



## Endogenous Tariffs in a Common-Agency Model: A New Empirical Approach Applied to India

*Las tarifas endógenas en un modelo de agencia común :  
Un nuevo enfoque empírico aplicado a la India*

### OLIVIER CADOT

HEC Lausanne, CERDI and CEPR.  
[olivier.cadot@unil.ch](mailto:olivier.cadot@unil.ch)

### LAURE DUTOIT

ECLAC, United Nations  
[laure.dutoit@cepal.org](mailto:laure.dutoit@cepal.org)

### JEAN-MARIE GREETHER

Université de Neuchâtel  
[jean-marie.grether@unine.ch](mailto:jean-marie.grether@unine.ch)

### MARCELO OLARREAGA

University of Geneva and CEPR  
[marcelo.olarreaga@unige.ch](mailto:marcelo.olarreaga@unige.ch)

### ABSTRACT

*This paper proposes a new method to test the Grossman-Helpman model of endogenous protection and lobby formation, which does not require data on lobby formation or contributions. It identifies politically organized industries using commonly available trade and production data, as well as the model's structural parameter estimates. Applied to India, it yields results that are qualitatively consistent with the model's predictions and that seem quantitatively more plausible than estimates given for the US by alternative methods. Our estimates imply that the weight put by the Indian government on contributions by politically organized sectors is a third of the weight it puts on (gross) social welfare, well above existing estimates for the United States.*

Keywords: India, Political Economy, Protection.

JEL classification numbers: F10, F11, F13

---

\* We are grateful to Jose Anson, Antonio Fatas, Caroline Freund, Hiau Looi Kee, Kishore Gawande, Pravin Krishna, Nicole Mathys, Jaime de Melo, Frederic Robert-Nicoud, Kamal Saggi, Maurice Schi, Daniel Traca, Timothy Van Zandt, participants at the *Political Economy of Integration* conference at the Southern Methodist University and seminars at INSEAD, University of Lausanne, and the World Bank for very helpful and constructive suggestions. The views expressed here are those of the authors and should not be attributed to the institutions to which they are affiliated.



## RESUMEN

*En este trabajo se propone un nuevo método para testear el modelo de Grossman-Helpman sobre protección endógena y la formación de grupos de interés (lobbies), que no requiere de datos respecto a la formación de los grupos de interés o de sus contribuciones. Identifica a las industrias políticamente organizadas a partir de datos de producción y comercio que son de acceso común, así como las estimaciones de los parámetros estructurales del modelo. Aplicado a la India, produce resultados que son cualitativamente consistentes con las predicciones del modelo, y que parecen cuantitativamente más plausibles que las estimaciones para los EE.UU obtenidas por métodos alternativos. Nuestras estimaciones implican que la importancia que el gobierno de la India asigna a las contribuciones de los sectores políticamente organizados es una tercera parte de la importancia que le asigna al bienestar social (bruto), muy por encima de las estimaciones existentes para los Estados Unidos.*

Palabras clave: India, Economía Política, Protección.

Código JEL: F10, F11, F13.

## I. INTRODUCTION

Drawing on the pioneering work of Stigler (1971), Peltzman (1976) and Becker (1983) on regulation, one strand of the political-economy literature that has gained prominence in the area of trade policy views the policymaking process as an economic exchange between politicians selling policies and lobbies willing to buy them. Grossman and Helpman (1994), who contributed to make this approach popular, treated the two-way relationship between political support and government favors as a common-agency game a la Bernheim and Whinston (1986a, 1986b). Compared to previous modeling approaches, Grossman and Helpman's takes the theory of endogenous protection closer to 'first-principles' microeconomics and relates equilibrium trade policy to measurable variables such as import-penetration ratios, elasticities, and so on.

In spite of the progress made, a number of puzzles remain. First, Rodrik (1995) pointed out that, according to the logic of the Grossman-Helpman model (henceforth GH), if exporting sectors have larger domestic outputs

than import-competing ones –as specialization according to comparative advantage would imply—they should also lobby more aggressively, leading to more export subsidies than import tariffs, a prediction that is difficult to reconcile with evidence.<sup>1</sup> A related point is that import-competing sectors with larger domestic outputs and hence (ceteris paribus) lower import-penetration ratios should also lobby harder and consequently get higher tariffs in equilibrium. However the evidence on this is, at best, mixed. A number of empirical studies (e.g. Marvel and Ray 1983, Baldwin 1985, or Lee and Swagel 1997) found a *positive* –instead of negative— relationship between import-penetration ratios and the level of protection in the United States.<sup>2</sup> The logic behind this result was straightforward: high import-penetration ratios proxy for industries where the United States has a comparative disadvantage, and those industries tend to lobby harder than others for trade protection.

Goldberg and Maggi (1999, GM henceforth) and Gawande and Bandopadhyay (2000, GB henceforth) offered a way out of the puzzle. The starting point was to observe that the GH model predicts a different relationship between equilibrium protection and the ratio of domestic output to imports for organized sectors vs. unorganized ones. For the former, the relationship is positive (hence negative between protection and import penetration, as noted above); for the latter, it is the reverse. In order to account for this distinction in the estimation, GB and GM interacted the output/imports ratio with a binary variable equal to one when a sector was politically organized and zero otherwise. Regressing the level of protection, proxied by non-tariff barriers incidence, on this interaction term (for organized sectors) and on the non-interacted output/imports ratio (for unorganized ones and welfare effects), both divided by the elasticity of import demand, yielded parameter estimates in accordance with the model's predictions. Both papers also showed that GH's parsimonious specification fared well compared to a traditional endogenous-protection regression in which a wide net was cast to catch a variety of politically-related effects (employment, industry concentration, etc.).

These studies and more recent ones (Gawande, Sanguinetti and Bohara, 2004, Mitra, Thomakos and Ulubasoglu, 2002, and McCalman, 2004)

1. For a discussion of the argument and caveats, see Gawande and Krishna (2003). It is worth noting in particular that a higher domestic output always raises the return to lobbying (as a given tariff increase is spread over larger quantities). Thus, by Hotelling's lemma, any model in which lobbying intensity depends on the price derivative of the profit function will yield this prediction, whether or not based on common agency.
2. It should be noted however that the relationship between import penetration and endogenous protection turned out to be less clearcut when the endogeneity of imports was explicitly taken into account, as in Ray (1981) or Trefler (1993).

provide evidence in favor of the common-agency approach to endogenous trade policy. However they have also raised, as a by-product, a second puzzle. Structural estimation yields more information than just the validation of qualitative relationships: if the model is to be taken seriously, its estimation should also yield quantitatively plausible estimates of key parameters. In particular, the weight given to welfare in the government's objective function (parameter “ $a$ ” in GH's formulation) can be retrieved algebraically from regression estimates. This weight matters not just as a measure of government preferences but also –perhaps more importantly— as a factor of proportionality between contributions and the welfare distortions generated by tariffs. As noted by GB, in a stripped-down setting without general equilibrium interaction, the common agency framework degenerates into a collection of “parallel” principal-agent relationships in which lobbies compensate the government for the welfare distortions implied by the tariffs they are asking for. The rate of conversion between welfare distortions and monetary contributions is then  $a$ . Values of  $a$  above one, for instance, imply that lobbies must contribute *multiples* of the welfare distortions their protectionist demands imply. The values of  $a$  estimated in the papers mentioned above were, from this perspective, very high,<sup>3</sup> “enough”, in Gawande and Krishna’s terms, “to cast doubt on the value of viewing trade policy determination through this political economy lens” (Gawande and Krishna 2003, p. 20).

One way out of the puzzle was recently suggested by Damania and Frederiksson (2007), who extended GH to a multi-agent multi-principal model in which trade-policy decisions are made by consensus among a number of government agencies. Each of these agencies is targeted for contributions by lobbies, but each has a probability  $\gamma$  of being replaced before having had a chance to implement the promised policy. “Replacements” are fresh politicians assumed initially impervious to lobbying. In such a setting, the parameter measured empirically is a function of the true weight of welfare in the agents' objective function and, inter alia (the model is richer than the description given here) of the probability of early replacement of any government agent. As it turns out, the true  $a$  is below the observed one, and simulations suggest that it may easily be so by a wide margin.

Although this approach provides a nice answer to the “high- $a$  puzzle”, it still implies that the ratio of equilibrium contributions to implied welfare distortions is proportional, by a factor  $1-\gamma$  (where  $\gamma$  is the probability of any

3. Parameter estimates from Goldberg and Maggi's (1999) basic specification imply values of  $\hat{a}$  lying between 51.93 and 70.43. Other estimates of  $a$  range from 43:41 (McCalman, 2004) to somewhere between 76 and 104 (Mitra et al., 2002) to 3175 (Gawande and Bandyopadhyay, 2000).

trade-policy official being replaced) to the measured  $a$  (what Damania and Frederiksson call  $a^e$ ).<sup>4</sup> Thus, as long as the estimated value of  $a$  is high, part of the empirical puzzle (high implied contributions) remains, even though the measured  $a$  may not truly reflect the weight on welfare in the government's objective function.

An additional issue has to do with the identification of sectors organized into active trade-policy lobbies. GB and GM relied on outside information, looking at Political Action Committee (PAC) contributions and choosing a cutoff above which they considered industries as organized. There are two problems with this. First, PAC contributions are a noisy signal of trade-related influence activities. On one hand, they accounted for only half the campaign contributions in the US 1997-98 electoral cycle, the other half being so-called "soft money" (contributed to national parties rather than individual lawmakers). Influence activities take other forms as well: the turnover of lobbying firms registered under the 1995 Lobbying Disclosure Act was \$1.46 billion in the same electoral cycle (CRP 2001). On the other hand, PAC contributions are not necessarily trade-related as lobbies also try to influence domestic policies. Thus, PAC contributions both understate and overstate trade-related influence activities and this may affect the ranking of sectors and hence the cutoff between organized and unorganized ones. Because the distinction between organized and unorganized sectors is so crucial empirically, this is a potential problem. Second and perhaps more importantly, relying on PAC contribution data precludes using the method in countries other than the US where no data is available on political contributions. Confined to the US, tests of the theory would rapidly hit diminishing returns (in addition, serious doubts have recently been raised by Imai et al. (2005) about the adequacy of NTB-based tests; but performing tests outside of the US requires new methods.

Several such methods have been proposed recently. Gawande, Sanguinetti and Bohara (2004) analyzing Mercosur's trade protection, assume that industries in which imports are above the sample mean are politically organized into protectionist lobbies. Mitra, Thomakos and Ulubasoglu (2002) analyzing Turkey's trade protection, make use of membership data from the Turkish Industrialist and Businessmen Association to determine which sectors are organized. The authors then statistically validate their choice of organized sectors using discriminant analysis methods. McCalman (2004) analyzes tariff changes in Australia between the late sixties and early nineties and uses the fact that tariff changes required Tariff Board inquiries, ninety percent of

---

4. We are grateful to Per Fredriksson for a helpful explanation on this.

which were undertaken at industry's initiative. He thus takes the initiation of a Tariff Board inquiry as evidence of an industry's political organization.

We propose an alternative method which, instead of drawing on outside information, uses information generated by the tariff data itself and can thus, in principle, be applied to any setting. While being close in spirit to the recent literature, it differs in two key respects. First, as mentioned, we endogenously derive from the model a classification of industries into organized vs. unorganized ones through an iterative procedure. In the first stage, we estimate a standard GH equation without differentiating between organized and unorganized sectors. This regression determines endogenous tariffs as functions of, *inter alia*, import penetration ratios. In the second stage, we use the first equation's residuals to rank industries, those with high residuals being, in some sense, more likely to be organized than others. On the basis of this ranking, we then set an arbitrary cutoff value above which industries are considered to be organized. Finally, we run a grid search over different cutoff values.

Second, we refine on GB's modeling of input-output linkages. Treating those linkages explicitly is important both conceptually and empirically. Conceptually, the gist of the common-agency approach is that good policies (small departures from free trade in a trade-policy context) result not just from governments being impervious to influence activities, but also from the balance of conflicting lobbying pressures. Counter-lobbying against protection of an industry by its down-stream users is one such countervailing force and is likely to be more effective when they are industrial users rather than final ones. Recognizing this leads to sharper predictions and hence more powerful tests of the theory. We depart from GB in our use of input-output data to determine jointly the protection of final and intermediate goods.<sup>5</sup> We also draw from Cadot, de Melo and Olarreaga (2003) to include the effect of duty-drawbacks on lobbying incentives.

The methodology is applied to India, a country that has, for a variety of reasons, enjoyed a large degree of independence in the denition of its trade policy. In addition, trade protection in India largely takes the form of tariffs, so it is not subject to the same critique of the power of empirical tests of the GH model on US data, where non-tariff barriers are relatively more important.<sup>6</sup>

---

5. A similar approach is followed by Cadot et al. (2004) and Gawande and Krishna (2005) who incorporate counter-lobbying by downstream users of intermediate inputs.

6. According to estimates by Kee, Nicita and Olarreaga (2009), non-tariff barriers explain 75 percent of the trade restrictiveness in the United States, but less than 20 percent in India.

The empirical results provide strong evidence in favor of the common-agency approach and are encouraging for further applications of the method. Direct producer lobbying, counterlobbying by users of intermediate inputs, as well as the counter-lobbying dilution effect introduced by tariff exemption schemes for exporters are all identified in the data. Based on parameter estimates, the weight given to political contributions is 31 percent higher than the weight granted to (other elements of) social welfare, well above the problematic estimates found in earlier studies, where governments seemed to be close to maximizing social welfare. Out of 80 ISIC sectors, we identify 13 as organized for trade policy purposes.

## II. PROTECTION AND LOBBYING: A BASIC FRAMEWORK

In this section, we present the basic model guiding our empirical estimation exercise. One can think of the setup as similar to the one in Mitra (1999), where owners of specific factors in import-competing industries first decide whether to organize into lobbies or not, after which trade policy is determined by a Grossman-Helpman (1994) common-agency game.<sup>7</sup> Here we focus on the second stage (i.e., trade policy determination) taking the decision to politically organize as given.

Consider a small open economy with  $n + 1$  tradable sectors, in which good zero serves as numeraire. Individuals have different endowments but identical tastes represented by a utility function:

$$U = c_0 + \sum_{j=1}^n u(c_j), \quad (1)$$

where  $c_0$  is consumption of the numéraire good,  $c_j$  is the consumption of non-numeraire good  $j$ , and  $u$  satisfies the usual properties.

All goods produced in the economy are potential inputs in other sectors, and all industries are perfectly competitive. In all sectors  $j = 1, \dots, n$  except the numeraire, technology is Leontief between intermediate consumptions and value added; thus, value added is nested in the Leontief production function and is created using a specific factor  $\kappa_j$  (“capital”) and a mobile factor  $\ell_j$  (“labor”) under a general constant-returns to scale technology  $f^j$ . Let  $a_{ij}$  be

---

7. We will use indifferently the terms “common agency” and “menu auction” to describe Grossman-Helpman's application to trade policy of the theoretical framework developed by Bernheim and Whinston (1986a, b). For our purposes, the two are mathematically equivalent.

the requirement of good  $i$  necessary to produce one unit of good  $j$ , and let  $x_{ij}$  be sector  $j$ 's demand for good  $i$  as an intermediate input;

$$y_j = \min \left\{ f^j(\kappa_j, \ell_j); \frac{x_{0j}}{a_{0j}}; \dots; \frac{x_{nj}}{a_{nj}} \right\}, \quad j = 1, \dots, n \quad (2)$$

We will henceforth omit  $\kappa_j$  as an argument of production and profit functions. The numeraire good is produced using labor only under constant returns to scale, so that the wage rate  $w$  is fixed.

Let  $p_j^*$  be good  $j$ 's world price and  $t_j$  an ad-valorem import tariff (subsidy if it is negative); good  $j$ 's domestic price is thus  $p_j = p_j^* (1 + t_j)$ . Let  $v_j(\mathbf{t})$  be the indirect utility function of the owners of specific capital in sector  $j$ , where  $\mathbf{t}$  is the  $n$ -dimensional vector of tariffs on imported goods (all goods are tradeable). Let  $\alpha_j$  be the share of sector  $j$ 's shareowners in the population. Under quasilinear preferences,  $v_j$  is the sum of income and consumer surplus. Income is the sum of profits  $\pi_j$  plus  $\alpha_j$  times economywide tariff revenue  $T(\mathbf{t})$ . Consumer surplus is  $j$  times economywide consumer surplus  $S(\mathbf{t})$ . Thus,  $v_j(\mathbf{t}) = \pi_j(\mathbf{t}) + \alpha_j T(\mathbf{t}) + \alpha_j S(\mathbf{t})$

Let  $L$  be the set of politically organized industries (determined endogenously in the first stage) and  $I_j$  an indicator function equal to one when  $j \in L$  and zero otherwise. Lobbies representing the owners of specific capital in those industries bid simultaneously for protection with 'truthful' contribution schedules  $C_j(\mathbf{t}) = \max \{0; v_j(\mathbf{t}) - b_j\}$  for some nonnegative constant  $b_j$ . Faced with such contributions, the government chooses best-response tariffs that maximize

$$G(\mathbf{t}) = \sum_{j=1}^n I_j C_j(\mathbf{t}) + aW(\mathbf{t}), \quad (3)$$

where  $W(\mathbf{p})$  is social welfare and  $a$  is the weight of social welfare. Therefore tariffs satisfy the FOC:

$$\frac{\partial G(\mathbf{t})}{\partial t_i} = \sum_{j=1}^n I_j \frac{\partial v_j}{\partial t_i} + a \frac{\partial W}{\partial t_i} = 0. \quad (4)$$

After tedious but straightforward derivation detailed in the mathematical appendix, solving (4) for  $t_i = t_i / [p_i^* (1 + t_i)]$ , the first order condition becomes:



$$\tilde{t}_i = \frac{I_i - \alpha_L}{a + \alpha_L} \frac{z_i}{|\varepsilon_i|} - \sum_{j=1}^n \frac{I_j - \alpha_L}{a + \alpha_L} (1 - \lambda_j) \frac{a_{ij} z_j}{|\varepsilon_i|}. \quad (5)$$

where  $\alpha_L = \sum_{j=1}^n I_j \alpha_j$ ;  $z_j = y_j/m_i$  is the output of sector  $j$  over imports of sector  $i$ ,  $\varepsilon_i$  is the import demand elasticity in sector  $i$  and  $\lambda_j$  is the share of sector  $j$  output that is exported. As in GB and GM, the first term in (5) shows that equilibrium tariffs are an increasing function of the output/import ratio for politically organized sectors (and the reverse for unorganized ones). The second term reduces tariffs due to counter-lobbying pressure by organized downstream sectors. However, due to the duty drawback scheme, this counter-lobbying effect is weaker the larger is the exported share of the downstream sector's output.

### III. INDIA'S TRADE POLICY

India is an interesting case study for several reasons. First, it is undoubtedly one of the countries in the world with the highest trade barriers. Indeed, in his review of the World Trade Organization's Trade Policy Review of India in 1998, Panagariya (1999) warned that "a reader unfamiliar with India's past is bound to wonder how a trade-policy regime such as the one about to be described can be characterized as having undergone serious reforms".<sup>8</sup> The average tariff was around 35 percent in 1997-1998. This compares badly with the average for developing countries in the late 1990s (around 13 percent overall and for manufacturing).

More importantly given the efforts in this paper to model input-output linkages, significant tariff escalation is present in India's tariff structure. Tariffs on unprocessed goods average 25 percent whereas fully processed goods average 37 percent. As it has been shown in the previous section, this can be explained by downstream counter-lobbying on upstream tariffs.

The tariff structure is further complicated by numerous and spreading exemptions (Panagariya, 1999 and Srinivasan, 2001), including those targeting exporters, such as the Duty Draw-back scheme which compensates exporters for tariffs paid on imported inputs upon export of the finished product; the Duty Exemption scheme, which offers large exporters duty

8. For a review of the extensive Indian reforms in the early 1990s, see Pursell (1996) or Srinivasan (2001). As an indication, the average tariff in 1990-1991 was around 128 percent, with an import-weighted average of 87 percent.

exemption on imported inputs prior to export of the finished product; and the Export Promotion Capital Goods scheme, which provides exporters with access to foreign capital goods at reduced tariff rates.<sup>9</sup> Use of these exporter exemptions schemes has tended to grow significantly since the early 1990s (Panagariya, 1999). These schemes will tend to dilute the counter-lobbying by downstream users of intermediate goods as shown in the previous section.

#### IV. EMPIRICAL METHODOLOGY

As explained in the introduction, we base the estimation of equilibrium import tariffs on a careful distinction between organized and unorganized sectors, using an approach that derives the classification of sectors into organized vs. unorganized ones from the data and the model itself. Before describing the empirical methodology, let us rewrite equation (5) as:

$$\frac{|\varepsilon_i| \tilde{t}_i}{z_i} = \frac{I_i - \alpha_L}{a + \alpha_L} - \sum_{j=1}^n \frac{I_j - \alpha_L}{a + \alpha_L} (1 - \lambda_j) \frac{a_{ij} y_j}{y_i} \quad (6)$$

The advantage of rewriting (5) as (6) is twofold. First, sending  $z_i$  and  $\varepsilon_i$  to the left-hand side (LHS) eliminates the need to correct for measurement errors in  $\varepsilon_i$  and to instrument for  $z_i$  and  $\varepsilon_i$  which are both potentially endogenous. Second, the theoretical equation to be estimated now has a constant, which facilitates the interpretation of some results and avoids biasing the estimates in the presence of missing variables.

##### IV.1. Estimating structural parameters

The estimation problem is one of estimating simultaneously the parameters of a switching-regression model and the true classification of observations into two possible regimes: organized vs. unorganized, recognizing that (i) the true classification is unknown and (ii) it results from choices that are endogenous to the equilibrium level of protection. To deal with such problems, several approaches are possible, including the EM algorithm (McLachlan and Krishnan 1997) and grid-search procedures (Goldfeld and Quandt 1973), all suitably adapted to deal with the endogeneity problem and therefore all complex. We adopt a grid-search procedure in which the variable serving to define the cutoff between the two regimes is generated as part of the estimation procedure. This classification procedure is justified in Appendix 2.

9. For a complete description of these schemes, see Table III.4 in WTO (1998).

As argued in the introduction, our procedure has the advantage of bypassing the use of out-of-model proxies such as PAC contributions which can bias classifications and parameter estimates. It has four steps.

*Step 1.* A stochastic version of (6) is estimated where all  $I_i$  are set equal to zero (no information):<sup>10</sup>

$$\frac{|\varepsilon_i| \tilde{t}_i}{z_i} = \gamma_0 + \gamma_1 \sum_{j=1}^n a_{ij} \frac{y_j}{y_i} + \gamma_2 \sum_{j=1}^n \lambda_j a_{ij} \frac{y_j}{y_i} + \mu_i \quad (7)$$

where  $\mu_i$  is the error term. Although the algebra implies that  $\gamma_1 = -\gamma_0 = -\gamma_2 = \alpha_L / (a + \alpha_L) > 0$ ; no constraint is imposed at this stage other than  $I_i = 0 \forall i$ .

*Step 2.* Residuals are retrieved from the estimation of (7) and their magnitude is taken to indicate how successful each lobby was in obtaining protection. Let  $\sigma_\mu$  be the standard deviation of the error term. Let also  $\rho$  be a parameter, at this stage assigned an arbitrary value. The political-organization variable  $I_i$  is determined by the following rule:

$$I_i = \begin{cases} 1 & \text{if } \mu_i > \rho \sigma_\mu, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

In words, whenever the error term for observation  $i$  is algebraically higher than  $\rho$  times the standard deviation, sector  $i$  is deemed organized and  $I_i$  is set equal to one.

*Step 3.* The vector  $\mathbf{I} = (I_1, \dots, I_n)$  constructed according to (8) is introduced into a stochastic, unconstrained version of (6) which is then re-run. Because  $I_i$  appears in (6)'s first term while other elements of  $\mathbf{I}$  (the  $I_j$ ) appear in its third term, all terms involving counter-lobbying (terms in  $a_{ij}$ ) need to be recalculated, since in equilibrium they depend on whether sectors  $j = 1, \dots, n$  are organized or not.<sup>11</sup> In order to facilitate the interpretation of parameter estimates, we separate them from terms involving counter-lobbying dilution through duty drawbacks (terms involving  $a_{ij}$  and  $\lambda_j$ ).

10. Note that setting all initial  $I_i$  to one yields the same equation; only the initial parameters interpretation would be different.

11. For the same reason instrumenting for  $I$  using a polynomial and non-linear method to estimate the tariff equation is not possible as we would run out of degrees of freedom. A panel dataset would allow for this alternative method.

Before introducing constraints on the parameters, the equation to be estimated is then:

$$\begin{aligned} \frac{|\varepsilon_i| \tilde{t}_i}{z_i} &= \beta_0 + \beta_1 I_i + \beta_2 \sum_{j=1}^n \frac{a_{ij} y_j}{y_i} + \beta_3 \sum_{j=1}^n \frac{\lambda_j a_{ij} y_j}{y_i} \\ &+ \beta_4 \sum_{j=1}^n \frac{I_j a_{ij} y_j}{y_i} + \beta_5 \sum_{j=1}^n \frac{I_j \lambda_j a_{ij} y_j}{y_i} + \mu_i. \end{aligned} \quad (9)$$

The algebra implies that  $\beta_2 = \alpha_L / (a + \alpha_L) = -\beta_0 = -\beta_3 > 0$  and  $\beta_5 = \beta_1 = -\beta_4 = 1 / (a + \alpha_L) > 0$ . Thus, the estimation of (9) yields five estimated coefficients ( $\beta$ s) for only two unknown structural parameters:  $a$  and  $\alpha_L$ .

In order to later be able to retrieve these, we also proceed with the estimation of a constrained version of (9) using again (7) as a first step:

$$\begin{aligned} \frac{|\varepsilon_i| \tilde{t}_i}{z_i} &= \beta_0 \left[ 1 - \sum_{j=1}^n (1 - \lambda_j) \frac{a_{ij} y_j}{y_i} \right] \\ &+ \beta_1 \left[ I_i - \sum_{j=1}^n I_j (1 - \lambda_j) \frac{a_{ij} y_j}{y_i} \right] + \mu_i \end{aligned} \quad (10)$$

where  $\beta_0 = -\alpha_L / (a + \alpha_L) < 0$  and  $\beta_1 = 1 / (a + \alpha_L) > 0$ .

Note that selectivity is involved in the classification of industries into organized vs. unorganized ones because the decision to organize is endogenous to the level of protection that can be obtained in equilibrium (see Mitra 1999 for a theoretical treatment and GB 2000 for empirical implications). Taking the selectivity into account calls for a treatment-effect estimation procedure. We use the two-step estimator described in Maddala (1983) which consists of augmenting regression (9) or (10) –whichever is used– with the estimated hazard rate retrieved from an auxiliary probit regression of the binary variable  $I$  (itself retrieved from Step 2) on a set of instruments including the shares of capital and skilled labor in value added as well as the number of

---

11. For the same reason instrumenting for  $I$  using a polynomial and non-linear method to estimate the tariff equation is not possible as we would run out of degrees of freedom. A panel dataset would allow for this alternative method.

firms in each sector. This procedure yields consistent estimates of regression coefficients and standard errors.<sup>12</sup>

*Step 4.* A grid search is used to find the  $\rho$  that minimizes the sum of squared residuals. The final  $\mathbf{I}^*$  vector can then be retrieved.

Together, these four steps provide a theoretically consistent method to determine which sectors are politically organized to influence trade policy when data on political contributions is not available. The model's structural parameters can then be algebraically retrieved from the estimates of  $\beta_0$  and  $\beta_1$ . Solving for  $a$  and  $\alpha_L$  gives

$$a = \frac{1 + \beta_0}{\beta_1} \quad (11)$$

$$\alpha_L = -\beta_0/\beta_1. \quad (12)$$

## V. EMPIRICAL RESULTS

The construction of variables is described in detail in Table 1. It provides constrained regression estimates with and without input-output coefficients (first and second columns respectively) as well as unconstrained ones (third column). Constrained regression estimates are for  $\rho = 0,3$  with input-output coefficients. Both parameter estimates have the expected sign and  $\hat{\beta}_1$  is significant at the 1% level, a remarkably sharp signal for such a small sample (80 observations) and given the noisiness of input-output coefficients.

Using (11)-(12) and constrained regression estimates, we retrieved numerical values for  $a$  and  $\alpha_L$ . The implied estimate for  $a$ , the weight of welfare in the government's objective function, is  $\hat{a} = 3.09$ . A test of the non-linear constraint (11) suggest that  $\hat{a}$  is significantly different from zero at the one percent level. This estimate is well below estimates previously found for the United States, which were in the hundreds or even thousands, and arguably more plausible, although still too high in the sense that it implies that lobbies must contribute multiples of the monetary value of the distortions their demands imply.

The estimate for  $\alpha_L$ , the share of the population employed in industries organized politically for trade-related purposes, is  $\hat{\alpha}_L = 0.036 / 0.310 = 0.12$ . Note that whereas the denominator of this expression (equal to  $\hat{\beta}_L$ ) is estimated

12. Whether this type of procedure is really useful on a small sample like ours is debatable. As it turns out, the hazard rate's coefficient in the augmented regression was significant in all specifications, suggesting that the selectivity bias ought not to be ignored.

**Table 1**  
**Explaining India's tariffs<sup>a</sup>**

	Constrained with IO linkages	Constrained without IO linkages.	Unconstrained
Net lobbying ( $\beta_1$ )	0.310*** (2.99)		
Net welfare ( $\beta_0$ )	-0.036 (1.15)		
Lobbying ( $\beta_1$ )		0.311*** (2.63)	0.275*** (3.13)
Welfare ( $\beta_0$ )		0.003 (0.42)	
Int. Sales on welfare ( $\beta_2$ )			0.001 (0.05)
Duty Dr. on welfare ( $\beta_3$ )			0.189 (1.378)
Counter-lobbying ( $\beta_4$ )			-0.334* (1.69)
Duty Dr. on lobbying ( $\beta_5$ )			0.422 (1.21)
Hazard rate	-0.104* (1.82)	-0.107* (1.66)	-0.088* (1.78)
Implied $a^b$	3.09	3.23	
Implied $a^c$	0.12	-0.01	
R2 adjusted	0.56	0.51	0.60
Observations	80	80	80
# org. sectors	13	11	13
$\rho$ parameter	0.3	0.4	0.3

<sup>a</sup>Estimated using the algorithm described in section 4.1.

<sup>b</sup>See equation (11).

<sup>c</sup>See equation (12).

\*\*\* stands for statistical significance at the 1% level;

\*\* for significance at the 5% level, and \* for significance at the 10% level.

precisely, its t-statistic being 2.99, its numerator is not significantly different from zero. This estimate of the organized share of the population, which is thus itself not significantly different from zero (a test of the non-linear constraint (12) fails to reject the hypothesis that  $\alpha_L$  is zero), is, in our view, more plausible than those found previously, some of which were as high as 70% on the basis of US data.

Results for the constrained regression without input-output linkages (i.e. without counter-lobbying by downstream users) are shown in the second

column of Table 1 and are not very different from those of the first column. Thus, the inclusion of input-output linkages is not what drives the sharp reduction in the estimated value of  $a$  that we report here.

The results of the unconstrained regression suggest that producer lobbying ( $I_i$ ) has a positive and statistically significant influence (at the 1% level) on tariffs. Similarly, counter-lobbying by organized downstream sectors ( $\sum_j I_j a_{ij} y_j / y_i$ ) tends to reduce tariffs as predicted by the model and is also statistically significant at the 10% level. The counter-lobbying dilution effect ( $\sum_j I_j \lambda_j a_{ij} y_j / y_i$ ) introduced by duty drawback schemes granted to exporters has the right sign, but it is not statistically significant. The non-lobbying terms are all statistically insignificant. This suggests that we may be asking too much information from a relatively small and noisy sample, which implies that introducing the theoretically valid constraints makes empirical sense.

Table 2 provides a list of organized and unorganized sectors. The number of organized sectors is low (13 out of 80). The identity of organized vs. unorganized sectors makes sense. Sectors derived as unorganized in India include those which are typically organized for protectionist lobbying in industrial countries: practically all the textile and clothing industry, footwear, furniture, and steel. Thus, broadly speaking, the predicted pattern of political organization appears consistent with the notion that “losers” (sectors in which a country has a comparative disadvantage) are more likely to organize themselves for political action than “winners”.

## VI. ROBUSTNESS

### VI.1 Classification of industries

One key selling point of the method proposed in this paper is its ability to bypass the usual problem of identifying which sectors are organized, because identification is obtained as a result of the estimation itself. It is therefore important to assess how stable are the method's predictions as to who is organized and who is not. In order to shed some light on this, we performed Monte Carlo experiments on simulated data in order to generate a sampling distribution for the vector of organized sectors. Instead of starting from a purely arbitrary data set, we used Gawande and Bandyopadhyay's US data and constructed an initial vector of organized sectors using our method.<sup>13</sup> We then generated a thousand alternative data sets by adding a white-noise term

---

13. We are grateful to them for providing us the data.

**Table 2**  
**Estimated organization dummy(I) <sup>a</sup>**

Unorganized Sectors		Organized Sectors			
3111	Meat prep.	3513	Synthetic resins	3212	Made-up textile goods
3112	Dairy prod.	3521	Paints	3233	Leather ex. Footwear
3113	Canned fruit and veg.	3522	Drugs and medec.	3511	Chemicals
3114	Canned fish and crust	3523	Soap and cleaning	3530	Petroleum refineries
3115	Veg and an. oils fats	3529	Chemical prod. Nec.	3720	Non-ferrous metal.
3116	Grain mill prod.	3540	Misc of petroleum	3823	Metal and wood machinery
3117	Bakery prod.	3551	Tyre and tube indus.	3824	Special ind. Machinery
3118	Sugar refineries	3559	Rubber prod. n.e.c.	3825	Off. and comput. eq.
3119	Cocoa confect.	3560	Plastic prod. n.e.c.	3829	Equip. except. electr.
3121	Food prod. n.e.c.	3610	"Pottery, china"	3832	Radio, TV, and comm.
3122	Prep. Animal feeds	3620	Glass	3851	Scientific equipment
3131	Alcoholic bev.	3691	Structural clay	3852	Optical goods
3132	Wine industries	3692	Cement, lime	3901	Jewellery
3133	Malt liquors and malt	3699	Non-metallic nec		
3134	Soft drinks	3710	Iron and steel		
3140	Tobacco manuf.	3811	Cutlery, hand-tools		
3211	Spinning textiles	3812	Metal Furniture		
3213	Knitting mills	3813	Metal Products		
3214	Manuf. Carpets and rugs	3819	Metal prod. n.e.c.		
3215	Cordage, rope	3821	Engines		
3219	Textiles nec	3822	Agricul. Machinery		
3220	Apparel, ex. Footwear	3831	Electr. ind. mach.		
3231	Tanneries and leather	3833	Electrical appliances		
3232	Fur dressinganddyeing	3839	Electrical eq. Nec		
3240	Footwear	3841	Ship building		
3311	Sawmill sandwood mills	3842	Railroad equipment		
3312	Wooden and c. cont.	3843	Motor vehicles		
3319	Wood and c. prod. n.e.c.	3844	Motorcyc. and bic.		
3320	Furnitures	3845	Manuf. of aircraft		
3411	Paper and Paper board	3849	Transport equip. n.e.c.		
3412	Paper boxes	3853	Watches and clocks		
3419	Pulp and paper	3902	Musical inst.		
3420	Printing and publ.	3903	Sporting goods		
3512	Fertilizers and pest.	3909	Manuf. prod. n.e.c.		

<sup>a</sup>These are the estimated organized sectors for trade purposes using the four steps of our empirical methodology described in section 4.

to the endogenous variable and, each time, re-estimating the model using our procedure. Finally, we calculated sample statistics for the thousand vectors of the predicted organizational dummy variable. Formally, our procedure ran as follows.

The first step consisted in adding a shock to the dependent variable. The shock was a normally-distributed random variable with mean zero and a standard error equal to 20% of the sample estimate of the dependent variable's standard error. This initial step, however, was less than straightforward given that the dependent variable used by Gawande and Bandyopadhyay ( $n_i / (1 + n_i)$  where  $n_i$  is the coverage ratio of non-tariff barriers for sector  $i$ )



is left- and right-censored (at zero and one half respectively). It can be shown (see e.g. Greene, 1997) that adding white noise to a censored variable raises its mean. We corrected for this using a first-order approach.<sup>14</sup> In addition, adding white noise to censored values does not make much sense, so instead we used the predicted value of the dependent variable for censored observations, added noise to this predicted value, and then censored when appropriate.

The next step was to apply our estimation method using a procedure based on Smith and Blundell's exogeneity test (Smith and Blundell 1986).<sup>15</sup> That is, in the first step we regressed the organizational dummy  $I_i$  on all the instruments we obtained from Gawande and Bandyopadhyay, including squared and interaction terms (see their paper for details). In the second step, residuals from this auxiliary regression were used as a RHS variable in the usual regression with non-tariff barriers as the dependent variable (we actually divided the dependent variable by  $z_i/e_i$  in order to eliminate the latter from the RHS and thus avoid simultaneity and measurement-error issues).<sup>16</sup> The second equation was estimated by Tobit to take care of the censoring. Next, we retrieved the residuals from this second regression and used them to classify sectors as organized or not based on an arbitrary cutoff value, and then searched for the cutoff value minimizing the sum of squared errors. Finally, we recorded the predicted value  $\hat{I}_i$  of  $I_i$ : This procedure was then replicated a thousand times. The results of this experiment are quite striking. Let  $\hat{\mathbf{I}}_0 = (I_1, \dots, I_n)'$  be the vector of dummy variables marking the organization of sectors  $i = 1, \dots, n$  in the initial data set,  $k = 1, \dots, 1000$  index replications, and  $\hat{\mathbf{I}}_k$  the predicted value of  $\hat{\mathbf{I}}_0$  at replication  $k$ . Let also  $\rho_k = \text{corr}(\hat{\mathbf{I}}_k, \hat{\mathbf{I}}_0)$ . The average value of  $\rho_k$  across the thousand replications was  $\bar{\rho} = 0.745$  (0.711 without the first-order correction, labeled FOC). Calling "type-I errors" sectors for which  $\hat{I}_k = 1$  whereas  $\hat{I}_{ik} = 0$  and vice-versa for type-II, the percentage of type-I errors was 2.15% and that of type-II errors 0.93% (3.67% and

14. Formally, consider two latent variables  $y^* \sim N(\mu, \sigma^2)$  and  $z^* = y^* + \varepsilon$  where  $\varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$ . Let  $\mu_z$  and  $\sigma_z^2$  denote the mean and variance of  $z$ . Suppose that we observe

$$y = \begin{cases} y^* & \text{if } y^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and similarly for  $z$ . That is, both are censored at zero. Finally let  $\lambda = \Phi/\phi$  be the inverse Mills ratio. Then it can be shown that  $E(y) = (\mu + \lambda\sigma) \Phi(\mu/\sigma)$  and that

$$E(z) \simeq E(y) \text{ if } \mu_\varepsilon = -\lambda(\sqrt{\sigma^2 + \sigma_\varepsilon^2} - \sigma)$$

This is the correction we used. Right-censoring was ignored as it concerned only three observations out of 242.

15. We used Smith and Blundell's procedure rather than the one we used for India's tariffs (a variant of Heckman's selection model) because the latter is unsuitable to censored data.

16. See Mitra et al. (2006) for a similar approach.

0.76% respectively without the first-order correction). The total percentage of “wrong” predictions was thus 3.07% with the first-order correction and 4.42% without. The distribution of optimal cutoff values was centered on 1.4 (1.5 without the FOC), for a “true” (initial) value of 1.5. Finally, the average error on the estimated value of  $a$  was 4.4% (13.88% without the FOC).

## 6.2 Consistency across estimators

As a further check, we estimated the specification without IO linkages using three different estimators: our four-step grid search, a standard maximum-likelihood grid-search estimator (Goldfeld and Quandt, 1958),<sup>17</sup> and a maximum likelihood EM algorithm. Results are shown in Table 3.

**Table 3**  
Explaining India's tariffs (without IO linkages)  
using different estimators

	Grid Search 4 steps	Grid Search ML	EM algorithm
Lobbying ( $\beta_1$ )	0.311*** (2.63)	0.169*** (2.46)	0.287*** (2.31)
Welfare ( $\beta_0$ )	0.003 (0.42)	0.004*** (40.4)	0.009*** (3.70)
Implied $a^b$	3.23	5.78	3.37
Implied $\alpha^c$	-0.01	-0.02	-0.03
Observations	80	80	80
# org. sectors	11	19	11

\*\*\* stands for statistical significance at the 1% level; \*\* for significance at the 5% level, and \* for significance at the 10% level.

<sup>b</sup>See equation (11).

<sup>c</sup>See equation (12).

The three estimation methods provide estimates of  $\beta_1$  that do not differ significantly from each other, and the resulting three estimates of  $a$  all lie within a reasonable range. Results using our four-steps grid search approach are particularly close to those obtained using the EM algorithm.

It would also have been interesting to estimate the specification *with* IO linkages using the ML grid search and EM algorithm, but we do not believe

17. See also Hotchkiss (1991) for a recent application.

that this is feasible, because when IO linkages are introduced the problem is no longer just one of unknown sample split (between organized and not-organized sectors), but of recalculating some of the regressors at each step on the basis of provisional sample-split estimates. Nevertheless, the fact that our four-step grid search—which circumvents this problem—yields results fairly similar to those obtained from two standard estimators gives us confidence in its robustness and consistency.

## VII. CONCLUDING REMARKS

The objective of this paper was to provide an empirical method to identify jointly, on the basis of the Grossman-Helpman approach, what are the driving forces behind observed patterns of trade protection and which sectors find it profitable to organize themselves for trade policy influence.

This endeavour is important for two reasons. First, outside of the United States, no information is available on the activity of special-interest groups and on their degree of organization. Taking Grossman-Helpman outside of the US, in particular to emerging countries where influence via monetary contributions is most likely to be prevalent, requires an indirect method such as ours. Second, our approach bypasses the problem of disentangling the share of contributions directed at trade-policy influence from the share directed at domestic policies.

Beyond methodological issues, our approach provides further vindication of the common-agency approach to trade-policy determination, yielding plausible results on the forces that shape India's trade policy and on the pattern of political organization across tradeable sectors. We explore trade-policy determination in a formulation embodying vertical linkages through the use of an input-output matrix, so that all tariffs are determined and estimated simultaneously. We also include the effect of duty-drawback schemes whereby exporters recover duties paid on imported intermediate inputs. These schemes reduce the incentive to lobby against upstream protection. We find that the cross-industry pattern of protection relates to import penetration and price elasticities of import demands in the way predicted by the theory, and that resistance to upstream protection is to some extent diluted by duty drawbacks, although this last effect is not statistically significant.

The weight on welfare in the government's objective function implied by our estimates is 3.09, well below recent estimates ranging between forty and three thousand. This number is still implausibly high in that it implies that

a lobby should contribute three rupees to the government for each rupee of deadweight loss. Given the size of the deadweight losses estimated by empirical studies of the cost of trade protection (see e.g. Gawande and Krishna's 2003 survey) this would put the price tag of protection at a prohibitive level. On that criterion, however, our estimate appears "closer to the truth" than previous ones by a substantial margin. As for the pattern of political organization, we find that organized industries include sectors in which India has a comparative disadvantage (e.g., machinery), the pattern of protection and lobbying being, in some sense, the mirror image of that which prevails in industrial countries.

In the spirit of the political-economy literature, our results are positive rather than normative. However they yield two direct policy implications. Protection can be expected to go down, in India as elsewhere, only if it becomes a less attractive political proposition. Reducing the political attractiveness of protection could be achieved in two ways. Either downstream users could be encouraged to get organized to lobby against protection in upstream industries and possibly even assisted in it. Or, alternatively, existing exemption schemes for exporters could be eliminated in order to make political organization more attractive for them. Put differently, as a policy tool, duty drawbacks which in a traditional analysis would appear to mitigate the inefficiency effects of import protection may end up being counterproductive if they neutralize a group of concentrated users who could otherwise constitute a politically powerful counterlobbying force.

### VIII. REFERENCES

- Baldwin, R. (1985). "The Changing Nature of US Trade Policy Since World War II". In R. Baldwin and A. Krueger, eds., *The Structure and Evolution of Recent US Trade Policy*. University of Chicago Press for NBER.
- Becker, G. (1983). "A Theory of Competition among Pressure Groups for Political Influence". *Quarterly Journal of Economics*, 98(3), 371-400.
- Bernheim, D. and Whinston M. (1986a). "Menu Auctions, resource allocation, and economic influence". *Quarterly Journal of Economics*, 101(1), 1-31.
- Bernheim, D. and Whinston M. (1986b). "Common agency". *Econometrica*, 54(4), 923-942.
- Cadot, O., de Melo, J., and Olarreaga M.(2003), "The protectionist bias of duty drawbacks: evidence from Mercosur", *Journal of International Economics* 59 (1), 161-182.

- Cadot, O., de Melo, J., and Olarreaga M. (2004). "Lobbying, counterlobbying and the structure of tariff protection in poor and rich countries". *World Bank Economic Review*, 18 (3), 1-22.
- Damania, R., and Frederiksson P. (2007). "Trade Policy: What's Welfare Got to Do With It?". *Economic Letters*, 96(1), 64-69.
- Gawande, K., and Bandyopadhyay, S. (2000). "Is protection for sale? A test of the Grossman-Helpman Theory of Endogenous Protection". *Review of Economics and Statistics*, 82(1), 139-152.
- Gawande, K., and Krishna, P. (2005). "Lobbying Competition Over US Trade Policy". NBER Working Paper 11371.
- Gawande, K., and Krishna, P. (2003). "The political economy of trade policy: empirical approaches". In J. Harrigan, ed., *Handbook of International Trade*, New York: Basil Blackwell.
- Gawande, K., Sanguinetti P., and Bohara, A. (2004). "Trade diversion and declining tariffs: evidence from Mercosur". *Journal of International Economics* 64(1), 65-88.
- Goldberg, P., and Maggi G. (1999). "Protection for Sale: an Empirical Investigation". *American Economic Review*, 89(5), 1135-1155.
- Goldfeld, S., and Quandt R. (1973). "The estimation of structural shifts by switching regressions". *Annals of Economic and Social Measurement*, 2(4), 475-485.
- Grossman, G., and Helpman E. (1994). "Protection for sale". *American Economic Review*, 84(4), 833-850.
- Hotchkiss, J. (1991). "The Denition of Part-Time Employment: A Switching Regression Model with Unknown Sample Selection". *International Economic Review* 32(4), 899-917.
- Kee, H. L., Nicita A. and Olarreaga M. (2009). "Estimating trade restrictiveness". *Economic Journal*, 119 (534), 172-199.
- Kee, H. L., Nicita A. and Olarreaga M. (2008). "Import demand elasticities and trade distorsions". *Review of Economics and Statistics*, 90(4), 666-682.
- Lee, J.W., and Swagel P. (1997). "Trade Barriers and Trade Flows Across Countries and Industries". *Review of Economics and Statistics*, 79(3), 372-382.
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, Econometric Series Monograph.
- Marvel, H., and Ray, E. (1983). "The Kennedy Round: Evidence on the Regulation of International Trade in the United States". *American Economic Review*, 73(1), 190-197.

- McCalman, P. (2004). "Protection for Sale and Trade Liberalization: an Empirical Investigation". *Review of International Economics*, 12(1), 81-94.
- McLachlan, G., and Krishnan T. (1997). *The EM Algorithm and Extensions*, Wiley.
- Mitra, D. (1999). "Endogenous Lobby Formation and Endogenous protection: A Long-Run Model of Trade Policy Determination"; *American Economic Review*, 89 (5), 1116-1134.
- Mitra, D., Thomakos D., and Ulubasoglu M. (2002). "Protection for sale in a developing country: democracy versus dictatorship". *Review of Economics and Statistics*, 84 (3), 497-508.
- Mitra, D., Thomakos D., and Ulubasoglu M. (2002). "Can We Obtain Realistic Parameter Estimates for the 'Protection for Sale' Model?". *Canadian Journal of Economics* 39(1), 187-210.
- Panagariya, A. (1999). "The WTO Trade Policy Review of India, 1998". *World Economy*, 22(6), 799-824.
- Peltzman, S. (1976). "Toward a More General Theory of Regulation". *Journal of Law and Economics*, 19(2), 211-240.
- Pursell, G. (1996). "Indian trade policies since the 1991/1992 reform". Mimeo. The World Bank.
- Ray, E. J. (1981). "The Determinants of Tariff and Nontariff Trade Restrictions in the United States". *Journal of Political Economy*, 89(1), 105-121.
- Rodrik, D. (1995). "Political economy of trade policy". In G. Grossman and K. Rogoff, eds., *Handbook of International Economics*, vol. III, North Holland.
- Smith, R. J., and Blundell R. (1986). "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply". *Econometrica*, 54(3), 679-686.
- Stigler, G. (1971). "The Theory of Economic Regulation". *Bell Journal of Economic Management and Science*, 2(1), 3-21.
- Srinivasan, T.N. (2001). "India's reform of external sector policies and future multi-lateral trade negotiations". Economic Growth Center discussion paper # 830, Yale University.
- Trefler, D. (1993). "Trade Liberalization and the Theory of Endogenous Protection: An Econometric Study of US Import Policy". *Journal of Political Economy* 101(1), 138-160.
- WTO (1998). "Trade Policy Review of India, 1998". World Trade Organization, Geneva.

**APPENDIX 1**

We now calculate the derivatives in (4) term by term. Suppose that a duty drawback (DD) scheme is in place whereby import duties paid on inputs used by sector  $j$ 's firms when producing for exports are reimbursed.<sup>18</sup> Let  $\lambda_j$  be the share of good  $j$ 's production that is exported. As import-competing domestic producers align their prices on the tariff-ridden price of imported goods, on the cost side it does not matter whether intermediates are imported or sourced domestically.<sup>19</sup> The unit cost of intermediate good  $i$  to user sector  $j$  is given by

$$\phi_{ij} = (1 - \lambda_j)p_i^*(1 + t_i) + \lambda_j p_i^* = p_i^*[1 + t_i(1 - \lambda_j)].$$

Sector  $j$ 's profits are

$$\begin{aligned} \pi_j(\mathbf{t}) &= p_j^*(1 + t_j)y_j - \sum_{i=1}^n \phi_{ij}a_{ij}y_j = \\ & \left[ p_j^*(1 + t_j) - \sum_{i=1}^n p_i^*[1 + t_i(1 - \lambda_j)]a_{ij} \right] y_j, \end{aligned}$$

so

$$\frac{\partial \pi_j}{\partial t_i} = \begin{cases} p_i^*[1 - (1 - \lambda_i)a_{ii}]y_i & \text{for } j = i \\ -p_i^*(1 - \lambda_j)a_{ij}y_j & \text{otherwise.} \end{cases}$$

Let  $m_i$  be imports of good  $i$ . Aggregate tariff revenue net of duty-drawback refunds is

$$T = \sum_i p_i^* t_i \left( m_i - \sum_{j=1}^n \lambda_j a_{ij} y_j \right)$$

so in the absence of cross-price effects on either supply or demand sides,

18. See Cadot, de Melo and Olarreaga (2003) who showed in a similar setting that in equilibrium the optimal level of reimbursement is full reimbursement of import duties.

19. Given that only imported intermediates are eligible for the DD, in equilibrium all intermediates used in the production of goods for export are imported. This implies a set of constraints of the form  $\sum_j \lambda_j x_{ij} \leq m_i$  for all  $i$ . In our data set these constraints are verified for most sectors except where special regimes apply, e.g. for alcohols.

$$\frac{\partial T}{\partial t_i} = p_i^* \left[ m_i - \sum_{j=1}^n \lambda_j a_{ij} y_j + t_i \left( m_i' - \sum_{j=1}^n \lambda_j a_{ij} \frac{\partial y_j}{\partial \tilde{p}_j} \frac{\partial \tilde{p}_j}{\partial t_i} \right) \right] \quad (13)$$

where  $\tilde{p}_j = p_j^*(1 + t_j) - \sum_{i=1}^n a_{ij} p_i^* [1 + t_i(1 - \lambda_j)]$  is the “net price” of good  $j$ . Letting  $y'_j$  stand for  $\partial y_j / \partial \tilde{p}_j$  and noting that

$$\frac{\partial \tilde{p}_j}{\partial t_i} = \begin{cases} p_i^* [1 - a_{ii}(1 - \lambda_i)] & \text{if } j = i \\ -p_i^* a_{ij}(1 - \lambda_j) & \text{otherwise,} \end{cases}$$

we have

$$\begin{aligned} & \sum_{j=1}^n \lambda_j a_{ij} \frac{\partial y_j}{\partial \tilde{p}_j} \frac{\partial \tilde{p}_j}{\partial t_i} \\ &= p_i^* \left\{ \lambda_i a_{ii} [1 - a_{ii}(1 - \lambda_i)] y'_i - \sum_{j \neq i} \lambda_j a_{ij}^2 (1 - \lambda_j) y'_j \right\} \\ &= p_i^* \left[ \lambda_i a_{ii} y'_i - \sum_{j=1}^n a_{ij}^2 \lambda_j (1 - \lambda_j) y'_j \right]. \end{aligned}$$

$$\text{Let } \xi_i = p_i^* \left[ \lambda_i a_{ii} y'_i - \sum_{j=1}^n a_{ij}^2 \lambda_j (1 - \lambda_j) y'_j \right]$$

Substituting this into (13) gives:

$$\frac{\partial T}{\partial t_i} = p_i^* \left[ m_i - \sum_{j=1}^n \lambda_j a_{ij} y_j + t_i (m_i' - \xi_i) \right].$$

Consumer surplus is

$$S = \sum_i u(c_i) - p_i^*(1 + t_i)c_i$$

$$\frac{\partial S}{\partial t_i} = u'(c_i)c_i' - p_i^*c_i - p_i^*(1 + t_i)c_i' = -p_i^*c_i.$$



Combining these gives

$$\frac{\partial v_i}{\partial t_i} = p_i^* \left\{ \alpha_i \left[ m_i - \sum_{j=1}^n \lambda_j a_{ij} y_j + t_i (m'_i - \xi_i) - c_i \right] + [1 - (1 - \lambda_i) a_{ii}] y_i \right\}.$$

Noting that  $m_i - c_i = \sum_j a_{ij} y_j - y_i$ ,

$$\frac{\partial v_i}{\partial t_i} = p_i^* \left\{ \alpha_i \left[ t_i (m'_i - \xi_i) + \sum_{j=1}^n (1 - \lambda_j) a_{ij} y_j - y_i \right] + [1 - (1 - \lambda_i) a_{ii}] y_i \right\} \tag{14}$$

and

$$\frac{\partial v_j}{\partial t_i} = p_i^* \left\{ \alpha_j \left( t_i (m'_i - \xi_i) + \sum_{j=1}^n (1 - \lambda_j) a_{ij} y_j - y_i \right) - (1 - \lambda_j) a_{ij} y_j \right\} \tag{15}$$

Adding up (14) and (15), aggregating and letting  $\alpha_L = \sum_{j=1}^n I_j \alpha_j$  be the proportion of the population belonging to organized lobbies gives

$$\sum_{j=1}^n I_j \frac{\partial v_j}{\partial t_i} = p_i^* \left\{ \alpha_L \left[ t_i (m'_i - \xi_i) + \sum_{j=1}^n (1 - \lambda_j) a_{ij} y_j \right] + (I_i - \alpha_L) y_i - \sum_{j=1}^n I_j (1 - \lambda_j) a_{ij} y_j \right\}$$

Similar calculations for aggregate welfare give

$$\frac{\partial W}{\partial t_i} = \frac{\partial T}{\partial t_i} + \frac{\partial S}{\partial t_i} + \sum_{j=1}^n \frac{\partial \pi_j}{\partial t_i} = p_i^* t_i (m'_i - \xi_i). \quad (16)$$

Combining these, simplifying and rearranging gives the following FOC:

$$\begin{aligned} \frac{1}{p_i^*} \frac{\partial G(\mathbf{t})}{\partial t_i} &= (a + \alpha_L) t_i (m'_i - \xi_i) + (I_i - \alpha_L) y_i \\ &\quad - \sum_{j=1}^n (I_j - \alpha_L) (1 - \lambda_j) a_{ij} y_j = 0, \end{aligned}$$

or, after isolating  $t_i$  on the LHS and simplifying,

$$t_i = \frac{(I_i - \alpha_L) y_i}{-(a + \alpha_L)(m'_i - \xi_i)} - \sum_{j=1}^n \frac{(I_j - \alpha_L) (1 - \lambda_j) a_{ij} y_j}{-(a + \alpha_L)(m'_i - \xi_i)}. \quad (17)$$

In order to convert this expression into elasticities, let  $\varepsilon_i$  be the own-price elasticity of good  $i$ 's import demand (in algebraic, not absolute value; i.e.  $\varepsilon_i < 0$ ). In order to limit the demands on data, we will suppose that 'net-price' supply elasticities are all zero.<sup>20</sup> Letting  $\tilde{t}_i = t_i / [p_i^* (1 + t_i)]$  and  $z_j = y_j / m_j$ , (17) can be converted into elasticities:

$$\tilde{t}_i = \frac{I_i - \alpha_L}{a + \alpha_L} \frac{z_i}{|\varepsilon_i|} - \sum_{j=1}^n \frac{I_j - \alpha_L}{a + \alpha_L} (1 - \lambda_j) \frac{a_{ij} z_j}{|\varepsilon_i|}. \quad (18)$$

## APPENDIX 2

Let  $y_1$  and  $y_2$  be two latent (unobserved) random variables corresponding to two regimes of an observed variable  $y$ . That is, indexing individual observations by  $i = 1 \dots n$ ,

$$y_i = \begin{cases} y_{i1} & \text{if observation belongs to regime 1} \\ y_{i2} & \text{if observation belongs to regime 2} \end{cases}$$

Let  $w_i$  be another latent random variable such that

$$w_i = \begin{cases} 1 & \text{if observation belongs to regime 1} \\ 0 & \text{if observation belongs to regime 2} \end{cases}$$

---

20. Attempts to estimate supply elasticities from the data proved unconvincing.

and let  $\pi$  be the probability that observation  $i$  belongs to regime 1, which we assume independent of  $i$ ; that is,

$$\pi = \text{prob} (w_i = 1)$$

Suppose that  $y_1$  and  $y_2$  are both normally distributed with means  $\mu_1$  and  $\mu_2$  respectively and with common variance  $\sigma$  and denote their densities by  $f_1$  and  $f_2$  respectively. An example of this setting is a model where  $u_{i1}$  and  $u_{i2}$  are i.i.d. normally distributed white-noise terms and

$$y_{i1} = \beta_1 + u_{1i} \tag{19}$$

$$y_{i2} = \beta_1 + \beta_2 + u_{2i} \tag{20}$$

Suppose that  $\beta_1 > 0$  and  $\beta_2 > 0$  in (19)-(20) and let  $\hat{\beta}_1$  and  $\hat{\beta}_2$  be their OLS estimates, based on the true sample split. Sort observations so that  $i = 1 \dots n_1$  belong to regime 1 and  $i = n_1 + 1 \dots n$ , to regime 2 and let

$$\begin{aligned} b_1 &= \hat{\beta}_1 = \bar{y}_1, \\ b_2 &= \hat{\beta}_1 + \hat{\beta}_2 = \bar{y}_2. \end{aligned}$$

Let also  $b = \bar{y}$  be an “average” estimator based on the whole sample. Obviously,  $b$  is a biased estimator of either  $\beta_1$  or  $\beta_2$ . Suppose, without loss of generality, that  $\alpha_1 < \alpha_2$ . Obviously,

$$E (b_1) < E (b) < E (b_2) .$$

Then, let  $E (\bar{e}_i) = E (y_i - b)$  .

If  $i \leq n_1$ ,  $E (\bar{e}_i) < 0$ ; if  $i > n_1$ ,  $E (\bar{e}_i) > 0$ . To see this, simply observe that if  $i \leq n_1$ ,  $E (y_i - b)$  is negative and conversely if  $i > n_1$ . Thus, residuals from the first regression give a correct classification rule. This observation carries over to subsequent regressions where observations are classified either in regime 1 or in regime 2, by establishing that regime-1 observations misclassified in regime 2 will have negative residuals and conversely for regime-2 observations misclassified in regime 1. To see this, let

$$\tilde{e}_i = \begin{cases} y_i - b_2 & \text{if } i \leq n_1 \\ y - b & \text{if } i > n_1 . \end{cases}$$

That is,  $\tilde{e}_i$  is the residual when observations are misclassified. Then  $E (\tilde{e}_i) \leq 0$  if  $i \leq n_1$ , as  $E (\tilde{e}_i) = (\alpha_1 - \alpha_2) x_i < 0$  if  $i \leq n$  and conversely if  $i > n$ . Thus, residuals are negative for regime-1 observations (unorganized sectors) misclassified in regime 2 (organized) and positive for

regime-2 observations (organized) misclassified in regime 1 (unorganized). It follows that the repeated use of a criterion assigning large residuals to regime 2 and small ones to regime 1 yields a valid proxy for the definition of the classification cutoff.<sup>21</sup>

## DATA APPENDIX

Tariff data is for the year 1997 and its source is India's WTO notification to the Integrated Database System of the WTO. The data comes originally at the six digit of the Harmonised System (5112 tariff line). It was converted to 4 digit of the ISIC classification for manufactures (81 sectors) using a filter developed at the World Bank and which is available from the authors upon request. Output and other industry-type data (employment, number of firms, etc...) are for the years 1993-1995 (average) and its source is UNIDO's Industrial database. Because output is measured at domestic prices, whereas imports are measured at world prices, in order to construct the import penetration ratio ( $z_j$ ), output was divided by  $(1+t_j)$  before dividing it by imports, so that they are both measured at world prices. Trade data is also for the years 1993-1995 and its source is United Nation's Comtrade. It comes originally at the six digit level of the Harmonized System (HS) and it was filtered using the same concordance as for the tariff data. Import demand elasticities were estimated at the ISIC 4 digit level for more than 100 countries by Kee, Nicita and Olarreaga (2008).

The input-output matrix is for the year 1994 and its source is the social accounting matrix of GTAP. The GTAP commodity classification differs from the ISIC classification, but a concordance exists to the 3 digit of the ISIC (available upon request).<sup>22</sup> Because our tariff and industry level data is at the 4 digit of the ISIC, we inflate the input-output components of the Social accounting matrix assuming that intermediate sales to GTAP category  $j$  are allocated to ISIC 4 digit sectors in GTAP category  $i$  according to output shares. The value and the share of intermediate sales in total output for each ISIC 4 digit sector is then calculated using this "inflated" input-output matrix. Capital stocks are calculated using historic data on gross fixed capital formation from UNIDO, using the permanent inventory method and a 10 percent annual depreciation rate.

21. The proof relies on an approximation, as  $\hat{\alpha}_2$  is not exactly given by (22) when a regime-1 observation is misclassified as belonging to regime 2 since the summations will then include the wrongly-classified observation. Using the exact formula for the "wrong" estimator complicates the notation but does not change the argument.

22. There are five sectors in the GTAP categories that have no correspondance in the ISIC classification, such as services for example.