A Quantile-based Test of Protection for Sale Model When Political Organization is Unknown

Susumu Imai, Queen's University

Hajime Katayama, University of Sydney

Kala Krishna, Princeton University, Pennsylvania State University and NBER

May 28, 2007

Abstract

We propose and conduct a simple test of the protection for sale model that does not require information on political organization. Our approach exploits the following prediction of the PFS model: given the inverse import penetration ratio and other control variables, industries with higher protection are more likely to be politically organized, and thus for those industries, we should expect a positive relationship between the inverse import penetration ratio and the protection measure. The quantile regression and quantile IV results are, if anything, the opposite of the above prediction. This cast some doubt on the empirical validity of the PFS model. On the other hand, a similar estimation exercise on the simulated data on a simple non optimizing model meant to loosely capture the institutional setup in the US (the surge protection model proposed in Imai et. al. (2006)) seems much more consistent with the quantile regression/IV results in the data.

1 Introduction

Recently there has been much interest in political economy aspects of trade policy. This growing interest is in part triggered by the theoretical framework in the Grossman and Helpman (1994) "Protection for Sale" model (hereafter the PFS model). Empirical studies such as Goldberg and Maggi (1999) and Gawande and Bandyopadhyay (2000) found that the data on trade protection are consistent with predictions by the PFS model. In particular, their results show that as predicted by this framework, protection is positively related to the import penetration ratio for politically unorganized industries, while negatively related for politically organized ones.

An important issue in these empirical studies is how to classify industries into politically organized and unorganized ones. When classifying industries, Goldberg and Maggi (1999) and Gawande and Bandyopadhyay (2000) have encountered the following problem: while only politically organized industries are assumed to make campaign contributions in the PFS model, their data indicate that all industries make Political Action Committees' (PAC) contributions.

Thus, if they were to follow the assumption in the model, all industries would be classified as politically organized. To overcome this problem, Goldberg and Maggi (1999) and Gawande and Bandyopadhyay (2000) used somewhat arbitrary rules to categorize industries. More recently, a second generation of empirical studies has taken a different approach to reconciling theory and the data. For example, Ederington and Minier (2006) extend the PFS model by hypothesizing that industries can lobby for both trade and domestic policies. In their model, it is possible that some industries are politically unorganized for trade policies but make contributions for domestic policies. Matschke (2006) takes a similar approach. Since the models by Ederington and Minier (2006) and by Matschke (2006) are more comprehensive than the PFS model, the authors impose additional assumptions to make the models tractable for estimation.

This paper proposes a new approach to testing the PFS model. Unlike most previous studies, our approach does not require classification of industries into organized and unorganized ones. This is important both because of the above mentioned problems in such a classification and because political contribution data itself is not available for most countries. In this manner, our approach can expand the realm of application of such models.

Our approach exploits the following prediction of the PFS model: politically organized industries should have higher protection than unorganized ones given the inverse import penetration ratio and other control variables. This suggests that industries with higher protection are more likely to be politically organized, and thus for those industries, we should expect a positive relationship between the inverse import penetration ratio and the protection measure.

We provide a formal proof of the above argument within the framework of recent work in quantile regressions and quantile IV's. To empirically test this implication, we use estimation techniques such as quantile regression (Koenker and Bassett, 1978) and instrumental variable quantile regression (Chernozhukov and Hansen, 2004a; 2004b). Our estimation results cast serious doubt on the validity of the PFS framework: contrary to the PFS prediction, the coefficient on the inverse import penetration ratio decreases, as the quantile increases. Note this is the exact opposite of what the PFS model suggests.

We also consider the alternative "Surge Protection" model (hereafter the SP model) proposed by Imai et. al. (2006). This is a simple non optimizing model meant to loosely replicate the institutional setup in the U.S. The idea is that today, most countries have signed the GATT and joined the WTO. in doing so, they have bound their tariffs and committed to limits on their ability to change trade policy. As a result, the main scope for trade policy lies in the safeguard or escape clause realm where temporary protection may be afforded an industry that is under stress and is organized enough to lobby for protection. This is modelled as follows. Politically organized industries can obtain a limit on imports if imports increases above a specified threshold. Data is then generated from a calibrated version of this SP model and the PFS test is conducted on this generated data. They find that even this data seems to fit the PFS model, casting doubt on the ability of the standard tests to distinguish between the PFS model and the SP model.

When quantile regressions are run on the data generated from the SP model, we find that the coefficient on the inverse import penetration ratio is negative, and if we add some randomness to the quota level, decreases as the quantile increases as occurs in our quantile regressions on real data as well. Thus, the evidence suggests that the SP model is, perhaps, more consistent with the data than the PFS model.

The remainder of the paper is organized as follows. In Section 2, we briefly review the PFS model and past empirical studies. Section 3 details our approach to testing the PFS model and then discusses the estimation results. In Section 4, we first describe how artificial data is generated from the SP model and then provide the estimation results. In Section 5, we briefly discuss the robustness of our results to heteroskedastic errors. Section 6 concludes.

2 The PFS Model and Its Estimation in the Literature

2.1 The PFS Model

The exposition in this section relies heavily on Grossman and Helpman (1994). There is a continuum of individuals, each of infinitesimal size. Each individual has preferences that are linear in the consumption of the numeraire good and are additively separable across all goods. As a result, there are no income effects and no cross price effects in demand which comes from equating marginal utility to own price. On the production side, there is perfect competition in a specific factor setting: each good is produced by a factor specific to the industry, k_i in industry i, and a mobile factor, labor, L. Thus, each specific factor is the residual claimant in its industry. Some industries are organized, and being organized or not is exogenous to the model. Tariff revenue is redistributed to all agents in a lump sum manner. Owners of the specific factors in organized industries can make contributions to the government to try and influence policy if it is worth their while.

Government cares about both social welfare and contributions made to it and puts a relative weight of α on social welfare. The timing of the game is as follows: first, lobbies simultaneously bid contribution functions that specify the contributions made contingent on the trade policy adopted (which determines domestic prices). The government then chooses what to do to maximize its own objective function. In this way, the government is the common agent all principals (organized lobbies) are trying to influence. Such games are known to have a continuum of equilibria.[1] By restricting agents to bids that are "regret free" equilibrium bids have the same curvature as welfare, and a unique equilibrium

[1] Given the bids of all other lobbies, each lobby wants a particular outcome to occur, namely, the one where it obtains the greatest benefit less cost. This can be attained by offering the minimal contribution needed for that outcome to be chosen by the government. However, what is offered for other outcomes (which is part of the bid function) is not fully pinned down as given other bids, it is irrelevant. However, bids at other outcomes affect the optimal choices of other lobbies and as their behavior affects yours, multiplicity arises naturally. Uniqueness is obtained by pinning down the bids at all outcomes to yield the same payoff as at the desired one, i.e., the bids are "regret free".

can be obtained.[2] The equilibrium outcome, thus, is as if the government was maximizing weighted social welfare ($W(p)$ where p is the domestic price and equals the tariff vector plus the world price vector, p^*) with a greater weight on the welfare of organized industries. Thus, equilibrium tariffs can be found by maximizing

$$G(p) = \alpha W(p) + \sum_{j \epsilon J_0} W_j(p),$$

where J_0 is the set of politically organized industries and the welfare of agents in industry j is

$$W_j(p) = \pi_j(p_j) + l_j + \frac{N_j}{N}\left[T(p) + S(p)\right],$$

where $\pi_j(p_j)$ is producer surplus in industry j, l_j is labor employed in industry j, wage is unity, $\frac{N_j}{N}$ is the share of workers employed in the jth industry, while $T(p) + S(P)$ is the sum of tariff revenue and consumer surplus in the economy.

Differentiating $W_i(p)$ with respect to p_j gives[3]

$$x_j(p_j)\delta_{ij} + \alpha_i\left[-x_j(p_j) + (p_j - p_j^*)m_j'(p_j)\right]$$

where so $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, α_i is the share of labor employed in industry i, $m_j'(p_j)$ is the derivative of the demand for imports, and $x_j(p_j) = \pi_j'(p_j)$ denotes supply of industry j. Differentiating $W(p)$ with respect to p_j gives

$$(p_j - p_j^*)m_j'(p_j).$$

[2] For a detailed discussion of this concept, see Bernheim and Whinston (1986).

[3] This follows from the derivative of consumer surplus from good j with respect to p_j being equal to $-d_j(p_j)$, where $d_j(p_j)$ is the demand for good j.

Hence, maximizing $G(p)$ with respect to p_j gives

$$\alpha\left[(p_j - p_j^*)m_j'(p_j)\right] + \sum_{i\in J_0}\left[x_j(p_j)\delta_{ij} + \alpha_i\left[-x_j(p_j) + (p_j - p_j^*)m_j'(p_j)\right]\right] = 0.$$

Now $\sum_{i\in J_0}\alpha_i = \alpha_L$, the employment share of organized industries and $\sum_{i\in J_0}\delta_{ij} = I_j$ is unity if j is organized and zero otherwise. Therefore, this equation can be reduced to

$$x_j(p_j)(I_j - \alpha_L) + (p_j - p_j^*)m_j'(p_j)(\alpha + \alpha_L) = 0.$$

If we further use the fact that $(p_j - p_j^*) = (t_j)\,p_j^*$, it can be also expressed as

$$\frac{t_j}{1+t_j} = \left(\frac{I_j - \alpha_L}{\alpha + \alpha_L}\right)\left(\frac{z_j}{e_j}\right)$$

where $z_j = \frac{x_j(p_j)}{m_j(p_j)}$ and $e_j = -m_j'(p_j)\frac{p_j}{m_j(p_j)}$. This is the basis of the key estimating equation. Note that protection is predicted to be positively related to $\frac{z_j}{e_j}$ if the industry is organized, but negatively related to it if the industry is not organized, and that the sum of the coefficients is predicted to be positive.

2.2 A Problem in Estimation — the Classification of Industries

For the key equation to be estimable, an error term is added in a linear fashion:

$$\frac{t_j}{1+t_j} = \gamma\frac{z_j}{e_j} + \delta I_j\frac{z_j}{e_j} + \varepsilon_j. \tag{1}$$

The error term is interpreted as the composite of variables potentially affecting protection that may have been left out and the measurement error of the dependent variable. To deal with the fact that a significant fraction of industries have zero protection in the data, the PFS equation can be modified as follows:

$$\frac{t_j}{1+t_j} = Max\left\{\gamma\frac{z_j}{e_j} + \delta I_j\frac{z_j}{e_j} + \epsilon_j, 0\right\}. \qquad (2)$$

The PFS model provides the following predictions on the coefficients on $\frac{z_j}{e_j}$ and $I_j\frac{z_j}{e_j}$: $\gamma < 0$, $\delta > 0$ and $\gamma + \delta > 0$.[4] To test these predictions, equations (1) and (2)(hereafter called the PFS equations) have been estimated in a number of previous studies (e.g., Goldberg and Maggi (1999), Gawande and Bandyopadhy (2000), McCalman (2001)).

Although data on the measure of trade protection, the import penetration ratio, and the import-demand elasticities are often available, it is harder to define whether an industry is politically organized or not. To deal with this problem, Goldberg and Maggi (1999) use data on campaign contributions at the three-digit SIC industry level. An industry is categorized to be politically organized if the campaign contribution exceeds a specified threshold level. Gawande and Bandyopadhyay (2000) used a different procedure for classification. They run a regression where the dependent variable is the log of the corporate Political Action Committee (PAC) spending per contributing firm relative to value added and the regressors include the interaction of the import penetration from

[4] Goldberg and Maggi (2000) and others note that $\gamma < 0$, $\delta > 0$ and $\gamma + \delta > 0$ are only necessary conditions for the validity of the PFS specification. However most empirical research in the political economy of trade claim that the right sign of the coefficients of the PFS equation gives strong empirical support of the PFS paradigm. Recently, Imai et al. (2006) criticize them by pointing out that even when estimating the PFS equation on an artificial data simulated from a simple quota model that has no PFS element, one will obtain the parameter estimates consistent with the PFS hypothesis.

five countries into the sub industry and the two-digit SIC dummies. The idea is that in organized industries, an increase in contributions would likely occur when import penetration increased. Thus, industries are classified as politically organized if any of the coefficients on its five interaction terms are found to be positive.

Note that both these two procedures are questionable. The procedure used in Goldberg and Maggi (1999) implicitly assumes that all the contributions are directed towards influencing trade policies. Moreover, any non zero cutoff level of contributions as indicating organization seems relatively arbitrary. In addition, the procedure does not control for other variables that potentially influence political clout such as industry size and electoral districts where the industry is concentrated. The procedure used by Gawande and Bandyopadhyay (2000) might have a potential identification problem, since a function of the import penetration is used to classify industries and the import penetration is concurrently used as a regressor in the PFS equation.

Recently, Cadot, Grether, and Olarreaga (2006) propose a different approach that that does not require any data on political organization. Instead of deriving the political organization dummy in an ad-hoc manner, they proposed to recover it as a by-product of the estimation process. Specifically, they initially set the political organization dummy to zero for every industry. Then, they estimate the PFS equation and obtain the error terms. If the error term of an industry is greater than some threshold value, its political organization dummy is set to be one. The idea is that such industries do not fit the not organized category.

Using the generated political organization dummies, they again estimate the PFS equation and obtain the error terms. They repeat the procedures of generating the political organization dummies and estimating the PFS equation until the parameters converge. Their method is attractive since information that has been used to classify industries (e.g., contributions) is unavailable in many countries. However, their approach by construction creates a positive correlation between the error term and the generated political organization dummies, which cannot be overcome by any conventional instruments.

3 A Proposed Approach

3.1 Quantile Regression

In this section, we detail our approach to testing the PFS model. The advantage of our approach is similar to that of Cadot, Grether and Olarreaga (2006) (C-G-O for short) in the sense that the approach allows us to test the PFS model without using data on political contributions, directly as in G-M or indirectly as in G-B or iteratively as in C-C-O, to construct an organization dummy. However, our approach substantially differs from theirs: instead of classifying industries as organized or not in some manner, our estimation procedure relies heavily on the relationship between observables implied by the PFS model.

Equation (2) and the restrictions on the coefficients have at least two implications. First, as has been discussed in the literature, $\frac{z_j}{e_j}$ has a negative effect on the level of protection for politically unorganized industries while it has a positive effect for politically organized industries. Second, given $\frac{z_j}{e_j}$, politically organized industries have higher protection. These implications lead to the following claim: given $\frac{z_j}{e_j}$, high protection industries are more likely to be politically organized and thus effect of an the increase in $\frac{z_j}{e_j}$ on protection tends to be that of politically organized industries. The relevant theorem, and proof, can be found in Appendix 1. The proposition essentially states that at the quantile close to $\tau = 1$, the coefficient on the inverse import penetration ratio should be close to $\gamma + \delta$. To empirically examine this, we use part of the data used in Gawande and Bandyopadhyay (2000)[5]. The data consist of 242 four-digit SIC industries in the U.S. See Gawande and Bandyopadhyay (2000) for a description of the variables.

Let $T_j = \frac{t_j}{1+t_j}$ and $Z_j = \frac{z_j}{e_j}$. Using quantile regression (Koenker and Bassett, 1978), we estimate the following equation:

$$Q_T(\tau|Z_j) = \alpha(\tau) + \beta(\tau) Z_j/10000. \tag{3}$$

where it is expected that $\beta(\tau)$ converges to $(\gamma + \delta) > 0$ as the quantile, τ, approaches its highest level of unity from below, if the PFS model is correct.[6] As the table indicates, $\beta(\tau)$ starts from zero at the quantile $\tau = 0.4$ (since there are a large number group of unprotected industries for whom the coverage ratio is zero) and **decreases** as τ goes from 0.4 to 0.9. Note that this is the **opposite** of what the PFS model predicts, casting doubt on the validity of the PFS model.

[5] We are grateful to Kishore Gawande for kindly providing us with the data.

[6] The estimation results are presented in Table 1. The estimation is done by using a MATLAB code written by Christian Hansen (available at http://faculty.chicagogsb.edu/christian.hansen/research).

It is fair to say that our argument here relies on the point estimates. The estimated standard errors are rather large and none of the $\beta(\tau)$'s is significantly different from zero. If the reader is not satisfied with our argument based on the point estimates, the evidence should be interpreted as suggesting that there is no strong evidence in favor of the PFS model (This applies to also to evidence from the instrumental variable quantile regression presented in the next subsection). Two aspects of the results are worth mentioning. First, α and β are estimated to start from zero at the 0.4 quantile, suggesting the corner solution ($T_j = 0$) greatly affect the estimates at the lower quantiles. From this evidence, it is conjectured that the existence of corners also affects the estimates at the mean. Thus, findings based on the linear model in Gawande and Bandyopadhyay (2000), Bombardini (2005), and others are likely to be subject to bias due to such corners. To address this issue, several studies (e.g., Goldberg and Maggi (1999), Biesebroeck et. al., (2004)) estimate a system of equations: equation (2), explicitly allowing for such truncation of protection at zero, as well as an import penetration equation, and an equation for political organization. On the other hand, the assumption of normality of the error terms is usually made and this may affect the estimation results. In contrast, our estimation results are unlikely to be subject to the corner solution problem, since we focus mainly on the higher quantiles where the effect of corner solution is minimal. Second, our results are not driven by the parametric assumption on the error term; the quantile regression does not require them.

Table 1: Quantile Regression

τ	$\alpha(\tau)$	Std. Error	$\beta(\tau)$	Std. Error
0.4	0.000	0.001	0.000	0.011
0.5	0.002	0.006	-0.003	0.139
0.6	0.028	0.016	-0.046	0.366
0.7	0.077	0.017	-0.126	0.825
0.8	0.157	0.034	-0.258	1.872
0.9	0.308	0.041	-0.505	2.436

One might wish to control for various factors as well in the quantile regression above. Following Gawande and Bandyopadhyay (2000), we control for tariff of intermediate goods ($INTERMTAR$) and NTB coverage of intermediate goods ($INTERMNTB$). Table 2 presents these results. Standard errors are in parentheses. Our main findings do not change; $\beta(\tau)$ decreases (for the most part) from zero to a negative value with the increase in τ, contrary to what the PFS model would predict. α and β are found to be be zero at the 0.1 and 0.2 quantiles, suggesting the importance of corner solution.

Table 2: Quantile Regression with Control Variables

τ	$\alpha(\tau)$	$\beta(\tau)$	$INTERMTAR$	$INTERMNTB$
0.1	0.000 (0.003)	0.000 (0.013)	0.000 (0.037)	0.000 (0.003)
0.2	0.000 (0.013)	0.000 (0.082)	0.000 (0.239)	0.000 (0.010)
0.3	−0.026 (0.008)	−0.099 (0.095)	0.459 (0.144)	0.028 (0.015)
0.4	−0.029 (0.006)	−0.020 (0.098)	0.554 (0.119)	0.028 (0.021)
0.5	−0.026 (0.015)	−0.032 (0.263)	0.617 (0.308)	0.028 (0.090)
0.6	−0.053 (0.023)	−0.082 (0.443)	1.166 (0.693)	0.213 (0.183)
0.7	−0.044 (0.019)	−0.125 (0.646)	1.157 (0.676)	0.397 (0.199)
0.8	−0.046 (0.015)	−0.145 (1.024)	0.823 (0.456)	0.756 (0.180)
0.9	−0.001 (0.048)	−0.225 (1.547)	0.581 (0.479)	0.961 (0.166)

3.2 Instrumental Variable Quantile Regression

In the quantile regression, Z_j is assumed to be an exogenous variable. However, Z_j is likely to be endogenous as discussed in the literature and hence the parameter estimates of the quantile regression are likely to be inconsistent.[7] Next, we present some results obtained by using quantile instrumental variable estima-

[7] We are relatively unconcerned about endogeneity of the political organization dummy. for the followig reasons. First, Goldberg and Maggi (1999) did not find any evidence of the endogeneity of the political organization dummy. Notice also that the primary reason they allowed for the econometric endogeneity is because of the endogeneity of campaign contribution, from which they constructed the political organization dummy. In our case, since we do not measure political organization dummy, their source of endogeneity is not an issue for us.

tion techniques developed by Chernozhukov and Hansen (2006). However, prior to this we we show that the main prediction of the PFS model in terms of our quantile approach does not change even in the presence of endogeneity. In other words, proposition 2, is the analogue of proposition 1, when Z_j is endogenous. This is done in Appendix 1.

To test the prediction in the presence of possible endogeneity of Z, we estimate the following equation using the Quantile IV technique,

$$P\left(T_j \leq \alpha\left(\tau\right) + \beta\left(\tau\right) Z_j/10000|W_j\right) = \tau^8$$

where W_j are the instruments. Our choice of instruments is guided by Gawande and Bandyopadhyay (2000). They used 34 distinct instruments, their quadratic terms, and some of the two-term cross products. We use 17 of their instruments (called IV Set 1)[9]. The quantile IV estimation is done by using a MATLAB code written by Christian Hansen. Standard errors are computed based on 200

[8] An extension for future work would be to estimate the protection equation where we allow the probability of political organization to depend on some of the exogenous variables $V_j \in W_j$. Then, the Quantile IV equation becomes

$$P\left(T \leq F_{\epsilon}^{-1}\left(\tau'\left(V_j\right)\right) + \left(\gamma + \delta\right) Z_j|W_j\right) = \tau.$$

Therefore, the linearized quantile IV equation to be estimated becomes.

$$P\left(T_j \leq \alpha_1\left(\tau\right) + \alpha_2\left(\tau\right) V_j + \beta\left(\tau\right) Z_j/10000|W_j\right) = \tau$$

[9] These are the basic instruments used by Bombardini (2006). We are grateful to Bombardini for providing us the programs for her PFS estimation and some helpful comments and suggestions.

bootstrap replications. Table 3 reports the estimation results. The estimated slope coefficients are negative, except at the lower quantiles and at 0.8 quantile. However, PFS would require the coefficient to be positive for the higher quantiles, which clearly not observed.

Table 3: Quantile IV Regression (IV Set 1)

τ	$\alpha(\tau)$	$\beta(\tau)$	$INTERMTAR$	$INTERMNTB$
0.1	0.000 (0.003)	0.000 (0.407)	0.000 (0.053)	0.000 (0.004)
0.2	0.000 (0.012)	0.000 (0.402)	0.000 (0.208)	0.000 (0.012)
0.3	−0.025 (0.011)	−0.370 (0.357)	0.473 (0.204)	0.024 (0.014)
0.4	−0.028 (0.009)	−0.200 (0.621)	0.550 (0.171)	0.027 (0.025)
0.5	−0.031 (0.023)	−0.270 (1.395)	0.702 (0.433)	0.048 (0.094)
0.6	−0.053 (0.023)	−0.080 (2.153)	1.168 (0.619)	0.215 (0.161)
0.7	−0.044 (0.015)	−0.130 (2.403)	1.159 (0.615)	0.397 (0.185)
0.8	−0.046 (0.016)	0.020 (2.722)	0.823 (0.661)	0.755 (0.216)
0.9	−0.002 (0.044)	−0.230 (3.572)	0.584 (0.569)	0.961 (0.173)

Table 4 reports the results where we use the 17 instruments, their quadratic terms, and the quadratic terms of $INTERMTAR$ and $INTERMNTB$ (called IV Set 2). Table 5 presents the results where, as in Gawande and Bandyopadhyay (2000), we also include some interaction terms in the set of instruments (called IV Set 3). Even after correcting for endogeneity, the main findings in the quantile regression survive: the slope coefficients remain negative by and large except at the lowest quantiles where there are corners. Note that the slope coefficient at the highest quantile is not positive as required by PFS. Hence, these results seem inconsistent with the PFS framework.

Table 4: Quantile IV Regression (IV Set 2)

τ	$\alpha(\tau)$	$\beta(\tau)$	$INTERMTAR$	$INTERMNTB$
0.1	0.000 (0.002)	0.000 (0.270)	0.000 (0.037)	0.000 (0.002)
0.2	0.000 (0.011)	0.000 (0.369)	0.000 (0.187)	0.000 (0.011)
0.3	−0.026 (0.011)	−0.370 (0.287)	0.473 (0.206)	0.024 (0.014)
0.4	−0.029 (0.009)	−0.200 (0.421)	0.550 (0.161)	0.027 (0.021)
0.5	−0.026 (0.023)	−0.270 (1.091)	0.702 (0.416)	0.048 (0.088)
0.6	−0.053 (0.024)	−0.080 (1.184)	1.168 (0.636)	0.215 (0.159)
0.7	−0.044 (0.014)	−0.130 (1.611)	1.159 (0.622)	0.397 (0.182)
0.8	−0.046 (0.014)	0.020 (1.826)	0.823 (0.645)	0.755 (0.211)
0.9	−0.001 (0.042)	−0.230 (3.383)	0.584 (0.617)	0.961 (0.179)

Table 5: Quantile IV Regression (IV Set 3)

τ	$\alpha(\tau)$	$\beta(\tau)$	$INTERMTAR$	$INTERMNTB$
0.1	0.000 (0.003)	0.000 (0.604)	0.000 (0.054)	0.000 (0.005)
0.2	0.000 (0.012)	0.000 (0.384)	0.000 (0.207)	0.000 (0.012)
0.3	−0.027 (0.011)	−0.010 (0.383)	0.513 (0.200)	0.026 (0.014)
0.4	−0.028 (0.009)	−0.020 (0.636)	0.540 (0.167)	0.030 (0.024)
0.5	−0.031 (0.022)	−0.030 (1.103)	0.701 (0.422)	0.048 (0.092)
0.6	−0.052 (0.024)	−0.080 (1.557)	1.163 (0.622)	0.213 (0.158)
0.7	−0.044 (0.015)	−0.130 (1.820)	1.157 (0.616)	0.397 (0.185)
0.8	−0.046 (0.015)	−0.150 (2.290)	0.824 (0.651)	0.756 (0.217)
0.9	−0.001 (0.044)	−0.260 (2.989)	0.585 (0.556)	0.960 (0.171)

4 "Surge Protection" Model

Imai et. al. (2006) proposed a simple institutionally based "Surge Protection" model as discussed above. In this section, we examine the validity of the SP model. In particular, we conduct the following exercise. We first simulate artificial data from a calibrated version of the SP model. We then estimate the quantile regression and the IV quantile regression on this data. We ask whether these estimates resemble those we obtained above. If the SP model is valid, then these patterns might be expected to be similar.

As in Imai et. al. (2006), we choose the parameters to match the aggregate statistics of the U.S. data. Using the artificial data simulated from the SP model as described in Appendix 2, we estimate the PFS equation via quantile regression. Table 6, Case 1 presents the results. We can see that at lower quantiles, the estimated coefficients on the inverse import penetration are close to zero, because industries at lower quantiles have zero protection. The estimated slope coefficients at 0.5-0.9 quantiles all become negative and statistically significant at 5%, which is consistent with the pattern we see in the data. Unlike the results of the actual data, however, there is a pattern to the coefficients which increases with the quantile.

Now, we slightly extend the SP model by allowing the quota to be stochastically determined. Specifically, we add some randomness to the quota, i.e. $\hat{Q}_{ij} = \widehat{Q} + \varsigma, \varsigma\tilde{\ } N(0,1)$.The results from doing so are presented in Case 2 of Table 6. The coefficients on the inverse import penetration ratio are found to be zero at lower quantiles, and thereafter decrease with quantile, which is consistent with the results of the actual data. The quantile IV results are presented in Table 7, where exogenous demand and supply shocks are used as instruments. Findings are similar with those of the quantile regression. The results overall suggest that the qualitative feature of the SP model is more consistent with the actual data than is the PFS model. The intuition behind the negative coefficient estimate of the surge protection model is simple. A surge in imports, which increases the import penetration ration, tends to result in the quota being binding, which corresponds to an increase in the NTB coverage ratio. Hence, the negative relationship between the inverse import penetration ratio and the NTB coverage ratio. The magnitude of the coefficients, however for the SP model are much larger than those of the data. Our conjecture is that this discrepancy is due to the fact that in our simulation, the variance of the demand shocks, which represents the "surge" is unrealistically high, because they are derived by matching the cross sectional variation of the model to the data (see Imai et. al. (2006) for details). In our future work, we plan to construct a panel data to estimate the parameters of the SP model. Then, the variance of the surge can be estimated to a realistic value, which we believe results in making the magnitude of the quantile slope coefficients closer to those of the data.

Table 6: Quantile Regression Estimates of Surge Protection Model

	Case 1		Case 2	
τ	$\alpha(\tau)$	$\beta(\tau)$	$\alpha(\tau)$	$\beta(\tau)$
0.1	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
0.2	0.000 (0.000)	0.000 (0.091)	0.000 (0.000)	0.000 (0.030)
0.3	0.000 (0.006)	0.000 (2.658)	0.000 (0.002)	−0.045 (0.696)
0.4	0.020 (0.066)	−6.672 (5.798)	0.006 (0.006)	−1.973 (1.985)
0.5	0.042 (0.007)	−11.333 (2.641)	0.020 (0.006)	−5.125 (1.759)
0.6	0.044 (0.001)	−9.615 (1.618)	0.033 (0.006)	−6.686 (1.721)
0.7	0.046 (0.001)	−7.841 (1.479)	0.049 (0.006)	−7.854 (2.022)
0.8	0.046 (0.000)	−6.076 (1.388)	0.072 (0.008)	−8.666 (2.469)
0.9	0.047 (0.000)	−4.276 (1.186)	0.111 (0.013)	−9.214 (3.103)

Table 7: Quantile IV Regression Estimates of Surge Protection Model

	Case 1		Case 2	
τ	$\alpha(\tau)$	$\beta(\tau)$	$\alpha(\tau)$	$\beta(\tau)$
0.1	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
0.2	0.004 (0.008)	−2.000 (4.755)	0.000 (0.031)	−0.093 (0.009)
0.3	0.031 (0.015)	−15.330 (7.286)	0.000 (0.002)	−2.931 (0.292)
0.4	0.042 (0.004)	−17.000 (2.153)	0.006 (0.670)	−7.210 (0.718)
0.5	0.043 (0.001)	−14.240 (1.613)	0.018 (1.921)	−8.934 (0.891)
0.6	0.044 (0.000)	−11.540 (1.349)	0.038 (1.836)	−9.866 (0.985)
0.7	0.044 (0.000)	−9.450 (1.141)	0.053 (2.162)	−10.832 (1.082)
0.8	0.045 (0.000)	−7.430 (1.138)	0.073 (2.618)	−11.078 (1.109)
0.9	0.045 (0.003)	−5.390 (1.074)	0.110 (3.342)	−10.531 (1.078)

5 Alternative Explanations

We would like to stress that we are not arguing that the PFS model has been rejected. Rather, we would like to think of our approach as looking for some further relatively simple predictions of the PFS model. These, do not seem to be in the data. We ask next, if this could be for econometric reasons. The obvious culprit is heteroskedasticity. If the error term has higher variance when the industry is politically unorganized, i.e.,

$$\epsilon_j = w_j + (1 - I_j)\,\varsigma_j \tag{4}$$

then politically unorganized industries would have error terms with much higher variance. As a result, they would be the ones that dominate in high quantiles as well as in low quantiles, whereas the politically organized industries would be found mostly around the median. Hence, at high quantiles, the negative quantile regression coefficients correspond to γ, which is negative, and not $(\gamma + \delta)$. This might explain the presence of negative slope coefficients in the higher quantiles. This cannot be completely ruled out: however, it is worth pointing out that G-M do test for heteroskedasticity and find that they can reject it. [10]

6 Conclusion

In this paper, we propose an estimation strategy for the PFS model where data on political organization is not required. It is based on the idea that industries with higher protection measures are more likely to be politically organized. Then, the PFS model predicts that at high quantiles the coefficient on the inverse import penetration ratio should be larger than the other industries. We

[10] If equation (4) is indeed the error structure, then one should estimate the modified PFS equation:

$$\frac{t_j}{1+t_j} = \gamma \frac{z_j}{e_j} + \delta \frac{z_j}{e_j} I_j + \varsigma_j (1 - I_j) + w_j.$$

Importantly, the modified equation has an additional regressor $1-I_j$ with a random coefficient ς_j. Not only does this suggest that the PFS equation estimated in the literature might have been misspecified, but also the original lobbying model needs to be substantially modified so that its PFS equation results to the modified equation above. Then, it would be unclear whether findings in past studies (i.e., $\gamma < 0$, $\delta > 0$, and $\gamma + \delta > 0$) can be interpreted as being in support of the PFS paradigm.

use the quantile regression and IV quantile regression techniques to see whether this is true in the data. The results are not supportive of the PFS model. We run the same regression on simulated data arising from the SP model proposed in Imai et. al. (2006). The coefficients from this regression are more in line with those found in the actual data suggesting that this may be what is going on. The findings overall seem to suggest that the SP model is qualitatively consistent with the data, while the PFS is not so.

7 Appendix 1

Let $T_j = \frac{t_j}{1+t_j}$ and $Z_j = \frac{z_j}{e_j}$.

Proposition 1 ***(Quantile Regression)*** *Assume that (1) Z_j is bounded below by a positive number, i.e. there exists $\underline{Z} > 0$ such that $Z_j \geq \underline{Z}$, (2) ϵ has a smooth density function which has support that is bounded from above and below, (3) ϵ is independent of both Z_j and and I_j, and (4) $\delta > 0$. Then, for τ sufficiently close to 1, τ quantile conditional on Z_j can be expressed as*

$$Q_T\left(\tau|Z_j\right) = F_\epsilon^{-1}\left(\tau'\right) + (\gamma+\delta)Z_j \tag{5}$$

where

$$\tau' = \frac{\tau - P\left(I_j = 0\right)}{P\left(I_j = 1\right)}. \tag{6}$$

Proof. For any $0 < \tau < 1$, for any $\overline{T} > 0$,

$$P\left(T_j \leq \overline{T}|Z_j\right) = P\left(\epsilon_j \leq \overline{T} - \gamma Z_j\right)P\left(I_j = 0\right) + P\left(\epsilon_j \leq \overline{T} - (\gamma+\delta)Z_j\right)P\left(I_j = 1\right). \tag{7}$$

Let

$$\overline{T} = F_{\epsilon}^{-1}\left(\tau'\right) + (\gamma+\delta)Z_j \tag{8}$$

where

$$\tau' = \frac{\tau - P\left(I_j = 0\right)}{P\left(I_j = 1\right)}, \text{ or } \tau = P\left(I_j = 0\right) + \tau' P\left(I_j = 1\right). \tag{9}$$

From equation (9), we can see that for $\tau \nearrow 1$, $\tau' \nearrow 1$ as well. Hence, for τ sufficiently close to 1, we have τ' close enough to 1 such that

$$F_{\epsilon}^{-1}\left(\tau'\right) + \delta Z_j \geq F_{\epsilon}^{-1}\left(\tau'\right) + \delta \underline{Z} > F_{\epsilon}^{-1}\left(1\right).$$

Hence,

$$\overline{T} = F_{\epsilon}^{-1}\left(\tau'\right) + (\gamma+\delta)Z_j > F_{\epsilon}^{-1}\left(1\right) + \gamma Z_j$$

and

$$P\left(\epsilon_j \leq \overline{T} - \gamma Z_j\right) \geq P\left(\epsilon_j \leq F_{\epsilon}^{-1}\left(1\right)\right) = 1$$

which results in

$$P\left(\epsilon_j \leq \overline{T} - \gamma Z_j\right) = 1. \tag{10}$$

Substituting equations (8), (9), and (10) into (7), we obtain

$$\begin{aligned} P\left(T_j \leq \overline{T}|Z_j\right) &= P(I_j = 0) + P\left(\epsilon_j \leq F_{\epsilon}^{-1}\left(\tau'\right)\right) P\left(I_j = 1\right) \\ &= P(I_j = 0) + \tau - P\left(I_j = 0\right) = \tau. \end{aligned}$$

Therefore, for τ sufficiently close to 1,

$$Q_T\left(\tau|Z_j\right) = \overline{T} = F_{\epsilon}^{-1}\left(\tau'\right) + (\gamma+\delta)Z_j.$$

■

We make two remarks on the assumptions. First, we assume that ϵ has bounded support (assumption 2). This assumption is reasonable since the protection measure is usually derived from the NTB coverage ratio (e.g., Goldberg and Maggi (1999), Gawande and Bandyopadhyay (2000)) and therefore it is clearly bounded above and below. Second, we assume that ϵ is independent of both Z_j and and I_j (assumption 3). This is rather a strong assumption and will be relaxed later when quantile IV's are discussed.

When we introduce IV's we show that $\beta(\tau) \rightarrow (\gamma+\delta) > 0$ as $\tau \nearrow 1$.

Assume the model is as follows:

$$\begin{aligned} T_j^* &= \gamma Z_j + \epsilon_j \text{ if } I_j = 0 \\ T_j^* &= (\gamma+\delta) Z_j + \epsilon_j \text{ if } I_j = 1 \end{aligned}$$

where

$$Z_j = g(W_j, v_j).$$

W_j is an instrument vector and v_j is a random variable independent of W_j. Let us define u_j as follows:

$$\epsilon_j = E[\epsilon_j|v_j] + u_j, \quad u_j \equiv \epsilon_j - E[\epsilon_j|v_j],$$

where u_j is assumed to be i.i.d. distributed. Furthermore,

$$T_j = \max\left\{T_j^*, 0\right\}.$$

Then, for $I_j = 0$ the model satisfies the assumptions A1-A5 of Chernozhukov and Hansen (2006). Similarly for $I_j = 1$. Therefore, from Theorem 1 of Chernozhukov and Hansen (2006), it follows that

$$P\left(T \leq F_{\epsilon}^{-1}(\tau) + \gamma Z_j | W_j\right) = \tau \text{ for } I_j = 0,$$

and

$$P\left(T \leq F_{\epsilon}^{-1}(\tau) + (\gamma + \delta) Z_j | W_j\right) = \tau \text{ for } I_j = 1.$$

Proposition 2 ***(Quantile IV)*** *Assume that Z_j is bounded below by a positive number, i.e. there exists $\underline{Z} > 0$ such that $Z_j \geq \underline{Z}$. Then, for τ sufficiently close to 1,*

$$P\left(T \leq F_{\epsilon}^{-1}(\tau') + (\gamma + \delta) Z_j | W_j\right) = \tau,$$

where $\tau' = \frac{\tau - P(I_j = 0)}{P(I_j = 1)}$.

Proof.

$$\tau' = \frac{\tau - P(I_j = 0)}{P(I_j = 1)}, \text{ or } \tau = P(I_j = 0) + \tau' P(I_j = 1).$$

Then,

$$\begin{aligned}
& P\left(T_j \leq F_{\epsilon}^{-1}(\tau') + (\gamma + \delta) Z_j | W_j\right) \\
= \; & P\left(\epsilon_j + \gamma Z_j \leq F_{\epsilon}^{-1}(\tau') + (\gamma + \delta) Z_j | W_j\right) P(I_j = 0) \\
& + P\left(\epsilon_j + (\gamma + \delta) Z_j \leq F_{\epsilon}^{-1}(\tau') + (\gamma + \delta) Z_j | W_j\right) P(I_j = 1) \\
= \; & P\left(\epsilon_j \leq F_{\epsilon}^{-1}(\tau') + \delta Z_j | W_j\right) P(I_j = 0) + P\left(\epsilon_j \leq F_{\epsilon}^{-1}(\tau') | W_j\right) P(I_j = 1) \\
= \; & P\left(\epsilon_j \leq F_{\epsilon}^{-1}(\tau') + \delta Z_j | W_j\right) P(I_j = 0) + \tau' P(I_j = 1)
\end{aligned}$$

From the definition of τ', for $\tau \nearrow 1$, $\tau' \nearrow 1$ as well. Hence, for τ sufficiently close to 1, we have τ' close enough to 1 such that

$$F_{\epsilon}^{-1}(\tau') + \delta \underline{Z} > F_{\epsilon}^{-1}(1).$$

Hence,

$$P\left(\epsilon_j \leq F_\epsilon^{-1}\left(\tau'\right) + \delta Z_j | W_j\right) = 1.$$

Therefore,

$$P\left(T_j \leq F_\epsilon^{-1}\left(\tau'\right) + \left(\gamma+\delta\right) Z_j | W_j\right) = P(I_j = 0) + \tau' P\left(I_j = 1\right) = \tau.$$

It follows that for τ *sufficiently close to* 1*,*

$$P\left(T \leq F_\epsilon^{-1}\left(\tau'\right) + \left(\gamma+\delta\right) Z_j | W_j\right) = \tau.$$

■

8 Appendix 2

In what follows, we detail the SP model. Our procedure follows Imai et. al. (2006) and is explained in Appendix 2. First, consider the domestic and foreign goods equilibrium without quota. For each industry i and subindustry j, there are two types of goods: domestic and foreign goods. To make matters simple, we assume that each good's demand depends only on its own price and random shocks and that home is the only source of demand. Let x_{ij}^H be the equilibrium quantity of home goods in industry i subindustry j, and let p_{ij}^H be its equilibrium price.

The equilibrium is described by the demand and supply equations. The demand for industry i subindustry j of the home good depends on a constant, the price of the good, and random terms as follows:

$$\ln x_{ij}^{Hd} = ahd_1 + ahd_2 \ln p_{ij}^{H} + xhd_i + uhd_{ij}.$$

Similarly, the supply of the same good follows the supply equation:

$$\ln x_{ij}^{Hs} = ahs_1 + ahs_2 \ln p_{ij}^{H} + xhs_i + uhs_{ij}.$$

The random terms xhd_i and xhs_i are industry specific demand and supply shocks, and hence, common across all subindustries, while uhd_{ij} and uhs_{ij} are subindustry specific demand and supply shocks and are idiosyncratic to each subindustry. All shocks are assumed to be i.i.d. with mean zero normal distributions with standard errors σ_{xhd}, σ_{xhs}, σ_{uhd}, and σ_{uhs}, respectively. Equilibrium satisfies

$$x_{ij}^{Hd} = x_{ij}^{Hs} = x_{ij}^{H}.$$

Similarly, let import demand be given by

$$\ln x_{ij}^{Md} = amd_1 + amd_2 \ln p_{ij}^{M} + xmd_i + umd_{ij}$$

and supply by

$$\ln x_{ij}^{Ms} = ams_1 + ams_2 \ln p_{ij}^{M} + xms_i + ums_{ij}.$$

As before, the random terms xmd_i, xms_i, umd_{ij}, and ums_{ij} are industry and subindustry specific demand and supply shocks. They are distributed i.i.d. normally with means zero and standard errors σ_{xmd}, σ_{xms}, σ_{umd}, and σ_{ums} respectively. Equilibrium satisfies

$$x_{ij}^{Md} = x_{ij}^{Ms} = x_{ij}^{Me}.$$

We assume that there are $n_t = 250$ industries and each industry has $n_j = 6$ subindustries. Each subindustry ij is politically organized with probability Po_i.

We simulate the output and prices of each subindustry by first drawing n_t industry demand and supply shocks xmd_i and xms_i for $i = 1, ..., n_t$ and for each industry i, drawing n_s subindustry demand and supply shocks umd_{ij} and ums_{ij} for $j = 1, ..., n_s$. Then, given these shocks and parameters of the demand and supply equations, we compute the equilibrium price and quantities for each subindustry ij.

We then simulate the political organization for each subindustry and introduce a uniform quota level $\hat{Q}$ for all politically organized subindustries. That is, the quota becomes binding in subindustry ij if the equilibrium output for the foreign goods exceeds $\hat{Q}$. Let d_{ij}^q be the indicator for a binding quota. That is, if x_{ij}^{Me} for subindustry ij exceeds $\hat{Q}$, then actual imports, x_{ij}^M, equal $\hat{Q}$ and $d_{ij}^q = 1$. Otherwise, $x_{ij}^M = x_{ij}^{Me}$ and $d_{ij}^q = 0$. One way of interpreting this is that there is a trigger level of imports, $\hat{Q}$, above which the relevant agency would restrict imports if asked, but only politically organized agencies ask for such protection. In other words, that there are provisions for preventing a surge of imports, but only organized subindustries can actually make use of these provisions perhaps because they can overcome the usual free rider problems.

Next we aggregate subindustry output to the industry level. Total industry equilibrium output is computed as

$$X_i^H = \sum_{j=1}^{n_j} x_{ij}^H$$

for home goods and

$$X_i^M = \sum_{j=1}^{n_j} x_{ij}^M$$

for foreign goods.

We then generate the variables that we use in the estimation as follows. First, we compute the coverage ratio C_i of industry i to be:

$$C_i = \frac{\sum_{j=1}^{n_j} x_{ij}^M d_{ij}^q}{X_i^M}.$$

That is, coverage ratio is the fraction of industry output i where quota is binding. Furthermore, the inverse import penetration ratio, z_i, for industry i is the ratio of domestic production to imports or

$$\frac{X_i^H + X_i^M}{X_i^M} = 1 + z_i.$$

We also derive the political organization dummy of industry i, I_i, as:

$$\begin{aligned} I_i &= 1 \; if \sum_{j=1}^{n_j} I_{ij} > \frac{n_j}{2} \\ &= 0 \;\; otherwise. \end{aligned}$$

9 References

Bombardini, M. (2005) "Firm Heterogeneity and Lobby Participation." *Mimeo. University of British Columbia.*

Cadot Olivier, Jean-Marie Grether and Marcelo Olarreaga (2006) "Endogenous Tariffs in a Common-Agency Model: A New Empirical Approach Applied to India." *mimeo. World Bank*

Chernozhukov, V. and Hansen, C. (2006) "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics, Vol.* 132*, pp.* 491 − 525.

Ederington Josh and Jenny Minier (2005) "Reconsidering the Empirical Evidence on the Grossman-Helpman Model of Endogenous Protection" *University of Kentucky Working Paper.*

Gawande Kishore and Usree Bandyopadhyay (2000) "Is Protection for Sale? Evidence on the Grossman-Helpman Theory of Endogenous Protection." *Review of Economics and Statistics*, *Vol.* 82, *No.* 1, *pp.* 139 − 152.

Goldberg Pinelopi K. and Giovanni Maggi (1999) "Protection for Sale: An Empirical Investigation." *American Economic Review*, *Vol.* 89, *No.* 5, *pp.* 1135 − 1155.

Grossman Gene and Elhanan Helpman (1994) "Protection for Sale." *American Economic Review*, *Vol.* 84, *No.* 4, *pp.* 833 − 50.

Imai Susumu, Hajime Katayama, and Kala Krishna (2006) "Protection for Sale or Surge Protection?" *NBER Working Paper No.* 12258.

Matschke Xenia (2006) "Costly Revenue-Raising and the Case for Favoring Import-Competing Industries" *University of Connecticut Working Paper.*

Shiells, C.R., Stern, R.M. and Deardorff, A.V. (1985) "Estimates of the Elasticities of Substitution between Imports and Home Goods for the United States," *Weltwirtschaftliches Archiv,* 122*, pp.* 497 − 519.