaken while Bruno Caprettini was visiting the International Economics Section (IES) at Princeton University: he thanks Gene Grossman, Stephen Redding and everybody in the International Trade Section for their very kind hospitality. Bruno Caprettini gratefully acknowledges financial support from the Private Enterprise Development in Low-Income Countries.

1 Introduction

Most production processes are sequential and involve a chain of firms that exchange inputs. In such production processes, distortions to trade between firms along the production chain can have large effects on aggregate productivity because production losses may cumulate along the production chain. Models with production chains have therefore recently been used to reexamine the aggregate implications of firm-level distortions in growth and development (e.g. Jones (2011)), business cycle economics (e.g. Acemoglu et al. (2012)) and international trade (e.g Antràs and Chor (2013)). In this paper we use a model with production chains for a quantitative study of the productivity and welfare losses of distortive taxation. In particular we are interested in quantifying the productivity and welfare gains of a tax reform that eliminates distortive turnover taxes, which are still used in many developing countries (according to the World Bank, 45 countries were still using turnover taxes in 2012).

Turnover taxes are taxes levied on the full value of firms' revenues (or turnover). Because turnover taxes do not allow firms to deduct the cost of intermediate inputs from the tax base, they raise the cost of purchasing intermediate inputs from other firms relative to producing inputs in-house. This is well-known to reduce the aggregate efficiency of production (Diamond and Mirrlees (1971a)) which is why international economic institutions have been recommending that turnover taxes should be abolished in favor of less distortive taxes (e.g. Keen (2009)). Our objective is to quantify the productivity and welfare effects of substituting a turnover tax with a non-distortionary tax when there are production chains. To do so we develop a model with production chains and calibrate it using information on input and output flows reported in input-output tables. Two other aspects of the model we develop are important. First, the model can accommodate production chains of any length. This allows us to quantify the productivity and welfare effects of turnover taxes for production chains of different length. Second, the model reduces to standard models of input-output linkages used by Jones (2011, 2013) and Antràs et al. (2012) when production chains become infinitely long. Thus, the model we develop is easy to relate to standard models of input-output linkages.

One important issue that arises when calibrating a model with production chains using input-output tables is that such tables do not contain enough information to calibrate both the intermediate input intensity and the number of stages of production. This point can be seen in a simple example. Suppose that we know that there are two firms in a sector and that firms in the sector have bought a total value of 10 million Euros of goods from other firms in the sector. This data is consistent with firm A buying 10 million Euro of goods from firm B. In this case goods are exchanged between firms once. But the data is also consistent with firm A buying 2.5 million Euro of a good produced by firm B, adding

5 million of value through processing and then selling 7.5 million Euro of the processed good to firm B. In this second case goods are exchanged between firms twice. Because the total value of transactions is the same in the two cases, knowing the value of goods exchanged across firms is not enough to know the number of stages of production. This is an important limitation of the data in input-output tables, as generally the productivity and welfare effects of turnover taxes will depend on the intermediate input intensity of production as well as the number of stages of production. We deal with this limitation of the data in input-output tables in the following way. We first set the number of stages of production in the model as an exogenous parameter. Next, we calibrate the remaining parameters to match the data in input-output tables. This allows us to derive the production gains of eliminating a tax on turnover as a function of the number of stages of production.

We calibrate the model to Brazilian input-output tables and use it to examine the productivity and welfare effects of the conversion of a turnover tax of 3.65% into a value added tax — a tax reform that Brazil actually undertook in 2002. A main conclusion of our analysis is that such gains are much larger in the case with long production chains than the case with short production chains. For example, we find that the reform results in productivity gains that are 4.1 times larger when the production process has more than 11 stages of production than when it has a single stage of production. However, for the intermediate input intensities observed in Brazil, the productivity gains of such a tax reform remain small even when production chains are infinitely long. For example, eliminating the tax on turnover increases national income by at most 0.05% in the baseline case. For the productivity effects of the Brazilian turnover tax to become large, intermediate input intensities would have to be substantially larger than what we observe from the input-output tables. For example, with infinitely long production chains, the productivity gain of the tax reform would be around 1% (0.95%) of national income if the overall intermediate input intensity of the Brazilian economy – the value of inputs over gross output – was 70% instead of the observed 43%. And the productivity gain of the tax reform would be around 1.33% of value added if the overall intermediate input intensity of the Brazilian economy was 95%.

Our baseline calibration for Brazil is based on an input-output matrix with 100 sectors. Because such a detailed input-output matrix will be unavailable for many developing countries, we also ask what productivity effect of the 2002 tax reform we would have found if our calibration had treated Brazil as a single-sector economy (with an overall intermediate inputs intensity of production of 43%). Somewhat surprisingly, we find almost the same effects as in the case of the input-output matrix with 100 sectors. Hence, the number of production stages appears to matter more for the quantitative results of the tax reform than the detail of sectoral disaggregation.

Our main results are obtained assuming an elasticity of substitution between intermediate inputs and labor equal to one . For robustness we examine how results depend on the elasticity of substitution. We find lower productivity gains brought by the reform when firms have a very low or very high elasticity of substitution between intermediate inputs and labor. The productivity effects of turnover taxes are low when the elasticity of substitution between intermediate inputs and labor is low because firms' technologies are such that they cannot alter their production decisions in response to the tax. In the opposite case when the elasticity of substitution between labor and inputs is very high, firms respond to the tax by changing significantly their input mix. However, because labor is a good substitute of intermediate inputs in this scenario, the productivity distortions of producing in-house rather than buying from other firms will be small. With numerical simulations we show that the elasticity that yields the highest welfare gain of the Brazilian tax reform depends on the number of stages of production. For example, we find that the largest welfare gains are obtained for an elasticity of substitution equal to 50 when the production process involves 3 stages of production. In this scenario the welfare gains of the reform would be slightly larger than 1% of national income. When the production process requires an infinitely long chain instead, the largest welfare gain of the reform are obtained for an elasticity of substitution equal to 37, when the welfare gains of the reform are equal to 0.87% of national income.

The results of the paper have implications for tax design, especially in developing countries. Taxes on turnover have been long known to reduce productivity (Diamond and Mirrlees (1971a,b)). However, it has also been argued that this type of taxes can be more desirable than other tax schemes. Taxes on turnover do not allow to deduct the cost of intermediate inputs from the tax base. Thus, firms cannot reduce their taxes by inflating the cost of inputs and for this reason they are harder to evade than value added taxes or taxes on profits. Best et al. (2013) present evidence consistent with this argument, by showing a large positive effect of turnover taxes on tax compliance in Pakistan, where firms are taxed on either turnover or profits, depending on which tax liability is larger.[1,2] The large informal sectors observed in developing countries, combined with the lower evasion guaranteed by turnover taxes may explain why these taxes are still very popular among

[1] The presence of an informal sector also creates theoretical reasons for the optimality of input taxation. When intermediate inputs are produced by informal firms, taxing inputs of production is optimal because it allows to raise revenue on at least part of the product of the informal sector (Newbery (1986); Emran and Stiglitz (2005)).

[2] There are also reasons to believe that a well enforced value added tax increases tax compliance. The general consensus is that the credit method used to collect value added taxes create "chains of firms" that exchange inputs of production and that are either all formal or all informal (Keen (2009); De Paula and Scheinkman (2010)). The creation of such production chains may increase the effectiveness of tax audits, as they create incentives for audited firms to demand receipts from their suppliers (Pomeranz (2013)). However, this implies that value added taxes may raise the returns of a given level of tax enforcement, but there is nothing in the literature that suggests that VAT increase tax compliance *per se*.

countries with limited tax capacity (Besley and Persson (2013)). As of 2012, 45 low and middle income countries had some form of tax on turnover in place (World Bank (2006, 2012)), despite the fact that the conversion of taxes on turnover into modern value added taxes is part of the standard reform package promoted by the International Monetary Fund and the World Bank (Keen (2009)).

Related literature

The paper contributes to several literatures. In public finance, the evaluation of the social costs of taxes was pioneered by Harberger (1962). Following his tradition, several authors have evaluated the efficiency costs of taxes that cumulate along the production chain (Bovenberg (1987); Gottfried and Wiegard (1991); Piggott and Whalley (2001)). These works evaluate the welfare loss caused by the inefficient design of value added taxes, and they do not analyze the welfare cost of input taxation that affects all the sectors of the economy. The closest precedent to our result on the welfare gain of eliminating a turnover taxes is Keen (2013), who analyzes production inefficiencies generated by turnover taxes. He derives an approximation of the deadweight loss of a turnover tax in two special cases. First, when there is a single stage of production and arbitrary elasticity of substitution between inputs. Second, when there is an arbitrary number of stages but every stage before the last one produces output according to a Leontief production function. In both cases, the turnover tax distorts decisions of a single stage of production, when intermediate inputs can be substituted with labor. Relative to his work, we assume that firms produce with either a Cobb-Douglas, or a general constant elasticity of substitution technology that combines labor and intermediate inputs. These assumptions allow us to study processes in which turnover taxes distort production decisions several times along a production chain, rather that a single time at the end of it. The model is flexible enough to analyze formally the interaction between the length of the production chain and the elasticity of substitution between intermediate inputs and labor in determining the welfare cost of turnover taxes.

In the literature of input-output economics several authors have pointed out that input-output linkages can magnify distortions or changes in productivity that happen in few sectors (Leontief (1936); Hirschman (1958); Hulten (1978); Long and Plosser (1983); Basu (1995) and more recently Ciccone (2002); Gabaix (2011); Jones (2011); Oberfield (2011)).[3] Similar ideas have been developed in the international trade literature, that has shown how the effect of tariffs cumulates along the production chain in the same way as a tax on inputs does (Corden (1966); Yi (2003); Fally (2012); Antràs et al. (2012); Baldwin and Venables (2013)).

[3]Bartelme and Gorodnichenko (2014) are to our knowledge the only authors who explored these ideas empirically. They use a panel of input-output tables from several countries to show that the process of economic growth is associated with an increase in the importance of input-output linkages.

Finally, some of the results we present are relevant for the literature on misallocation. In particular, Jones (2011) shows that firm-level TFP may be amplified by input-output linkages and can have very large aggregate output effects. As Jones, we find that a tax on turnover distorts the way resources are allocated, and its effect is magnified by the input-output structure of the economy. However, we also find that the deadweight loss implied by the turnover tax is relatively small. The source of the difference between our estimates and those of Jones, is the consideration of tax revenues. In his paper Jones considers wedges that reduce the productivity of different sectors and are entirely wasteful.[4] When the distortion to relative prices comes from a tax however, part of the increased cost of production is collected by the fiscal authority in the form of tax revenue, and should be accounted in welfare calculations.

The paper proceeds as follows. Section 2 develops the main ideas with a simple example. Section 3 presents background on the 2002-3 Brazilian tax reform. Section 4 derives a general model that delivers explicit expressions for the welfare gain of the tax reform. Section 5 presents an extension of the model where we relax the Cobb-Douglas assumption and assume that production takes place with a generic constant elasticity of substitution function. Section 6 concludes. Explicit derivation of analytical results are in the appendix.

2 An example

We now illustrate the main idea of the paper. Consider an economy where firms produce goods using labor and goods that are produced by other firms (intermediate inputs). Suppose that a tax on the full value of the good is applied every time a firm buys or sells a good to another firm or to final consumers. To estimate the deadweight loss of such a tax one would have to observe a detailed input-output table for each good and firm. This table would have to report the value of production for each good and firm as well as the value of each input, the firms from which inputs are bought and the good for which inputs are used. With such data, plus the price of each good, one could keep track of every transaction and evaluate the deadweight loss of the tax. In practice, such detailed information is never available. What is usually available however is a sectoral input-output table. In this paper we therefore investigate whether such input-output tables can be used to evaluate the deadweight loss of the tax.

[4] In this sense the way sectoral distortions show up in aggregate TFP in Jones (2011) differs to the way distortions to the prices faced by individual firms reduce aggregate production in models of heterogeneous firms such as Banerjee and Duflo (2005), Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). The easiest way to see the difference is to assume that symmetric (i.e. identical) positive wedges throughout the economy. In models of misallocations of factors across firms symmetric wedges create no output loss. In Jones (2011) a positive wedge create output loss, even when it is the same for every sector.

The main limitation of input-output tables when evaluating the deadweight loss of turnover taxes is that they specify the value of goods exchanged within and across sectors, but contain no direct information on the number of times goods are exchanged. Moreover, it is impossible to use the information in input-output tables to calibrate both the intermediate input intensities of production and the number of stages of production. This can be seen by considering the situation of the two firms exchanging 10 million Euro of goods discussed in the introduction. In that case knowing the value of the goods exchanged was not enough to determine the number of transaction between them. Although this information can never be recovered from input-output tables, it matters for the welfare gains of a reform that eliminates a tax on turnover, because these are generally higher when firms exchange inputs more often. We now illustrate this in a simple (calibrated) example.

The example is a one sector model. Hence, the corresponding input-output table would only tell us the total value of goods produced by firms in the sector and the total value of goods that firms in the sector bought from other firms in the sector. Suppose that within the sector firms produce different varieties that are either used as intermediate inputs or consumed. We assume the simplest possible structure and consider two versions of the example that differ in the production structure. In the first version, the economy produces two varieties and in the second version the economy produces three varieties. In both cases production is organized sequentially. To emphasize this point, we refer to the number of different varieties as to "stages of production". The objective of the example is to illustrate that for a given observed input-output table, the production parameters and the deadweight loss of a turnover tax depend on the number of stages of production.

Two production stages. In this version the economy produces two varieties of the same good, variety 1 and variety 2. Variety 1 is produced by firm 1 with labor only and it is entirely sold to firm 2, which uses it to produce variety 2. Variety 2 is then sold to households as consumption good.

Three production stages. In the version with three production stages the economy produces three varieties, variety 1, variety 2, and variety 3. Variety 1 is produced by firm 1 with labor only and sold as an intermediate input to firm 2 to produce variety 2. Variety 2 is then sold to firm 3 which uses it to produce variety 3. Variety 3 is sold to households as consumption good.

We assume that firms operate under perfect competition and that all production that combines labor and intermediate inputs takes place with Cobb-Douglas production functions. The two different production structures and the production functions are described in figure 1. It can be seen that the example with three stages of production simply adds a stage of production to the example with two stages of production. An important aspect of this setup is that the only parameter that requires calibration in both cases is the intermediate input intensity of production (which is called ϕ in the first case and ψ in the second).

Preferences. In both versions of the example, a representative household suppies labor inelastically and consumes a privately produced good C and a public service supplied by the government. Her utility is:

$$U = C \cdot G^{\alpha}$$

where G is a public service supplied by the government and C is the quantity of the privately produced consumption good.

Government. The government collects taxes and uses them to hire workers and produce G units of the public service, which is produced using labor only.

The government can choose to raise taxes in two ways. It can tax the total revenues of every firm using a turnover tax of rate t. This tax makes labor cheaper than intermediate inputs for all firms that use both factors in production and therefore it generally distorts production. Alternatively, the government can impose a tax of rate v on the consumption of the household. This tax does not affect the price of labor relative to intermediate inputs and therefore it does not distort production (in our model this tax does not distort any economic decision and is therefore equivalent to a lump-sum tax).[5]

Input-output tables. The information in input-output tables does not allow us to see whether the production structure corresponds to the model with two or three stages of production. However, it does allow us to calibrate the intermediate inputs intensity of production in the two cases. To see this notice that the total value of production in the economy (V) and the value of intermediate inputs used (X) are in the two cases respectively:

$$\begin{aligned} X &= P(1)Q(1) \\ V &= P(1)Q(1) + P(2)Q(2) \end{aligned} \qquad \text{and} \qquad \begin{aligned} X &= P(1)Q(1) + P(2)Q(2) \\ V &= P(1)Q(1) + P(2)Q(2) + P(3)Q(3) \end{aligned}$$

The assumption of Cobb-Douglas production allows to solve for ϕ in the two stages case, and for ψ in the three stages case. These are the solutions of the following two equations:

$$\frac{X}{V} = \frac{\phi}{1+\phi} \qquad \text{and} \qquad \frac{X}{V} = \frac{\psi + \psi^2}{1 + \psi + \psi^2}$$

[5] From the point of view of production, a consumption tax is equivalent to a tax on value added. We study the welfare gains of a reform that converts a tax on turnover into a consumption tax because the latter is easier to work with than a value added tax.

Tax reform. Once we have used the input-output tables to calibrate the intermediate input intensity of production, we can evaluate the welfare gains of a reform that substitutes a turnover tax with a consumption tax. Here as in the rest of the paper we consider a tax reform that does not affect the level of tax revenues. Thus, the tax rate of the consumption tax is set to the level that guarantees the level of revenues needed to finance the level of public service provided before the reform.

Taking the wage to be the numeraire ($W = 1$), the inelastic supply of labor implies that welfare can always be written as $U = \left(\overline{L}/P_c\right) \cdot G^{\alpha}$, where $\overline{L}$ is the total amount of labor the household supplies, P_c is the price paid by households for one unit of the consumption good. Because we maintain the provision of public service fixed, in the example the welfare change is equal to the log change of the price of the consumption good. Thus, when the production process involves S separate stages of production, the welfare gains of the reform is:

$$\Delta \log U = \log \left(\frac{P^{tt}(S)}{P^{v}(S)} \right)$$

where $P^{tt}(S)$ and $P^{v}(S)$ are respectively the prices paid by the household for one unit of consumption good when there is a turnover tax and the price of the same good when there is a consumption tax.

As stated before, here as in the rest of the paper we keep tax revenues fixed to a given value in every tax scheme. Thus, we obtain the tax rate of the turnover tax t (consumption tax v) as the tax rate that raises the tax revenue we observe. In the example we assume that the government needs to raise an amount of taxes equal to 4.6% of GDP. This is what turnover taxes raised as a share of Brazilian GDP in the two years before the tax reform. We then set the "effective" turnover tax in our example equal to the value that ensures these tax revenues. Formally this means that the tax rates in the example with two production stages and in the example with three production stages solve:

$$0.046 = \frac{t(2)}{1+t(2)} \left(1 + \frac{\phi}{1+t(2)} \right)$$
$$0.046 = \frac{t(3)}{1+t(3)} \left[1 + \frac{\psi}{1+t(3)} + \left(\frac{\psi}{1+t(3)} \right)^2 \right]$$

Where $t(2)$ is the effective tax rate in the economy with two stages of production and $t(3)$ is the effective tax rate in the economy with three stages. Because $\phi = \psi + \psi^2 \equiv x$, the two equations can be written as:

$$\begin{aligned}0.046 &= \frac{t(2)}{1+t(2)}\left(1+\frac{x}{1+t(2)}\right)\\ 0.046 &= \frac{t(3)}{1+t(3)}\left(1+\frac{x}{1+t(3)}\right)-\frac{1}{1+t(3)}\left(\frac{t(3)\psi}{1+t(3)}\right)^2\end{aligned}$$

Since the last term of the second equation is negative, it follows that $t(3) > t(2)$ which means that the effective turnover tax must be higher in the economy with 3 stages than in the economy with 2 stages. As a result, tax reform yields a larger welfare gains in the economy with 3 stages of production, because in this case the effective turnover tax that the reform abolishes is higher. This is illustrated in figure 2, which plots the welfare gain of the reform for different observed input share X/V in the version of the example that involves three stages of production (in green) and in the version of the example that involves two stages of production (in blue). Comparing the two gains for $X/V = 43\%$ (the observed input share for Brazil in 2005, 2 years after the reform) one can see that the calibrated welfare gain of the reform is 3.72 times larger in the production process with three stages than in the production process with only two stages.

Infinite number of production stages. It is instructive to look at the case in which the final good is produced after an infinitely long production chain. There are two reasons to discuss this special case. First, because the welfare gain of the tax reform grows with the number of stages of production, this case provides an upper bound for the gains in a model with a single final good. Second, when the number of stages of production grows to infinity, the model with a single final good becomes identical to a model in which a single sector produces the final good using some of its own output as input of production. This is a type of model that is close to standard models used in the growth literature (e.g. Jones (2011)), and it is interesting to see how the model we develop here relates to these standard models.

When there are an infinite number of stages of production, the contribution to total output of the initial stage of production becomes negligible. Because every stage after the initial one is assumed to produce with an identical Cobb-Douglas function, in this special case the observed input share of the economy equals the input elasticity parameter of the production function. We call this input elasticity $X/V = o$. Solving the model with an infinite number of stages yields the following equations for the welfare gains and the effective tax rate as a function of the input elasticity o, tax revenues T and national income $W\overline{L} = \overline{L}$:

$$\lim_{S\to\infty} \Delta \log U(S) = \log\left[(1+t(S))^{\frac{1}{1-o}}\left(1-\frac{T}{\overline{L}}\right)\right]$$
$$\lim_{S\to\infty} t(S) = \frac{T}{\overline{L}-T}(1-o)$$

Figure 3 shows the welfare gain of the reform as a function of the observed input share when tax revenues over national income are kept constant at 4.6%. The blue and the green lines show the welfare gains of the reform in the two and three stages economies. The red line represents the welfare gain of the reform when the number of stages of production goes to infinity. The plot shows that the welfare gains are always larger in the economy with an infinitely long production chain. However, even as the observed input share approaches 100%, the welfare gains do not become infinitely large. This result is in stark contrast with what happens when the tax rate is fixed (and equal to the statutory tax rate, say) and tax revenues are allowed to adjust as input share changes. In this case, tax revenues collected on an infinitely long production chain grow without bounds as the input share approaches 100%. For this reason, in this case also the welfare gain of eliminating a turnover tax grows to infinity when the observed input share approaches unity. When tax revenues are kept constant instead, the effective tax rate needed to raise a given amount of tax revenues becomes smaller as the observed input share grows. In the limit, this tax rate approaches zero as the observed input share approaches unity. A lower effective tax rate partly compensate for the larger distortions created by a turnover tax in an economy that produces with a larger input elasticity. As a result, the welfare gain of eliminating such tax does not become indefinitely large as input share tends to one.

Finally, this simple example allows to ask what would be the welfare gain of the reform if the turnover tax was raising a higher share of revenues relative to total income than what we observe for Brazil in 2002. The interest of this exercise lays in the possibility that, although the Brazilian reform brought only modest gains, it was nonetheless a necessary step towards a more efficient tax system that would allow the government to raise more revenues. In figure 4 we plot the welfare gains of the reform as a function of the observed input share when the tax revenue raised by the turnover tax before the reform is equal to 10%. This is on the upper end of what developed countries usually raise by means of value added taxes: according to OECD data cited by Keen and Lockwood (2006), in 2005 only Hungary and Iceland were raising more than 10% of GDP with this type of tax. In the graph, we plot the welfare gain for an economy that produces with two stages (in blue), with three stages (in green) and with an infinite number of stages (in red). For the input share observed in Brazil in 2005 (43%), the welfare gains of the reform are around 0.24%

of total income even when there is an infinite number of stages. Figure 5 repeats the same exercise assuming that the turnover tax that is eliminated was raising revenues equal to 15% of total income, a level that is higher than the level observed anywhere today for VATs. In this case, for the observed input share of Brazil the welfare gain of the reform would be at most 0.56% of national income, when production requires an infinite number of stages.

3 The 2002-3 Brazilian tax reform

In this section, we provide background information on the 2002 Brazilian tax reform and few relevant aspects of the Brazilian fiscal system. In Brazil, every business has to pay several different taxes and social contributions.[6] This paper focuses on two of the social contributions paid by firms: the *Programa de Integração Social* (PIS: contribution for the social integration programme) and the *Contribuição para Financiamento da Seguridade Social* (COFINS: contribution for the funding of social security).[7] Until 2002 these contributions were levied on the total value of turnover, with no deductions for the cost of intermediate inputs. The tax rate of PIS was 0.65%, while the tax rate of COFINS was 2% between 1991 and 1998, and 3% since 1998. Except for the different tax rate, until 2002 the two taxes were very similar in terms of taxpayers, tax base and exceptions.

Between December 2002 and December 2003 the Brazilian Federal Congress passed two separate laws that modified the regime of both PIS and COFINS. The first reform, in December 2002, modified the regime of PIS and it allowed to deduct the cost of intermediate inputs from the the tax base, effectively converting PIS into a VAT. The law also increased the tax rate of PIS by 1 percentage point, to 1.65%, in order to avoid a fall in fiscal revenues. The second reform, in December 2003, modified the regime of COFINS, and it allowed to deduct the cost of intermediate inputs from the tax base of this tax too. The new tax rate of COFINS was set to 7.6%, a number that was chosen by applying to the existing tax rate the same multiplier adopted to adjust the rate of PIS in the VAT regime (roughly 2.53; see Werneck (2006) for a discussion of this choice).[8] In general, the

[6]Brazilian companies have to pay a corporate income tax on their profits (the *Imposto de Renda de Pessoa Jurídica*, IRPJ), and two different VATs on the value of their shipments (the *Imposto sobre Produtos Industrializados*, IPI, paid to the Federal Government, and the *Imposto sobre a Circulação de Mercadorias e Serviços*, ICMS, paid to the State where the firm operates). They are also responsible to pay a contribution for a social integration programme (*Programa de Integração Social*, PIS), a contribution for public servant assets (*Programa de Formação do Patrimônio do Servidor Público*, PASEP), a contribution to finance social security (*Contribuição para Financiamento da Seguridade Social*, COFINS), an union contribution, a retirement contribution for each employee, a contribution to finance education and municipal taxes according to local laws.

[7]PIS was introduced with the complementary law no. 7 on the 7th of September 1970. COFINS was created by the complementary law no. 70 on the 31st of December 1991, and it replaced the existing *Fundo de Investimento Social* (FINSOCIAL).

[8]For the PIS reform, see law no. 10637 passed on the 30th of December 2002, especially clause 3. For

two reforms aimed at maintaining PIS and COFINS similar taxes: the law that modified PIS contained the express provision for the Government to convert COFINS into a VAT within a year, and the COFINS reform is almost identical to the PIS reform in content, structure and wording. For this reasons in this paper we treat the two reforms as a single one.

The stated objective of these reforms was to increase competitiveness of the Brazilian manufacturing sector, and the policies were strongly encouraged by the International Monetary Fund (IMF) as part of the agreements to obtain recovery loans in the aftermath of the capital account crisis of 1998 (IMF (2002) and IMF (2003)).

In the next two sections we develop a formal framework to evaluate the welfare gain of this reform under different assumptions regarding the structure of production.

4 Cobb-Douglas production

The example discussed in section 2 is intuitive and simple to analyze. However it is special in that it assumes that production can be organized in only two ways. Moreover, it abstracts from input-output linkages between sectors producing different goods, and imposes a special functional form on production. In this section and in the next section, we develop two extensions of the example. In both models, we consider production processes that involve an arbitrary number of stages. In section 4.1 we consider an economy with a realistic input-output structure in which several sectors exchange intermediate inputs. In section 5 we consider a single-sector economy, but relax the Cobb-Douglas assumption.

4.1 The model

Setup. In this model, a representative household values the consumption of two types of goods. The first good is a public service that the government finances with taxes and provides to the household free of charge. The second good is an aggregate good made of a bundle of N differentiated products.

Production of public service only requires labor, and the government hires workers with the revenues collected with taxes. Production of the N differentiated products that compose the second good is carried out by private firms working in N separated sectors of the economy. Within each of the N sectors of the economy there are many firms that produce gross output using labor and intermediate goods produced by firms operating in all other sectors of the economy. In the input-output table of this economy we observe how much the firms of each sector buy from other firms, and how much they sell to other firms in the economy. As in the simple example developed in section 2 however,

the COFINS reform, see law no. 10833 passed on the 29th of December 2003, especially clause 3.

these aggregate flows may be generated by firms exchanging goods an arbitrary number of times, and this number will not be observable. As in the single-sector example, we will think of the inputs sold by the same sector at different stages as different varieties of the same good.

The household supplies labor inelastically. In the model, prices in the private sectors are determined by technology and perfect competition among many firms that produce and sell the same varieties. Given homothetic preferences, consumption shares are then uniquely determined by prices. The level of consumption of private goods depends on the total labor supply which determines the income of the household. Given the quantity of goods that the household wants to consume, the quantity produced by every sector is then pinned down by the structure of production.

Households. The representative consumer has the following Cobb-Douglas utility:

$$U = C \cdot G^{\alpha}$$

The arguments of the utility function are G, the amount of public service received from the government, and C, a composite of the N private goods purchased by households. The composite private good is a Cobb-Douglas aggregate of the goods produced in the N sectors of the economy:

$$C = \prod_{i=1}^{N} C_i^{\beta_i} \tag{1}$$

In (1), C_i is consumption of the i-th good and $\sum_i \beta_i = 1$.

The household supplies inelastically $\overline{L}$ hours of labor. At the given hourly wage W, she earns a total income of $W\overline{L}$ that she can spend on the consumption of the private goods C_i. Given Cobb-Douglas preferences, consumption of each good i is equal to $\beta_i W\overline{L}$.

Firms. There are N separate sectors where private firms produce differentiated final goods. For each of the N final goods produced and sold to consumers there are S different varieties produced. For every final good i, the first variety is produced using only labor:

$$Q_i(1) = A_i L_i(1)$$

In the production function of firms producing variety "1" in sector i, A_i is a sector-specific Hicks-neutral productivity parameter and $L_i(1)$ is the amount of labor used. The other $S - 1$ varieties in the same sector are produced with labor and intermediate inputs according to a sector-specific production function:

$$Q_i(s) = A_i L_i(s)^{1-\omega_i} \left[\prod_{j=1}^{N} M_{ij}(s)^{\omega_{ij}} \right] \qquad s = 2, 3, \ldots, S$$

where $\omega_i = \sum_j \omega_{ij}$. We assume that in every sector i, Hicks-neutral productivity parameter and input elasticities are sector-specific. However we do not allow these parameters to vary across varieties of the same sector. In other words, we assume that $\omega_{ij}(s) = \omega_{ij}(s') = \omega_{ij}$ and $A_i(s) = A_i(s') = A_i$ for all $s \neq s'$.

In order to derive a tractable solution to the model we make the following assumption:

Assumption 1 *The output of variety s of every sector is entirely absorbed by the production of the varieties $s+1$ of the economy. Only the output of varieties "S" is directly consumed. Formally, for every sector i, the following two conditions are assumed to be always true:*

$$\begin{aligned} Q_i(s) &= \sum_{j=1}^{N} M_{ji}(s+1) \qquad && s = 1, 2, \ldots, S-1 \\ Q_i(S) &= C_i \end{aligned}$$

Although extreme, this assumption allows to solve the model and to derive a formula for the aggregate tax distortion. Moreover, it allows to order varieties according their position along the production chain, and it justifies calling the different varieties "stages of production".

Firms are owned by the household, but perfect competition and constant return to scale drive their profits to zero. Zero profit imply that prices equal unit costs. Thus prices are:

$$\begin{aligned} P_i(1) &= \frac{W}{A_i} && s = 1 \\ P_i(s) &= \frac{1}{A_i} \left(\frac{W}{1-\omega_i} \right)^{1-\omega_i} \left[\prod_{j=1}^{N} \left(\frac{P_j(s-1)}{\omega_{ij}} \right)^{\omega_{ij}} \right] && s = 2, 3, \ldots, S \end{aligned} \tag{2}$$

Given the wage W, (2) is a system of $N \times (S+1)$ equations in $N \times (S+1)$ unknown prices. Prices at stage "1" are pinned down by wage and technology alone. Then, all prices from stage "2" through "S" are determined by wage, technology and the prices of the varieties produced by the stage immediately upstream. In appendix 6 we derive explicitly the solution of the price vector at every stage of production.

Government. The government produces a single service with a constant return to scale technology that requires only labor:

$$G = L_G$$

The government provides to the household an exogenous level of this public service $\overline{G}$. The public good is provided free of charge, and it is financed with tax revenues.

We assume that the government can tax the activities of private firms. It can do so in one of two ways. First, it can impose a turnover tax equal to $t\%$ of the value of revenues earned by every firm in the economy. Because in this economy some firms earn revenues by selling intermediate inputs to other firms, a turnover tax is distortionary, because it raises the price of intermediate inputs relative to the price of labor. If the government chooses to raise taxes in this way it will collect tax revenues equal to:

$$T^{tt} = t \sum_{i=1}^{N} \sum_{s=1}^{S} MC_i^{tt}(s) Q_i^{tt}(s)$$

where $Q_i^{tt}(s)$ is the quantity produced of variety s in sector i when the turnover tax is in place and $MC_i^{tt}(s)$ is the marginal cost of producing this variety.

Second, the government can also decide to collect taxes by imposing a consumption tax equal to $v\%$ of the value of purchases of final consumers. Because such a consumption tax does not affect firms producing intermediate inputs, it does not distort production decisions. As a result, the allocation of resources within the private sector will be optimal. If the government chooses to raise revenues with the consumption tax it will collect:

$$T^{v} = v \sum_{i=1}^{N} MC_i^{v}(S) Q_i^{v}(S)$$

where $Q_i^v(S)$ is the quantity of the variety S produced by sector i and sold to final consumers when the private sector is taxed with the consumption tax and $MC_i^v(s)$ is the marginal cost of producing this variety.

Equilibrium. We solve the model as follows. First, we take the wage to be the numeraire and fix it to 1. For a given tax scheme, before-tax prices are given by equations (2), which depend on the wage, technological parameters and the tax scheme. Because labor supply is fixed, the household can spend an income equal to $\overline{L}$. Total income, along with the price of the final goods fixed by equations (2) determines the quantity of each good i the household demands.

All product markets must clear. Thus, demand of consumption by the households determines the quantity produced at the last stage of production. Production of the $N \times S$ varieties produced at stages $S-1, S-2, \ldots, 1$ is then determined recursively by the structure of production implied by assumption 1. Recall that assumption 1 states that the whole production of stage s, $Q_i(s)$ is absorbed in the production of stage $s+1$. Given Cobb-Douglas production, this implies that $V_i(s)$, the value of output of sector i at stage s, satisfies the following relation: $V_i(s) = \omega_{1i}V_1(s+1) + \ldots \omega_{Ni}V_N(s+1)$. Collecting the $V_i(s)$ into the vector $\boldsymbol{V}(s)$, and the ω into the input-output matrix Ω we can write:

$$\boldsymbol{V}(s) = \Omega' \boldsymbol{V}(s+1) \tag{3}$$

I start from stage S, where final demand determines production: $V_i(S) = \beta_i(S)\overline{L}$. From there, the system of equations (3) can be solved backwards for all stages of production $S-1, S-2, \ldots, 1$.

The government can only choose how much of the public service it wants to provide to the household. We do not model this decision explicitly, and simply assume that the government wants to provide a fixed amount of services, equal to $\overline{G}$. Given this decision, the government needs to pay for L_G hours of work, which require tax revenues equal to $T = L_G$. Given the decisions of the firms, this total level of revenues uniquely determines the tax rate. This is t if it chooses to rely on a turnover tax or v if it chooses to rely on the consumption tax.

In this way, the model uniquely determines quantities G, T, L_G, $\{C_i,\ Q_i(s),\ L_i(s), M_{ij}(s)\}$, prices $W, \{P_i(s)\}$ and the tax rate (either t or v).

Welfare gain of the reform. In this section we derive the welfare gain of a tax reform that converts a turnover tax t into a consumption tax with rate v. We assume that the government always wants to raise the same amount of revenues when it switches from the turnover to the consumption tax. Thus, when the turnover tax is converted into a consumption tax, the new tax rate v is fixed so that the total value of revenues collected is not affected:

$$t \sum_{i=1}^{N} \sum_{s=1}^{S} MC_i^{tt}(s) Q_i^{tt}(s) = v \sum_{i=1}^{N} MC_i^{v}(S) Q_i^{v}(S) \tag{4}$$

When the wage is the numeraire, and the government provides a level of public service equal to $\overline{G}$ the indirect utility of the household is equal to:

$$U(P_C) = \left(\frac{\overline{L}}{P_C}\right) \overline{G}^{\alpha}$$

where P_C is the price index of the Cobb-Douglas aggregate (1):

$$P_C = \prod_{i=1}^{N} \left[\frac{P_i(S)}{\beta_i}\right]^{\beta_i} \tag{5}$$

Thus, the welfare gain of the reform can be calculated as the percentage change in consumption of the final goods aggregate. This is summarized by the percentage change in the price index of the aggregate final good P_C:

$$u^v - u^{tt} = p_C^{tt} - p_C^v$$

where a lowercase letter denotes the logarithm of a variable: $x \equiv \log X$ and we use tt and v to label equilibrium variables when taxes are collected on turnover or on consumption respectively. In the appendix we show that given the price equations (2) and the revenue neutrality of the reform (4) the welfare gain of the reform can be written as:

$$\begin{aligned} u^v - u^{tt} = \sum_{i=1}^{N} \beta_i \cdot & \left[\log(1+t) \sum_{j=1}^{N} \lambda_{ij} \right] \\ & + \log \left[1 - \frac{t}{1+t} \sum_{i=1}^{N} \beta_i \left(\sum_{j=1}^{N} \lambda_{ji}^{tt} \right) \right] \end{aligned} \tag{6}$$

where β_i are the consumption share parameters of the Cobb-Douglas aggregator (1) and λ_{ij} and λ_{ij}^{tt} are the ij-th elements of the following two matrices:

$$\Lambda = \left[\sum_{s=0}^{S-1} \Omega^s \right] \qquad \text{and} \qquad \Lambda^{tt} = \left[\sum_{s=0}^{S-1} \left(\frac{1}{1+t} \Omega' \right)^s \right]$$

The first matrix is a measure of how much input was used along the whole chain of production. In equation (6), this measure controls the extent to which the tax increases the cost of all the inputs exchanged along the production line for a given final good i, and through this channel how much it increases its final price relative to its undistorted price. The second matrix summarizes the revenues that can be raised along the production chain.

Discussion. This model allows to extend some of the intuitions introduced in section 2. First, the underlying parameters of production are not directly observable in standard input-output tables. In particular, the input elasticities of the different sectors, the ω_{ij} in the production functions, depend on the input-output flows reported in input-output tables as well as on the total number of stages of production S, which is not observed. Intuitively, for a given observed input share, longer production processes imply that the same inputs must be used across more stages. Thus, for a given observed input share, the input share underlying the production of longer processes must be smaller. In section 4.3 we explain how we calibrate the input elasticities to match the input shares observed in the input-output tables.

Second, it is interesting to note that as the number of stages S approaches infinity, the formula of the matrix that governs the price distortion Λ becomes $[I-\Omega]^{-1}$. The term $[I-\Omega]^{-1}$, appears also in Fally (2012) and Antràs et al. (2012). Fally (2012) in particular, calls it the "length of the production chain" and relates it to the total cumulative transport costs that are associated to trade when there is a constant transport cost on every transaction. The similarity should not surprise, because the turnover tax acts as a constant transport cost in our model.

4.2 Special case: a single final good

Before explaining how we calibrate to the Brazilian economy the model described in the previous section it is useful to study the special case in which $N = 1$ and the economy produces a single final good. This special case allows to gain some intuition over the mechanism that makes the distortion of a turnover tax grow as the production process becomes longer. The special case with a single final good is also a natural extension of the examples developed in section 2, where the production process is not restricted to have either two, three or an infinite number of production stages.

When private firms produce a single final good $Q(S) = C$, household's utility equals:

$$U = \frac{\overline{L}}{P(S)} \cdot G^{\alpha}$$

where $P(S)$ is the price of good $Q(S)$. Production structure in this case is a simplified version of the structure described in section 4.1, and variety $s = 1, 2, \ldots, S$ are produced with the following technologies:

$$\begin{aligned} Q(1) &= AL(1) && s = 1 \\ Q(s) &= AL(s)^{1-\omega} M(s)^{\omega} && s = 2, 3, \ldots, S \end{aligned}$$

Also in this case we simplify the structure of the model and assume the following:

Assumption 2 *The output of variety s is entirely absorbed by the production of the variety $s+1$. Only the output of variety "S" is directly consumed. Formally:*

$$\begin{aligned} Q(s) &= M(s+1) && s = 1, 2, \ldots, S-1 \\ Q(S) &= C \end{aligned}$$

The welfare gain of a reform that converts a turnover tax of rate t into a consumption tax of rate v is in this case:

$$u^v - u^{tt} = p^{tt}(S) - p^v(S)$$

where $p^{tt}(S)$ and $p^v(S)$ are the logarithm of the price of the final good with the turnover tax and with the consumption tax respectively. Using the condition that the government fixes the new tax rate v in order to raise the same amount of revenues collected with the turnover tax, the formula for the welfare gain (6) in this case becomes:

$$u^v - u^{tt} = \log(1+t) \cdot \lambda + \log\left[1 - \frac{t}{1+t}\lambda^{tt}\right] \tag{7}$$

where λ and λ^{tt} are a function of the input elasticity of the only sector of the economy ω and the tax rate of the turnover tax t:

$$\lambda = \left[\sum_{s=0}^{S-1} \omega^s\right] \qquad \text{and} \qquad \lambda^{tt} = \left[\sum_{s=0}^{S-1} \left(\frac{\omega}{1+t}\right)^s\right]$$

Equation (7) allows to develop some intuition over how the length of the production chain S affects the welfare gain of a reform that eliminates a tax on turnover. The equation shows that the overall gain depends on two terms. The first term captures by how much the price of the final good grows as a result of having a tax that makes input of production more expensive. This term depends on λ. In the case of single final good it is easy to show that this parameter is uniquely determined by the input share in the whole economy observed when the turnover tax has been removed and it is not affected by the number of stages of production. In other words, observing the total value of inputs over the gross value of production after the reform has eliminated the turnover tax is sufficient to calibrate the parameter λ, without further information on the number of stages of production S. To see this, write:

$$\begin{aligned} o &= \frac{\sum_{s=2}^{S} X^v(s)}{\sum_{s=1}^{S} V^v(s)} \\ &= \frac{\sum_{s=1}^{S-1} \omega^s}{\sum_{s=0}^{S-1} \omega^s} \\ &= \frac{\lambda - 1}{\lambda} \end{aligned}$$

The second term of the welfare gain captures the revenues that are raised by the turnover tax along the production chain, and it depends on λ^{tt}. Contrary to what happens with λ, for a given observed input share o, the parameter λ^{tt} does depend on the number of stages of production S. In particular, numerical simulations show that in the case of a single final good for a given o, the parameter λ^{tt} is a decreasing function of the number of stages of production S.[9] Because total tax revenue raised with the turnover tax is equal to $T = t\left(\overline{L}\lambda^{tt}\right)/(1+t)$, this implies that the tax base of the turnover tax is a decreasing function of the number of stages of production. Thus, looking at an economy with an observed input share o, a given tax rate of t raises more revenues when the production process has fewer stages. Alternatively, the same level of tax revenue T can be raised with lower tax rates in an economy that has fewer production stages. Both exercises yield the same conclusion: namely, that for a constant input share o the welfare gain of a reform that converts a turnover tax into a consumption tax is greater in longer production processes.

The intuition for the result is the following. Longer production processes allow production to be more specialized across different units of production. This gives greater opportunity to substitute the taxed input with labor, and allows longer production processes to shift production towards the last stage of production to a greater extent. This in turn reduces the overall tax base, and it is the source of the greater welfare gain that we estimate for longer production processes in the next section.

4.3 Welfare gains

In this section we use the model developed in section 4.1 to estimate the welfare gains of the Brazilian 2002-3 tax reform. First, we present the source of the data we use. Next, we explain how we calibrate the parameters of the model to the data and present the estimates for the welfare gain.

Data. We calibrate the main parameters of the model presented in section 4.1 using the Brazilian input-output tables. The tables are compiled by the *Instituto Brasileiro de Geografia e Estatística* (IBGE) based on national accounts statistics (IBGE (2008))[10] and report total value of production, total value of inputs used and total value of goods sold to domestic and foreign consumers for 110 commodities and 55 industries.

We use the Brazilian input-output tables for the year 2005. The original tables are prepared following the international standards and report "rectangular" matrices that allow for the production of many goods by a single sector. The "make" matrix reports total production of each of 110 commodities carried out by 55 different industries, while the "use" matrix reports the consumption of the 110 commodities as inputs by the same 55 industries. We use the "make" matrix (table 1) and the domestic "use" matrix at base

[9] It seems to be possible to prove this result analytically. We did not have time to include this formal proof in the current version of the paper.

[10] See `http://www.ibge.gov.br/home/estatistica/economia/matrizinsumo_produto/default.shtm`.

price (table 3) to compute a "square" matrix of direct requirement coefficients. We construct a 100×100 commodity-by-commodity matrix of direct requirement coefficients by first constructing a 110×110 input-output table and then aggregating the 11 agricultural commodities reported in the table.[11] We follow the "Industry Technology Assumption" (Guo et al. (2013)) and compute the typical entry of matrix O, the value of inputs j used to produce 1 Real of product i (o_{ij}) as:

$$o_{ij} = \sum_{k=1}^{N} \left(\frac{X_{kj}}{\sum_{j'=1}^{N} V_{kj'}} \cdot \frac{V_{ki}}{\sum_{k'=1}^{N} V_{k'i}} \right)$$

where V_{ki} is the value of output of commodity i produced by industry k from the "make" matrix, $\sum_{j'=1}^{N} V_{kj'}$ is the total value of output in industry k, $\sum_{k'=1}^{N} V_{k'i}$ is the total value of the production of commodity i by any industry and X_{kj} is the value of domestically-produced commodity j purchased by industry k as an input from the "use" matrix. In words, the "Industry Technology Assumption" construct the share of inputs j in the production of good i as the weighted share of inputs used by all industries k that produce good i, with weights equal to the share of production of commodity i that is accounted by each industry k. Consumption shares of good i, β_i, are computed using data on the value of goods sold to domestic and foreign consumers.

Data on aggregate revenues collected with COFINS and PIS as a percentage of Brazilian GDP come from the Brazilian Ministry of Finance.[12]

Calibration. With the data from the Brazilian input-output tables, we now use the model of section 4.1 to evaluate the productivity gains of the 2002-3 Brazilian tax reform.

The reform eliminated two turnover taxes with statutory tax rates equal to 3.65% in total. We estimate the welfare gains of this reform using equation (6). In this equation, the consumption shares of the N sectors, β_i, are directly observable from national accounts. However, neither the number of stages of production S, nor the input elasticities collected in the matrix Ω are directly reported in official statistics. Moreover, although in principle the statutory tax rate t is observable, in practice the average tax rate effectively enforced may be lower due to tax evasion. Because the actual welfare gain of the reform depends on the tax rate that is effectively enforced before the policy change, it is important to estimate the gains while allowing the effective tax rate to be different from the statutory one.

In what follows we use the information contained in standard input-output tables to

[11] These are: rice, maize, wheat, sugarcane, soy, manioc, tobacco, cotton, citrus, coffee and other agricultural products.

[12] See `http://www.receita.fazenda.gov.br/Historico/EstTributarios/Estatisticas/default.htm`.

calibrate the input elasticities in Ω. We also use the input-output tables and the observed value of tax revenues collected to calibrate the effective tax rate for the same production processes. Because input-output tables and tax revenues are not enough to calibrate both the input elasticities, the tax rate and the number of stages of production, we proceed in the following way. First, we set the number of stages of production in our model as an exogenous parameter. Next, we calibrate the remaining parameters (Ω and t) to match the data contained in the input-output tables and in the tax revenue record. This allows us to derive the production gains of eliminating a tax on turnover as a function of the number of stages of production. We now illustrate the calibration of Ω and t.

The Brazilian input-output tables report information about "O": the matrix collecting the "direct coefficients" of the N sectors in the economy. For any sector i in the economy, the direct coefficient o_{ij} contains information about the total value of inputs purchased by sector i from sector j per each dollar of production of sector i. Calling V_i the gross value of production of sector i and X_{ij} the total value of inputs that sector i purchases from sector j, o_{ij} is:

$$o_{ij} = \frac{X_{ij}}{V_i} \tag{8}$$

The procedure described here shows how, once the number of stages S is fixed, the production parameters in Ω can be calculated from the matrix of direct coefficients O.

In the model of section 4.1, inputs are purchased at every stage of production after the first one. At the same time, output is sold at every stage. Thus, equation (8) is equal to:

$$o_{ij} = \frac{\sum_{s=2}^{S} X_{ij}(s)}{\sum_{s=1}^{S} V_i(s)}$$

where $X_{ij}(s)$ is the value of inputs purchased by firms producing at stage s in sector i from firms producing at stage $s-1$ in sector j and $V_i(s)$ is the gross output of firms producing at stage s in sector i. Because ω_{ij} are assumed to be identical across stages, $X_{ij}(s)/V_i(s) = \omega_{ij}$ for all stages after the first one. Thus, the last equation can be written as:

$$\begin{aligned} o_{ij} &= \omega_{ij}\left(\sum_{s=2}^{S} \frac{V_i(s)}{V_i}\right) \\ &= \omega_{ij}\left(1 - \frac{V_i(1)}{V_i}\right) \end{aligned}$$

Equation (9) makes clear that the observed direct coefficients are equal to the underlying input elasticities times the share of gross output produced by stages that use intermediate inputs in production. Intuitively, the direct coefficients reported in the input-output matrix are equal to a weighted average of the input elasticities of every stage of production, with weights equal to the share of production of every stage. Because stage 1 only uses labor, its input elasticities are all equal to 0. Moreover, because within a sector, input elasticities are identical across the following S stages of production, direct coefficients are equal to the underlying input elasticities times the share of output produced by the stages following the first one. For production processes that involve a large number of stages, the share of output contributed by stage 1 becomes negligible, and observed direct coefficients become identical to the true input elasticities.

Collecting equations (9) into matrices, the true parameters ω of the model must solve the following system of equations:

$$\begin{bmatrix} o_{11} & o_{12} & \dots & o_{1N} \\ o_{21} & o_{22} & \dots & o_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ o_{N1} & o_{N2} & \dots & o_{NN} \end{bmatrix} =$$

$$\begin{bmatrix} 1-\theta_1(1) & 0 & \dots & 0 \\ 0 & 1-\theta_2(1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1-\theta_N(1) \end{bmatrix} \cdot \begin{bmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1N} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N1} & \omega_{N2} & \dots & \omega_{NN} \end{bmatrix} \tag{9}$$

where $\theta_i(1) \equiv V_i(1)/V_i$ is the share of production of the first stage in sector i.

There are two challenges to solve the system (9). First, we must express the shares of output at the first stage, $\theta_i(1)$, in terms of the parameters of the model. This is made possible by the sequential structure imposed to the model by assumption 1, that implies that the vector of output of the N sectors at stage s can be written as $\boldsymbol{V}(s) = \Omega' \boldsymbol{V}(s+1)$ for all stages from 1 through $S-1$. Solving the system recursively, the vector of output at stage 1 is: $\boldsymbol{V}(1) = (\Omega')^{S-1} \boldsymbol{V}(S)$. This in turn means that the vector $\boldsymbol{\theta}(1)$ of shares of output at stage 1 depends on the input elasticities and on the number of stages of production:

$$\boldsymbol{\theta}(1) = (\Omega')^{S-1} \boldsymbol{\iota} \tag{10}$$

where $\boldsymbol{\iota}$ is a $N \times 1$ vector of ones.

The second issue to address is that for large input-output matrices, the system (9) is hard to solve numerically. We use Brazilian input-output tables with 100 industries: this result in a system of 10000 non-linear equations in 10000 unknowns. We make the system tractable by exploiting the fact that the ratio of any two input elasticities of a given sector i depends only on observables direct coefficients. Cobb-Douglas production in fact implies:

$$\frac{\omega_{ij}}{\omega_{ip}} = \frac{o_{ij}}{o_{ip}} \tag{11}$$

Thus, if we select an input j that is used by every other sector in the economy (i.e. such that $o_{ij} \neq 0$ for all sectors i), we can re-write the matrix Ω as:

$$\begin{bmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1N} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N1} & \omega_{N2} & \dots & \omega_{NN} \end{bmatrix} =$$

$$\begin{bmatrix} \omega_{1j} & 0 & \dots & 0 \\ 0 & \omega_{2j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_{1j} \end{bmatrix} \cdot \begin{bmatrix} \frac{o_{11}}{o_{1j}} & \frac{o_{12}}{o_{1j}} & \dots & \frac{o_{1N}}{o_{1j}} \\ \frac{o_{21}}{o_{2j}} & \frac{o_{22}}{o_{2j}} & \dots & \frac{o_{2N}}{o_{2j}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{o_{N1}}{o_{Nj}} & \frac{o_{N2}}{o_{Nj}} & \dots & \frac{o_{NN}}{o_{Nj}} \end{bmatrix} \tag{12}$$

This way of expressing matrix Ω allows us to write system (9) as a system of only 100 non-linear equations in 100 unknowns. This system is solvable if there is at least one sector whose inputs are used by every other sector in the economy: this is necessary in order to avoid that any of the entries in the second matrix of equation (12) has a denominator equal to 0. We select sector "information services", and estimate the input shares of this input for all 100 sectors of the economy. This sector is one of the 7 sectors whose output is used by every other sector in the economy as input. Next, we use these estimates to back out all the other 9900 parameters of the Ω matrix, using equations (11).

To summarize, we use the structure of the model in section 4.1 and calibrate the input elasticities of production so that they match the observed direct coefficient reported in the input-output matrix of Brazil. Thus, the ω_{ij} are the solutions to the system (9), where the $\theta_i(1)$ are defined by (10). We solve the system numerically on a 100×100 matrix by substituting the matrix Ω with the two matrices on the right-hand side of (12): this reduces the complexity of the model and allows us to solve it for only 100 unknowns. Finally, we use equations (11) to retrieve all the other unknowns of system (9).

I now explain how we calibrate the effective tax rate of the turnover tax. From the public records made available by the Brazilian Ministry of Finance we observe the value of revenues raised by COFINS and PIS as a share of the Brazilian GDP in the years before the reform. In the years 2001 and 2002 total revenues from these two turnover tax was equal to 4.63% and 4.62% respectively. Thus, we set $T/\overline{L}$ in the model to 0.046, and set the effective tax revenue of the turnover tax to hit this aggregate value. Because we still do not have enough information to calibrate the number of stages of production we proceed as for the calibration of Ω. First, we set the number of stages of production as an exogenous parameter. Next, we use the procedure outlined above to back out the implied input elasticities ω_{ij}. Finally, we use the definition of tax revenues in our model, and find the effective tax rate t that raises 4.6% of national income in turnover taxes, for the assumed number of stages. This requires solving the following equation in t:

$$\begin{aligned} 0.046 &= \sum_{s=1}^{S} MC_i^{tt}(s) Q_i^{tt}(s) \\ &= \frac{t}{1+t} \left[\sum_{i=1}^{N} \beta_i \left(\sum_{j=1}^{N} \lambda_{ji}^{tt} \right) \right] \end{aligned}$$

where λ_{ij}^{tt} is the ij-th entry of the matrix Λ^{tt} defined in section 4.1 and depends on the number of stages of production S.

Results. Once we calibrate Ω and t, we use the formula (6) to estimate the welfare gains of the reform. We do this for several exogenous production lengths, in order to evaluate how the number of stages affects the overall welfare gains of the reform. Baseline results are shown in figure 6, where we plot on the horizontal axis the number of stages on which we calibrate Ω and t, and on the vertical axis the welfare gain of converting a turnover tax that raises revenues equal to 4.6% of national income into a consumption tax. The graph shows that the welfare gain of the reform is 4.1 times larger when the production process has an infinite number of stages compared to a production process in which there are only two stages of production. Most of this increase happens relatively quickly: the welfare gain of the reform is 3.95 times larger when the production process involves 8 stages compared to when the production process involves only two stages. Overall, for the input-output table observed in Brazil, the productivity gains of the tax reform remain small even when production chains are infinitely long. For example, eliminating the tax on turnover increases aggregate production by at most 0.05%, when the production chain is infinitely long.

This baseline estimate is obtained by calibrating an input-output with 100 sectors. An interesting question is what would have been the welfare gain of the reform had we treated Brazil as a single-sector economy with an overall intermediate inputs intensity ω equal to the input share observed in the aggregated input-output matrix. This is an interesting exercise because many of the developing countries that make use of turnover taxes do not produce detailed input-output matrices. To answer this question, we adapt equation (6) to the special case in which there is a single final good produced, and calculate the welfare gain of the reform using equation (7). As for the model with many sectors, we set the number of stages as an exogenous parameter, and calibrate the input elasticity ω so that the observed input share in the undistorted economy o is equal to 43%, the observed input share in Brazil in 2005. For the same number of stages, we also set the effective tax rate t equal to the tax rate that would raise revenues equal to 4.6% of national income.

The results of this exercise are shown in figure 7. The figure plots for any given number of stages of production (on the horizontal axis) the welfare gain of the reform for a single-sector economy with an overall intermediate inputs intensity of production of 43% (on the vertical axis). To ease comparison with the previous exercise, we plot in the same figure the welfare gain obtained calibrating the model with 100 sectors shown in figure 6. The graph shows that the welfare gain of the reform are several times larger in longer production processes: in the single-sector model, the welfare gains of the reform are 7.1 times larger when the number of stages converges to infinity than when inputs are exchanged a single time. Interestingly, for any given production length, the welfare gain of the reform are always larger when we calibrate the model using the information from the full input-output matrix. However the difference between the two estimates is small: when the number of stages converges to infinity, the welfare gain calculated using the information from the full input-output matrix is only 1.14 times larger than the welfare gains obtained assuming that Brazil is an economy that produces a single final good.

5 Constant elasticity of substitution production

In the model of this section we relax the assumption of Cobb-Douglas production. The objective is to evaluate how the elasticity of substitution between labor and intermediate inputs affects the welfare gains of the reform.

5.1 The model

The model developed in this section is identical to the model of section 4.2, except for the production function, which is assumed to have a more general structure.

Households. Preferences are identical to the ones derived in the special case with a single final good in section 4.2. The representative consumer has preferences defined over a single final good C supplied by private firms and a public good G provided by the government:

$$U = C \cdot G^{\alpha}$$

The household supplies inelastically $\overline{L}$ hours of work, receiving an hourly wage of W. Because we set the wage to be the numeraire, total income in the economy is $\overline{L}$.

Firms. The model has a single sector that produces S varieties.[13] Each variety is produced under perfect competition. Production of the first variety requires only labor. The following varieties are produced combining labor and intermediate inputs using a generalized constant elasticity of substitution (CES) function. Thus, technology is described by:

$$\begin{aligned} Q(1) &= BL(1) && s = 1 \\ Q(s) &= B\left[L(s)^{\frac{\eta-1}{\eta}} + \delta M(s)^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} && s = 2, 3, \ldots, S \end{aligned} \tag{13}$$

In (13) B is a Hicks-neutral productivity parameter, $L(s)$ and $M(s)$ are labor and intermediate inputs used by firms producing variety s, η is the elasticity of substitution between the two factors of production and δ is a distribution parameter that governs the importance of intermediate input relative to labor. We assume that the productivity parameter B is identical across all varieties and that both the distribution parameter δ and the elasticity of substitution η are the same across all varieties that use intermediate inputs.

In order to solve the model we make the following assumption:

Assumption 3 *The output of every variety s is entirely used as an intermediate input by the production of variety $s+1$. Only the output of variety S is consumed. Formally:*

$$\begin{aligned} Q(s) &= M(s+1) && s = 1, 2, \ldots, S-1 \\ Q(S) &= C \end{aligned}$$

As assumption 1, assumption 3 gives to the production process a clear sequential structure. It also allows to pin down the price of every variety produced. The price of variety 1 is determined by the wage alone. The price of the following varieties is then determined recursively with the wage and the price of the variety produced immediately upstream. Formally this can be written:

$$\begin{aligned} P(1) &= B^{-1}W \\ P(s) &= B^{-1}\left[W^{1-\eta} + \frac{\gamma}{A}P(s-1)^{1-\eta}\right]^{\frac{1}{1-\eta}} && s = 2, 3, \ldots, S \end{aligned} \tag{14}$$

where we define $A \equiv B^{1-\eta}$ and $\gamma \equiv \delta^{\eta}$.

[13] Extending the theory to allow for N sectors, each producing many varieties combining labor and inputs from all other sectors with arbitrary elasticity of substitution is relatively easy. However, the calibration of the model to the observed input-output matrix presents some challenges that we have still not been able to solve. In section 4.3 we have shown that considering more than one sector does not have a large effect on the estimated welfare gains of the reform. In this section we explore whether maintaining the single-sector assumption and allowing the elasticity of substitution to vary has larger effects on the welfare gains of the reform.

Government. As in section 4.1, the government produces a single service G with the following constant return to scale technology:

$$G = L_G$$

Of this service, the government provides the household an exogenous level $\overline{G}$, free of charge. The government finances the production of this services with tax revenues.

As in the previous model, we assume that the government can chose to raise tax with either a turnover tax or a tax on consumption. A turnover tax with rate t would raise total revenues equal to:

$$T^{tt} = t \sum_{s=1}^{S} MC^{tt}(s) Q^{tt}(s)$$

A tax on consumption with a rate of v would raise total revenues equal to:

$$T^{v} = vMC^{v}(S)Q^{s}(S)$$

In the following we consider welfare gains of a reform that converts a turnover tax into a consumption tax, while leaving the total revenue raised by the tax unaffected. Before that, we briefly explain how the model solves.

Equilibrium. Income is equal to $W\overline{L} = \overline{L}$. The prices of the varieties produced at every stage are uniquely determined by the wage, the structure of production and the type of taxation. Thus, the system of prices (14) determines the price of the consumption good, which is the variety produced at the last stage of production. Given this price and the income in the economy, the household's preferences will determine the demand of the good produced at the last stage of production. Once demand of the last variety is determined, the market clearing conditions and the structure of production determine the quantity produced at every stage of production.

Welfare gain of the reform. We derive here the expressions to calculate the welfare gain of a reform that converts a turnover tax into a consumption tax. As in the exercises of section 4, we assume that the reform leaves the overall level of revenues unaffected:

$$t \sum_{s=1}^{S} MC^{tt}(s) Q^{tt}(s) = vMC^{v}(S)Q^{v}(S)$$

This also means that the government will be able to produce and provide the same amount of public good before and after the reform. Because income in units of labor is unaffected by the reform, the welfare gain of the reform can be written as a function of the change in price of the final good:

$$u^v - u^{tt} = p^{tt}(S) - p^v(S)$$

In the appendix 6 we show that the welfare gain in this model can be written as:

$$\begin{aligned} u^v - u^{tt} &= \log\left[\frac{(1+t)MC^{tt}(S)}{(1+v)MC^v(S)}\right] \\ &= \log\left\{(1+t)\cdot\frac{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^S}{1-\frac{\gamma}{A}(1+t)^{1-\eta}}\cdot\frac{1-\frac{\gamma}{A}}{1-\left(\frac{\gamma}{A}\right)^S}\times\right. \\ &\left.\left[1-\frac{t}{1+t}\left(1+\sum_{s=1}^{S-1}\prod_{k=s+1}^{S}\frac{\omega^{tt}(k)}{1+t}\right)\right]\right\} \end{aligned} \tag{15}$$

In (15) $\omega^{tt}(k)$ is the cost share of intermediate inputs at stage k:

$$\omega^{tt}(k) = \frac{\gamma\left[P^{tt}(k-1)\right]^{1-\eta}}{1-\gamma\left[P^{tt}(k-1)\right]^{1-\eta}} \tag{16}$$

In the appendix 6 we also show that given the way prices are set in this economy, the sum of products in the second line of equation (15) can be written as a function of the parameters of the model:

$$\begin{aligned} \sum_{s=1}^{S-1}\prod_{k=s+1}^{S}\frac{\omega^T(k)}{1+t} =& \frac{\left[\frac{\gamma}{A}(1+t)^{-\eta}\right]^S}{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^S}\times \\ & \left\{\frac{\left[\frac{\gamma}{A}(1+t)^{-\eta}\right]^{1-S}-1}{1-\frac{\gamma}{A}(1+t)^{-\eta}}+\frac{1+t-(1+t)^S}{t}\right\} \end{aligned} \tag{17}$$

In the next section we explain how we calibrate the parameters of this model to the Brazilian economy and use equation (15) to calculate the welfare gains of the 2002-3 reform.

5.2 Welfare gains

Calibration. In this section we explain how we estimate the welfare gain of the 2002-3 Brazilian tax reform using equation (15) and data from the Brazilian economy.

As in section 4.3, we use two observable moments to calibrate the model. The first one is the intermediate input share of the Brazilian economy, as observed after the 2002-3 reform. The second one is the revenues raised by turnover taxes before the reform as a share of Brazilian GDP. As it was the case for the model described in section 4, these variables are not sufficient to calibrate all the parameters of the model: the number of stages of production S, the elasticity of substitution between intermediate inputs and labor η, the distribution parameter δ, the productivity B and the effective tax rate t. For this reason, we extend the approach adopted in section 4.3 and proceed as follows. First, we set the number of stages of production and the elasticity of substitution as exogenous parameters. Next, we use information on the intermediate input share coming from the input-output tables along with the structure of the model to calibrate the parameters of the production function. This procedure does not allow to identify separately the distribution parameter δ and the technology parameter B. Instead, it allows to set the joint parameter $\gamma/A = \delta^\eta/B^{1-\eta}$. Because in equations (15) and (17) δ and B always enter as $\delta^\eta/B^{1-\eta}$, this joint parameter is sufficient to calculate the welfare gain of the reform. Finally, we use information about the tax revenue collected before the reform to calibrate the effective tax rate t. This approach is equivalent to the one we adopted in section 4.3, where we also set the tax rate in order to match the available data on tax revenues. The procedure builds on the idea that the information on the total value of tax revenues can be observed with greater precision than the tax rate effectively paid by Brazilian firms before the reform.

We start by describing the procedure to calibrate γ/A. Since the economy produces a single final good, we use the information on the aggregate share of gross output that is used as intermediate input by any sector of the economy after the tax reform. If V is gross output and X is gross value of intermediate inputs, this share is:

$$\begin{aligned} o &= \frac{X}{V} \\ &= \frac{\sum_{s=2}^{S} X(s)}{\sum_{s=1}^{S} V(s)} \end{aligned}$$

Multiplying and dividing every term in the numerator by the value of of all varieties produced in the following stages, we can write the observed input share as a function of the underlying intermediate input shares at all stages of production, $\omega(k)$:

$$\frac{o}{1-o} = \sum_{s=1}^{S-1} \prod_{k=s+1}^{S} \omega(k) \tag{18}$$

When the production function has constant elasticity of substitution, the terms $\omega(k)$ on the right-hand side of equation (18) are:

$$\omega(k) = \frac{\gamma}{A} \frac{1-\left(\frac{\gamma}{A}\right)^{k-1}}{1-\left(\frac{\gamma}{A}\right)^{k}}$$

Plugging this expression of $\omega(k)$ back into equation (18), it is possible to write the input share o observed after the reform as a function of the parameters of the model (see also the derivations in appendix 6):

$$\frac{o}{1-o} = \frac{\left(\frac{\gamma}{A}\right)^S}{1-\left(\frac{\gamma}{A}\right)^S}\left[\frac{\left(\frac{\gamma}{A}\right)^{S-1}-1}{1-\frac{\gamma}{A}} - S + 1\right] \tag{19}$$

Equation (19) is the key expression used to calibrate the parameters γ/A in the constant elasticity of substitution model. Because in the 2005 Brazilian input-output tables we observe an aggregate input share equal to 43%, we set γ/A equal to the value that solves (19) when $o = 0.43$. We do this for several different values for the elasticity of substitution and the number of stages.

In the second step of the calibration, we use the values of γ/A that we have found in the first step of the procedure along with the tax revenue observed before the reform to calibrate the effective tax rate t. We do so by writing down the value of tax revenue as a share of total income as a function of the cost shares of every stage of production:

$$\begin{aligned}
\frac{T}{L} &= t\sum_{s=1}^{S} MC^{tt}(s)Q^{tt}(s) \\
&= \frac{t}{1+t}\left[1+\sum_{s=1}^{S-1}\prod_{k=s+1}^{S}\frac{\omega^{tt}(k)}{1+t}\right]
\end{aligned}$$

Using the expression for $\omega^{tt}(k)$ reported in (16), the last equation can be written as (see also the derivations in appendix 6):

$$\begin{aligned}
\frac{T}{L} =& \frac{t}{1+t}\left\{1+\frac{\left[\frac{\gamma}{A}(1+t)^{-\eta}\right]^S}{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^S}\times\right. \\
&\left.\left[\frac{\left[\frac{\gamma}{A}(1+t)^{-\eta}\right]^{S-1}-1}{1-\frac{\gamma}{A}(1+t)^{-\eta}}+\frac{1+t-(1+t)^S}{t}\right]\right\}
\end{aligned} \tag{20}$$

Equation (20) is the key expression used to calibrate the effective tax rate t in the constant elasticity of substitution model. Because Brazilian turnover taxes collected around 4.61% and 4.63% of GDP in the two years leading to the tax reform, we set the left hand side of equation (20) equal to 0.046, and for every number of stages, elasticity of substitution and corresponding γ/A, we set the effective tax rate t equal to the value that solves equation (20).

Results. Once we calibrate γ/A and the tax rate, we use equation (15) to evaluate the welfare gains of the 2002-3 tax reform. As in section 4.3, we perform this exercise for several different values of production length and elasticity of substitution, in order to analyze how these two parameters affect the welfare gains.

The baseline results of this exercise are shown in figure 8, where we plot on the horizontal axis the number of stages of production and on the vertical axis the welfare gains implied by equation (15). In the figure, we plot the welfare gains for three different values of the elasticity of substitution η: 0.5 (in blue), 1 (in green) and 5 (in red). An elasticity of substitution equal to 0.5 corresponds to a case in which labor and the taxed input are complement in production. The other two values of η imply that labor and the intermediate input are substitutes in production. The unit-elasticity case reproduces the single-sector Cobb-Douglas model discussed in section 4.2, while the case with $\eta = 5$ allows for greater substitutability than in the Cobb-Douglas model.

Figure 8 shows that the welfare gains are larger the higher the elasticity of substitution. When the number of stages goes to infinity, an elasticity of substitution equal to 0.5 implies welfare gains that are half of those obtained in the Cobb-Douglas model (50.11%). When the elasticity of substitution is equal to 5 instead, the welfare gains of the reform in an economy that produces with an infinitely long production process are 4.8 times larger than with the Cobb-Douglas technology, and equal to 0.23% of national income.

The welfare gain of the reform becomes larger as the elasticity of substitution grows because the easier it is to substitute intermediate inputs with labor, the more a tax on turnover distorts production decisions. In the extreme case in which production at all stages after the first one is carried out with a Leontief technology ($\eta = 0$) a tax on turnover does not distort decisions, and for this reason it creates no loss at all. On the other hand, the relationship is not increasing for all values of elasticity of substitution and also when the elasticity is very high, production losses of a turnover tax turn out to be small. This is because when intermediate inputs are easy to substitute with labor, the two inputs have a similar effect on production, and the change in the input mix caused by a tax on turnover does not affect overall production very much. This inverse-U shaped relationship between elasticity of substitution and welfare gains is shown in figure 9. In this figure we plot on the horizontal axis the elasticity of substitution, while on the vertical axis we show how the welfare gain changes as a function of it. The figure shows the relationship between these two variables for three different production structures: one in which production is carried out in 3 stages (in blue), one in which production requires 5 stages (in green) and one in which production requires an infinite number of stages (in red). The inverse-U shape pattern appears in the three production structures. Interestingly, the elasticity of substitution that maximizes the welfare gains depends on the production structure, and it is larger for shorter production processes. In particular, the welfare gains of the reform are highest in the 3-stages economy when the elasticity of substitution between labor and intermediate inputs is equal to 50. In this case, the welfare gain are much higher than in the Cobb-Douglas case, and equal to 1.02% of national income. The highest welfare gain is reached for lower elasticities in the economies that produce with 5 and an infinite number of stages stages: 39 and 37 respectively.

Overall, this exercise suggests that the elasticity of substitution plays an important role in determining the welfare gains of the reform. In particular, if labor and intermediate inputs are highly substitutable, the welfare gains calculated with the Cobb-Douglas model may significantly underestimate the actual effects of the reform.

6 Discussion

In this paper we use a tax reform in Brazil to study the welfare gain of eliminating distortionary turnover taxes. The main challenge when quantifying the welfare gain of such a reform is that when the production process is sequential, the welfare gain of removing turnover taxes depends on the number of stages of production. Our approach is to develop a model of sequential production and calibrate it to the Brazilian economy. We find that when we consider production processes that have 11 or more stages of production the productivity gains of removing the turnover tax are around 4 times larger than when we consider a production process with a single stage. But overall, even with infinitely long production processes, gains from the reform are modest and equal to around 0.05% of national income.

The results of the paper have implications for tax design, especially in developing countries. Taxes on turnover are popular in countries where the informal sector is large and tax capacity low, because they are easier to administer and enforce than more complex tax schemes such as taxes on profits (Best et al. (2013)). Since taxes on turnover are known to be inefficient (Diamond and Mirrlees (1971a,b),Keen (2009)), their diffusion must reflect a belief among policy makers that the benefits in terms of higher tax compliance more than offset the productivity loss created by these taxes. Our results suggest that productivity losses are modest, and they may justify maintaining these schemes in cases in which informality is a major concern. Moreover, the gain we estimated for Brazil may be considered as an upper bound for the productivity loss of a tax on turnover for two reasons. First, Brazil in the early 2000s was relatively developed, and had a diversified economy in which sectors were connected in relatively complex production chains. To the extent that the economies of poorer countries have simpler production structures with fewer input-output linkages, the productivity loss of a tax on turnover are likely to be smaller. Second, the combined tax rates of PIS and COFINS in Brazil in 2002 were 3.65%. This tax rate is high compared to turnover taxes observed elsewhere in the world, where tax rates for turnover taxes usually do not exceed 1%. For these reasons, the welfare gain of eliminating turnover taxes in other developing countries is likely to be smaller than the one calculated here for Brazil.

These conclusions have two limitation. First, in the model we do not consider the potential effects of turnover taxes on vertical integration, which in turn could have effects on productivity.[14] This would be an interesting avenue of future research. Second, the model derives all results by assuming perfect competition at every stage of production. Relaxing this assumption and allowing different market structures would be another promising avenue for future research.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Acemoglu, D., Griffith, R., Aghion, P., and Zilibotti, F. (2010). Vertical integration and technology: theory and evidence. *Journal of the European Economic Association*, 8(5):989–1033.

Acemoglu, D., Johnson, S., and Mitton, T. (2009). Determinants of vertical integration: financial development and contracting costs. *Journal of Finance*, 64(3):1251–1290.

Antràs, P. and Chor, D. (2013). Organizing the global value chain. *Econometrica*, 81(6):2127–2204.

Antràs, P., Chor, D., Fally, T., and Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review, Papers & Proceedings*, 102(3):412–416.

Antràs, P. and Helpman, E. (2004). Global sourcing. *Journal of Political Economy*, 112(3):552–580.

Baldwin, R. and Venables, A. J. (2013). Spiders and snakes: offshoring and agglomeration in the global economy. *Journal of International Economics*, 90(2):245–254.

Banerjee, A. V. and Duflo, E. (2005). Growth Theory through the Lens of Development Economics. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 1 of *Handbook of Economic Growth*, chapter 7, pages 473–552. Elsevier.

[14]The literature here is extensive: theoretical contributions are: Grossman and Hart (1986); Hart and Moore (1990); Antràs and Helpman (2004). Empirical contributions include: Woodruff (2002); Acemoglu et al. (2009, 2010); Macchiavello (2012).

Bartelme, D. and Gorodnichenko, Y. (2014). Linkages and economic development. Mimeo.

Basu, S. (1995). Intermediate goods and business cycles: Implications for productivity and welfare. *American Economic Review*, 85(3):512–31.

Besley, T. and Persson, T. (2013). Taxation and development. In Alan J. Auerbach, Raj Chetty, M. F. and Saez, E., editors, *Handbook of Public Economics, Vol. 5*, pages 51 – 110. Elsevier.

Best, M. C., Brockmeyer, A., Kleven, H. J., Spinnewijn, J., and Waseem, M. (2013). Production vs revenue efficiency with limited tax capacity: Theory and evidence from pakistan. Mimeo.

Bovenberg, A. L. (1987). Indirect taxation in developing countries: A general equilibrium approach. *Staff Papers-International Monetary Fund*, 34(2):333–373.

Ciccone, A. (2002). Input chains and industrialization. *Review of Economic Studies*, 69(3):565–587.

Corden, W. (1966). The structure of a tariff system and the effective protective rate. *Journal of Political Economy*, 74:221–237.

De Paula, A. and Scheinkman, J. A. (2010). Value-added taxes, chain effects, and informality. *American Economic Journal: Macroeconomics*, 2(4):195–221.

Diamond, P. A. and Mirrlees, J. A. (1971a). Optimal taxation and public production I: Production efficiency. *American Economic Review*, 61(1):8–27.

Diamond, P. A. and Mirrlees, J. A. (1971b). Optimal taxation and public production II: Tax rules. *American Economic Review*, 61(3):261–78.

Emran, M. S. and Stiglitz, J. E. (2005). On selective indirect tax reform in developing countries. *Journal of Public Economics*, 89(4):599–623.

Fally, T. (2012). Production staging: Measurement and facts. Mimeo.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772.

Gottfried, P. and Wiegard, W. (1991). Exemption versus zero rating: A hidden problem of VAT. *Journal of Public Economics*, 46(3):307–328.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719.

Guo, J., Lawson, A. M., and Planting, M. A. (2013). *From make-use to symmetric IO tables: an assessment of alternative technology assumptions*. BiblioGov.

Harberger, A. C. (1962). The incidence of the corporation income tax. *Journal of Political Economy*, 70(3):215–240.

Hart, O. D. and Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98(6):1119–1158.

Hirschman, A. O. (1958). *The strategy of economic development*. Yale University Press New Haven.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4):1403–1448.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *Review of Economic Studies*, 45(3):511–518.

IBGE (2008). *Matriz de Insumo-Produto — Brasil 2000/2005*, volume 23 of *Contas Nacionais*. Instituto Brasileiro de Geografia e Estatística - IBGE, Rio de Janeiro, Brazil.

IMF (2002). IMF approves US$ 30.4 billion stand-by credit for Brazil. Press Release 02\40.

IMF (2003). IMF approves 15-month extension, and US$6.6 billion augmentation of Brazil's stand-by credit. Press Release 03\217.

Jones, C. I. (2011). Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomics*, 3(2):1–28.

Jones, C. I. (2013). Misallocation, economic growth, and input-output economics. In Acemoglu, D., Arellano, M., and Dekel, E., editors, *Advances in Economics and Econometrics, Vol. II*, pages 419–454. Cambridge University Press.

Keen, M. (2009). What do (and don't) we know about the value added tax? A review of Richard M. Bird and Pierre-Pascal Gendron's the VAT in developing and transitional countries. *Journal of Economic Literature*, 47(1):159–70.

Keen, M. (2013). *Targeting, Cascading, and Indirect Tax Design*. International Monetary Fund.

Keen, M. and Lockwood, B. (2006). Is the VAT a money machine? *National Tax Journal*, 59(4):905–28.

Leontief, W. W. (1936). Quantitative input and output relations in the economic systems of the United States. *Review of Economic Statistics*, 18(3):105–125.

Long, J. B. and Plosser, C. I. (1983). Real business cycles. *Journal of Political Economy*, 91(1):39–69.

Macchiavello, R. (2012). Financial development and vertical integration: theory and evidence. *Journal of the European Economic Association*, 10(2):255–289.

Newbery, D. M. (1986). On the desirability of input taxes. *Economics Letters*, 20(3):267–270.

Oberfield, E. (2011). Business networks, production chains, and productivity: A theory of input-output architecture. Working paper, Federal Reserve Bank of Chicago.

Piggott, J. and Whalley, J. (2001). VAT base broadening, self supply, and the informal sector. *American Economic Review*, 91(4):1084–1094.

Pomeranz, D. (2013). No taxation without information: Deterrence and self-enforcement in the value added tax. Technical Report w19199, National Bureau of Economic Research.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Werneck, R. L. F. (2006). An evaluation of the 2003 tax reform effort in Brazil. *Revista de Economia Política*, 26(1):75–94.

Woodruff, C. (2002). Non-contractible investments and vertical integration in the Mexican footwear industry. *International Journal of Industrial Organization*, 20(8):1197–1224.

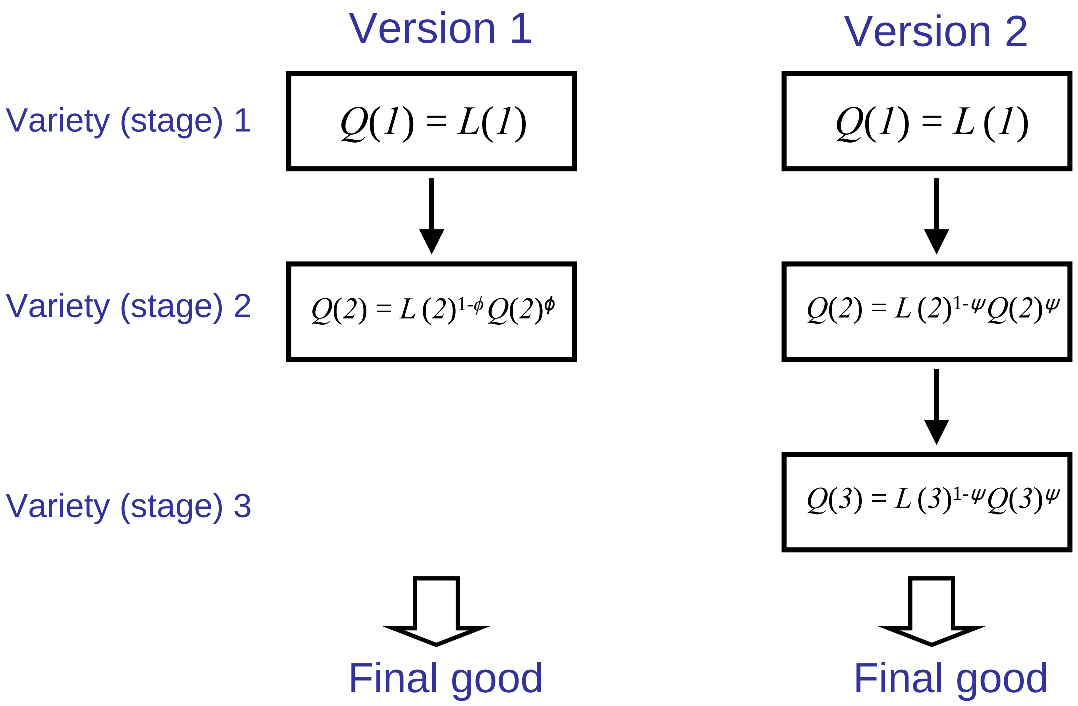
World Bank (2006). *Doing business 2007: How to reform.* World Bank Publications.

World Bank (2012). *Doing business 2013: smarter regulations for small and medium-size enterprises.* World Bank Publications.

Yi, K.-M. (2003). Can vertical specialization explain the growth of world trade? *Journal of Political Economy*, 111(1):52–102.

Figures and tables

Figure 1
Two versions of the one sector economy



Notes: The figure illustrates the basic structure of the two versions of the economy described in section 2. On the left, it is represented the production process when the economy produces 2 varieties (stages). In this case, the observed input flow X is equal to the output of variety 1 sold as intermediate input to the producers of variety 2. The observed gross output V is equal to the output of variety 1 plus the output of variety 2. On the right, it is represented the production process when the economy produces 3 varieties (stages). In this case, the observed input flows X are equal to the output of variety 1 sold as intermediate input to the producers of variety 2 plus the output of variety 2 sold as intermediate input to the producers of variety 3. The observed gross output V is equal to the sum of the output of varieties 1, 2 and 3.

Analytical derivations of Cobb-Douglas model

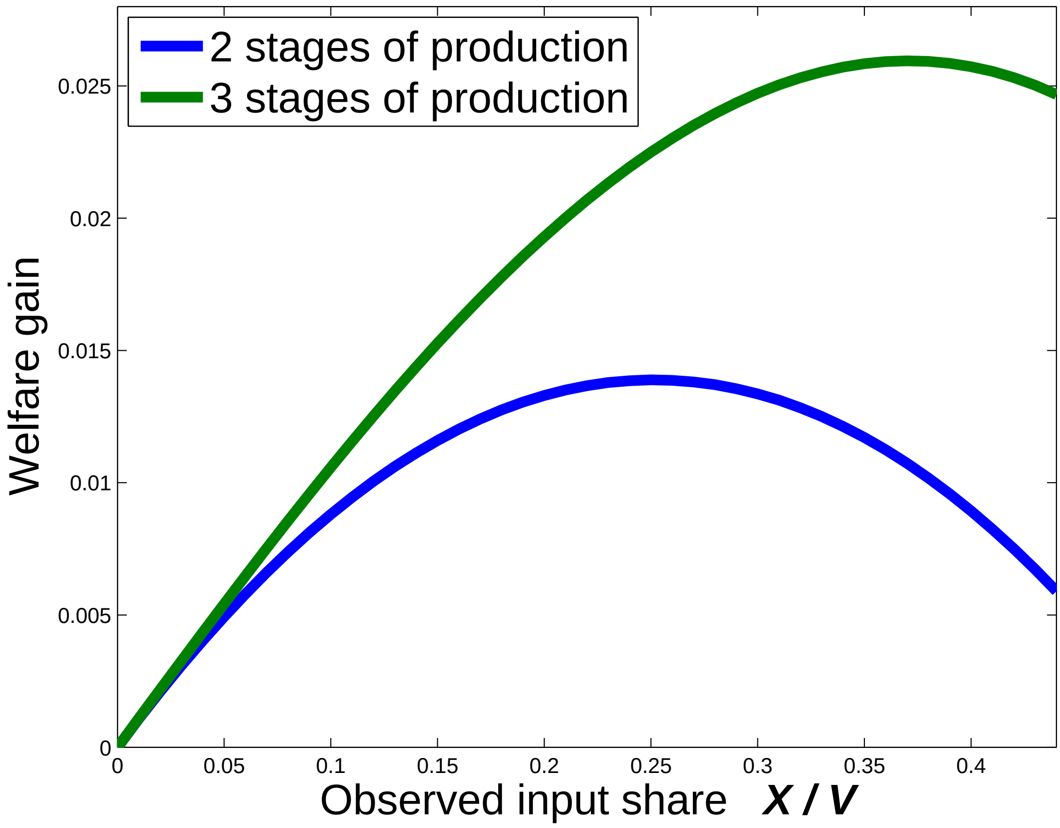
In this appendix we derive explicitly equation (6): the welfare gains of a reform that converts a turnover tax into a consumption tax in a model with a full input-output matrix.

The indirect utility formula derived in section 4.1 implies that the relative change in

Figure 2

Welfare gain of the tax reform in the one sector example

Welfare gain with two and three stages of production

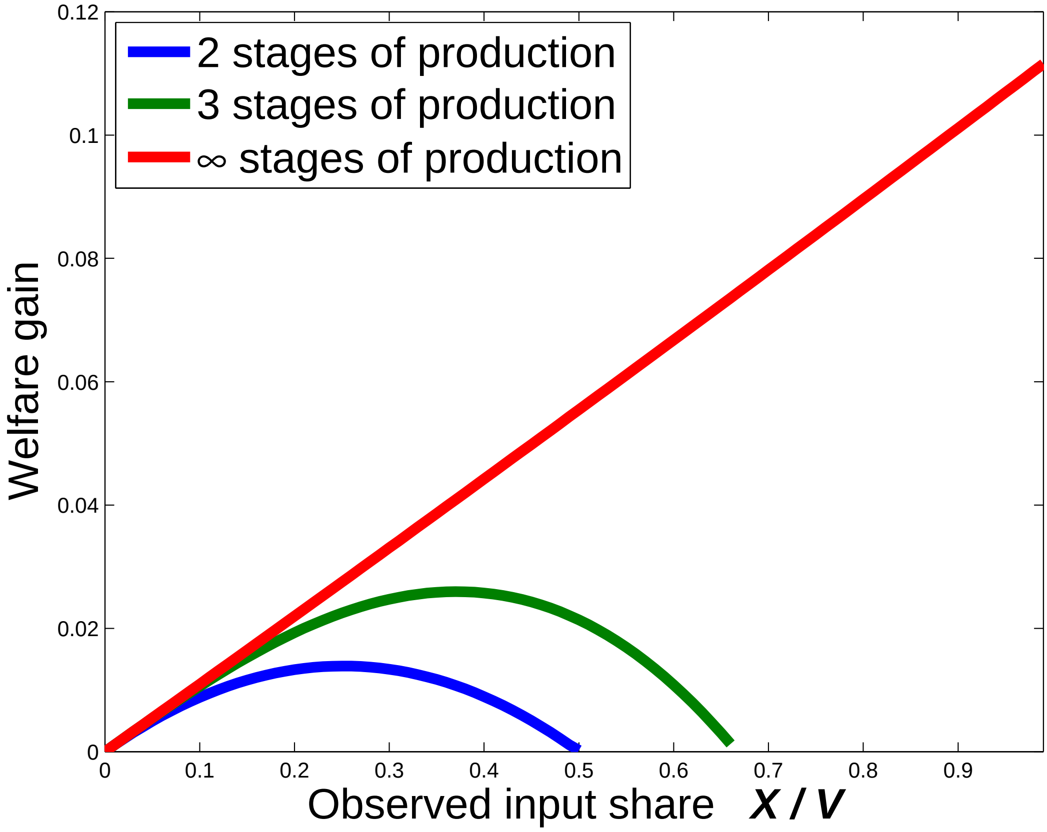


Notes: The figure shows the welfare gain (in percentage points) of removing a turnover tax as a function of the observed input share of an economy in the first two versions of the economy described in section 2. The blue line shows the welfare gain of the reform when the economy produces in two stages, and the green line shows the welfare gain when the economy produces in three stages. Both lines are drawn using equation (7) when the number of stages is either two or three. For every assumed production length, the input elasticities ω are calibrated to the observed input share shown on the horizontal axis, while the tax rate is set to match a value of tax revenue over national income equal to 4.6% (the value that is observed in Brazil before the 2002-3 reform).

Figure 3

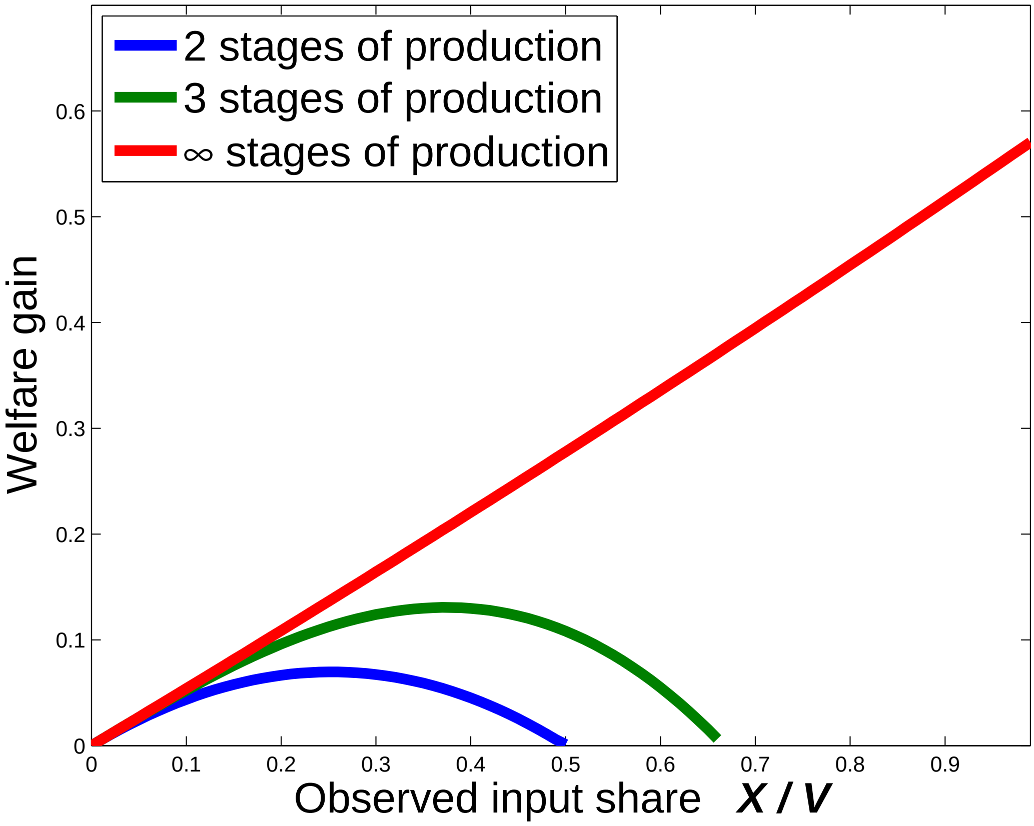
Welfare gain of the tax reform in the one sector example

Welfare gain with two, three and an infinite number of production stages



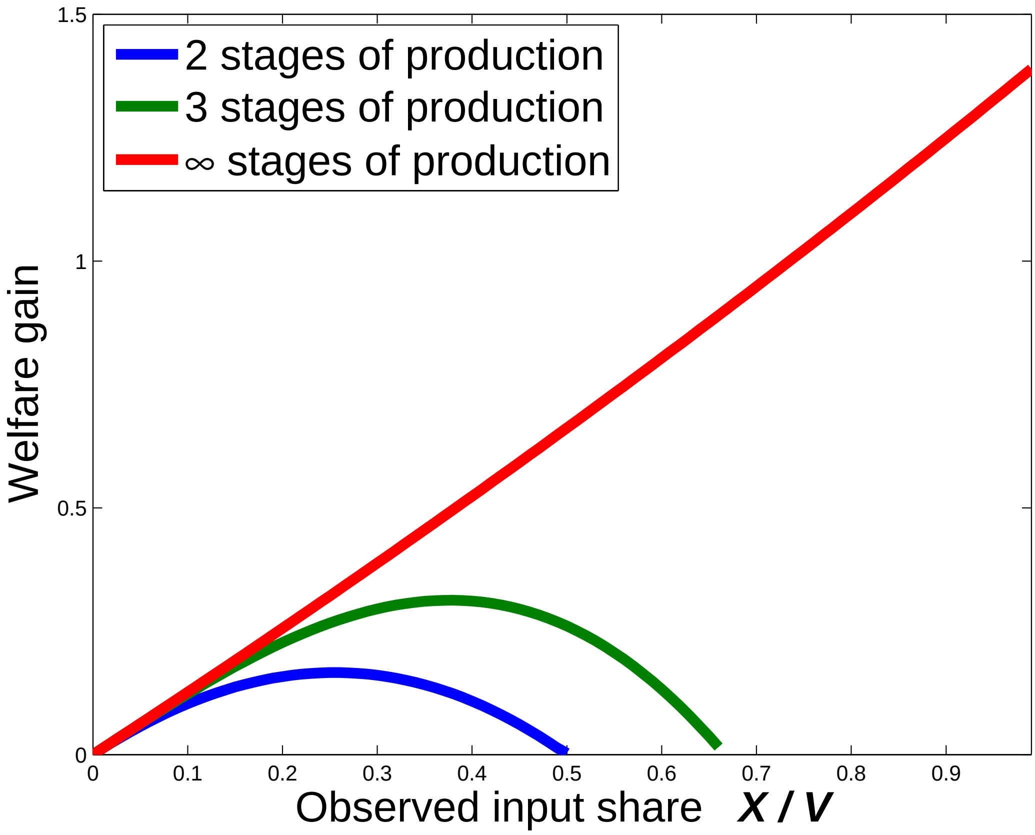
Notes: The figure shows the welfare gain (in percentage points) of removing a turnover tax as a function of the observed input share of an economy in the three versions of the economy described in section 2. The blue line shows the welfare gain of the reform when the economy produces in two stages, the green line shows the welfare gain when the economy produces in three stages and the red line shows the welfare gain when the economy produces with an infinite number of stages. All lines are drawn using equation (7) when the number of stages is either two, three or it approaches infinity. For every assumed production length, the input elasticities ω are calibrated to the observed input share shown on the horizontal axis, while the tax rate is set to match a value of tax revenue over national income equal to 4.6% (the value that is observed in Brazil before the 2002-3 reform).

Figure 4
Welfare gain of the tax reform in the one sector example
Tax revenues equal to 10% of national income



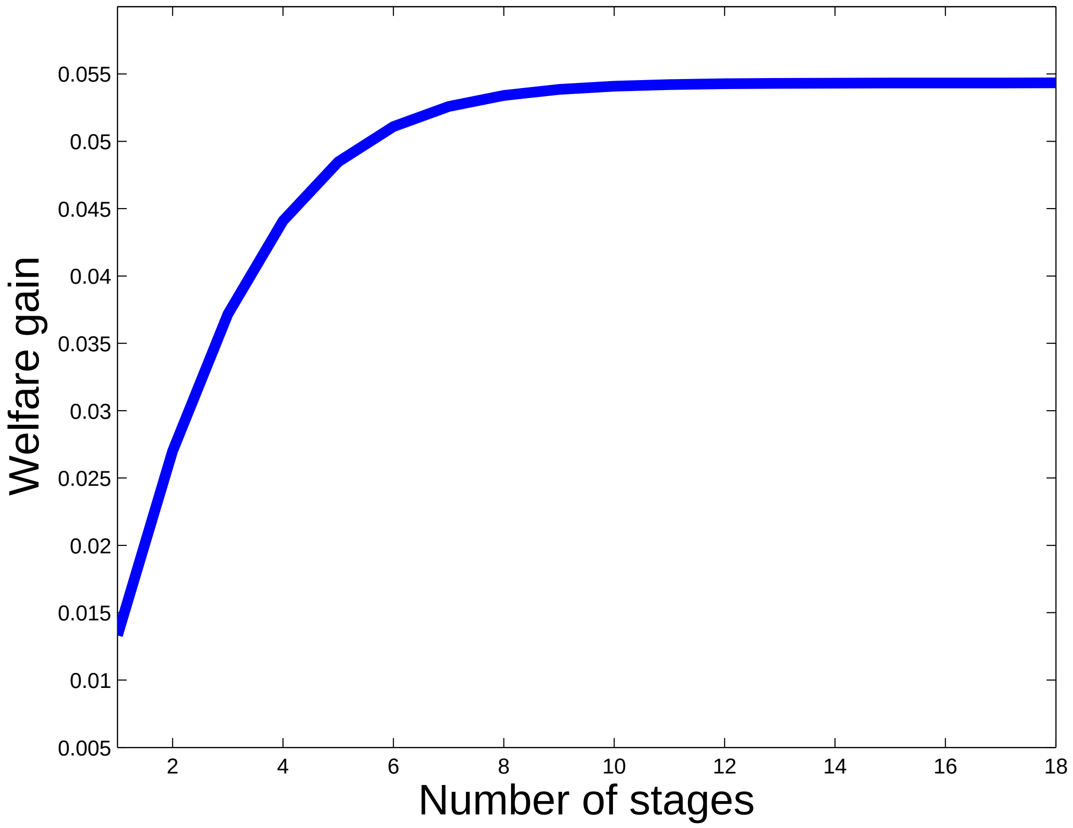
Notes: The figure shows the welfare gain (in percentage points) of removing a turnover tax as a function of the observed input share of an economy in the three versions of the economy described in section 2. The blue line shows the welfare gain of the reform when the economy produces in two stages, the green line shows the welfare gain when the economy produces in three stages and the red line shows the welfare gain when the economy produces with an infinite number of stages. All lines are drawn using equation (7) when the number of stages is either two, three or it approaches infinity. For every assumed production length, the input elasticities ω are calibrated to the observed input share shown on the horizontal axis, while the tax rate is set to match a value of tax revenue over national income equal to 10%.

Figure 5
Welfare gain of the tax reform in the one sector example
Tax revenues equal to 15% of national income



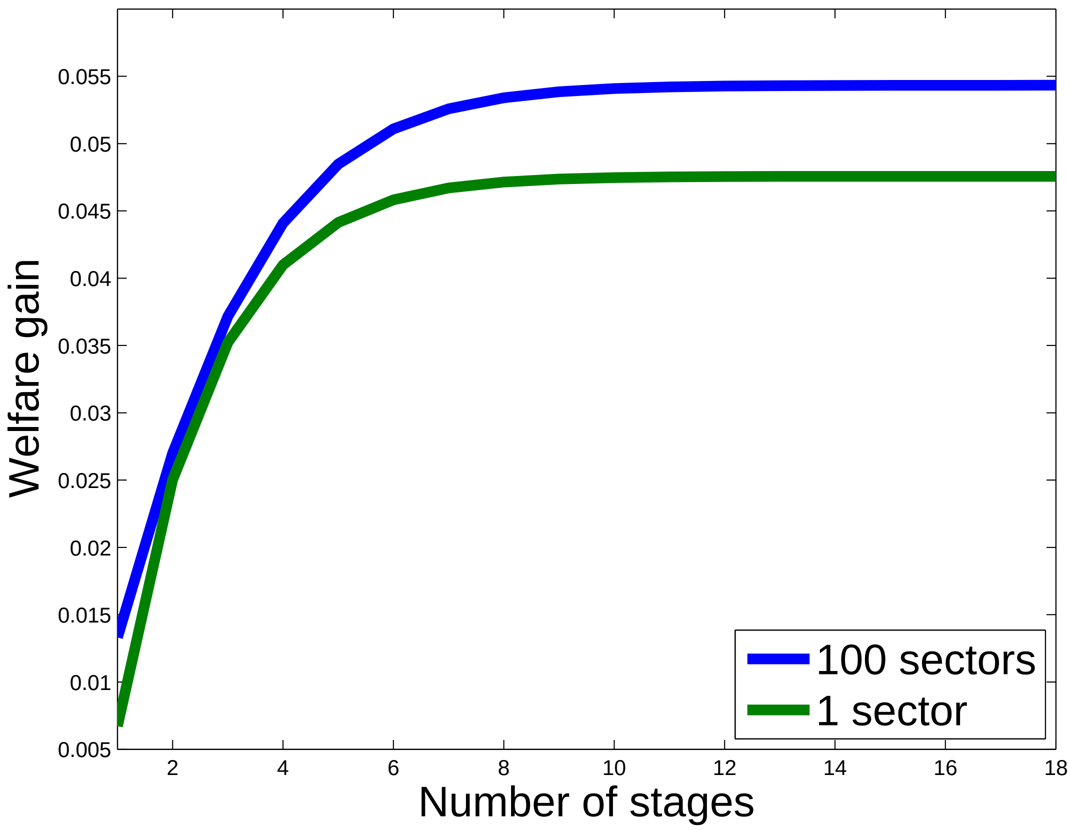
Notes: The figure shows the welfare gain (in percentage points) of removing a turnover tax as a function of the observed input share of an economy in the three versions of the economy described in section 2. The blue line shows the welfare gain of the reform when the economy produces in two stages, the green line shows the welfare gain when the economy produces in three stages and the red line shows the welfare gain when the economy produces with an infinite number of stages. All lines are drawn using equation (7) when the number of stages is either two, three or it approaches infinity. For every assumed production length, the input elasticities ω are calibrated to the observed input share shown on the horizontal axis, while the tax rate is set to match a value of tax revenue over national income equal to 15%.

Figure 6
Welfare gain of the tax reform in the 100 sectors economy



Notes: The figure shows the welfare gain (in percentage points) of the Brazilian 2002-3 reform as a function of the number of stages of production. The line is drawn using equation (6) adapted to an economy that produces 100 final goods. For every assumed production length, the input elasticities of the 100 $\times$ 100 matrix ω_{ij}, are calibrated to match the observed Brazilian input-output matrix in 2005, using the procedure described in section 4.3. The tax rate is set to match a value of tax revenue over national income equal to 4.6% (the value that is observed in Brazil before the 2002-3 reform) using the procedure described in the same section.

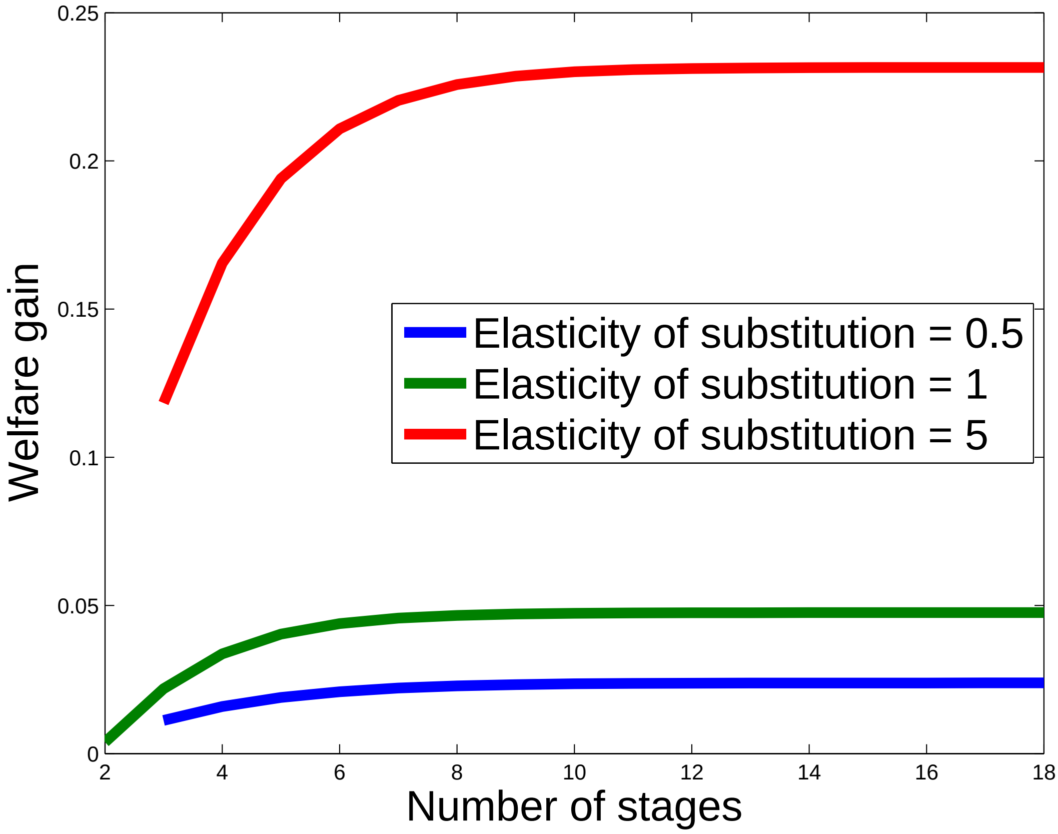
Figure 7
Welfare gain of the tax reform in the 100 sectors economy
Comparison with welfare gain in a single sector economy



Notes: The figure shows the welfare gain (in percentage points) of the Brazilian 2002-3 reform as a function of the number of stages of production. For every given production length, it plots two lines. The blue line is drawn using equation (6) adapted to an economy that produces 100 final goods. The green line is drawn using equation (7): the welfare gain of the reform in an economy that produces a single final good. In the first case, the input elasticities of the 100 $\times$ 100 matrix ω_{ij}, are calibrated to match the observed Brazilian input-output matrix in 2005 using the procedure described in section 4.3. In the second case the input elasticity of the Cobb-Douglas production function is calibrated to 0.43: the input share observed in the aggregated Brazilian input-output matrix of 2005. In both cases, the tax rate is set to match a value of tax revenue over national income equal to 4.6% (the value that is observed in Brazil before the 2002-3 reform) using the procedure described in section 4.3.

Figure 8

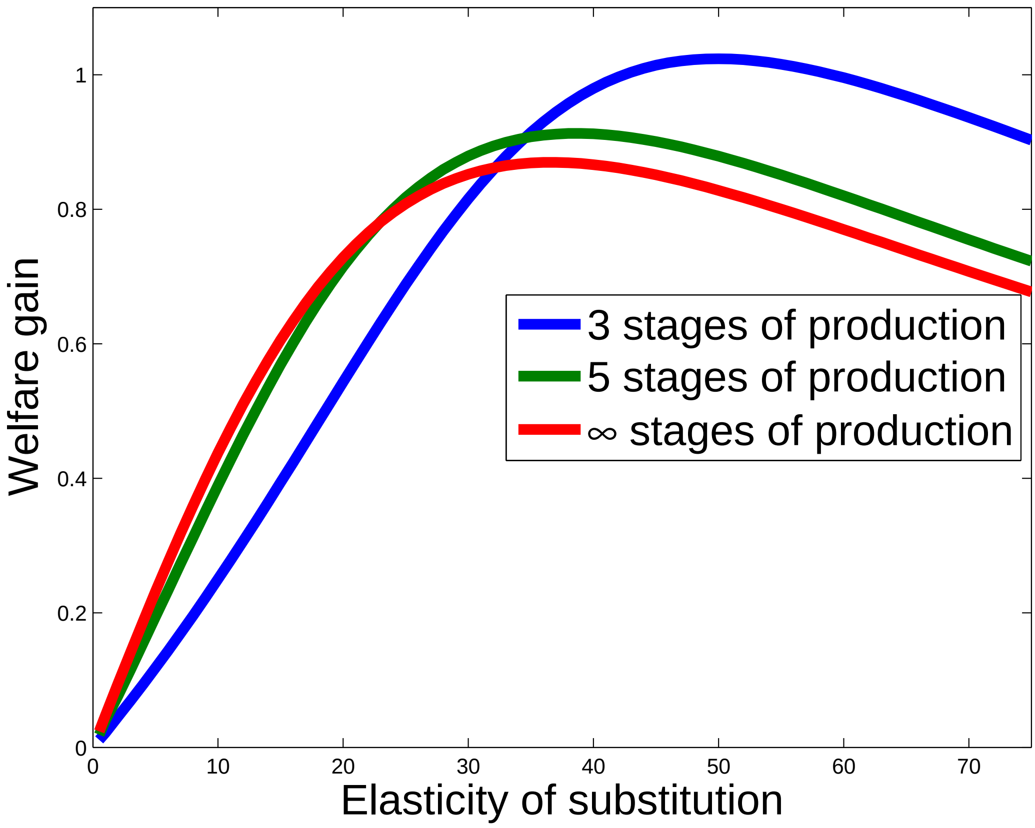
Welfare gain of the tax reform in the constant elasticity of substitution model **Number of stages and welfare gain**



Notes: The figure shows the welfare gain (in percentage points) of the Brazilian 2002-3 reform as a function of the number of stages of production. Welfare gain are calculated using equation (15): the welfare gain of the reform in the constant elasticity of substitution model. For every given production length, the figure shows three lines. The blue line is drawn assuming an elasticity of substitution equal to 0.5. The green line is drawn assuming an elasticity of substitution equal to 1: because this corresponds to the Cobb-Douglas case with a single final good, this line reproduces the green line shown in figure 7. The red line is drawn assuming an elasticity of substitution equal to 5. For every assumed production length and elasticity of substitution, the joint parameter of the production function γ/A is calibrated using the procedure described in section 5.2 to match 0.43: the input share observed in the aggregated Brazilian input-output matrix of 2005. The tax rate is set to match a value of tax revenue over national income equal to 4.6% (the value that is observed in Brazil before the 2002-3 reform) using the procedure described in section 5.2.

Figure 9

Welfare gain of the tax reform in the constant elasticity of substitution model **Elasticity of substitution and welfare gain**



Notes: The figure shows the welfare gain (in percentage points) of the Brazilian 2002-3 reform as a function of the elasticity of substitution between labor and intermediate input. For every given production length, it plots three lines. All lines are drawn using equation (15): the welfare gain of the reform in the constant elasticity of substitution model. The blue line is drawn assuming a production process of 3 stages. The green line is drawn assuming a production process of 5 stages. The red line is drawn assuming a production process with an infinite number of stages. For every assumed production length and elasticity of substitution, the joint parameter of the production function γ/A is calibrated using the procedure described in section 5.2 to match 0.43: the input share observed in the aggregated Brazilian input-output matrix of 2005. The tax rate is set to match a value of tax revenue over GDP equal to 4.6% (the value that is observed in Brazil before the 2002-3 reform) using the procedure described in the same section.

utility after the reform can be written as:

$$u^v - u^{tt} = p_C^{tt} - p_C^v$$

where a lowercase letter denotes the logarithm of a variable: $x \equiv \log X$ and we use tt and v to label equilibrium variables when taxes are collected on turnover or on consumption.

Under both scenarios P_C is the price index of the Cobb-Douglas aggregate (5), which is a function of the prices of the N goods sold at the last stage of production. With a turnover tax equal to t, the price of consumption good i is: $P_i^{tt}(S) = (1+t)MC_i^{tt}(S)$, where $MC_i^{tt}(S)$ is the marginal cost of variety S produced by sector i when there is a turnover tax. With a consumption tax equal to v, the price of consumption good i is: $P_i^v(S) = (1+v)MC_i^v(S)$, where $MC_i^v(S)$ is the marginal cost of variety S produced by sector i when there is no distortion at any point in the production chain. Thus, using the price index of equation (5), the welfare gain of the reform can be written as:

$$u^v - u^{tt} = \left[\sum_{i=1}^{N} \beta_i \left(\log(1+t) + mc_i^{tt}(S) - mc_i^v(S)\right)\right] + \log\left(\frac{1}{1+v}\right) \tag{21}$$

To derive a solution to (21) we proceed in two steps. First, we use technology and market structure to find the solution of the sum in brackets. Next, we use assumption of revenue neutrality to derive an expression for $\log[1/(1+v)]$.

As argued in section 4.1, under perfect competition marginal costs and prices of every variety in every sector are uniquely determined by input-output parameters and the tax scheme. This can be derived explicitly starting from stage 1, where the log of the marginal cost of sector i are the same under the two tax schemes: $mc_i^v(1) = mc_i^{tt}(1) = w - a_i$. Given the technology and the market structure, the vector $\boldsymbol{mc}(s)$ that collects the logarithm of marginal costs of all sectors at stage $s > 1$ can be written as:

$$\boldsymbol{mc}^v(s) = \boldsymbol{k} + (\mathbf{1} - \boldsymbol{\omega})w + \Omega\ \boldsymbol{mc}^v(s-1) \qquad s = 2, 3, \ldots, S \tag{22}$$

when the tax is collected only from firms producing at stage S as a consumption tax. When there is a turnover tax equal to t, the same vector becomes:

$$\boldsymbol{mc}^{tt}(s) = \boldsymbol{k} + (\mathbf{1} - \boldsymbol{\omega})w + \Omega\ \boldsymbol{mc}^{tt}(s-1) + \Omega\ \boldsymbol{\iota}\log(1+t) \qquad s = 2, 3, \ldots, S \tag{23}$$

In both equations $\boldsymbol{k}$ is a vector with typical element equal to: $a_i + (1-\omega_i)\log(1-\omega_i) - \sum_{j=1}^{N} \omega_{ij} \log \omega_{ij}$, $\boldsymbol{\omega}$ is a vector collecting ω_i and $\boldsymbol{\iota}$ is a vector of ones. Solving systems (22) and (23) from stage 1 forward, yields the solution for $\boldsymbol{mc}^{tt}(S) - \boldsymbol{mc}^{v}(S)$:

$$\boldsymbol{mc}^{tt}(S) - \boldsymbol{mc}^{v}(S) = \left[\sum_{s=1}^{S} \Omega^s\right] \boldsymbol{\iota} \log(1+t) \tag{24}$$

Thus, each term in the sum inside the brackets of equation (21) can be written as:

$$\beta_i \log(1+t) \sum_{j=1}^{N} \lambda_{ij}$$

where λ_{ij} the typical entry of the matrix Λ:

$$\Lambda = \left[\sum_{s=0}^{S-1} \Omega^s\right]$$

Next, we derive an expression for $\log\left[1/(1+v)\right]$. The condition that the new tax scheme should collect the same amount of tax revenues T^{tt} collected with the turnover tax implies that $\left[v/(1+v)\right]\overline{L} = T^{tt}$. Hence:

$$\begin{aligned} \frac{1}{1+v} &= 1 - \frac{T^{tt}}{\overline{L}} \\ &= 1 - \frac{t}{\overline{L}} \sum_{i=1}^{N} \sum_{s=1}^{S} MC_i^{tt}(s) Q_i^{tt}(s) \end{aligned} \tag{25}$$

Or, in matrix notation:

$$\frac{t}{\overline{L}} \sum_{i=1}^{N} \sum_{s=1}^{S} MC_i^{tt}(s) Q_i^{tt}(s) = \frac{t}{(1+t)\overline{L}} \sum_{s=1}^{S} \boldsymbol{\iota}' \boldsymbol{V}^{tt}(s)$$

where $\boldsymbol{V}^{tt}(s)$ is the vector collecting the value of sales of the firms producing variety s in the N sectors of the economy and $\boldsymbol{\iota}$ is a $N \times 1$ vector of ones. We use 1 to write the value of production at stage s as:

$$\boldsymbol{V}^{tt}(s) = \frac{1}{1+t} \Omega' \, \boldsymbol{V}^{tt}(s+1)$$

Thus, the revenues collected with the turnover tax can be written as:

$$\frac{t}{\overline{L}} \sum_{i=1}^{N} \sum_{s=1}^{S} MC_i^{tt}(s) Q_i^{tt}(s) = \frac{t}{(1+t)\overline{L}} \boldsymbol{\iota}' \left[\sum_{s=0}^{S-1} \left(\frac{1}{1+t} \Omega'\right)^s\right] \boldsymbol{V}^{tt}(S)$$

Given constant consumption shares implied by the Cobb-Douglas aggregator of final goods, $\boldsymbol{V}^{tt}(S) = \boldsymbol{\beta}\overline{L}$. Re-arranging equation (25), the tax rate $\log\left[1/(1+v)\right]$:

$$\log\left(\frac{1}{1+v}\right) = \log\left[1 - \frac{t}{1+t}\sum_{i=1}^{N}\beta_i\left(\sum_{j=1}^{N}\lambda_{ji}^{tt}\right)\right] \tag{26}$$

where λ_{ij}^{tt} is the generic element of the following matrix:

$$\Lambda^{tt} = \left[\sum_{s=0}^{S-1}\left(\frac{1}{1+t}\Omega'\right)^s\right]$$

Putting together (24) and (26), one obtains the welfare gain (6).

Analytical derivations of CES model

In this appendix we derive equation (15): the welfare gain of a reform that converts a turnover tax into a consumption tax in the constant elasticity of substitution model.

We start from the indirect utility function of the household. Because the level of the public good provided by the government and the total number of hours worked by the household are assumed to be unaffected by the reform, the welfare gain of the reform can be written as the log change of the price of the consumption good:

$$u^v - u^{tt} = p^{tt}(S) - p^v(S)$$

Because the price of the consumption good is $P^{tt}(S) = (1+t)MC^{tt}(S)$ and $P^v(S) = (1+v)MC^v(S)$ with a turnover tax and a consumption tax respectively, the welfare gain of the reform can be written as:

$$u^v - u^{tt} = \log\left[(1+t)\frac{MC^{tt}(S)}{MC^{tt}(S)}\right] + \log\left[\frac{1}{1+v}\right] \tag{27}$$

We solve for the two terms of this equations in turn.

The system of prices (14) allows to write the marginal cost of the variety produced at stage S with the turnover tax and with the consumption tax as:

$$\begin{aligned} MC^{tt}(S) &= \frac{1}{B}\left[\sum_{s=0}^{S-1}\left(\frac{\gamma}{A}(1+t)^{1-\eta}\right)^s\right]^{\frac{1}{1-\eta}} \\ MC^v(S) &= \frac{1}{B}\left[\sum_{s=0}^{S-1}\left(\frac{\gamma}{A}\right)^s\right]^{\frac{1}{1-\eta}} \end{aligned}$$

Solving these sums, the first term in equation (27) can be written as:

$$\log\left((1+t)\frac{MC^{tt}(S)}{MC^{v}(S)}\right) = \log\left[(1+t)\cdot\frac{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^{S}}{1-\frac{\gamma}{A}(1+t)^{1-\eta}}\cdot\frac{1-\frac{\gamma}{A}}{1-\left(\frac{\gamma}{A}\right)^{S}}\right] \tag{28}$$

Next, we use the assumption that total revenues are not affected by the reform to express the second term of equation (27) as a function of the parameters of the model. Setting $T^{tt} = T^{v}$ implies:

$$\begin{aligned}\frac{1}{1+v} &= 1-\frac{T^{tt}}{\overline{L}} \\ &= 1-\frac{t}{\overline{L}}\left[\sum_{s=1}^{S} MC^{tt}(s)Q^{tt}(s)\right]\end{aligned}$$

By multiplying and dividing every term in the sum in brackets by the after-tax value of all varieties produced in the stages following it, it is possible to write:

$$\begin{aligned}MC^{tt}(s)Q^{tt}(s) &= \left[\prod_{k=s+1}^{S}\frac{\omega^{T}(k)}{1+t}\right]\cdot\frac{P^{tt}(S)Q^{tt}(S)}{1+t} \\ &= \left[\prod_{k=s+1}^{S}\frac{\omega^{T}(k)}{1+t}\right]\cdot\frac{\overline{L}}{1+t} \qquad (29) \\ & \qquad (30)\end{aligned}$$

where the second line follows from the budget constraint and $\omega^{tt}(k)$ is the cost share of intermediate inputs in the production of variety k, as defined in equation (16). Using the price equations (14), these cost shares can be written as a function of parameters:

$$\omega^{tt}(k) = \frac{\gamma}{A}(1+t)^{1-\eta}\cdot\frac{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^{k-1}}{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^{k}}$$

Plugging this expressions for the cost share into (29), it is possible to write each term in the sum inside the brackets of equation (15) as:

$$\prod_{k=s+1}^{S}\frac{\omega^{tt}(k)}{1+t} = \left[\frac{\gamma}{A}(1+t)^{-\eta}\right]^{S-s}\left[\frac{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^{s}}{1-\left[\frac{\gamma}{A}(1+t)^{1-\eta}\right]^{S}}\right]$$

Summing these terms for all stages of production allows to write (17). Combining this equation with equation (28) derived above we get a formula for the welfare gain of the reform that is a function only of the parameters of the model.