Trade Agreements as Endogenously Incomplete Contracts

Henrik Horn (Research Institute of Industrial Economics, Stockholm)
Giovanni Maggi (Princeton University)
Robert W. Staiger (Stanford University and Wisconsin)
Introduction

Real-world trade agreements display an interesting combination of *rigidity* and *discretion*. Consider the GATT/WTO:

1. Trade instruments bound; domestic instruments largely left to discretion, but must satisfy National Treatment, and now (WTO) regulation of subsidies.

2. Bindings rigid, but with “escape clauses” (e.g. GATT Article XIX).

3. Bindings stipulate ceilings, so governments have downward discretion.

Why? Conjecture: Can be understood from incomplete contracts perspective.
A number of possible sources of incompleteness. We focus on two features of fundamental importance to trade negotiators:

(1) Wide array of trade-relevant policies: border measures but also internal measures. Controlling opportunism requires *comprehensive policy coverage*.

(2) Uncertainty about future economic/political conditions. This calls for agreements that are *highly contingent*.

Trade-law literature emphasizes contracting implications of costs associated with these features:
“...The standard trade policy rules could deal with the common type of trade policy measure governments usually employ to control trade. But trade can also be affected by other “domestic” measures, such as product safety standards, having nothing to do with trade policy. It would have been next to impossible to catalogue all such possibilities in advance.” (Hudec, 1990).

“...Many contracts are negotiated under conditions of considerable complexity and uncertainty, and it is not economical for the parties to specify in advance how they ought to behave under every conceivable contingency ... The parties to trade agreements, like the parties to private contracts, enter the bargain under conditions of uncertainty. Economic conditions may change, the strength of interest group organization may change, and so on.” (Schwartz and Sykes, 2002).
We introduce *contracting costs* explicitly into economic analysis of trade agreements, and study their implications for the structure of the optimal (incomplete) agreement. We argue that contracting costs can help explain some of the core features of the GATT/WTO.

**The Model**

Two countries, H and F. Two non-numeraire goods, 1 and 2, plus numeraire: partial-equilibrium analysis.

H a natural importer of good 1/exporter of good 2. Sectors 1 and 2 are mirror-image, so focus on sector 1. Demand: \( D(p) = \alpha - \beta p; \) \( D^*(p^*) = \alpha^* - \beta^* p^* \). Supply: \( X(q) = \lambda q; X^*(q^*) = \lambda^* q^* \).

H chooses tariff \( \tau \), separate consumption taxes on domestic and foreign products (\( t_h \) and \( t_f \)), production subsidy (\( s \)). F does not intervene in this sector.
The following must hold: \( q^* = p^*; \ p^* = p - \tau - t_f; \ q = p - t_h + s. \)

More compactly: \( p = p^* + T; \ q = p^* + T + S, \) where \( T \equiv \tau + t_f \) and \( S \equiv s - t_h. \)

Market clearing/arbitrage: \( p = p(T, S); \ q = q(T, S); \ p^* = q^* = p^*(T, S). \)

Importing country H experiences a negative consumption externality equal to \(-\gamma D\) with \( \gamma > 0. \) (Later consider production externality/pol.ec. forces).

Governments maximize welfare, so (still maintaining our focus on sector 1):

\[
W = CS + PS + T \cdot M - S \cdot X - \gamma D; \quad W^* = CS^* + PS^*.
\]
Efficient and Nash equilibrium policies

Globally efficient policies maximize $W^G \equiv W + W^*$, yielding

$$T^{eff} = \gamma; \quad S^{eff} = -\gamma.$$

Nash equilibrium policies:

$$T^{NE} = \gamma + \frac{E^*}{\beta^* + \lambda^*} = \gamma + \frac{p^*}{\eta^*}$$

$$S^{NE} = -\gamma.$$  

Note: $T^{NE} > T^{eff}; \quad S^{NE} = S^{eff}$. Nash trade taxes inefficiently high; Nash domestic instruments set at efficient levels.
Uncertainty

To simplify, focus for now on one-dimensional uncertainty.

Consider two possible sources of uncertainty: consumption externality ($\gamma$) and import demand level ($\alpha$).

Timing: (1) The agreement is drafted; (2) Uncertainty is resolved; (3) Policies are chosen subject to the constraints set by the agreement.

Let $\Omega \equiv EW^G(\cdot)$ denote expected gross-of-contracting-costs global welfare.
The costs of contracting

We focus mostly on *instrument-based* (not *outcome-based*) agreements.

Key idea: more detailed agreements are more costly (similar to Battigalli and Maggi, 2002).

\[ c_p: \text{ cost of including a policy variable } (\tau, t_f, s, t_h). \quad c_s: \text{ cost of including a state variable } (\gamma, \alpha). \]

Cost of writing an agreement: \( C = c_s \cdot n_s + c_p \cdot n_p, \)

with \( n_s \) (\( n_p \)) the number of state (policy) variables in the agreement.

Note: Basic results go through if \( C \) is increasing in \( n_s \) and \( n_p \). Modeling of \( C \) is slightly more general than Battigalli and Maggi.

Let \( c_p \equiv c, \ c_s \equiv k \cdot c. \) Fix \( k. \) Then \( C = c \cdot (n_p + k \cdot n_s) \equiv c \cdot m, \)

where \( m \) is rough measure of complexity of the agreement.
Optimal agreements

An *optimal agreement* maximizes expected net global welfare, $\omega \equiv \Omega - C$.

Before imposing more structure on uncertainty, three basic results.

First (Lemma 1): nec. and suff. conditions for an agreement to be optimal for a range of contracting costs. Second (Lemma 2): complexity of optimal agreement rises as level of contracting cost falls. Figure 1.
Figure 1
To state third result, recall: $T = \tau + t_f; S = s - t_h$. Hence $T$ and $S$ the relevant policy variables, with cost $2c$ for each.

**Proposition 1**: An agreement that constrains the effective subsidy $S$ (even in a state-contingent way) while leaving the import tax $T$ to discretion cannot improve over the Nash equilibrium, and therefore cannot be an optimal agreement.

Broad intuition: contracting over $S$ alone is useless because inefficiency in the NE concerns $T$, not $S$.

In world of costless contracting, Proposition 1 irrelevant. But as we next show, gains relevance when contracting costly.
Uncertainty about the consumption externality $\gamma$

Assume that $\gamma$ can take one of two values: $\bar{\gamma} - \Delta \gamma$ or $\bar{\gamma} + \Delta \gamma$.

For now consider only agreements that impose separate equality constraints on $T$ and $S$ (e.g. $(T = \gamma)$ or $(S = 10)$). Call this set $A_0$.

Note: $\{FB\}$ agreement is $\{T = \gamma; S = -\gamma\}$, which costs $(4 + k) \cdot c$.

If $c$ is small enough $\{FB\}$ optimal; for large $c$ empty agreement (yielding NE payoffs) is optimal. What happens between these two extremes?

Two ways to save on contracting costs relative to $\{FB\}$: agreement can be rigid (i.e. non-contingent); and/or it can leave some policies to discretion.
By Proposition 1, can focus on three kinds of agreement (aside from \( \{FB\} \) and \( \{\emptyset\} \): \( \{T, S\}\); \( \{T(\gamma)\}\); and \( \{T\}\).

Note: \( \{T, S\}\) rigid; \( \{T(\gamma)\}\) discretion; \( \{T\}\) both rigid and discretionary.

Basic trade-off: rigid agreement prevents TOT manipulation, but Pigouvian intervention only “on average;” discretion allows implicit state-contingency but creates scope for manipulating TOT.

Optimal sequence of agreements as \( c \) increases from zero? By Lemma 2, must be a subsequence of \((\{FB\}, \{T, S\}, \{T(\gamma)\}, \{T\}, \{\emptyset\})\).

But \( \{T, S\}\) and \( \{T(\gamma)\}\) can’t be part of the same optimal sequence.

Why? Cost of discretion lower in presence of rigidity, as *discretion introduces implicit state-contingency into a rigid agreement*, so rigidity and discretion are *complementary*; this plus Lemma 1 implies Figure 2a/b.
Figure 2a

Figure 2b
Proposition 2: Assume that only $\gamma$ is uncertain. There exist scalars $c_1$, $c_2$, $c_3$ and $c_4$ with $0 < c_1 \leq c_2 \leq c_3 \leq c_4 < \infty$ such that the optimal agreement in $A_0$ is:

(a) the $\{FB\}$ agreement for $c \in (0, c_1)$;
(b) of the form $\{T, S\}$ for $c \in (c_1, c_2)$;
(c) of the form $\{T(\gamma)\}$ for $c \in (c_2, c_3)$;
(d) of the form $\{T\}$ for $c \in (c_3, c_4)$; and
(e) the empty agreement for $c > c_4$.
Moreover, either $c_2 = c_1$ or $c_3 = c_2$ (or both).

Implications. Non-empty trade agreements must always include commitments over $T$, but not necessarily $S$. As $c$ rises, several possible paths: may first introduce rigidity, then add discretion; or, may first introduce discretion, then add rigidity; or, could skip rigidity completely; but can’t “oscillate.”

As $c$ rises, which path taken? Depends on the cost of discretion versus the cost of rigidity.

Determinants of cost of discretion and rigidity reflect simple economic intuitions.

First, discretion costly to gross global welfare when H’s monopoly power in trade is high, and when $S$ a close substitute for $T$ for TOT manipulation.

Second, rigidity costly to gross global welfare when level of uncertainty is high.

So, for example, if monopoly power high (high $\alpha$) and/or uncertainty low (low $\Delta\gamma$), then as $c$ rises, first introduce rigidity, then consider discretion; if $S$ a poor substitute for $T$ (low $\lambda$), then as $c$ rises, first introduce discretion, then consider rigidity.
Comparative Statics: Focus first on $\alpha$. Recall that, through the monopoly power effect, increasing $\alpha$ increases the cost of discretion.

**Proposition 3:** As the import demand level $\alpha$ increases (holding all other parameters fixed): (i) The optimal degree of discretion decreases, in the sense that the number of policy instruments specified in the optimal agreement increases (weakly); (ii) The optimal degree of rigidity decreases (weakly).

At a broad level, Proposition 3(i) suggests: (1) possible explanation for new WTO regulation of domestic subsidies; and (2) possible benefit of “special and differential treatment” for small/developing countries when it comes to contracting over domestic subsidies.

Proposition 3(ii) is due to the fact that rigidity and discretion are *complementary*, hence they tend to move together.
Next focus on the impact of uncertainty. As the degree of uncertainty $\Delta_\gamma$ increases, the optimal agreement tends to become more contingent:

**Proposition 4:** As the degree of uncertainty $\Delta_\gamma$ increases (holding all other parameters fixed), the optimal agreement may switch from a rigid agreement to a contingent agreement, but not vice-versa.

By itself not surprising. But also a more subtle feature: concerns uncertainty over a state variable that is directly relevant for the setting of $T$ in the $\{FB\}$ agreement. As we next demonstrate, effects of increasing uncertainty over state variables (such as $\alpha$) that are not directly relevant for the setting of $T$ in the $\{FB\}$ can be very different.
Uncertainty about the import demand level $\alpha$

Preview: $\alpha$ has no impact on the first-best levels of $T$ or $S$. Consequently, implications of uncertainty over $\alpha$ very different from uncertainty over $\gamma$.

Assume $\alpha$ can take one of two possible values: $\bar{\alpha} + \Delta_\alpha$ or $\bar{\alpha} - \Delta_\alpha$.

Now the $\{FB\}$ agreement is $\{T = \bar{\gamma}; S = -\bar{\gamma}\}$. Note, this is not a contingent agreement.

We can focus on two types of agreement (aside from $\{FB\}$ and $\{\emptyset\}$): $\{T(\alpha)\}$; and $\{T\}$. 
New possibility: \( \{T(\alpha)\} \), with \( T(\alpha) \) an increasing function. This has flavor of an “escape clause,” but rationale is novel: to manage distortions in domestic instruments introduced by trade agreements.

Intuition: an escape clause is appealing when the agreement leaves discretion over \( S \); allowing for a higher \( T \) mitigates the incentive to distort \( S \), and this is more valuable when \( \alpha \) (and hence underlying import volume) is higher.

Novel feature: rigidity and discretion now \textit{substitutes} (cost of discretion higher in presence of rigidity) because rigidity precludes escape clause. Hence, nature of interaction between rigidity and discretion depends on source of uncertainty.

Next, What is the optimal sequence of agreements as \( c \) increases?
Proposition 5: Assume that only $\alpha$ is uncertain. There exist $c_1$, $c_2$, and $c_3$ with $0 < c_1 \leq c_2 \leq c_3 < \infty$ such that the optimal agreement in $A_0$ is:
(a) the $\{FB\}$ agreement $\{T = \bar{\gamma}; S = -\bar{\gamma}\}$ for $c \in (0, c_1)$;
(b) of the form $\{T(\alpha)\}$ for $c \in (c_1, c_2)$;
(c) of the form $\{T\}$ for $c \in (c_2, c_3)$;
(d) the empty agreement for $c > c_3$.

Note: if $k$ sufficiently low and $\lambda$ sufficiently low (but strictly positive), then escape-clause-type agreement $\{T(\alpha)\}$ is optimal for range of $c$; optimal contract is contingent on the “wrong” state variable.

Finally, note: cost of discretion convex in $\alpha$, so as $\Delta \alpha$ rises, optimal agreement may switch from contingent ($\{T(\alpha)\}$) to rigid ($\{T, S\}$); could never happen for $\Delta \gamma$, so source of uncertainty matters.
The role of the National Treatment (NT) clause

We now extend the feasible set of agreements by allowing for an NT clause, that is a constraint $t_h = t_f$, costing $2c$. (For simplicity, return to $\gamma$ uncertainty).

Definition: an NT-based agreement is an agreement that includes the NT clause. Let $\mathcal{A}_{NT}$ denote the class of NT-based agreements.

Note: for NT-based agreements, the price relationships are: $p = p^* + \tau + t$; $q = p^* + \tau + s$. Recall for non-NT: $p = p^* + T$; $q = p^* + T + S$.

An agreement $\{NT, \tau, s\}$ costs less than $\{FB\}$ and ties down the producer price wedge $q - p^*$ while leaving the consumer price wedge $p - p^*$ to discretion. This type of discretion is not possible with non-NT agreements.
Under what conditions is an NT-based agreement strictly optimal for a range of $c$?

**Proposition 6:** Consider the agreement class $\mathcal{A}_0 \cup \mathcal{A}_{NT}$. If $\Delta \gamma > 0$ and $\beta$ is sufficiently small (so that the consumption tax $t$ is a poor substitute for the tariff $\tau$), then there is an intermediate range of $c$ for which the optimal agreement includes the NT clause.

NT-based agreement strictly optimal if low substitutability between $t$ and $\tau$, and $c$ lies in intermediate range. Describes world in which NT-based agreement gets close to first best ($\{t^{\text{eff}} = \gamma, \tau^{\text{eff}} = 0, s^{\text{eff}} = 0\}$) by exploiting implicit state-contingency associated with discretion over internal taxes.
The role of weak bindings

A weak binding is a constraint of the type $T \leq \bar{T}(\cdot)$, which allows for downward discretion. Can this be an optimal way to save on contracting costs?

Note: weak bindings can potentially improve on strong bindings only if they are rigid, i.e. of the type $T \leq \bar{T}$ (e.g. a clause of the type $T \leq \bar{T}(\gamma)$ cannot improve on $T = \bar{T}(\gamma)$), because downward discretion allows for some implicit state-contingency which can then be valuable.

**Proposition 7:** Suppose the optimal agreement in $A_0 \cup A_{NT}$ includes some rigid strong binding. Then this agreement can be weakly improved upon by replacing the rigid strong bindings with rigid weak bindings, and the improvement is strict for some configurations of parameters.
Production externalities and political-economy motives

A production externality in $H$ equal to $\sigma X$ with $\sigma > 0$: the $\{FB\}$ agreement takes the form $\{T = 0; S = \sigma\}$, so that now only $S$ is state-contingent.

Implication: appeal of $\{T(\sigma)\}$ similar to $\{T(\alpha)\}$, not $\{T(\gamma)\}$.

A political economy extension with extra weight $\xi$ on producer surplus: the $\{FB\}$ agreement is $\{T = 0; S = S(\xi, \alpha)\}$, where $S(\xi, \alpha)$ is increasing in both arguments.

Implication: $\{T(\alpha)\}$ agreement of particular interest here. A tariff response to trade shocks in the presence of distributional concerns, but for the purpose of mitigating distortions in the discretionary use of $S$. 
Conclusion

• Our paper takes a first step in the analysis of trade agreements as endogenously incomplete contracts. We have shown that this perspective provides a novel explanation for:

  – the emphasis on border measures in real world trade agreements;

  – the appeal of “escape clauses” in the presence of surging import demand;

  – the appeal of the National Treatment provision in GATT/WTO;

  – the emphasis on weak bindings.
Possible directions for future research:

- consider a broader class of feasible agreements, and in particular outcome-based agreements (e.g. imposing constraints on trade volumes or prices);

- consider export-sector policies;

- consider a multi-country setting, to examine the potential appeal of the MFN rule and exceptions for FTAs/CUs;

- consider a commitment role for trade agreements;

- consider the potential appeal of a dispute settlement body, as a mechanism to “complete” the incomplete contract;

- more explicit modeling of contract costs.