Economics Bulletin

Volume 38, Issue 3

Parallel Imports, Product Quality and Endogenous Trading Bloc Formation

Sunandan Ghosh
Centre For Development Studies

Abstract

This paper analyzes the role of parallel imports (PIs) in determining both optimal product quality and optimal trade policy. We consider a three country world economy, where poor (P) and middle-income (M) countries import a vertically differentiated good produced and exported by a patent holder monopolist in the rich country (R), when PIs are allowed. In presence of both inter and intra-country taste diversity, PIs lower the level of innovation irrespective of tariff regimes vis-à-vis when PIs are not allowed. The optimal tariff levels imposed by the importing countries are lower under PIs irrespective of the tariff regimes. Formation of a Free Trade Area (FTA) by R is feasible with side-payments to its FTA partner given intra-country taste diversity. A customs union (CU) among the importing countries turns out to be globally Pareto superior given both inter and intra-country taste diversity. The level of innovation is highest under such CU as compared to all other tariff regimes.

The author is grateful to the suggestions and inputs from Rajat Acharyya, Rajit Biswas and two anonymous referees on an earlier draft. The usual disclaimer, however, applies.
Citation: Sunandan Ghosh, (2018) "Parallel Imports, Product Quality and Endogenous Trading Bloc Formation", *Economics Bulletin*, Volume 38, Issue 3, pages 1415-1427
Contact: Sunandan Ghosh - sunandan@cds.edu.
Submitted: October 18, 2017. **Published:** July 18, 2018.

1. Introduction

Parallel imports (PIs) are legitimately produced goods imported legally into a country without the authorization of the original manufacturer (or producer) holding a trademark, copyright, or patent. The legal doctrine governing the permissibility of PIs is exhaustion, or the point of distribution at which rights to control further distribution are exhausted[1]. At the level of national policy making, there exists a conflict between implementing strict International Patent Regime (IPR) and allowing for PI. The impact of PIs on the level of innovation remains the cornerstone of this debate on allowing PIs vis-à-vis enforcing IPR. The argument against PIs is that strict implementation of IPR will encourage research and development (R&D) and hence, PIs would adversely affect innovation [Valletti (2006), Li and Maskus (2006)][2]. The argument in favor of PIs claims that PIs curb monopoly pricing of the patent holder and hence, can be welfare improving [Malueg and Schwartz (1994), Jelovac and Bordoy (2005), Valletti (2006), Acharyya and Garcia-Alonso (2008)]. However, PIs reduce optimal tariffs chosen by the importing countries. This reduction in tariffs helps the exporting patent holder monopolist. The impact of PIs on the welfare of the importing nation is not unambiguous. On the one hand, PIs increase welfare through increased consumer surplus and on the other hand PIs reduce welfare through reduction in tariff revenue. One set of literature does focus on the role of trade policy on the optimality of PIs [Maskus (2000), Knox and Richardson (2002) and Hur and Riyanto (2006)] but does not analyze impact of PIs on product innovation. We argue that it is important to analyze both these aspects together as PIs affect both innovation and trade policy. We analyze the role of PIs in determining both optimal product quality and optimal trade policy in a three country world economy where poor (P) and middle-income (M) countries import a vertically differentiated good produced and exported by a patent holder monopolist in the rich country (R) when PIs are allowed. We analyze the impact of PIs on product innovation and in turn, on the welfare maximizing trade policies of the importing nations. The later part has been captured through endogenous trading bloc formation. The contribution of this paper lies in endogenous trading bloc formation under a vertically differentiated monopoly when PIs are allowed.

2. The Model

We consider a three country world. The countries are labeled P, M and R with population sizes being N_P, N_M and N_R respectively. A patent holder monopolist in R produces a vertically differentiated good. The quality of this good, indexed by *s*, is developed by the monopolist by investing an amount *c* in R&D. This R&D cost is sunk in nature and is convex in the level of quality being developed[3]:

$$c = \frac{1}{2}s^2 \tag{1}$$

All consumers in the j^{th} country have identical marginal willingness-to-pay (MWP) for quality which is captured by the taste parameter α_j $(j = P, M, R)$. We assume that there is a positive association between income[4] and taste parameter such that $\alpha_P < \alpha_M < \alpha_R$. Each buyer buys, if

[1] National policies pertaining to PIs vary widely across countries. USA pursues only national exhaustion in all fields of intellectual property (like trademarks, patents and copyrights) while the EU pursues a policy of regional exhaustion. Countries like Japan and Australia allow for international exhaustion and hence, PIs.

[2] However, Li and Robles (2007) and Grossman and Lai (2008) argue that PIs can accelerate the pace of innovation.

[3] Similar to Mussa and Rosen (1978) and Rochet and Stole (2002).

[4] The implicit assumption is $Y_R > Y_M > Y_P$ where, Y_j is the GDP of the j^{th} country $(j = P, M, R)$.

at all, only one unit of the good and derives gross utility of $u(\alpha_j, s)$. Following the literature on endogenous quality choice [Mussa and Rosen (1978), Tirole (1989), Choi and Shin (1992), Acharyya (1998), Rochet and Stole (2002)], the "net" utility (V_j) derived from consuming the good is additively separable in quality and price.

$$V_j = \alpha_j s - A_j \tag{2}$$

where, A_j is the price charged by the monopolist in country-j.

A buyer of type α_j will participate in the market if $V_j \geq 0$. We assume that if a buyer is indifferent between buying and not buying, then she buys the good.

Given that all countries allow PIs, it might not be profitable for the monopolist to serve all markets (universal coverage). The monopolist may, in fact, price-out the poor country, depending on the cross-country demand dispersion captured through differences in taste parameters. It can be verified that the monopolist will opt for universal coverage if[5]

$$\frac{3N_R+N_M}{2(N_R+N_P)} < \frac{\alpha_P}{\alpha_M} \tag{3}$$

Given "(3)", the monopolist cannot extract the entire consumer surplus in M and R in presence of PIs from P and sets price such that $V_j > 0$ in M and R markets. Note that, we here assume costless arbitrage which means full price convergence across all markets. On the one hand, such a pricing implies $V_P = 0$. On the other hand, such pricing also ensures the monopolist that units of the good sold in the j^{th} market will be N_j $(j = P, M, R)$.

3. Benchmark Case: No Trade Bloc

When the monopolist opts for universal coverage, it can ensure the profit for itself in the benchmark case (as noted by B in the superscript) as

$$\pi_R^B = [(1-t_P)\alpha_P N_P + (1-t_M)\alpha_P N_M + \alpha_P N_R]s - \frac{1}{2}s^2 \tag{4}$$

where t_P and t_M denote the ad-valorem tariff rates that countries P and M impose on their respective imports from R. The tariff rates are chosen unilaterally and simultaneously. Countries allow PI of the innovated product from the low-price country. For any given set of tariffs the monopolist chooses the optimal quality level (s_B^*) so as to maximize the profit level which yields the quality level as

$$s_B^* = [(1-t_P)\alpha_P N_P + (1-t_M)\alpha_P N_M + \alpha_P N_R] \tag{5}$$

Since the monopolist extracts all surpluses of buyers in P, the national welfare level of P is equal to the tariff revenue (TR) accrued to the national government. However, the national welfare of M consists of both TR and consumer surplus (CS) accruing to the consumers in M (since $\alpha_P < \alpha_M$).

$$W_P^B = t_P \alpha_P N_P s_B^* \tag{6a}$$

$$W_M^B = t_M \alpha_P N_M s_B^* + N_M(\alpha_M - \alpha_P)s_B^* \tag{6b}$$

The national welfare levels depend on own as well as other country's tariff level since from "(5)", $s_B^* = s(t_P, t_M, \alpha_P, \alpha_M, N_P, N_M)$. This means the unilaterally optimum tariff levels are inter-dependent and the best-response unilateral tariffs are inversely related[6]. The welfare maximizing unilaterally optimum Nash equilibrium (NE) tariff levels can be calculated as

[5] Appendix A1 provides the solution for the monopolist's profit maximization problem. Also note that if the country sizes are equal, that is, $N_P = N_M = N_R$, then the above condition will not hold as $\alpha_P < \alpha_M$.

[6] An increase in t_M, for example, lowers the level of innovation. The optimum response for P is then to lower its

$$t_P^B = \frac{1}{3\alpha_P N_P}\{K + N_M(\alpha_M - \alpha_P)\} \quad (7a)$$

$$t_M^B = \frac{1}{3\alpha_P N_M}\{K - 2N_M(\alpha_M - \alpha_P)\} < t_P^B \quad (7b)$$

where, $K = \alpha_P \sum N_j \ \ \forall\, j = P, M, R.$

The optimal tariff levels as obtained in "(7a)" and "(7b)" are strictly less than the optimal tariff levels $[t_i^B = \frac{1}{3}(\frac{\sum_{j=1}^{3} \alpha_j N_j}{\alpha_i N_i}); i = P, M]$ imposed by P and M on imports from R given no PIs. This result is exactly similar to the findings of Knox and Richardson (2002).

Substitution of "(7a)" and "(7b)" in "(4)" yields the optimal level of quality to be

$$s_B^* = \frac{1}{3}\{K + N_M(\alpha_M - \alpha_P)\} < \bar{s}_B^* \quad (8)$$

where, $\bar{s}_B^* = \frac{1}{3}\{\sum \alpha_j N_j\} \ \forall \ j = P, M, R$ is the level of quality under similar tariff regime but without PI. Note that the quality level depends both on size of the markets (N_j) and the MWP in P (α_P). Finally, substitution of "(7a)", "(7b)" and "(8)" in "(6a)" and "(6b)" yields national welfare levels as

$$W_P^B = W_M^B = (s_B^*)^2 \quad (9a)$$

$$W_R^B = \frac{1}{2}(s_B^*)^2 + N_R(\alpha_R - \alpha_P)(s_B^*) \quad (9b)$$

4. Trading Bloc Options

The different trading blocs among the countries possible under this framework are FTA_{RM} (Free Trade Area between countries R and M), FTA_{RP} (FTA between countries R and P) and CU_{PM} (CU between countries P and M)[7]. We also consider the case of global free trade (GFT) for welfare comparisons. However, relevant calculations reveal that neither FTA_{RP} is a feasible option, nor FTA_{RM} given $\frac{1}{4}K > N_M(\alpha_M - \alpha_P) + N_R(\alpha_R - \alpha_P)$. Neither of P and M finds it optimal to form an FTA with R[8]. This happens because the TR accruing to j^{th} country (j=P and M) comes down to zero as it joins FTA_{Rj} and such loss in welfare is more than the gain in national welfare of R. Hence, the only relevant trade bloc worth studying is CU_{PM}.

4.1. Customs Union between P and M

When P and M form a CU among themselves, they impose a joint welfare maximizing common external tariff (CET) on their imports from R –

$$t^{CU} = \frac{1}{2\alpha_P(N_P + N_M)}\{K - N_M(\alpha_M - \alpha_P)\} \quad (10)$$

The optimal level of quality under the CU_{PM} regime turns out to be

---

tariff level which mitigates the disincentive effect of the increase in t_M for the monopolist and such a "gain" in terms of smaller reduction in the quality level overcompensates the welfare loss for P from lower t_P.

[7] FTA_{PM} is analogous to the benchmark case discussed in section 2. Similarly, CU_{RP} and CU_{RM} are analogous to the cases of FTA_{RP} and FTA_{RM} respectively. This happens as we don't explicitly model the exports of P and M for keeping the analysis simple. We can assume that all the countries consume and export a homogeneous good X. This good is produced under constant returns-to-scale technology and perfectly competitive conditions. It can be treated as the numeraire good. The constant cost of producing this good is normalized to one. It is traded freely around the world with marginal utility equal to one.

[8] See appendix A2.

$$s^*_{CU} = \frac{1}{2}\{K + N_M(\alpha_M - \alpha_P)\} > s^*_B \quad (11a)$$

Here also we have[9] $s^*_{CU} < \bar{s}^*_{CU}$ (11b)

The joint welfare level of countries P and M turns out to be –

$$W^{CU}_{PM} = (s^*_{CU})^2 > W^B_P + W^B_M \quad (12)$$

As evident from "(12)", both P and M have incentives to form a CU among themselves. Welfare of R consists of CS accruing to the consumers in R and profit of the monopolist, which in turn depends on the optimal level of quality. The higher the level of quality, the higher will be the profit and hence, the welfare level of country R. In the case of CU_{PM} the welfare level of country R turns out to be

$$W^{CU}_R = \frac{1}{2}(s^*_{CU})^2 + N_R(\alpha_R - \alpha_P)(s^*_{CU}) > W^B_R \quad (13)$$

Hence, a CU between P and M is not only the feasible trading bloc option but welfare improving as well. Interestingly, such welfare improvement is Pareto superior to the benchmark case. The source of such welfare improvement lies in higher quality of the good.

Proposition 1: Formation of a CU between P and M raises the level of quality of the differentiated good developed by the monopolist in the non-member R. Such a CU is globally Pareto superior to unilateral protection regime.

Proof: Follows from "(9a)", "(9b)", "(12)" and "(13)". □

4.2. Global Free Trade (GFT)

Under GFT, the endogenous level of quality turns out to be the maximum –

$$s^*_{GFT} = K > s^*_{CU} > s^*_B \quad (14a)$$

The level of quality under GFT is lower in case of PIs vis-à-vis no PIs[10], that is,

$$s^*_{GFT} < \bar{s}^*_{GFT} \quad (14b)$$

The global welfare under GFT can be compared to that under the alternative regimes as

$$W^{GFT}_G = \frac{3}{2}(s^*_{GFT})^2 + N_R(\alpha_R - \alpha_P)(s^*_{GFT}) > W^{CU}_G > W^B_G \quad (15)$$

Under GFT national welfare levels of both P and M decline (as TRs disappear). Hence, GFT can only be sustained if R makes side-payments to P and M, which is not a plausible proposition.

Proposition 2: Parallel imports lower the level of innovation irrespective of the tariff regimes.

Proof: Follows from "(8)", "(11b)" and "(14b)". □

PIs undermine IPR protection for the monopolist. Hence, it is optimal for the patent holder monopolist to develop a lower quality level.

5. Robustness: Intra-country Taste Diversity

In this section we extend the benchmark model to the case of intra-country taste diversity. We, however, restrict ourselves to the case of two discrete types in each country. We assume, that there exist n_{1j} buyers with MWP α_{1j} and n_{2j} buyers with MWP α_{2j} such that $\alpha_{1j} < \alpha_{2j}$ and

[9] $\bar{s}^*_{CU} = \frac{1}{2}\{\sum \alpha_j N_j\}\ \forall\ j = P, M, R$ is the level of quality under the same CU regime but given no PI.

[10] $\bar{s}^*_{GFT} = \{\sum \alpha_j N_j\}\ \forall\ j = P, M, R$ is the level of quality under GFT regime when PIs are not allowed.

$n_{1j} + n_{2j} = N_j \ \forall \ \ j = P, M, R$. For the purpose of making comparisons with the case of no intra-country taste diversity, we make the following assumption[11] –

$$\alpha_{1j} n_{1j} + \alpha_{2j} n_{2j} = \alpha_j N_j \ \forall \ j = P, M, R \tag{16}$$

Again, the monopolist may charge higher price and serve only the consumers with higher MWP (partial coverage) or charge a lower price so as to serve both types of consumers (full coverage). With detailed derivations given in appendix A3, we find that the monopolist would cater to all consumers across all the countries (universal full coverage) if the following condition holds –

$$\frac{\alpha_{1P}}{\alpha_{2P}} < \frac{N_R}{n_{1P} + N_R} \tag{17}$$

5.1. Trading Bloc Options under Intra-country Taste Diversity

Given two types of buyers in each country the monopolist is able to extract the surpluses only from the consumers with lower MWP in P leaving the rest with positive surpluses. The national welfare of P and M now comprises of this CS and, as before, the TR. The optimal level of quality that would be developed by the monopolist turns out to be[12] –

$$\tilde{s}_B = \frac{1}{3}\{G + n_{2P}(\alpha_{2P} - \alpha_{1P}) + N_M(\alpha_M - \alpha_{1P})\} < \bar{\bar{s}}_B^* \tag{18}$$

where, $G = \alpha_{1P} \sum N_j$.

The welfare levels of the countries can easily be calculated to be

$$\widetilde{W}_j^B = (\tilde{s}_B)^2 \ \forall \ j = P, M. \tag{19a}$$

$$\widetilde{W}_R^B = \frac{1}{2}(\tilde{s}_B)^2 + N_R(\alpha_R - \alpha_{1P})(\tilde{s}_B) \tag{19b}$$

Now, the welfare levels for P and M given no PIs are $[(\bar{\bar{s}}_B^*)^2]$. As evident from "(18)" the welfare of P and M unambiguously decreases when PIs are allowed as compared to no-PI regime. This result supports that of Acharyya and Garcia-Alonso (2008).

The optimal level of quality under GFT turns out to be

$$\tilde{s}_{GFT} = G > \tilde{s}_B \tag{20a}$$

Similar to the case of no intra-country taste diversity, here also we have[13]

$$\tilde{s}_{GFT} < \bar{\bar{s}}_{GFT}^* \tag{20b}$$

5.1.1. Feasibility of FTAs with R

In contrast to no intra-country taste diversity, formation of FTAs with R becomes feasible under intra-country taste diversity[14]. Under the two different FTA possibilities of R (FTA_{RP} and FTA_{RM}), the optimum levels of quality under universal full coverage are obtained as

---

[11] Assumption "16" implies that $\alpha_j = \gamma_{1j}\alpha_{1j} + \gamma_{2j}\alpha_{2j}$ which in turn implies $\alpha_{1j} < \alpha_j < \alpha_{2j}$ where, $\gamma_{1j} = \frac{n_{1j}}{N_j}$ is the population share of buyers with lower MWP and $\gamma_{2j} = \frac{n_{2j}}{N_j}$ is the population share of buyers with higher MWP.

[12] $\bar{\bar{s}}_B^* = \frac{1}{3}\{\sum \alpha_{1j} N_j + n_{2P}(\alpha_{2P} - \alpha_{1P}) + n_{2M}(\alpha_{2M} - \alpha_{1M})\} \ \forall \ j = P, M, R$ is the level of quality under the benchmark case with intra-country taste diversity but given no PI.

[13] $\bar{\bar{s}}_{GFT}^* = \sum \alpha_{1j} N_j \ \forall \ j = P, M, R$ is the quality level under GFT given no PI

[14] The feasibility of formations of FTAs under intra-country taste diversity arises due to higher CS accruing to the high type consumers. This was not possible in the benchmark case. Hence, R has to make lesser side payments to its partner (who foregoes TR) to form the FTA.

$$\tilde{s}_{RP} = \frac{1}{2}\{G + N_M(\alpha_M - \alpha_{1P})\} \quad (21a)$$

$$\tilde{s}_{RM} = \frac{1}{2}\{G + n_{2P}(\alpha_{2P} - \alpha_{1P})\} \quad (21b)$$

Now, comparison with the analogous cases under no PI yields

$$\tilde{s}_{Ri} < \bar{\bar{s}}_{Ri} \quad (21c)$$

where, $\bar{\bar{s}}_{Ri} = \frac{1}{2}\left\{\sum \alpha_{1j} N_j + n_{2j}\left(\alpha_{2j} - \alpha_{1j}\right)\right\} \forall\, j = P, M; i = P, M; i \neq j$

However, calculations reveal that if $n_{2P} < N_M$, then, we have

$$\tilde{s}_{RP} > \tilde{s}_{RM} > \tilde{s}_B \quad (22)$$

and given equal sizes of M and P ($N_M = N_P$) we have

$$\tilde{t}_P^{RM} > \tilde{t}_M^{RP} \quad (23)$$

Further calculations yield the national welfare levels as

$$\widetilde{W}_j^{Ri} = (\tilde{s}_{Ri})^2 > \widetilde{W}_j^B \quad \forall\ j = P, M;\ \ i = P, M; i \neq j \quad (24a)$$

$$\widetilde{W}_P^{RP} = (\tilde{s}_{RP}) n_{2P}(\alpha_{2P} - \alpha_{1P}) \quad (24b)$$

$$\widetilde{W}_M^{RM} = (\tilde{s}_{RM}) N_M(\alpha_M - \alpha_{1P}) \quad (24c)$$

$$\widetilde{W}_R^{Ri} = (\tilde{s}_{Ri}) N_R(\alpha_R - \alpha_{1P}) + \frac{1}{2}(\tilde{s}_{Ri})^2 \ \forall\ \ i = P, M \quad (24d)$$

Combining "(22)" and "(24d)", we have –

$$\widetilde{W}_R^{RP} > \widetilde{W}_R^{RM} > \widetilde{W}_R^B \quad (25)$$

Hence, FTAs of R can be agreed upon with P and M with the possibility of some side-payments made by R to its FTA partner. Further, the gain, net of side-payments, is higher for R when it negotiates an FTA with P than with M making FTA_{RP} more feasible than FTA_{RM}.

Proposition 3: Given the assumption that the high-type population in P is less than the total population in M, R will always prefer to form an FTA with P than with M.

Proof: See appendix A4. □

The monopolist can, however, opt for technology transfer or FDI in order to jump tariff instead of FTA formation. As elaborated in appendix A5, as long as there is a positive fixed set up cost for FDI, country R would prefer to form an FTA.

5.1.2. Customs Union between P and M

When P and M form a CU, the CET that they impose on their imports from R, the optimal quality level [15] and welfare levels turn out to be –

$$\tilde{t}_{CU} = \frac{1}{2\alpha_{1P}(N_P + N_M)}\{G - N_{2P}(\alpha_{2P} - \alpha_{1P}) - N_M(\alpha_M - \alpha_{1P})\} \quad (26)$$

$$\tilde{s}_{CU} = \frac{1}{2}\{G + N_{2P}(\alpha_{2P} - \alpha_{1P}) + N_M(\alpha_M - \alpha_{1P})\} < \bar{\bar{s}}_{CU} \quad (27)$$

$$\widetilde{W}_{PM}^{CU} = \widetilde{W}_P^{CU} + \widetilde{W}_M^{CU} = (\tilde{s}_{CU})^2 \quad (28a)$$

$$W_R^{CU} = \frac{1}{2}(\tilde{s}_{CU})^2 + N_R(\alpha_R - \alpha_{1P})(\tilde{s}_{CU}) \quad (28b)$$

Note that, in this case as well, the CET and the joint welfare of P and M are lower than those under no PIs. Relevant calculations[16] involving national welfare levels reveal that even under intra-country taste diversity, both P and M have incentives to form a CU.

---

[15] $\bar{\bar{s}}_{CU} = \frac{1}{2}\left\{\sum \alpha_{1j} N_j + N_{2P}(\alpha_{2P} - \alpha_{1P}) + N_{2M}(\alpha_{2M} - \alpha_{1M})\right\}\ \forall\, j = P, M, R$ is the level of innovation under the CU regime with intra-country taste diversity given no PI.

[16] See appendix A6.

$$\widetilde{W}_{PM}^{CU} > \left(\widetilde{W}_{P}^{B} + \widetilde{W}_{M}^{B}\right) \tag{29}$$

Given both FTA_{RP} and FTA_{RM} being feasible, we compare the welfare levels of P and M under CU and FTA. We find that formation of CU among P and M is Pareto superior for both P and M compared to formation of FTA by P and M with R.

We also explore the consequences of partial coverage options by the monopolist (when condition "17" does not hold). As shown in appendix A7, allowing for partial coverage options by the monopolist does not drastically change the preferences over trade policy, particularly, the type (FTA or CU) and composition (member countries) of endogenous RTBs.

***Proposition 4*:** A mean-preserving taste dispersion does not alter the incentives for CU formation for P and M. The quality level is still larger as well.

Proof: See appendix A6. □

***Proposition 5*:** A mean-preserving taste dispersion does not alter the decline in the level of innovation under PI under different tariff regimes.

Proof: Follows from "(18)", "(20b)", "(21c)" and "(27)". □

6. Conclusion

This paper analyzes the role of PIs in determining both optimal product quality and optimal trade policy in a three country world economy. We analyze the impact of PIs on product innovation (as captured by the endogenous level of quality of the vertically differentiated good) done by a monopolist in the exporting country and in turn, on the welfare maximizing trade policies of the importing nations, which has been captured through endogenous trading bloc formation.

Given both inter and intra-country taste diversity, we find that PIs lower the level of innovation irrespective of tariff regimes as compared to when PIs are not allowed. PIs undermine IPR protection for the monopolist and hence, the patent holder monopolist endogenously chooses to develop a lower quality. Optimal tariff levels imposed by the importing countries are lower under PIs irrespective of the tariff regimes. This result supports the findings of Knox and Richardson (2002). PIs would increase the welfare of importing nations through higher consumer surplus and reduce welfare through reduced tariff revenues. In this model, the net effect of PIs on the welfare of the importing nations is negative. This result corroborate to the findings of Acharyya and Garcia-Alonso (2008).

The major contribution of this paper lies in the endogenous trading bloc formation. Formation of CU among the importing countries turns out to be globally Pareto superior to unilateral protection regime. Formation of CU raises the level of innovation by the monopolist vis-à-vis other tariff regimes. We also find that, allowing for partial coverage options by the monopolist does not drastically change the preferences over trade policy, particularly, the type (FTA or CU) and composition (member countries) of endogenous trading blocs.

References

1. Acharyya, R. (1998) "Monopoly and Product Quality: Separating or Pooling Menu?" *Economics Letters* **61**, 187-194
2. Acharyya, R and M. D. C. Garcia-Alonso (2008) "Parallel Imports, Innovations and National Welfare: The Role of the Sizes of Income Classes and National Markets for Health Care" *The Singapore Economic Review* **53**, 57-79.
3. Choi, CJ and H.S. Shin (1992) "A Comment on a Model of Vertical Product Differentiation" *The Journal of Industrial Economics* **40**, 229-231.
4. Grossman, GM and E.L.-C. Lai (2008). "Parallel Imports and Price Controls" *Rand Journal of Economics* **39**, 378-402.
5. Hur, J and Y.E. Riyanto (2006) "Tariff Policy and Exhaustion of Intellectual Property Rights in the Presence of Parallel Imports" *Oxford Economic Papers* **58**, 549-568.
6. Jelovac, I and C. Bordoy (2005) "Pricing and Welfare Implications of Parallel Imports in the Pharmaceutical Industry" *International Journal of Health Care Finance and Economics* **5**, 5-21.
7. Knox, D and M. Richardson (2002) "Trade Policy and Parallel Imports" *European Journal of Political Economy* **19**, 133-151.
8. Li, C and K. Maskus (2006) "The Impact of Parallel Imports on Investments in Cost-reducing Research and Development" *Journal of International Economics* **68**, 443-455.
9. Li, C and J. Robles (2007) "Product Innovation with Parallel Imports" *International Journal of Industrial Organization* **25**, 419–429.
10. Malueg, DA and M. Schwartz (1994). "Parallel Imports, Demand Dispersion and International Price Discrimination" *Journal of International Economics* **37**, 167–195.
11. Maskus, K. (2000) "Parallel Imports" *The World Economy* **23**, 1269-1284.
12. Mussa, M and S. Rosen (1978). Monopoly and Product Quality. *Journal of Economic Theory* **18**, 301-317.
13. Rochet, J-C and L.A. Stole (2002) "Nonlinear Pricing with Random Participation" *The Review of Economic Studies* **69**, 27-311.
14. Tirole, J. (1989) *The Theory of Industrial Organization*, 2nd Printing. MIT Press: Cambridge.
15. Valletti, T.M. (2006) "Differential Pricing, Parallel Imports and the Incentive to Invest" *Journal of International Economics* **70**, 314–324.

Appendices

A1. Condition for universal coverage (equation 3)

The optimal levels of quality can be worked out to be $s_1^* = \frac{1}{2}\alpha_R N_R$ when the monopolist serves only in R, $s_2^* = \frac{1}{2}\alpha_M(N_M + N_R)$ when it serves in M and R and $s_3^* = \frac{1}{3}[\alpha_P \sum N_j + (\alpha_M - \alpha_P)N_M]$ when it serves in all three markets (equations 4-8 provide the method of solution). Further, from "(4)" it is evident that the profit of the monopolist can be written as $\pi_i = \frac{1}{2}s_i^2$ for all the three cases as discussed.
Thus we can arrive at the following conditions by comparing the profit levels –

$\frac{3N_R+N_M}{2(N_R+N_P)} < \frac{\alpha_P}{\alpha_M}$ ensures that the monopolist prefers universal coverage over serving to M and R only and
$\frac{N_R}{N_P+N_M} < \frac{\alpha_P}{3\alpha_R-\alpha_P}$ ensures that the monopolist prefers universal coverage over serving to R only

Now, given assumption $\alpha_P < \alpha_M < \alpha_R$, $\frac{\alpha_P}{\alpha_M}$ is the smallest and $\frac{3N_R+N_M}{2(N_R+N_P)}$ is the biggest. Hence, if $\frac{3N_R+N_M}{2(N_R+N_P)} < \frac{\alpha_P}{\alpha_M}$ holds, then, the monopolist will prefer universal coverage over all other possible types of market coverage available to itself.

A2. Infeasibility of FTA$_{RM}$ and FTA$_{RP}$

From "(8)" we have $s_B^* = \frac{1}{3}\{K + N_M(\alpha_M - \alpha_P)\}$ where, $K = \alpha_P \sum N_j \ \ \forall\, j = P, M, R.$
From "(9a)" and "(9b)" we have $W_P^B = W_M^B = (s_B^*)^2$ and $W_R^B = \frac{1}{2}(s_B^*)^2 + N_R(\alpha_R - \alpha_P)(s_B^*)$
Again, welfare maximization under the FTA$_{RP}$ regime yields

$$W_R^{RP} = \frac{1}{2}(s_{RP}^*)^2 + N_R(\alpha_R - \alpha_P)(s_{RP}^*) \text{ where, } s_{RP}^* = \frac{1}{2}\{K + N_M(\alpha_M - \alpha_P)\}$$

Now, under FTA$_{RP}$ regime $W_P^{RP} = 0 < W_P^B = (s_B^*)^2$ and $W_R^{RP} > W_R^B$ as $s_{RP}^* > s_B^*$.
Again, $(W_R^B + W_P^B) - (W_R^{RP} + W_P^{RP}) > 0$ if $K + N_M(\alpha_M - \alpha_P) > 4N_R(\alpha_R - \alpha_P)$
Welfare maximization under the FTA$_{RM}$ regime yields

$$W_R^{RM} = \frac{1}{2}(s_{RM}^*)^2 + N_R(\alpha_R - \alpha_P)(s_{RM}^*)$$
$$W_M^{RM} = N_M(\alpha_M - \alpha_P)(s_{RM}^*)$$

where, $(s_{RM}^*) = \frac{1}{2}K$
Again, $(W_R^B + W_M^B) - (W_R^{RM} + W_M^{RM}) > 0$ if $\frac{1}{4}K > N_M(\alpha_M - \alpha_P) + N_R(\alpha_R - \alpha_P)$
Similar welfare levels of P and M under their respective FTAs with R can be compared with that under the customs union regime. Such welfare comparisons yield

$$(W_P^{RM} + W_M^{RM}) < (W_P^{CU} + W_M^{CU})$$
$$(W_P^{RP} + W_M^{RP}) = (W_P^{CU} + W_M^{CU})$$

Hence, formation of FTAs by P and M with R are not only infeasible given the parametric restrictions but such FTA formation may be Pareto inefficient compared to CU formation by P and M as well.

A3. Condition for universal full coverage (equation 17)

In each country we have two types of consumers with different MWP which we capture by $\alpha_{2j} > \alpha_{1j} \ \ \forall \ \ j = P, M, R$. The monopolist knows about the existence of two such types, however, he can't distinguish between them. As a result, he will have to offer a menu for which each type of consumer will reveal his/her type. The price that the monopolist may charge must be such that a consumer will buy the good. Hence, the monopolist must consider the following constraints while deciding about the level of quality and, in turn, the prices of each level of quality that the monopolist may produce.

Participation Constraint (each individual consumer must have a non-negative net utility by consuming the good) –

$$V_{ij} = \alpha_{ij} s_i - A_i \geq 0 \ \ \forall \ \ i = 1,2; j = P, M, R$$

Self-selection Constraint (the net utility that the high type consumer gets by purchasing the higher quality must be at least as that he gets by purchasing the lower quality) –

$$\alpha_{2j} s_2 - A_2 \geq \alpha_{2j} s_1 - A_1 \ \ \forall \, j = P, M, R$$

Given these, and the fact that the monopolist incurs only a sunk fixed cost while developing the quality, the profit function off the monopolist looks like –

(1) If the monopolist opts for a menu of high quality high price and low quality low price

$$\pi(s_1, s_2) = n_1 \alpha_1 s_1 - \tfrac{1}{2} s_1^2 + n_2(\alpha_2 s_2 - \alpha_2 s_1 + \alpha_1 s_1) - \tfrac{1}{2} s_2^2 \ \ldots\ldots \text{(since, } \alpha_1 s_1 = A_1)$$

(2) If the monopolist offers a single quality

$$\pi(\bar{s}) = (n_1 \alpha_1 + n_2 \alpha_2)\bar{s} - \frac{1}{2}\bar{s}^2$$

From these alternative profit functions we can calculate the profit maximizing quality levels as –

$$s_1^* = (n_1 \alpha_1 + n_2 \alpha_1 - n_2 \alpha_2)$$
$$s_2^* = (n_2 \alpha_2)$$
$$\bar{s}^* = (n_1 \alpha_1 + n_2 \alpha_2)$$

Now, using the above optimal levels of quality we can simply calculate the alternative profits levels as –

$$\pi(s_1^*, s_2^*) = \tfrac{1}{2}[(s_1^*)^2 + (s_2^*)^2]$$
$$\pi(\bar{s}^*) = \tfrac{1}{2}[(\bar{s}^*)^2]$$

Now, $[\bar{s}^* - (s_1^* + s_2^*)] = [(n_1 \alpha_1 + n_2 \alpha_2) - (n_1 \alpha_1 + n_2 \alpha_1 - n_2 \alpha_2) - (n_2 \alpha_2)]$

$= n_2(\alpha_2 - \alpha_1) > 0$ … [since, by assumption $\alpha_2 > \alpha_1$]

Hence, $(\bar{s}^*)^2 > [(s_1^* + s_2^*)]^2 \Rightarrow (\bar{s}^*)^2 > [(s_1^*)^2 + (s_2^*)^2] \Rightarrow \pi(\bar{s}^*) > \pi(s_1^*, s_2^*)$

Hence, the monopolist finds it optimal to develop a single quality and charge uniform price vis-à-vis developing two different qualities and charging separate prices. Now, once the monopolist finds it optimal to develop a single quality and charge uniform price, the question arises whether the monopolist will serve both types of consumers and charge a lower price (full coverage) or serve only the consumers with higher MWP and charge a higher price (partial coverage).

Given the framework developed in section 5 of the main text, the different types of market coverage possibilities available to the monopolist and the ensuing levels of quality are –

Types of market coverage by the monopolist	Levels of Quality
Universal full coverage	$\tilde{s}_B = \frac{1}{3}\{G + n_{2P}(\alpha_{2P} - \alpha_{1P}) + N_M(\alpha_M - \alpha_{1P})\}$
MNC sells to all but low-type in P	$\tilde{s}_{PC} = \frac{1}{3}\{\alpha_{2P}(N - n_{1P}) + N_M(\alpha_M - \alpha_{2P})\}$
MNC sells in M and R	$\hat{s}_{PC} = \frac{1}{2}\{\alpha_{1M}(N_M + N_R) + n_{2M}(\alpha_{2M} - \alpha_{1M})$
MNC sells in R and high-type in M	$\bar{s}_{PC} = \frac{1}{2}(n_{2M} + N_R)\alpha_{2M}$
MNC sells only in R	$\bar{\bar{s}}_{PC} = N_R\alpha_{1R}$

The monopolist's profit is directly proportional to the level of quality developed. Hence, universal full coverage will be preferred by the monopolist over other types of partial coverage if the following conditions are satisfied –

Universal full coverage will be preferred to	Conditions
MNC sells to all but low-type in P	$\frac{N_R}{N_{1P}+N_R} < \frac{\alpha_{1P}}{\alpha_{2P}}$
MNC sells in M and R only	$\frac{N_M+3N_R}{2(N_P+N_R)} < \frac{\alpha_{1P}}{\alpha_M}$
MNC sells in H and high-type in M	$\frac{N_{2M}+3N_R}{2(N_P+N_R)} < \frac{\alpha_{1P}}{\alpha_{2M}}$
MNC sells only in R	$\frac{N_R}{N_P+N_M} < \frac{\alpha_P}{3(\alpha_{1R}-\alpha_{1P})}$

Now, given assumption $\alpha_{1P} < \alpha_{2P} < \alpha_{1M} < \alpha_{2M} < \alpha_{1R} < \alpha_{2R}$, $\frac{\alpha_{1P}}{\alpha_{2P}}$ is the smallest and given $N = \sum N_j$ and $N_j = N_{1j} + N_{2j} \forall\, j = P, M, R$; $\frac{N_R}{n_{1P}+N_R}$ is the biggest. Hence, if $\frac{N_R}{n_{1P}+N_R} < \frac{\alpha_{1P}}{\alpha_{2P}}$ holds, then, the monopolist will prefer universal full coverage over all other possible types of partial coverage available.

A4. Proof of Proposition 3

From "(24b)" – "(24d)" we have

$$\widetilde{W}_P^{RP} = (\tilde{s}_{RP})N_{2P}(\alpha_{2P} - \alpha_{1P})$$
$$\widetilde{W}_M^{RM} = (\tilde{s}_{RM})N_M(\alpha_M - \alpha_{1P})$$
$$\widetilde{W}_R^{Ri} = (\tilde{s}_{Ri})N_R(\alpha_R - \alpha_{1P}) + \frac{1}{2}(\tilde{s}_{Ri})^2 \;\forall\; i = P, M$$

Using the above equations we get

$$\left(\widetilde{W}_R^{RM} - \widetilde{W}_M^{RM}\right) - \left(\widetilde{W}_R^{RP} - \widetilde{W}_P^{RP}\right)$$
$$= \frac{1}{2}(\tilde{s}_{RP})^2 - \frac{1}{2}(\tilde{s}_{RM})^2 + (\tilde{s}_{RP})\{N_R(\alpha_R - \alpha_{1P}) - N_{2P}(\alpha_{2P} - \alpha_{1P})\}$$
$$-(\tilde{s}_{RM})\{N_R(\alpha_R - \alpha_{1P}) - N_M(\alpha_M - \alpha_{1P})\}$$

From "(22)" we have $\tilde{s}_{RP} > \tilde{s}_{RM}$

Using these two in the above equation and assumption "(16)" we have

$$\left(\widetilde{W}_R^{RM} - \widetilde{W}_M^{RM}\right) - \left(\widetilde{W}_R^{RP} - \widetilde{W}_P^{RP}\right) > 0$$

A5. Preference of the monopolist for FTA over FDI

If the monopolist opts for FDI or technology transfer, then the import tariff would become zero (tariff jumping argument). However, for FDI, there is a fixed set up cost (say, *f*). Now, the monopolist would compare its profits under the two alternative situations – exporting without tariff under FTA (π_{FTA}) and producing in foreign country through FDI (π_{FDI}).

Without loss of generality, we analyze the situation when the monopolist considers between forming an FTA with M and producing in M through FDI (the analyses would be exactly the same if the monopolist considers country P instead of M). Assuming the case of universal coverage with intra-country taste diversity, we can write these two profit functions as –

$$\pi_{FTA} = [(1 - t_P)\alpha_{1P}N_P + \alpha_{1P}N_M + \alpha_{1P}N_R]s - \frac{1}{2}s^2$$

$$\pi_{FDI} = [(1 - t_P)\alpha_{1P}N_P + \alpha_{1P}N_M + \alpha_{1P}N_R]s - \frac{1}{2}s^2 - f$$

Welfare maximization tariff for country P turns out to be the same under both situations –

$$t_P = \frac{1}{2\alpha_{1P}N_P}[\alpha_{1P}N_P + \alpha_{1P}N_M + \alpha_{1P}N_R - N_{2P}(\alpha_{2P} - \alpha_{1P})]$$

Now, as evident from the above equations, optimal quality (*s*) would turn out to be the same under both the conditions –

$$s_{FTA} = s_{FDI} = \tfrac{1}{2}[\alpha_{1P}N_P + \alpha_{1P}N_M + \alpha_{1P}N_R + N_{2P}(\alpha_{2P} - \alpha_{1P})]$$

Hence, the fixed set up cost for FDI does not have any impact on the endogenously chosen quality level.

Now, welfare of country P depends on *s* and t_P and that of M depends on *s*. Given $s_{FTA} = s_{FDI}$ and same t_P, both P and M are indifferent between whether R forms and FTA with M or R produces in M via FDI channel. That is, both P and M would be indifferent between "FDI without FTA" and "import with FTA". Thus the entire decisions boils down to the choice made by the R country (which depends on the welfare levels of R under the two alternative situations)

As evident from the profit functions of the monopolist, given $s_{FTA} = s_{FDI}$, π_{FTA} would be higher than π_{FDI} by the amount of the fixed set up cost (*f*). Comparing the welfare levels, we find that $W_R^{FTA} - W_R^{FDI} = f$. As a result, the monopolist would prefer to export with no tariff under FTA over producing in M (tariff jumping) through FDI.

A6. Proof of Proposition 4

From "(19a)" we have $\widetilde{W}_j^B = (\tilde{s}_B)^2 \ \forall \ j = P, M$ where,

$$\tilde{s}_B = \frac{1}{3}\{G + N_{2P}(\alpha_{2P} - \alpha_{1P}) + N_M(\alpha_M - \alpha_{1P})\}$$

From "(28a)" we have $\widetilde{W}_{PM}^{CU} = \widetilde{W}_P^{CU} + \widetilde{W}_M^{CU} = (\tilde{s}_{CU})^2$ where,

$$\tilde{s}_{CU} = \frac{1}{2}\{G + N_{2P}(\alpha_{2P} - \alpha_{1P}) + N_M(\alpha_M - \alpha_{1P})\}$$

Let $\{G + N_{2P}(\alpha_{2P} - \alpha_{1P}) + N_M(\alpha_M - \alpha_{1P})\} = D$

Putting the value of D in "(19a)" and "(28a)" we get

$$\widetilde{W}_P^B + \widetilde{W}_M^B = 2(\tilde{s}_B)^2 = \frac{2}{9}D^2 \text{ and } \widetilde{W}_{PM}^{CU} = (\tilde{s}_{CU})^2 = \frac{1}{4}D^2$$

Therefore, $\widetilde{W}_{PM}^{CU} - \left(\widetilde{W}_P^B + \widetilde{W}_M^B\right) = \frac{1}{36}D^2 > 0$

A7. Partial coverage options and preferences regarding trade policy

We have evaluated the feasibility of endogenous trading blocs under different partial coverage situations and present the results below. It must be noted that under all other partial coverage options (apart from the situation where the monopolist serves all but the low-type in P), the monopolist in R does not serve consumers in P and hence, the possibilities of CU between P and M and FTA between R and P do not arise.

Feasibility of Trading Blocs under Partial Coverage

Type of Coverage	CU between P and M	FTA between R and P	FTA between R and M
The monopolist serves all but low type in P	Feasible	Feasible if $\delta < 4r$	Not feasible
The monopolist serves all in M and R	Possibility doesn't arise	Possibility doesn't arise	Feasible if $\frac{N_M + N_R}{N_{2M}} > \frac{\alpha_{2M} - \alpha_{1M}}{\alpha_{1M}}$
The monopolist serves in R and only high type in M	Possibility doesn't arise	Possibility doesn't arise	Feasible
The monopolist serves only in R	Possibility doesn't arise	Possibility doesn't arise	Possibility doesn't arise

$\delta = [\alpha_{2P}(N_{2P} + N_M + N_R) + N_M(\alpha_M - \alpha_{2P})]$ and $r = N_R(\alpha_R - \alpha_{2P})$