

THE ECONOMIC STRUCTURE OF INTERNATIONAL TRADE-IN-SERVICES AGREEMENTS*

Robert W. Staiger

Dartmouth College and NBER

Alan O. Sykes

Stanford University Law School

December 2016

Abstract

The existing economics literature on international trade agreements focuses on tariff agreements covering trade in goods, and offers an explanation for core features of the GATT. Tariffs play almost no role in services markets, however, and the existing models cannot account for the dramatically different approach to trade liberalization in agreements such as the WTO General Agreement on Trade in Services (GATS). We develop a model through which key features of GATS, including its emphasis on “deep integration” – sector-by-sector negotiations on behind the border policy instruments – can be understood. And we use this model to suggest that there may also be a middle ground for services trade liberalization between the GATS deep-integration approach and the traditional border-policy focused “shallow integration” approach of GATT.

*Staiger gratefully acknowledges financial support from the National Science Foundation (NSF Grant SES-1326940). We thank Kyle Bagwell, Petros C. Mavroidis, Alan Spearot and seminar participants at Stanford for useful comments.

1. Introduction

There is now an established literature on the economics of international trade agreements (see the relevant chapters in Bagwell and Staiger, eds, forthcoming, for recent comprehensive reviews). This literature has enjoyed success in illuminating many features of real-world trade agreements, but the insights have been developed almost entirely in the context of trade in goods. Given the historical unimportance of trade in services, until recently the literature's exclusive focus on trade in goods made sense. But the importance of trade in services has grown rapidly over the past several decades, bringing services to the top of the trade liberalization agenda; and as a consequence the need for the literature to consider trade-in-services agreements has become more pressing. In this paper we take a first step in filling this lacuna.

The World Trade Organization (WTO) includes agreements aimed at liberalizing both goods trade and trade in services. The General Agreement on Tariffs and Trade (GATT) is the central WTO agreement covering goods trade, while the General Agreement on Trade in Services (GATS) is the WTO agreement covering services trade. There are striking differences between GATT and GATS that cry out for explanation. Our paper catalogs these differences, and asks why they exist. Our answer builds from the terms-of-trade theory, a theory that can account for many of the core features of GATT (see, for example, Bagwell and Staiger, 2002). We show that the broad differences between GATT and GATS can be accounted for when this theory is augmented with a set of restrictions, motivated by salient features of services trade, on the policy instruments and trade data available to governments. This is the main positive message of our paper.

The GATT/WTO has been highly successful in facilitating the liberalization of goods trade among its member governments. GATS, on the other hand, has to date been largely unsuccessful in achieving liberalization of services trade (Francois and Hoekman, 2010). What explains this difference in success? A potential reason is the distinct nature of integration that each agreement has attempted. While both agreements seek to expand market access, GATT was designed with “shallow integration” in mind, achieving an expansion in market access primarily through negotiated reductions in tariffs; by contrast, the design of GATS reflects an orientation towards “deep integration,” whereby the negotiated change or removal of domestic regulations in the service sector is seen as the primary method of expanding market access. Clearly, the latter orientation raises significant challenges for negotiations seeking to expand market access

that do not arise with the former. Copeland and Mattoo (2008) put the point this way:

For goods trade, analysts typically distinguish between tariff and non-tariff barriers (NTBs) to trade. Tariffs are discriminatory taxes on trade. An import tax is a tax levied on foreign goods but not domestic goods. Tariffs tend to be easy to measure and are very transparent. It is therefore quite straightforward to design an agreement to liberalize trade via tariff reduction (provided governments have the will to open up their economies to trade). Successive rounds of GATT negotiations were successful in achieving broad-based reductions in trade barriers via across-the-board reductions in trade taxes.

NTBs, on the other hand, are much more complex...The pervasiveness of NTBs in the service sector means that trade liberalization in this sector is complex. Moreover, a major reason for the pervasiveness of NTBs is because of market imperfection in service sectors. Many trade barriers in the service sector are a side effect of domestic regulations that have legitimate purposes. For example, because of issues in asymmetric information, doctors must be certified to protect patients, engineers need certification to ensure that bridges they build do not collapse, and insurance companies have to be regulated to ensure their solvency. However, these same rules can be manipulated to protect local suppliers. For example, a rule requiring that an engineer graduate from a domestic university might ensure that quality standards are met, but would prevent consumers from having access to the services of highly qualified foreign engineers. The regulatory apparatus may therefore serve the dual purpose of responding to market failures and protecting local suppliers at the expense of consumers. A challenge for trade-policy analysis is to isolate the protective effect of regulatory policy from the beneficial effects, and to suggest rules for liberalization that provide the benefits of increased trade while ensuring that other legitimate policy objectives are achieved. (Copeland and Mattoo, 2008, p. 104)

Our augmented terms-of-trade model can help interpret the deep-integration focus of GATS, while at the same time clarifying the underlying problems that a trade-in-services agreement should be designed to solve. A clear understanding of the underlying problems to be solved can then inform the consideration of alternative design approaches for the agreement. We use

our augmented model to ask if a shallow-integration approach more in line with that taken by GATT might be possible in a trade-in-services agreement. Surprisingly, we find that a shallow-integration approach, suitably modified to fit the needs of the services trade context, may indeed be possible. This is the main normative message of our paper.

To provide some intuition up front for our results, it is helpful first to review the logic of shallow integration that emerges from the terms-of-trade theory when applied to agreements on trade in goods. That logic begins from the observation that a trade agreement must address an international inefficiency that exists under non-cooperative (Nash equilibrium) policy choices if the agreement is to generate mutual gains for the participants. The argument then proceeds by noting that terms-of-trade manipulation is the source of international inefficiencies in Nash policy choices; that trade taxes (import and export tariffs) are the first-best instruments for manipulating the terms of trade, and hence with Nash trade taxes unconstrained no other policies in the Nash equilibrium will be distorted for this purpose; and that therefore in the Nash equilibrium trade taxes are too high, making trade volumes too low, but there are no other policy distortions from an international perspective. From this starting point, it is then natural that a trade agreement might focus on lowering tariffs as a means of expanding market access and trade volumes to efficient levels, while putting in place various rules to prevent governments from back-sliding on the market access commitments implied by their negotiated tariff bindings with the substitution of new protective non-tariff (e.g., behind-the-border) measures. As we will describe more fully below, this logic fits nicely with the basic structure of GATT's shallow integration approach.

What can go wrong with this logic when it is applied to trade in services rather than trade in goods? An immediate issue arises for services whose importation necessitates the establishment by the foreign service provider of a commercial presence in the importing country, the kinds of services that are a central focus of GATS ("Mode 3" services in the terminology used by GATS) and the topic of our paper. The imports of such services do not cross international borders, and so import tariffs and export taxes/subsidies collected at the border are by definition simply unavailable to governments that might wish to intervene in the international trade of these services. There are other policy instruments that can in principle replicate the effects of trade taxes, and below we will exploit the existence of these alternative policies to derive benchmark results, but we argue that as a practical matter these alternatives too are unavailable to governments. This has two implications: first, without an import tariff to manipulate its

services terms of trade, an importing government will in the Nash equilibrium tend to spread protective distortions for this purpose widely across the policy instruments that it does wield in the service sector, thereby “contaminating” many of its Nash policies with internationally inefficient terms-of-trade motives; and second, without an export tax/subsidy, the exporting government must rely on policy adjustments from the importing government if it wishes to raise (in a targeted fashion) the incomes of its service exporters, introducing a new potential source of international inefficiency in the Nash equilibrium that an agreement can address. As we will demonstrate in a terms-of-trade model augmented to acknowledge the unavailability of such policies, these two implications can go a long way toward accounting for the broad differences between the structure of GATT and GATS. A related issue concerns the data on trade in (Mode 3) services that is typically available to governments. Because such trade does not cross international borders, the amount of detailed and accurate data on trade flows is limited. This has a further implication: the ability of governments to measure the trade volume impacts of policy changes in the service sector, an ability which as we explain below would be especially important for implementing any sort of shallow integration for services trade, is compromised. As we will argue, this too can help explain the broad differences between GATT and GATS.

Does this then mean that shallow integration is impossible for services? Not necessarily. Our augmented terms-of-trade model suggests that a shallow-integration approach may indeed be possible for services, but that to facilitate this approach to liberalization it would first be necessary to make some changes to the structure of GATS which move it closer to the structure of GATT. More specifically, we show that, in spite of the limited availability of policies in our augmented terms-of-trade model, if governments were to first agree to a set of blanket non-discrimination rules along the lines of those contained in GATT, they would then be induced to concentrate their internationally inefficient terms-of-trade motives into distorting a narrow set of fiscal – but not regulatory – measures. And with international policy inefficiencies concentrated in fiscal instruments, governments could then in principle use negotiations over these fiscal instruments as a means of establishing market access commitments in the service sector without the need to directly negotiate over domestic regulatory measures, much as GATT has used negotiated commitments on tariffs in the goods sector. Provided that the services data issues raised above can be addressed, we interpret this finding as pointing toward a possible way forward for achieving success in future efforts to liberalize trade in services.

The rest of the paper proceeds as follows. Section 2 reviews some stylized facts about

GATT and GATS, highlighting the substantial differences in their basic architecture. Section 3 introduces our benchmark model that assumes an expansive set of service-sector policy instruments available to governments, while results on the purpose and design of a trade-in-services agreement in our benchmark setting are contained in section 4. Section 5 then develops an augmented model under the assumption of a more limited set of policy options, shows that the broad differences between GATT and GATS can be understood from the perspective of this augmented model, and employs the augmented model to establish that a shallow-integration approach more in line with that taken by GATT might nevertheless be possible in a trade-in-services agreement. Section 6 concludes.

2. GATT versus GATS

The GATT was first negotiated in 1947. Its commitments were deepened and refined in a number of subsequent negotiating “rounds,” and its disciplines have been elaborated in several additional treaties within the WTO system. GATS is a more recent development that dates back only to 1994. It is thus no surprise that WTO goods market disciplines are more elaborate and “complete” than those in services sectors. Nevertheless, the legal structure of GATS looks dramatically different from both the early GATT and the more modern set of WTO disciplines in goods markets. This section highlights a number of key distinctions.

2.1. GATT

The GATT was a post-WWII response to the high tariff rates that had emerged globally following the Smoot-Hawley Tariff Act of 1930 in the United States. The central commitments are the tariff ceilings or “bindings” negotiated on a product-by-product basis pursuant to GATT Article II. Tariffs are not prohibited, and GATT members are under no general obligation to reduce tariffs, but they can choose to open their markets through reciprocal tariff reductions. Concurrently, GATT seeks to prevent the substitution of other forms of protection. The key obligations in this regard are the prohibition on quantitative restrictions in GATT Article XI, and the national treatment obligation of GATT Article III which, roughly speaking, prohibits domestic taxation and regulation that discriminates against imported goods. Additional disciplines limit the degree to which state monopolies and other state trading enterprises can engage in protectionist conduct. Discrimination among trading partners is targeted by the “most-favored nation” (MFN) obligation of GATT Article I. Finally, the drafters recognized

the incompleteness of the GATT contract by including a provision that allows members to seek redress if some change in domestic policy by an importing government, even though not specifically prohibited by GATT, nevertheless curtails trade in a manner that upsets the reasonable market access expectations associated with tariff commitments – the so-called “nonviolation doctrine.”

This basic approach of GATT may be termed “tariffication.” The effort was not simply to reduce tariffs, but to channel all remaining trade protection into tariffs by precluding the use of other policy instruments for protectionist objectives. Tariffication makes protection more transparent since tariffs are published and imposed at the border, and reduces the costs of trade negotiations by reducing the number of instruments in play (see Bagwell and Sykes, 2004). Tariffs also tend to minimize the deadweight costs of protection for any given local price in a protected market (see Schwartz and Sykes, 1996).

The basic approach of GATT may also be termed “shallow integration.” By shallow integration, we mean that detailed product-by-product negotiations are focused on a single border instrument – tariffs – and do not extend to behind-the-border measures (such as domestic taxes and regulations) that are unquestionably important for particular products or industries. Instead, behind-the-border measures are subject to across-the-board rules that apply in all goods markets, such as the national treatment obligation noted above. These across-the-board rules apply *whether or not* the goods in question are the subject of a tariff binding.

Goods market obligations have evolved considerably through the years, and now extend to several additional treaty instruments pertaining to policies such as the use of antidumping and countervailing duties, the use of government subsidies, and problems associated with regulatory or “technical barriers” to trade. Nevertheless, the general approach in goods markets remains overwhelmingly one of shallow integration.¹ For example, the WTO’s Agreement on Technical Barriers to Trade (TBT Agreement) requires that product regulations obey certain general principles, including an obligation to ensure that they are not more trade restrictive than necessary to achieve a legitimate regulatory objective. The WTO’s Agreement on Sanitary and Phytosanitary Measures (SPS Agreement) requires that such measures be based on a scientific risk assessment where possible. And the WTO’s Antidumping and Subsidies Agreements require that certain substantive and procedural rules be satisfied before a member imposes antidumping

¹An important exception arises under the WTO’s Agreement on Agriculture, pursuant to which product-by-product negotiations on subsidy levels have taken place.

or countervailing duties on any imported product.

2.2. GATS

The structure and approach of GATS is profoundly different. To aid in understanding the differences, it is useful to begin by setting out the “modes” of services trade and the attendant focus of GATS commitments within that framework.

GATS defines four modes of trade. “Mode 1” trade involves the cross border sale of a service from the exporting country to a consumer in the importing country – an American buys an insurance policy from a Swiss insurer, for example, receiving an insurance policy by mail or email while sending payment by mail or wire transfer. Such cross-border transactions are the closest analogue to ordinary trade in goods. The difference here lies in the fact that cross-border services transactions are often unobservable by importing governments because they do not pass through port facilities or come to the attention of customs inspectors. As a result, governments generally do little to tax or otherwise regulate Mode 1 trade.

“Mode 2” involves the consumption of a service in the exporting country by a national of another importing country – a tourist goes to France and gets a haircut. Governments make little effort to regulate Mode 2 trade.

To a great extent, therefore, the status quo ante for Mode 1 and Mode 2 trade was *de facto* free trade, and GATS has little effect on it. Although governments do make formal GATS commitments on Mode 1 and 2 trade, these commitments typically entail no more than a promise not to interfere with such trade, in effect preserving the free trade policy that prevailed for the most part before GATS.

“Mode 3” trade involves the establishment of a commercial presence in the importing nation by a foreign service provider – a foreign bank or law firm opens a domestic branch office, for example. Importing governments have much more capacity to restrict or regulate services trade when it involves the establishment of local offices within their territory, and such restrictions are commonplace.²

“Mode 4” trade occurs when a foreign supplier not only establishes a commercial presence in the importing nation, but also employs foreign nationals in its domestic offices. A foreign

²The fact that domestic branches of foreign service suppliers are typically staffed at least in part by domestic nationals may diminish the political demands for protection when services trade entails direct investment in a commercial presence (see generally Blanchard, 2007). Nevertheless, service providers seeking to establish a presence in a foreign market often confront formidable obstacles.

bank or law firm might wish to bring in senior personnel from its home country to oversee the domestic branch, for example. Mode 4 thus touches on immigration policy, where most nations also regulate extensively.

Accordingly, by far the bulk of the barriers to trade in service sectors prior to GATS involve restraints on Mode 3 or Mode 4. The focus of GATS negotiations, in turn, has been on Modes 3 and 4. Mode 4 commitments are largely limited to certain “horizontal” policies that allow the temporary presence of highly skilled and executive workers from abroad. Most of the sectoral commitments are for Mode 3. We therefore focus our discussion henceforth on Mode 3 services.

In contrast to GATT, however, tariffication is not an important dimension of GATS. As noted, conventional tariffs, imposed when a transaction involves cross-border exchange, play a negligible role in services markets. Indeed, with respect to Mode 3 (and Mode 4) trade, tariffs are unavailable by definition.

Moreover, unlike GATT, GATS does not try to channel trade protection into any particular policy instrument. There is no generally applicable analogue to the GATT Article XI prohibition on quantitative restrictions, or the GATT Article III national treatment obligation for domestic regulation and taxation. Instead, GATS members retain the ability to use any of these instruments for protective purposes. Nevertheless, they may *choose* to make commitments to limit the use of these instruments, much as they may choose to negotiate tariff bindings under GATT Article II.

Thus, with respect to each service sector under GATS,³ member governments elect whether to make any commitments or not. Absent commitments, their future policy choices are unconstrained beyond an MFN obligation. In sectors where commitments are made, members negotiate and memorialize their trade restrictive policies. These commitments are divided into schedules of market access restrictions and schedules of national treatment exceptions. And in the sectors where commitments are made, any restriction that is not memorialized in a member’s schedule is effectively “waived” and cannot be employed in the future.

The market access restrictions are embodied in each member’s GATS Article XVI schedules. Such restrictions include, among other things, limitations on the number of service providers in a sector, limitations on the value or quantity of transactions, ceilings on the number of persons that may be employed in a sector, restrictions on the type of legal entity that may

³Service sector classifications are part of the Central Product Classification (CPC) system, developed by the United Nations. The WTO created its own classification system based on the CPC, found in the Services Sector Classification List, MTN.GNS/W/120, July 10, 1991.

supply services, and limits on the percentage of foreign ownership in the sector as a whole or in individual service suppliers.

GATS Article XVII imposes a national treatment obligation *only for sectors where market access commitments are made*. The national treatment obligation is also subject to any enumerated exceptions that the importing member wishes to schedule. Thus, national treatment is also a subject of sector by sector negotiation.

In sum, GATS members retain complete freedom of action in sectors where they do not wish to make commitments, except for the MFN principle. In the sectors where commitments are made, the only fundamental obligation is transparency – members may still employ a wide array of quantitative restrictions, tax and regulatory policy instruments for protective purposes, but they must record those instruments in their schedules of commitments. Over time, Members may negotiate further changes in these schedules to enhance market access on a reciprocal basis.

Plainly, the approach of GATS goes beyond the predominantly shallow integration strategy of GATT. Generally applicable disciplines are few and, instead, members can and do negotiate over a variety of sector-specific behind-the-border measures, most commonly involving regulatory measures. In this respect, although still in its relative infancy, GATS may be characterized as a “deep integration” agreement by comparison to GATT.

3. A Benchmark Model of Trade in Services

In this section we construct a benchmark partial equilibrium terms-of-trade model of Mode 3 services trade between a domestic and a foreign country, recasting the model of Staiger and Sykes (2011) to apply to this context.⁴ We refer to this model as the Benchmark Model.

3.1. The Model

We assume for simplicity that the service under consideration is demanded only in the domestic country, and we represent demand by the linear demand curve $D = \alpha - P$ for $P \in [0, \alpha]$, with P the consumer price of this service in the domestic market. The service must be produced in the domestic country where it is consumed (mode 3), and to reflect the idea that there are widespread market imperfections in the service sector we assume that the consumption

⁴As we note earlier, the primary role for liberalization through GATS arises for Mode 3 and Mode 4 trade. We suppress the distinction between Modes 3 and 4 here, and simply model a market in which a foreign service supplier establishes a domestic commercial presence.

of this service generates a negative (“eye sore”) externality that does not effect production and that is not internalized by individual consumers (and hence does not impact demand for the product), but which detracts from aggregate national welfare in the domestic country (the externality does not cross borders). For ease of reference we will refer to this negative externality as a “pollutant” – for example, one could think of noise levels or dust levels associated with construction services – but it could correspond loosely to any of a variety of negative externalities that might be associated with the provision of services, such as specific health or safety risks imposed on the general public by sub-standard service suppliers in the construction, food or health services industries, or general risks to the domestic financial system associated with imprudent banking practices. We note also that as the service must be consumed where it is produced, it is immaterial for our purposes whether the externality arises from the process of production or from the act of consuming the service: again for ease of reference we assume it is consumption that generates the externality.

The domestic government can impose a regulatory standard s which specifies as a condition of entry into the domestic service market a (maximum) level of pollution $\theta(s)$ generated per unit of the service provided, and in principle a different standard may be applied to domestic and foreign service providers. We denote by r and ρ the particular standards imposed on domestic and foreign service providers respectively, with $\theta(r)$ and $\theta(\rho)$ the associated per-unit pollution levels generated by consumption of the services provided by domestic and foreign service providers under their respective standards. We assume that θ is a decreasing and convex function. A nondiscriminatory standard would satisfy $r = \rho$, while $\rho > r$ would indicate that the domestic government has imposed a discriminatory standard on foreign service providers in the domestic market. Many policy interventions in the service sector take the form of entry restrictions, some implicit (licensing or certification requirements for entry) and some explicit (numerical quotas on numbers of service suppliers that may enter), and our modeling of regulatory standards can be thought of as a shorthand for such policy interventions.

To meet the standard s , service providers must incur a per-unit compliance cost. These compliance costs – which could include both the cost of actually meeting the standard as well as the cost of establishing conformity with the standard – are not immutable, and can be reduced by government investment in the efficient design and implementation of a given standard. For example, to ensure a minimum quality standard for the provision of legal services, the government could simply require as a condition of entry that all providers of legal services

obtain a 3-year law degree from an accredited domestic law school. But the government could also with minimal additional effort provide a list of foreign law school degrees that would suffice, or with more substantial effort the government might alternatively invest in the design of a specific entrance exam that provides reliable information for the purpose of quality screening, or devise a system to monitor performance or a system to permit reliable self-monitoring of performance, any of which might reduce compliance costs for (domestic, or foreign, or both domestic and foreign) suppliers. Similar investments in the design and implementation of regulations applying to the financial or insurance sectors could reduce the costs of compliance with a particular standard for those services.

To capture this possibility, we assume that by investing $I \geq 0$ at a cost of $c \cdot I$ in the design and implementation of the standard s , the domestic government imposes the per-unit compliance cost $\phi(s, I) \equiv \kappa(s) - \lambda(I)$ on the service providers subject to the standard, where κ is increasing and convex in s and λ is increasing and concave in I with $\kappa(s) > \lambda(I) \geq 0$ for all s and I . We assume that separate investments must be made by the domestic government to reduce the costs for domestic and for foreign service providers in meeting their respective standards (even if the standards themselves are set at the same level), and we denote by I_d and I_f these investments: the per-unit cost of compliance for domestic service providers to meet the standard r is then given by $\phi(r, I_d)$, while the per-unit cost of compliance for foreign service providers to meet the standard ρ is given by $\phi(\rho, I_f)$. For any regulatory standards r and ρ , supply of domestic and foreign service providers are then given respectively by $S_d = q_d - \phi(r, I_d)$ for $q_d \geq \phi(r, I_d)$, and $S_f = q_f - \phi(\rho, I_f)$ for $q_f \geq \phi(\rho, I_f)$, where q_d and q_f are the prices received in the domestic market by the domestic and foreign service providers, respectively.⁵

Import tariffs and export taxes are central instruments of intervention for goods trade, but they are by definition not available in the context of Mode 3 services trade. Still, governments may have other fiscal instruments at their disposal. To reflect this possibility, we consider in this section a benchmark case in which an expansive list of non-tariff fiscal instruments is available. In particular, in addition to the regulatory standards, we assume that the domestic government

⁵Notice that we have assumed domestic and foreign service providers facing the same standard will generate the same per-unit level of pollution. And hence with our specification of the per-unit compliance cost $\phi(s, I)$, we have assumed as well that for fixed investment levels I_d and I_f the marginal impact on compliance cost of reducing the per-unit level of pollution with a slightly higher standard is the same across domestic and foreign service providers. However, domestic and foreign service providers will face different compliance costs for the same standard as long as the domestic government invests different amounts in reducing the costs of compliance for domestic and foreign service firms (i.e., if $I_d \neq I_f$). In this way our specification of $\phi(s, I)$ highlights the possibility that a given standard may impose very different costs on domestic and foreign service providers.

has at its disposal two distinct tax/subsidy instruments: a *nondiscriminatory* sales tax t that it levies uniformly on domestic and foreign service providers alike (tax if positive and subsidy if negative), and an additional *discriminatory* sales tax or surcharge t_f that it levies only on foreign service providers (tax if positive and subsidy if negative), each expressed in specific terms.⁶ Setting $t_f = 0$ would imply a nondiscriminatory tax (t) on the sales of domestic and foreign service suppliers in the domestic market, whereas $t_f > 0$ reflects a discriminatory tax (in the amount $t + t_f$) imposed by the domestic government on the domestic market sales of foreign service suppliers (discriminatory, because domestic service providers only pay the sales tax t on their domestic market sales). And for its part, we assume that the foreign government can levy a tax t_f^* on its service providers' sales in the domestic market (tax if positive and subsidy if negative, also expressed in specific terms).

Of course, a discriminatory sales tax imposed on the sales of foreign service providers in the domestic market is simply an import tariff (if imposed by the domestic government) or export tax (if imposed by the foreign government) by another name. And as with import tariffs and export taxes, there may be good reasons why governments do not or cannot use discriminatory sales taxes in the context of Mode 3 services trade. Consider for example the foreign government's tax t_f^* on the sales of foreign service providers in the domestic market: owing to the nature of Mode 3 services, the imposition of this sales tax would require the foreign government to monitor output and collect taxes in the jurisdiction of the domestic government, clearly something that is not likely to be observed in practice. But in contrast to the unavailability of import tariffs and export taxes in the context of Mode 3 services trade, the unavailability of such sales taxes does not follow by *definition*. And as will become clear below, assuming that these sorts of discriminatory sales taxes are available to governments provides an important benchmark with which to illuminate the essential issues associated with liberalizing Mode 3 trade in services, and from which we can then consider the implications of imposing more realistic limitations on the set of instruments available to governments in the context of Mode 3 service trade.

We assume that all taxes are set at non-prohibitive levels. With this, the price paid by domestic consumers and the prices received by domestic and foreign service providers must

⁶In light of the fact that any Mode 3 service that is consumed in the domestic market must also be produced there, production taxes are equivalent to sales/consumption taxes in our setting. Therefore, while for simplicity we refer to these taxes as sales taxes, thinking of them as production taxes would be equally valid and leave unchanged our analysis.

satisfy

$$q_d + t = P = q_f + t_f^* + t + t_f. \quad (3.1)$$

Note that services sell in the domestic country at the same price P regardless of the standard to which they are produced, reflecting the fact that individual consumers do not differentiate across units of the service on the basis of how much pollution it generates when it is consumed. We also define $q_w \equiv q_f + t_f^*$, the “world” price of the foreign service provided in the domestic market (i.e., the price outside the foreign market at which this service would be available for sale from foreign service providers). Note that (3.1) implies:

$$q_w = q_d - t_f. \quad (3.2)$$

The market-clearing condition $D = S_d + S_f$ determines equilibrium in this market. Using the explicit expressions for demands and supplies and the pricing relationships in (3.1)-(3.2), this condition determines the market-clearing world price for the service as a function of the tax and regulatory policies:

$$\tilde{q}_w = \frac{1}{3}[\alpha - 2t_f + t_f^* - t + \phi(r, I_d) + \phi(\rho, I_f)]. \quad (3.3)$$

With (3.1)-(3.2) we may also derive expressions for the market-clearing levels of each of the local prices in the domestic market as functions of the tax and regulatory policies:

$$\begin{aligned} \tilde{P} &= \frac{1}{3}[\alpha + (t_f + t_f^*) + 2t + \phi(r, I_d) + \phi(\rho, I_f)], \\ \tilde{q}_d &= \frac{1}{3}[\alpha + (t_f + t_f^*) - t + \phi(r, I_d) + \phi(\rho, I_f)], \text{ and} \\ \tilde{q}_f &= \frac{1}{3}[\alpha - 2(t_f + t_f^*) - t + \phi(r, I_d) + \phi(\rho, I_f)]. \end{aligned} \quad (3.4)$$

Following Staiger and Sykes (2011), we define as well the market-clearing foreign producer price of the “raw” *unregulated* service – prior to bringing it into compliance with the prevailing domestic regulatory standard – as a function of the tax and regulatory policies, and the associated world price of the foreign-produced unregulated service, by

$$\begin{aligned} \tilde{q}_f^0 &\equiv \tilde{q}_f - \phi(\rho, I_f) = \frac{1}{3}[\alpha - 2(t_f + t_f^*) - t + \phi(r, I_d) - 2\phi(\rho, I_f)], \text{ and} \\ \tilde{q}_w^0 &\equiv \tilde{q}_w - \phi(\rho, I_f) = \frac{1}{3}[\alpha - 2t_f + t_f^* - t + \phi(r, I_d) - 2\phi(\rho, I_f)]. \end{aligned} \quad (3.5)$$

As the second line of (3.5) indicates, for any ρ and I_f there is a one-to-one correspondence between \tilde{q}_w and \tilde{q}_w^0 , but in what follows it will prove useful to refer to \tilde{q}_w^0 rather than \tilde{q}_w as “the

terms of trade” in services. Notice that \tilde{q}_f^0 is also the market-clearing output of foreign service providers in the domestic market (S_f), and therefore the trade volume in Mode 3 services. As (3.5) indicates, while neither government can impose a tariff on this trade volume, each of the non-tariff policies that the governments do possess can be altered to reduce this trade volume and hence act as a non-tariff barrier (NTB) to trade in Mode 3 services.

We next define domestic and foreign welfare. The welfare level in the domestic country is calculated by subtracting from the usual partial equilibrium measure of consumer surplus plus producer surplus plus tax revenue the disutility of the consumption-generated pollution and the cost of investments in design and implementation of the standards. However, the producer surplus accruing to the domestic country is limited to that associated with domestic service suppliers: the producer surplus generated by foreign service suppliers in the domestic market accrues to the foreign country. Domestic consumer surplus (CS) and domestic producer surplus (PS) are given by

$$CS = \int_{\tilde{P}}^{\alpha} [\alpha - P] dP \equiv CS(\tilde{P}), \text{ and } PS = \int_{\phi(r, I_d)}^{\tilde{q}_d} [q - \phi(r, I_d)] dq \equiv PS(r, I_d, \tilde{q}_d).$$

With the pricing relationships above and the definition of \tilde{q}_w^0 , the tax revenue collected by the domestic government (TR) can be written as

$$\begin{aligned} TR &= [\tilde{P} - \tilde{q}_d] \cdot [\tilde{q}_d - \phi(r, I_d)] + [\tilde{P} - \tilde{q}_w^0 - \phi(\rho, I_f)] \cdot [(\alpha - \tilde{P}) - (\tilde{q}_d - \phi(r, I_d))] \\ &\equiv TR(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_w^0). \end{aligned}$$

And the utility cost of domestic pollution (Z) is given by

$$\begin{aligned} Z &= \theta(r) \cdot [\tilde{q}_d - \phi(r, I_d)] + \theta(\rho) \cdot [(\alpha - \tilde{P}) - (\tilde{q}_d - \phi(r, I_d))] \\ &\equiv Z(r, \rho, I_d, \tilde{P}, \tilde{q}_d). \end{aligned}$$

With these definitions, domestic welfare may now be expressed as

$$\begin{aligned} W &= CS(\tilde{P}) + PS(r, I_d, \tilde{q}_d) + TR(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_w^0) \\ &\quad - Z(r, \rho, I_d, \tilde{P}, \tilde{q}_d) - c[I_d + I_f] \\ &\equiv W(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_w^0). \end{aligned} \tag{3.6}$$

Notice that by the definition of $TR(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_w^0)$ and market clearing, it follows from (3.6) that $W_{\tilde{q}_d^w} = -[(\alpha - \tilde{P}) - (\tilde{q}_d - \phi(r, I_d))] = -\tilde{q}_f^0 < 0$ (where here and throughout a function

subscripted with a variable denotes the partial derivative of the function with respect to the variable). This reflects the domestic welfare loss associated with a terms-of-trade movement against the domestic country (i.e., a rise in \tilde{q}_w^0) holding fixed all regulatory standards, associated investments and domestic local prices. This loss is simply the income effect of the terms-of-trade deterioration for the domestic country, which amounts to the domestic sales volume of the foreign service providers.

Given the absence of foreign demand for the service under consideration and of foreign pollution, the welfare level in the foreign country is composed of just two components: the producer surplus accruing to foreign service providers operating in the domestic market, and tax revenue. However, while we have for simplicity (and without impact for the results we emphasize below) abstracted from political economy influences in characterizing the objectives of the domestic government, it will prove important for some of our later results to include the possibility of political economy influences in the foreign service sector when specifying the foreign government's objectives, which we do by placing a weight $\gamma^* \geq 1$ on foreign producer surplus in the foreign government objective function. More specifically, using the pricing relationships above and the definitions of \tilde{q}_f^0 and \tilde{q}_w^0 , foreign producer surplus (PS^*) and trade tax revenue (TR^*) can be defined as

$$\begin{aligned} PS^* &= \int_{\phi(\rho, I_f)}^{\tilde{q}_f^0 + \phi(\rho, I_f)} [q_f - \phi(\rho, I_f)] dq_f = \int_0^{\tilde{q}_f^0} q_f dq_f \equiv PS^*(\tilde{q}_f^0), \text{ and} \\ TR^* &= [\tilde{q}_w^0 - \tilde{q}_f^0] \cdot \tilde{q}_f^0 \equiv TR^*(\tilde{q}_f^0, \tilde{q}_w^0). \end{aligned}$$

With these definitions, foreign welfare may now be expressed as

$$\begin{aligned} W^* &= \gamma^* \cdot PS^*(\tilde{q}_f^0) + TR^*(\tilde{q}_f^0, \tilde{q}_w^0) \\ &\equiv W^*(\tilde{q}_f^0, \tilde{q}_w^0). \end{aligned} \tag{3.7}$$

Notice an implication of (3.7): as expressed by $W^*(\tilde{q}_f^0, \tilde{q}_w^0)$, foreign welfare does *not* depend directly on the standard ρ to which foreign service providers must comply when they sell in the domestic market, but only depends on ρ indirectly through the impact of ρ on \tilde{q}_f^0 and \tilde{q}_w^0 , the market-clearing producer price and world price of the foreign-produced unregulated service. Intuitively, and following Staiger and Sykes (2011), we have modeled production of the unregulated service as an increasing cost (upward-sloping supply) industry, while for a given standard level ρ and standard-writing investment I the per-unit cost of coming into compliance

with the standard is then constant (and equal to $\phi(\rho, I_f)$) regardless of how many units of the unregulated service must be altered to meet the standard. As a consequence, foreign producer surplus is impacted by the standard level ρ only to the extent that ρ impacts the market-clearing foreign supply decisions for the unregulated service (through \tilde{q}_f^0). Note as well that by (3.7) and the definition of $TR^*(\tilde{q}_f^0, \tilde{q}_w^0)$, it follows that $W_{\tilde{q}_w^0}^* = \tilde{q}_f^0 > 0$, reflecting the foreign welfare gain from a terms-of-trade improvement (i.e., arise in \tilde{q}_w^0) holding the foreign local price fixed. This gain is the income effect of the terms-of-trade improvement for the foreign country, which amounts to the domestic sales volume of the foreign service providers.

Finally, we develop an expression for the joint (sum of) domestic and foreign welfare. When we characterize efficient policies, we look for policy choices that maximize the sum of the welfare across the two countries.⁷ Using the market-clearing condition that the domestic demand for services ($\alpha - \tilde{P}$) must be satisfied by supply from domestic and foreign service providers $[(\tilde{q}_d - \phi(r, I_d)) + \tilde{q}_f^0]$, observe that the world price \tilde{q}_w^0 cancels from the sum of domestic and foreign tax revenue:

$$\begin{aligned} TR(r, \rho, \tilde{P}, I_d, I_f, \tilde{q}_d, \tilde{q}_w^0) + TR^*(\tilde{q}_f^0, \tilde{q}_w^0) &= [\tilde{P} - \tilde{q}_d] \cdot [\tilde{q}_d - \phi(r, I_d)] \\ &\quad + [\tilde{P} - \tilde{q}_f^0 - \phi(\rho, I_f)] \cdot [(\alpha - \tilde{P}) - (\tilde{q}_d - \phi(r, I_d))] \\ &\equiv g(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_f^0). \end{aligned}$$

This allows us to write

$$\begin{aligned} W + W^* &= W(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_w^0) + W^*(\tilde{q}_f^0, \tilde{q}_w^0) \\ &= CS(\tilde{P}) + PS(r, I_d, \tilde{q}_d) + \gamma^* \cdot PS^*(\tilde{q}_f^0) + g(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_f^0) \\ &\quad - Z(r, \rho, I_d, \tilde{P}, \tilde{q}_d) - c[I_d + I_f] \\ &\equiv G(r, \rho, I_d, I_f, \tilde{P}, \tilde{q}_d, \tilde{q}_f^0). \end{aligned} \tag{3.8}$$

Hence, while the world price \tilde{q}_w^0 enters into each country's welfare function, it is absent from the expression for joint welfare. This is because $W_{\tilde{q}_w^0}^* + W_{\tilde{q}_w^0} = \tilde{q}_f^0 - [(\alpha - \tilde{P}) - (\tilde{q}_d - \phi(r, I_d))] = 0$, so that movements in the world price represent pure (lump-sum) international transfers between countries.

⁷Implicitly, we are assuming that lump sum transfers are available to distribute surplus across the two countries as desired

3.2. Efficient and Noncooperative Policies

With our Benchmark Model described, we close this section by characterizing the jointly efficient and the noncooperative (Nash) policy choices. We begin with efficient policies.

To characterize efficient policies, recall that only local prices are relevant for joint welfare, as (3.8) indicates. But now observe from (3.3), (3.4) and (3.5) that, while world prices depend on the individual levels of both t_f and t_f^* , all local prices depend on t_f and t_f^* only through their sum. Therefore, in addition to the choices of t , r , ρ , I_d and I_f , efficiency ties down only the sum of t_f and t_f^* , not their individual levels.

With reference to the expression for joint welfare given in (3.8), there are then six first-order conditions that the efficient policy choices must satisfy.⁸ Evaluating these first-order conditions using the expressions in (3.3)-(3.7), denoting the efficient domestic-market sales of the domestic and foreign service providers respectively by S_d^E and S_f^E , and denoting the efficient policy choices by $t_f^E + t_f^{*E}$, t^E , r^E , ρ^E , I_d^E and I_f^E , the following conditions which must be satisfied by the efficient policy levels may be derived:

$$\begin{aligned} t_f^E + t_f^{*E} &= -(\gamma^* - 1) \cdot S_f^E, \\ t^E &= \theta(r^E), \\ \left[\left(-\frac{\partial \theta(r^E)}{\partial r} \right) - \frac{\partial \kappa(r^E)}{\partial r} \right] &= 0 \quad \text{and} \quad \left[\left(-\frac{\partial \theta(\rho^E)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^E)}{\partial \rho} \right] = 0, \\ \left[\frac{\partial \lambda(I_d^E)}{\partial I_d} \cdot S_d^E - c \right] &= 0 \quad \text{and} \quad \left[\frac{\partial \lambda(I_f^E)}{\partial I_f} \cdot S_f^E - c \right] = 0. \end{aligned} \tag{3.9}$$

The interpretation of (3.9) is intuitive. Notice first from the third line of (3.9) that efficiency requires that the standards for both domestic and foreign service providers are chosen to equate the marginal benefit of a slightly higher standard in terms of per-unit pollution reduction ($-\frac{\partial \theta(r^E)}{\partial r}$ and $-\frac{\partial \theta(\rho^E)}{\partial \rho}$) with the increase in marginal cost of service production from the slightly higher standard ($\frac{\partial \kappa(r^E)}{\partial r}$ and $\frac{\partial \kappa(\rho^E)}{\partial \rho}$), and hence efficiency implies $r^E = \rho^E$: given the symmetric technologies across domestic and foreign service suppliers in terms of both the externality that their services generate for a given standard ($\theta(s)$) and the sensitivity of the cost of compliance to changes in the standard ($\kappa(s)$), there is no efficiency reason to impose discriminatory standards across domestic and foreign service providers. And with the efficient standard $r^E = \rho^E$ in

⁸We assume throughout that policy choices correspond to interior solutions of the relevant maximization problems. The second-order conditions associated with the maximization problems considered here and throughout the paper are satisfied under our convexity/concavity assumptions for θ , κ and λ .

place, the efficient nondiscriminatory domestic sales tax (t^E) is simply the Pigouvian tax that internalizes the remaining externality imposed by the domestic and foreign service providers in the domestic market, as the second line of (3.9) indicates. Now consider the first line of (3.9): in the case where the foreign government is subjected to political economy influences ($\gamma^* > 1$), it is then also efficient to offset somewhat the sales tax imposed on foreign service providers in the domestic market with discriminatory tax rebates ($t_f^E + t_f^{*E} < 0$), reflecting the extra value that the foreign government places on redistributing surplus towards foreign service providers. And as the bottom line of (3.9) shows, it is efficient to make compliance-cost-reducing investments in the design and implementation of each standard to the point where the marginal benefit (the reduction in the total costs incurred by the impacted service providers in meeting the standard, $\frac{\partial \lambda(I_d^E)}{\partial I_d} \cdot S_d^E$ and $\frac{\partial \lambda(I_f^E)}{\partial I_f} \cdot S_f^E$) is equal to the marginal cost of such investments (c), implying that in general $I_d^E \neq I_f^E$.⁹

Notice that $I_d^E \gtrless I_f^E$ as $S_d^E \gtrless S_f^E$ for the intuitive reason that it is worth investing more in reducing the cost of compliance when the cost savings applies to a larger sales volume. For the case where $\gamma^* = 1$, it is straightforward to show that $S_d^E = S_f^E$ and hence $I_d^E = I_f^E$, but as noted above for $\gamma^* > 1$ it is efficient to lower the sales tax imposed on foreign service providers in the domestic market with the discriminatory tax rebate ($t_f^E + t_f^{*E} < 0$), and this stimulates the domestic sales of foreign service providers, leading to $S_d^E < S_f^E$ and therefore $I_d^E < I_f^E$.¹⁰

Next we characterize the noncooperative (Nash) policy choices. Facing a given foreign tax t_f^* on the sales of foreign service providers in the domestic country and using the domestic welfare expression in (3.6), the best-response policy choices of the domestic government are the choices

⁹As discussed in note 5, the possibility that the same standard imposes different compliance costs on domestic and foreign service providers is captured in our model by the level of domestic-provider-specific and foreign-provider-specific investments in compliance-cost reduction that the domestic government selects. For this reason, as (3.9) indicates, the efficient investment levels will typically differ across domestic and foreign service providers but the efficient standard level will not. Together these two features imply that at efficient policy choices the domestic and foreign service providers will face the same standard and generate the same level of per-unit pollution as we have observed, but may face different costs of compliance with the standard. As we discuss further below, this distinction in turn allows us to distinguish between the nondiscrimination obligations imposed by the WTO's NT rule on regulatory standards, and the obligation imposed by the WTO's TBT rule to implement regulations in a way that is no more trade restrictive than necessary.

¹⁰In our benchmark model, efficiency implies a greater domestic market share for foreign service providers than for domestic service providers whenever foreign political pressure is present. At the cost of extra notation, we could easily introduce into the model exogenous supply-curve shifters that delivered the arguably more realistic prediction that foreign service providers supply a relatively small share of the domestic market, with the implication that efficiency would then require greater investment in compliance-cost reduction for domestic service suppliers than for foreign service suppliers.

of t , t_f , r , ρ , I_d and I_f that satisfy the following six first-order conditions:

$$\begin{aligned}
W_{\tilde{P}} \frac{d\tilde{P}}{dt_f} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dt_f} + W_{\tilde{q}_w^0} \frac{d\tilde{q}_w^0}{dt_f} &= 0, \\
W_{\tilde{P}} \frac{d\tilde{P}}{dt} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dt} + W_{\tilde{q}_w^0} \frac{d\tilde{q}_w^0}{dt} &= 0, \\
W_r + W_{\tilde{P}} \frac{d\tilde{P}}{dr} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dr} + W_{\tilde{q}_w^0} \frac{d\tilde{q}_w^0}{dr} &= 0, \\
W_\rho + W_{\tilde{P}} \frac{d\tilde{P}}{d\rho} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{d\rho} + W_{\tilde{q}_w^0} \frac{d\tilde{q}_w^0}{d\rho} &= 0, \\
W_{I_d} + W_{\tilde{P}} \frac{d\tilde{P}}{dI_d} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dI_d} + W_{\tilde{q}_w^0} \frac{d\tilde{q}_w^0}{dI_d} &= 0 \text{ and} \\
W_{I_f} + W_{\tilde{P}} \frac{d\tilde{P}}{dI_f} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dI_f} + W_{\tilde{q}_w^0} \frac{d\tilde{q}_w^0}{dI_f} &= 0.
\end{aligned} \tag{3.10}$$

And facing given domestic choices of t , t_f , r , ρ , I_d and I_f and using the foreign welfare expression (3.7), the best-response foreign tax t_f^* must satisfy the following first-order condition:

$$W_{\tilde{q}_f^0}^* \frac{d\tilde{q}_f^0}{dt_f^*} + W_{\tilde{q}_w^0}^* \frac{d\tilde{q}_w^0}{dt_f^*} = 0. \tag{3.11}$$

Nash policy choices simultaneously satisfy the conditions in (3.10) and (3.11), ensuring that each country is adopting its best-response policy to the other country's policy choices.

Evaluating the first-order conditions contained in (3.10) and (3.11) that define the Nash policies using the expressions in (3.3)-(3.7), denoting the Nash domestic-market sales of the domestic and foreign service providers respectively by S_d^N and S_f^N and the Nash policy choices by t_f^N , t_f^{*N} , t^N , r^N , ρ^N , I_d^N and I_f^N , and reporting the sum $t_f^N + t_f^{*N}$ rather than t_f^N and t_f^{*N} separately in order to facilitate comparison with the efficient policies in (3.9), it follows that the Nash policy levels satisfy

$$\begin{aligned}
t_f^N + t_f^{*N} &= -(\gamma^* - 1) \cdot S_f^N + \frac{3}{2} S_f^N, \\
t^N &= \theta(r^N), \\
\left[\left(-\frac{\partial \theta(r^N)}{\partial r} \right) - \frac{\partial \kappa(r^N)}{\partial r} \right] &= 0 \text{ and } \left[\left(-\frac{\partial \theta(\rho^N)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^N)}{\partial \rho} \right] = 0, \\
\left[\frac{\partial \lambda(I_d^N)}{\partial I_d} \cdot S_d^N - c \right] &= 0 \text{ and } \left[\frac{\partial \lambda(I_f^N)}{\partial I_f} \cdot S_f^N - c \right] = 0.
\end{aligned} \tag{3.12}$$

With Nash and efficient policy choices characterized for our Benchmark Model of trade in services, we are ready to consider the purpose and design of trade-in-services agreements.

4. Trade-in-Services Agreements: Benchmark Results

In this section we use our Benchmark Model to investigate the purpose and design of a trade-in-services agreement. We show that the Benchmark Model fails to offer an explanation for the striking differences between GATT and GATS that we described above. This negative result is instructive, however, as it both points the way to an augmented model (which we develop in section 5) that can account for the differences between GATT and GATS and at the same time suggests changes in the existing GATS architecture (which we discuss in section 5) that might facilitate greater liberalizing success for services through an analog to the shallow integration approach that proved successful for GATT.

4.1. The Purpose of a Trade-in-Services Agreement

What problem must an agreement on trade in services solve if it is to move governments from inefficient Nash choices to the efficiency frontier? The answer is important, because it reveals the purpose of a trade-in-services agreement. Armed with the characterization of efficient policies in (3.9) and Nash policies in (3.12), we now turn to a comparison of these policies as a means of answering this question.

A first observation is immediate from a comparison of (3.9) with (3.12): the inefficiencies associated with Nash choices are not spread widely across the range of government interventions in the service sector, but are instead limited to a narrow range of policies. Specifically, conditional on the trade volume, the Nash choices of t , r , ρ , I_d and I_f described in (3.12) satisfy the same conditions as do the efficient choices of these policies described in (3.9), and it is only $t_f + t_f^*$, the sum of the discriminatory domestic and foreign sales taxes levied on foreign service providers in the domestic market, that under Nash choices is excessively protective and results in diminished market access and trade volume relative to the efficient choice, owing to the extra term $\frac{3}{2}S_f^N$ in the expression for $t_f^N + t_f^{*N}$.¹¹

¹¹That the sum of the Nash discriminatory taxes is higher than efficient follows from our focus on non-prohibitive intervention, which ensures that the Nash supply of foreign services to the domestic market S_f^N is strictly positive. We use the term “market access” informally here in order to tie in with the language present in trade agreements (e.g., GATT and GATS), but see Bagwell and Staiger (2002) for a formalization of this term within the context of the terms-of-trade theory. In essence, a market access commitment in a trade agreement is a commitment to conditions of competition between foreign exporting and domestic import-competing firms, and can be interpreted in the context of our Benchmark Model as a commitment to a given level of sales volume by foreign service providers in the domestic market (i.e., a given level of domestic import volume) at a given foreign-service-provider price.

A second observation then also follows: the difference between Nash and efficient sales taxes on foreign service providers is driven by each country's incentive to exert its market power over the terms of trade (\tilde{q}_w^0) with its unilateral tax choice and to impose its Johnson (1953-54) "optimal tariff" on the sales of foreign service providers in the domestic market. To see this, consider the expression for t_f^N , which it can be shown is given by $t_f^N = \theta(\rho^N) + S_f^N$, and note that the elasticity of foreign service supply into the domestic market can be written as $\frac{\partial S_f}{\partial \tilde{q}_w} \frac{\tilde{q}_w}{S_f} = \frac{\tilde{q}_w}{S_f}$. But then, dividing the expression for t_f^N by \tilde{q}_w to convert from a specific to an ad valorem tax yields $\frac{t_f^N}{\tilde{q}_w} = \frac{\theta(\rho^N)}{\tilde{q}_w} + \frac{S_f^N}{\tilde{q}_w}$, and hence the second term in this expression is simply the inverse of the foreign export supply elasticity, the Johnson (1953-54) optimal ad-valorem tariff term. A similar calculation can be performed for the Nash foreign tax, which is given by $t_f^{*N} = -(\gamma^* - 1) \cdot S_f^N + \frac{1}{2} S_f^N$, leading to an analogous interpretation.

Collecting these two observations, we may conclude that the difference between Nash and efficient service sector policies can be traced to each country's incentive to exert its market power over the terms of trade. To confirm this at a more general level and develop further its implications, we now follow Bagwell and Staiger (1999, 2001) and define *politically optimal* policies as those policies that would hypothetically be chosen by governments unilaterally if they did not value the terms-of-trade implications of their policy choices. We then ask whether politically optimal policies so defined are efficient when evaluated in light of the governments' actual objectives, and thereby consider whether the Nash inefficiencies identified above can in fact be given the terms-of-trade interpretation we have just outlined.

Formally, to characterize politically optimal policies we suppose hypothetically that the domestic government acts as if $W_{\tilde{q}_w^0} \equiv 0$ when choosing unilaterally its policies, while the foreign government acts as if $W_{\tilde{q}_w^0}^* \equiv 0$ when choosing unilaterally its policy, and we solve for the non-cooperative (Nash) policy choices that would emerge with these hypothetical objectives. It is easy to show that the politically optimal policies, which we denote by t_f^{*PO} , t_f^{PO} , t^{PO} , r^{PO} , ρ^{PO} , I_d^{PO} and I_f^{PO} , are efficient when evaluated in light of the governments' actual objectives, and take an intuitive form of the efficient policies described by (3.9). At the political optimum, the foreign government departs from laissez-faire and subsidizes its service providers' sales in the domestic market if and only if it is motivated by political economy considerations; and the domestic government sets efficient and nondiscriminatory standards ($r^{PO} = \rho^{PO}$) for domestic and foreign service providers in the domestic market while making efficient compliance-cost-reducing investments in the design and implementation of these standards, and imposes a

Pigouvian nondiscriminatory sales tax to internalize the remaining externality. Formally,

$$\begin{aligned}
t_f^{PO} &= 0 \quad \text{and} \quad t_f^{*PO} = -(\gamma^* - 1) \cdot S_f^{PO}, \\
t^{PO} &= \theta(r^{PO}), \\
\left[\left(-\frac{\partial \theta(r^{PO})}{\partial r} \right) - \frac{\partial \kappa(r^{PO})}{\partial r} \right] &= 0 \quad \text{and} \quad \left[\left(-\frac{\partial \theta(\rho^{PO})}{\partial \rho} \right) - \frac{\partial \kappa(\rho^{PO})}{\partial \rho} \right] = 0, \quad \text{and} \\
\left[\frac{\partial \lambda(I_d^{PO})}{\partial I_d} \cdot S_d^{PO} - c \right] &= 0 \quad \text{and} \quad \left[\frac{\partial \lambda(I_f^{PO})}{\partial I_f} \cdot S_f^{PO} - c \right] = 0,
\end{aligned} \tag{4.1}$$

where S_d^{PO} and S_f^{PO} denote the (efficient) politically optimal domestic-market sales of the domestic and foreign service providers respectively. Comparing (4.1) with (3.9), it is clear that politically optimal policies satisfy the conditions for efficiency.

Just as with trade in goods, therefore, if governments could be induced to make service sector policy choices free from motives reflecting terms-of-trade manipulation, there would be nothing left for a trade-in-services agreement to do. We thus arrive at an important conclusion: as long as governments can levy discriminatory sales taxes on foreign service providers in the domestic market as we have assumed in our Benchmark Model, the scope of the Nash policy inefficiencies in the service sector will be limited to these tariff-equivalent sales tax instruments; and the fundamental problem for a trade-in-services agreement to address is then to prevent terms-of-trade manipulation, and to thereby reduce the discriminatory sales taxes on foreign service providers and increase market access and service trade volumes without introducing distortions into the choice of domestic regulatory and tax policies. This is the same problem that arises with goods trade according to the terms-of-trade theory (see for example Bagwell and Staiger, 2001, and Staiger and Sykes, 2011), with the only difference being that discriminatory taxes on the sales of foreign service providers play the role of import tariffs and export taxes.

Evidently, while it is true that import tariffs and export taxes are unavailable in the context of (Mode 3) services trade as we have assumed, this does not *necessarily* imply that the fundamental nature of Nash policy inefficiencies in services trade is any different from the nature of Nash policy inefficiencies in the context of goods trade. We summarize with:

Proposition 1. *In the Nash equilibrium of the Benchmark Model, the domestic and foreign governments exert their power over the terms of services trade with discriminatory sales taxes imposed on foreign service providers in the domestic market that restrict market access and service trade volumes below efficient levels; governments leave all other policy instruments*

undistorted. The purpose of a trade-in-services agreement is then to remove the terms-of-trade driven distortions from discriminatory sales taxes and raise trade volumes.

4.2. A ‘GATT-like’ Agreement for Trade in Services

We now show that, according to our Benchmark Model, with only minor adjustments it should have been possible to apply the broad structure of GATT to the liberalization of trade in services. This result both sets the stage for our later normative results and frames the positive question that we later address: How can our benchmark terms-of-trade model be modified to account for the striking differences between GATT and GATS?

We begin with a basic point: while Proposition 1 indicates that the purpose of a trade-in-services agreement is to remove the terms-of-trade driven distortions from discriminatory sales taxes and raise trade volumes, this does not mean that an agreement over $t_f + t_f^*$ alone, without restrictions on other policies, could lead governments to the international efficiency frontier.

To see this, let us suppose for example that the domestic and foreign governments attempt to implement the politically optimal policies defined in (4.1) with an agreement to bind their discriminatory sales taxes at the politically optimal levels $t_f^{PO} = 0$ and $t_f^{*PO} = -(\gamma^* - 1) \cdot S_f^{PO}$, leaving all other (domestic) policy instruments unconstrained. Denoting best-response choices with a superscript “ R ,” and letting S_d^R and S_f^R denote the domestic market sales of domestic and foreign service providers, respectively, under the politically optimal discriminatory sales taxes and the best-response levels for the remaining policies of the domestic government, it is straightforward to show that t^R , r^R , ρ^R , I_d^R and I_f^R satisfy the following conditions:

$$\begin{aligned}
 t^R &= \frac{1}{2} \cdot [\theta(r^R) + \theta(\rho^R)] + \frac{1}{2} S_f^R, \\
 \left[\left(-\frac{\partial \theta(r^R)}{\partial r} \right) - \frac{\partial \kappa(r^R)}{\partial r} \right] &= \frac{1}{2 S_d^R} \cdot [S_f^R - (\theta(r^R) - \theta(\rho^R))] \cdot \frac{\partial \kappa(r^R)}{\partial r}, \\
 \left[\left(-\frac{\partial \theta(\rho^R)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^R)}{\partial \rho} \right] &= -\frac{1}{2 S_f^R} \cdot [S_f^R - (\theta(r^R) - \theta(\rho^R))] \cdot \frac{\partial \kappa(\rho^R)}{\partial \rho}, \\
 \left[\frac{\partial \lambda(I_d^R)}{\partial I_d} \cdot S_d^R - c \right] &= -\frac{1}{2} \cdot [S_f^R - (\theta(r^R) - \theta(\rho^R))] \cdot \frac{\partial \lambda(I_d^R)}{\partial I_d}, \text{ and} \\
 \left[\frac{\partial \lambda(I_f^R)}{\partial I_f} \cdot S_f^R - c \right] &= \frac{1}{2} \cdot [S_f^R - (\theta(r^R) - \theta(\rho^R))] \cdot \frac{\partial \lambda(I_f^R)}{\partial I_f}.
 \end{aligned} \tag{4.2}$$

Comparing (4.1) and (4.2), it is clear that an agreement that binds discriminatory sales taxes

at their politically optimal levels but leaves all other policy choices unconstrained will lead the domestic government to (i) impose a discriminatory standard ($\rho^R > r^R$) against foreign service providers in the domestic market, with a higher-than-efficient standard imposed on foreign service providers and a lower-than-efficient standard imposed on domestic service providers, (ii) make smaller-than-efficient compliance-cost-reducing investments in the design and implementation of the standard applied to foreign service providers (I_f^R) and higher-than-efficient compliance-cost-reducing investments in the design and implementation of the standard applied to domestic service providers (I_d^R), and (iii) set the domestic sales tax above the Pigouvian level.

How, then, can a trade-in-services agreement be structured to reduce sales taxes on foreign service providers and increase market access and service trade volumes from their Nash levels without introducing distortions into the choice of domestic regulatory and tax policies? One possible approach is to structure the agreement so as to place constraints directly on each policy at its efficient level, a deep form of integration that entails negotiations over all trade-impacting policies. To be effective, deep integration requires that governments are able to accurately distinguish in their negotiations between internationally efficient and inefficient levels of regulatory standards (and the same goes for domestic taxes). Here we consider the possibility of adopting an alternative, shallow integration, approach, analogous to the approach taken by GATT in the context of goods trade liberalization.

To this end, suppose that governments were to agree to some across-the-board rules that apply whether or not the services in question are subject to negotiated market access commitments. We focus on three such rules, analogs of which figure prominently in GATT.

Consider first the impact of governments adopting a *national treatment* (NT) rule which prohibits regulation – but *not* taxation – that discriminates against the sales of foreign service providers in the domestic market, implying the restriction $r \geq \rho$. In the presence of such a rule, if the two governments then agree to bind their discriminatory sales taxes at the politically optimal levels $t_f^{PO} = 0$ and $t_f^{*PO} = -(\gamma^* - 1) \cdot S_f^{PO}$, it is straightforward to show that t^R , r^R , ρ^R , I_d^R and I_f^R now satisfy the following conditions:

$$\begin{aligned}
t^R &= \theta(r^R) + \frac{1}{2}S_f^R, \\
\left[\left(-\frac{\partial \theta(r^R)}{\partial r} \right) - \frac{\partial \kappa(r^R)}{\partial r} \right] &= 0 \quad \text{and} \quad \left[\left(-\frac{\partial \theta(\rho^R)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^R)}{\partial \rho} \right] = 0, \\
\left[\frac{\partial \lambda(I_d^R)}{\partial I_d} \cdot S_d^R - c \right] &= -\frac{1}{2}S_f^R \cdot \frac{\partial \lambda(I_d^R)}{\partial I_d}, \quad \text{and} \quad \left[\frac{\partial \lambda(I_f^R)}{\partial I_f} \cdot S_f^R - c \right] = \frac{1}{2}S_f^R \cdot \frac{\partial \lambda(I_f^R)}{\partial I_f}.
\end{aligned} \tag{4.3}$$

Evidently, as a comparison of the second and third lines of (4.2) with the second line of (4.3) makes clear, by eliminating the possibility of setting discriminatory standards against foreign service providers, the NT rule works to ensure that the standards will in fact remain at their Nash, efficient, levels (as defined by the third line of (3.12)) when the negotiated constraints on discriminatory sales taxes are imposed, even though the choice of the level of the (nondiscriminatory) standards is left to the discretion of the domestic government.¹² Intuitively, when discriminatory sales taxes are bound by agreement and therefore unavailable for terms-of-trade manipulation and nondiscriminatory sales taxes become the only fiscal policy available for this purpose, the use of *discriminatory* regulatory standards becomes an attractive weapon to add to the terms-of-trade manipulation arsenal; but if the application of discriminatory standards is also unavailable as under the NT rule, the use of standards for this purpose loses its luster, and standards become immune to terms-of-trade manipulation motives.

It might be thought that the ability of the NT rule to mitigate against terms-of-trade motives in the unilateral setting of standards hinges on a sufficiently dominant presence of domestic service suppliers who, in the presence of the NT rule, must face any terms-of-trade driven distortions in standards that foreign service providers face. But in fact, as the second line of (4.3) indicates, the market share of domestic service providers plays no role in this argument: in the presence of the NT rule, the non-discriminatory sales tax simply dominates non-discriminatory standards for purposes of terms-of-trade manipulation, independent of the market share of domestic service providers.

Consider next the impact of governments adopting as well in their trade-in-services agreement a rule akin to aspects of the *Technical Barriers to Trade* (TBT) agreement, under which the governments are obligated to ensure that, whatever non-protectionist objectives they choose to pursue with their choice of standards, they do so with regulations that are no more trade restrictive than necessary to achieve these objectives.¹³ In terms of our Benchmark Model, we can impose and interpret a “TBT rule” as committing the domestic government to make efficient compliance-cost-reducing investments in the design and implementation of the standards that it chooses to put in place (that is, efficient choices for I_d and I_f).¹⁴ With the domestic gov-

¹²Staiger and Sykes (2011) make the analogous point in the context of goods trade.

¹³Under the TBT Agreement, governments have broad latitude to pursue non-protectionist objectives relating to health, safety, the environment, consumer protection, and the like. But they are subject to an array of across-the-board rules designed to ensure transparency, that the means chosen to achieve these objectives are reasonable, and that the rules do not impose unnecessary costs on trade.

¹⁴Nothing in the TBT Agreement creates an obligation in precisely these terms, but in broad brush this

ernment's choices of I_d and I_f then satisfying the conditions for efficiency under the TBT rule, it is straightforward to show that the NT rule continues to ensure that the domestic government's standards choices will remain at their Nash, efficient, levels when the two governments agree to bind their discriminatory sales taxes at the politically optimal levels, and only the nondiscriminatory domestic sales tax remains inefficiently high.

That is, if the trade-in-services agreement were to include an NT rule and a TBT rule as we have described these rules here, and if the domestic and foreign government were to agree to bind their discriminatory sales taxes at the politically optimal levels $t_f^{PO} = 0$ and $t_f^{*PO} = -(\gamma^* - 1) \cdot S_f^{PO}$, the remaining policies t^R, r^R, ρ^R, I_d^R and I_f^R would satisfy the following conditions:

$$\begin{aligned} t^R &= \theta(r^R) + \frac{1}{2}S_f^R, \\ \left[\left(-\frac{\partial \theta(r^R)}{\partial r} \right) - \frac{\partial \kappa(r^R)}{\partial r} \right] &= 0 \quad \text{and} \quad \left[\left(-\frac{\partial \theta(\rho^R)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^R)}{\partial \rho} \right] = 0, \\ \left[\frac{\partial \lambda(I_d^R)}{\partial I_d} \cdot S_d^R - c \right] &= 0, \quad \text{and} \quad \left[\frac{\partial \lambda(I_f^R)}{\partial I_f} \cdot S_f^R - c \right] = 0. \end{aligned} \tag{4.4}$$

Evidently, according to (4.4), if governments were to bind their discriminatory sales taxes at the politically optimal levels in the presence of NT and TBT rules, the spread of distortions in service sector policies would be limited to the remaining “commercial measures”/fiscal instruments of governments (t^R); there would be no induced distortions in regulatory choices (r, ρ, I_d and I_f).

Finally, suppose that in addition to the NT and TBT rules, a *non-violation* (NV) doctrine is adopted, under which a service-exporting government could seek redress if some change in domestic policy by an importing government, even though not specifically prohibited by the trade-in-services agreement, nevertheless curtails trade in a manner that upsets the reasonable market access expectations associated with discriminatory-sales-tax commitments. If we use the phrase “market access” to denote the domestic import volume at a given terms of trade (see also note 11), then we can think of the NV rule as implying that, once the domestic government

characterization captures the essence of what important aspects of the agreement seek to achieve. It seeks to ensure that regulatory standards are cost-justified and are “necessary” to the attainment of some legitimate regulatory objective, both as to the substance of regulation and the measures used to ensure conformity. And the Appellate Body has interpreted the “necessary” requirement (in a case involving GATT Article XX) to preclude an enforcement system that shifts most of the enforcement costs to foreigners (see *Korea – Measures Affecting Imports of Fresh, Chilled and Frozen Beef*, WT/DS161/AB/R, adopted January 10, 2001).

makes a market access commitment by binding its discriminatory sales tax, it will be dissuaded from making any subsequent changes to its full set of policies that together would have the effect of reducing the volume of service imports it demands at a given terms of trade.

Here we follow Staiger and Sykes (2011) and formalize the NV doctrine as a “market-access preservation” rule defined in terms of the *raw* (unregulated) service.¹⁵ A key observation is that, if the NV rule prevents the domestic government from making unilateral post-agreement changes in its policies in a way that would alter its demand for imported raw services at the terms of trade implied by its negotiated market access commitments, then the market-clearing level of imports of raw services \tilde{q}_f^0 and the terms of trade \tilde{q}_w^0 cannot be altered by any post-agreement changes in domestic policies allowable under the NV rule either. And without the ability to manipulate the terms of trade with its remaining (unconstrained) policy instruments, the incentive for the domestic government to introduce distortions in these policy instruments once its discriminatory sales tax is constrained in a market access agreement is removed.

Collecting these points, suppose, then, that the domestic and foreign governments negotiate a *shallow integration* services agreement, in which they agree to abide by the NT, TBT and NV rules, and where, beginning from the Nash policies defined in (3.12), the domestic and foreign governments then negotiate a market access agreement with the following features: (i) the foreign government agrees to set its sales tax on foreign service providers in the domestic market at the politically optimal level $t_f^{*PO} = -(\gamma^* - 1) \cdot S_f^{PO}$; and (ii) the domestic government agrees to set its sales tax on foreign service providers in the domestic market at a level \bar{t}_f defined by $\tilde{q}_f^0(\bar{t}_f + \bar{t}_f^{*PO}, t_d^N, r^N, \rho^N, I_d^N, I_f^N) = S_f^{PO}$, where using (3.5) we now express \tilde{q}_f^0 as a function of domestic and foreign policies. With these market access commitments implying a volume of imported raw services \tilde{q}_f^0 and a terms of services trade \tilde{q}_w^0 which are then fixed at their politically optimal level under the NV rule, it follows that, subsequent to their market access agreement, the domestic government will be induced to implement the politically optimal choices for its policies as defined in (4.1).¹⁶ In this way, according to our Benchmark Model

¹⁵More specifically, as emphasized by Staiger and Sykes (2011) in the context of goods trade and shared in the trade-in-services setting we analyze here, an important feature of the market-access preservation rule we describe in the text is that, to deliver its desirable impact, it must be defined with respect to the world price of the unregulated service \tilde{q}_0^w rather than with respect to \tilde{q}^w , the world price of the service actually being imported by the domestic country under its chosen standard. This is because a rule that prevents the domestic country from altering the foreign exporter price of the unregulated service as it considers various levels of domestic taxes and regulatory standards ensures that the domestic country pays the full price of higher standards and taxes and – as it also enjoys the full benefits – therefore makes internationally efficient policy choices.

¹⁶This follows by noting that the first-order conditions defining the politically optimal policies in (4.1), which

a ‘GATT-like’ shallow integration approach to services trade liberalization could in principle be used by governments to negotiate from inefficient Nash policies to the (efficient) political optimum.

Moreover, taking the perspective of the Nash equilibrium described in (3.12), from which according to the Benchmark Model governments would have considered the design of a trade-in-services agreement, it is natural that these governments might be drawn to the possibility of designing an agreement that borrowed heavily from the broad features of GATT, and that therefore focuses on lowering discriminatory sales taxes as a means of expanding market access and trade volumes in the service sector to efficient levels, while putting in place various rules to prevent governments from back-sliding on the market access commitments implied by their negotiated bindings with the substitution of new protective behind-the-border measures.

We summarize with:

Proposition 2. *According to the Benchmark Model, a ‘GATT-like’ shallow integration approach to services trade liberalization, which relies on across-the-board NT, TBT and NV rules combined with market access negotiations to bind the levels of discriminatory sales taxes on foreign service providers in the domestic market, could in principle be used by governments to negotiate from inefficient Nash policies to the efficiency frontier.*

Notice how the three rules – NT, TBT and NV – work in tandem in the Benchmark Model to facilitate a shallow integration approach to liberalization. With NT addressing the incentives to distort regulatory standards choices for purposes of terms-of-trade manipulation that arise once market access commitments are made, and with TBT addressing the potential distortion of compliance-cost-reducing investments in the design and implementation of these standards, a comparison of (3.12) and (4.4) confirms that the NV rule can then be aimed primarily at the more narrow task of preventing these market access commitments from being frustrated by the introduction of new “commercial measures”/fiscal instruments (t) as a second-best means of terms-of-trade manipulation.¹⁷

for the domestic government assume that it acts as if $W_{\tilde{q}_w^0} \equiv 0$ when making its politically optimal choices and amount to the first-order conditions in (3.10) with the last term on the left-hand side of each first-order condition removed, will be the same first order conditions that the domestic government faces when instead, as a result of the NV rule and evaluated at politically optimal trade volumes, it cannot alter \tilde{q}_w^0 .

¹⁷A secondary task of the NV rule would be to prevent terms-of-trade motivated distortions from spreading to regulatory choices once the NV rule had blocked the path of introducing new commercial measures for this purpose. See Staiger and Sykes (2013) on the observed/on-equilibrium use of non-violation claims in

According to Propositions 1 and 2, our Benchmark Model provides no reason to expect that either the purpose or the design of a trade-in-services agreement would differ in any substantial way from an agreement to liberalize goods trade, suggesting in turn that GATS should look much like GATT. And yet, as we have detailed in section 2, the structure of these two agreements is strikingly different, with GATT following a shallow integration approach much as we have described here and GATS instead adopting a deep integration approach that bears little resemblance to GATT. In the next section we describe an augmented terms-of-trade model of services trade that can account for these differences, and then revisit the possibility of shallow integration from the perspective of this augmented model.

5. The Implications of Limited Service-Sector Policy Options

Thus far we have assumed that an expansive set of policy options is available to the domestic and foreign governments for intervention in Mode 3 service trade. This set includes discriminatory sales taxes that both the domestic and the foreign government may impose on foreign service providers in the domestic country. However, as we discuss below, there are good reasons to believe that such taxes may be unavailable to governments in practice. We have also assumed that governments have access to the data they would need to implement an effective NV rule. This requires the ability to detect when a non-contracted policy measure is used in a way that denies market access, which ultimately comes down to being able to measure changes in trade volumes and prices with reasonable accuracy; but it is far from obvious that the necessary data on services trade are currently available. To reflect these and other possible limitations that were omitted from our Benchmark Model, we now consider the implications of imposing further constraints on the set of service-sector policy options available to governments.

5.1. The Discriminatory Domestic Sales Tax Instrument

While import tariffs are by definition not available in the context of Mode 3 services trade, we have derived our benchmark results under the assumption that the domestic government has access to a tariff-equivalent policy: a discriminatory tax on the sales of foreign service providers in the domestic market. In principal there is no reason that governments could not use these

GATT/WTO practice to address primarily the introduction of such commercial measures, and see Staiger and Sykes (forthcoming) for an analysis of the possible off-equilibrium importance of the non-violation doctrine with regard to regulatory measures as well.

kinds of tax instruments, and indeed in the context of goods trade GATT's Article III national treatment obligation explicitly rules out the use of such taxes, suggesting that in the context of goods trade governments might regularly resort to these kinds of taxes in the absence of an obligation not to do so. And yet, in the context of services such taxes are rarely if ever observed.

This is not to say that the sale of services cannot be or is not taxed. In the United States, for example, Hendrix and Zodrow (2003) describe the wide variation across states in the contemporary pattern of services taxation generally:

“Public utilities offer a good example of the variation in the tax treatment of services across states. Many states tax most utility services, including those provided to businesses, and all states tax at least some. ... Roughly three-quarters of the states tax local telephone services, with about half extending the sales tax to long-distance services. Cable television is taxed in roughly half of the states.”

“Sales taxation of other services varies to an even greater degree. Almost all states tax rentals of tangible personal property to some extent...reflecting the popularity of taxes that may be exported to nonresidents, and most states tax admissions to amusements. However, repairs of tangible personal property are taxable in only about half of the states, and only roughly a quarter of states tax repairs of real property. About half of the states tax custom programmed computer programs (although virtually all tax packaged software). Relatively few states tax professional services, especially medical services, which are exempt largely on grounds that they are a “merit good” that should remain untaxed. ...” (Hendrix and Zodrow, 2003, p 413 [footnotes omitted])

More generally, the broad-based feasibility of the taxation of services, including the exemption of select services from taxation that might arguably reflect Pigouvian “merit good” consideration, is well-illustrated by a proposal to broaden the taxation of services in the state of Florida.¹⁸ As Hendrix and Zodrow (2003) observe:

¹⁸By assumption, the Pigouvian considerations associated with the Mode 3 service in our model take the form of negative externalities, and hence call for Pigouvian taxes. But nothing in our formal analysis would change if, for the particular service under consideration, we allowed the externality to become positive under efficient standards, in which case $\theta(r^E)$ would be negative, the efficiently regulated service would exhibit features similar to a “merit good,” and the efficient sales tax would become a subsidy. To the extent that the sales tax reflects

...If the current proposal for state taxation of services in Florida were enacted, the state would be among the states with the most comprehensive coverage of services. The current proposal—like the law enacted in 1987—would extend the sales tax to a broad range of services. Services that would be taxed include advertising, legal services, accounting services, transportation and construction services, automobile repair, recreational services, personal grooming services, and investment counseling. In contrast to the situation in most states under which services are subject to taxation only if they are specifically enumerated as taxable, the Florida proposal takes the more comprehensive approach of specifying that services are taxable unless specifically exempted. Exempt services include medical and health services, educational and social services, communication services, and services that employees provide for their employers. (Hendrix and Zodrow, 2003, pp. 413-414).

The Florida proposal is comprehensive in its approach, suggesting the feasibility of services taxes generally. And interestingly, Hendrix and Zodrow do note in the first-quoted passage above that some service taxes are borne heavily by non-residents, making them an especially popular form of taxation. But none of these examples fit the description of *discriminatory* taxes against foreign service providers.¹⁹

These observations suggest that as a practical matter, the taxation of services is broadly feasible when applied to service providers on a nondiscriminatory basis, but sales taxes that discriminate against foreign service providers are probably best thought of as unavailable to the importing government (perhaps for reasons of high transaction costs).²⁰ Formally, we introduce

terms-of-trade motives, these motives would continue to push for a higher tax or lower subsidy than efficient levels, and an efficient agreement would then lower the tax or raise the subsidy to the efficient Pigouvian level (see also note 21).

¹⁹It is not uncommon to observe discriminatory quotas applied to foreign service providers, and one might invoke the logic of “tariff-quota equivalence” to argue that such quotas serve as the analog of discriminatory sales taxes. But these are typically quotas that limit the number of foreign service providers that can operate in a market, not the quantity of output that foreign service providers can produce, and so tariff-quota equivalence arguments would typically not apply.

²⁰There are a variety of reasons why the transaction costs associated with sales taxes that discriminate against foreign service providers may be high relative to the transaction costs associated with non-discriminatory sales taxes. For example, ownership structures can be altered after entry, and so in order to collect sales taxes at different rates from domestic and foreign service providers, the reported nationality of ownership would have to be certified on an ongoing basis, an administrative task that could be especially costly in light of the small scale of many service providers. More generally, discriminatory sales taxes would create additional opportunities for tax-avoidance and give rise to the associated costs of preventing such behavior.

this policy constraint into the Benchmark Model with the assumption that

$$t_f \equiv 0. \quad (\text{Assumption 1})$$

Under Assumption 1, the domestic government still has a rich set of service-sector policy instruments that include a nondiscriminatory sales tax, separate and potentially discriminatory regulations applied to domestic and foreign service providers, and distinct levels of compliance-cost-reducing investment in the implementation and design of each standard; but it no longer has access to a tariff-equivalent policy instrument for (Mode 3) service-sector intervention.

Recalling from (3.9) that only the sum of the domestic and foreign discriminatory taxes (but not their individual levels) is pinned down by the conditions for efficiency, it is clear that Assumption 1 does not alter the conditions that efficient policies must meet. And recalling that under politically optimal policies t_f is not utilized in any event ($t_f^{PO} = 0$), it is clear that the political optimum defined in (4.1) is still efficient under Assumption 1, hence the purpose of a trade-in-services agreement is unchanged under Assumption 1.

But now, without its tariff-equivalent policy instrument, the domestic government will be forced to rely on its other policies as second-best means to manipulate the terms of trade, and the Nash policies under Assumption 1 are defined by

$$\begin{aligned} t_f^{*N} &= -(\gamma^* - 1) \cdot S_f^N + \frac{1}{2} S_f^N, \\ t^N &= \frac{1}{2} \cdot [\theta(r^N) + \theta(\rho^N)] + \frac{1}{2} S_f^N, \\ \left[\left(-\frac{\partial \theta(r^N)}{\partial r} \right) - \frac{\partial \kappa(r^N)}{\partial r} \right] &= \frac{1}{2 S_d^N} \cdot [S_f^N - (\theta(r^N) - \theta(\rho^N))] \cdot \frac{\partial \kappa(r^N)}{\partial r}, \\ \left[\left(-\frac{\partial \theta(\rho^N)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^N)}{\partial \rho} \right] &= -\frac{1}{2 S_f^N} \cdot [S_f^N - (\theta(r^N) - \theta(\rho^N))] \cdot \frac{\partial \kappa(\rho^N)}{\partial \rho}, \\ \left[\frac{\partial \lambda(I_d^N)}{\partial I_d} \cdot S_d^N - c \right] &= -\frac{1}{2} \cdot [S_f^N - (\theta(r^N) - \theta(\rho^N))] \cdot \frac{\partial \lambda(I_d^N)}{\partial I_d}, \text{ and} \\ \left[\frac{\partial \lambda(I_f^N)}{\partial I_f} \cdot S_f^N - c \right] &= \frac{1}{2} \cdot [S_f^N - (\theta(r^N) - \theta(\rho^N))] \cdot \frac{\partial \lambda(I_f^N)}{\partial I_f}. \end{aligned} \quad (5.1)$$

Comparing (3.9) and (5.1), it is clear that operating under the policy constraint implied by Assumption 1, the domestic government will be led in the Nash equilibrium to impose a higher-than-efficient standard on foreign service providers and a lower-than-efficient standard on domestic service providers, to make smaller-than-efficient compliance-cost-reducing investments in the design and implementation of the standard applied to foreign service providers and

higher-than-efficient compliance-cost-reducing investments in the design and implementation of the standard applied to domestic service providers, and to set the domestic sales tax above the Pigouvian level.²¹ In short, under Assumption 1 the domestic government’s motives to manipulate the terms of trade will spread inefficiency to all of its Nash policies.

Two important points, one positive and one normative, follow from the Nash equilibrium conditions under Assumption 1 as reflected in (5.1). First, as a positive matter, if (5.1) describes the starting point from which governments would have considered the design of a trade-in-services agreement, the strategy of borrowing heavily from the broad features of GATT – and therefore focusing market access negotiations on a single policy instrument while putting in place various rules to prevent governments from back-sliding on the market access commitments implied by their negotiated bindings with the substitution of new protective behind-the-border measures – no longer seems like an obvious and natural, or even viable, way to proceed, certainly not to the same degree that would have been suggested if governments were starting from the Nash equilibrium conditions of the Benchmark Model in (3.12). Rather, with (5.1) as their starting point and facing evident behind-the-border policy distortions spread throughout the domestic service market, a decision to adopt a deep-integration approach to services liberalization seems almost inevitable. Hence, according to our Benchmark Model augmented with Assumption 1, the lack of an effective tariff or tariff-equivalent policy instrument for (Mode 3) service-sector intervention could go a long way in accounting for the striking differences in the architecture of GATS and GATT.

A second point that follows from (5.1) is normative: even when the Benchmark Model is augmented with Assumption 1, a modified shallow-integration approach to service-sector liberalization could still work. In particular, suppose that governments agreed to the across-the-board NT, TBT and NV rules described above. And with some abuse of notation, let us now use the superscript “ N ” to denote the “non-cooperative” policy choices under these across-the-board rules but prior to any negotiated market access commitments. Then under Assumption 1 and subsequent to the introduction of the NT, TBT and NV rules but prior to any negotiated market access commitments, it is straightforward to derive that the non-cooperative

²¹As described in note 18, if the service under consideration were assumed to exhibit features of a “merit good” under efficient regulation, then its consumption would carry with it a positive externality so that $\theta(r^E) < 0$. In this case, $[\theta(r^N) + \theta(\rho^N)]$ would also be negative provided r^N and ρ^N are not too far from r^E , and hence the expression for t^N in (5.1) could be positive, negative or zero depending on the strength of the motives for terms-of-trade manipulation captured by the term $\frac{1}{2}S_f^N$. Nevertheless, as we state in the text the term $\frac{1}{2}S_f^N$ ensures that t^N will be set above the Pigouvian level.

policy choices will be described as follows:

$$\begin{aligned}
t_f^{*N} &= -(\gamma^* - 1) \cdot S_f^N + \frac{1}{2}S_f^N, \\
t^N &= \theta(r^N) + \frac{1}{2}S_f^N, \\
\left[\left(-\frac{\partial \theta(r^N)}{\partial r} \right) - \frac{\partial \kappa(r^N)}{\partial r} \right] &= 0 \quad \text{and} \quad \left[\left(-\frac{\partial \theta(\rho^N)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^N)}{\partial \rho} \right] = 0, \\
\left[\frac{\partial \lambda(I_d^N)}{\partial I_d} \cdot S_d^N - c \right] &= 0 \quad \text{and} \quad \left[\frac{\partial \lambda(I_f^N)}{\partial I_f} \cdot S_f^N - c \right] = 0.
\end{aligned} \tag{5.2}$$

Similarly to our discussion around (4.3) and (4.4) and despite the lack of a discriminatory sales tax instrument implied by Assumption 1, the NT rule ensures that the non-cooperative regulatory standards will be set efficiently and the TBT rule ensures that the non-cooperative compliance-cost-reducing investments in the design and implementation of the standards will be set at efficient levels as well, and the inefficiencies in the non-cooperative service-sector policy choices will be limited to higher-than-efficient levels for the tax instruments t_f^{*N} and t^N . From here, and proceeding along similar lines to our earlier description of shallow integration, market access negotiations that bind t_f^{*N} and t^N at levels that would imply an efficient volume of services trade could, in combination with the NV rule to prevent back-sliding on these market access commitments with the substitution of new protective regulatory standards (see also note 17), allow governments to achieve the efficiency frontier.

We summarize with:

Proposition 3. *According to the Benchmark Model augmented by Assumption 1, when the domestic government lacks a discriminatory sales tax and therefore a tariff-equivalent policy for use in the service sector, it exerts its power over the terms of services trade in the Nash equilibrium by distorting all of its (behind-the-border) policies; the purpose of a trade-in-services agreement is to remove the terms-of-trade driven distortions from all of the domestic and foreign policies and raise trade volumes, and a deep-integration approach therefore seems natural. Nevertheless, a ‘GATT-like’ shallow integration approach to services trade liberalization, which relies on across-the-board NT, TBT and NV rules combined with market access negotiations to bind the levels of taxation in the domestic service market, could in principle be used by governments to negotiate from inefficient Nash policies to the efficiency frontier.*

5.2. The Foreign Tax/Subsidy Instrument

In deriving our benchmark results we have assumed that the foreign government can levy a tax t_f^* on its service providers' sales in the domestic market (tax if positive and subsidy if negative), and we have observed that this is simply an export tax/tariff by another name. And in fact, that the foreign government has access to this tariff-equivalent policy is required for the derivation of our benchmark results: no other policy instrument (e.g., a foreign profits tax, or a subsidy to foreign investment) will do.²² But in reality there *is* a distinction between an export tax and a policy such as t_f^* , because unlike an export tax which is collected at the foreign border, a policy such as t_f^* is available to the foreign government only if it can administer a program of sales tax/subsidies to its service firms within the jurisdiction of the domestic government. In practice, arrangements of this kind are rarely if ever observed, presumably for reasons relating to a combination of high transaction costs and national sovereignty issues. So it is important to consider how our benchmark results would be altered if the foreign government did not have access to this policy instrument.

Formally, to consider the implications of the policy constraint that intervention with t_f^* is not feasible for the foreign government, we introduce into our Benchmark Model the following assumption:

$$t_f^* \equiv 0. \quad (\text{Assumption 2})$$

Observe that under Assumption 2 the foreign government is now passive in the domestic service market under consideration. This is not crucial for the results we emphasize in our discussion below – all that is required is that the foreign government does not have access to a policy described by t_f^* – but it does make the discussion very transparent.

For simplicity, we will impose Assumption 2 but relax our prior Assumption 1. Relaxing Assumption 1 is convenient, because the efficiency frontier can then still be reached in the presence of Assumption 2, owing to the fact that, as we have emphasized, (3.9) implies that only the sum of the domestic and foreign discriminatory taxes (but not their individual levels) is pinned down by the conditions for efficiency. In particular, under Assumption 2 the efficient

²²More specifically, our benchmark results require that the foreign government has access to a tariff equivalent policy as long as the foreign government would utilize this policy instrument in the political optimum, which will be the case provided that (i) $\gamma^* > 1$, and (ii) the foreign government does not possess a lump-sum instrument with which to redistribute toward its service providers located in the domestic market.

policies are given by

$$\begin{aligned}
t_f^E &= -(\gamma^* - 1) \cdot S_f^E, \\
t^E &= \theta(r^E), \\
\left[\left(-\frac{\partial \theta(r^E)}{\partial r} \right) - \frac{\partial \kappa(r^E)}{\partial r} \right] &= 0 \quad \text{and} \quad \left[\left(-\frac{\partial \theta(\rho^E)}{\partial \rho} \right) - \frac{\partial \kappa(\rho^E)}{\partial \rho} \right] = 0, \\
\left[\frac{\partial \lambda(I_d^E)}{\partial I_d} \cdot S_d^E - c \right] &= 0 \quad \text{and} \quad \left[\frac{\partial \lambda(I_f^E)}{\partial I_f} \cdot S_f^E - c \right] = 0.
\end{aligned} \tag{5.3}$$

We can now make an initial observation: in the presence of Assumption 2, and as long as the foreign government objectives include political economy motivations, the problem for a trade-in-services agreement to solve can no longer be given a terms-of-trade interpretation. To establish this, we need only demonstrate that the politically optimal policies are not efficient under Assumption 2. Note that, as only the domestic government is policy active under Assumption 2, the politically optimal policies are now defined by supposing hypothetically that the domestic government acts as if $W_{\tilde{q}_w^0} \equiv 0$ when choosing unilaterally its policies, taking $t_f^* \equiv 0$ as given. This yields the following expressions for politically optimal policies under Assumption 2:

$$\begin{aligned}
t_f^{PO} &= 0, \\
t^{PO} &= \theta(r^{PO}), \\
\left[\left(-\frac{\partial \theta(r^{PO})}{\partial r} \right) - \frac{\partial \kappa(r^{PO})}{\partial r} \right] &= 0 \quad \text{and} \quad \left[\left(-\frac{\partial \theta(\rho^{PO})}{\partial \rho} \right) - \frac{\partial \kappa(\rho^{PO})}{\partial \rho} \right] = 0, \\
\left[\frac{\partial \lambda(I_d^{PO})}{\partial I_d} \cdot S_d^{PO} - c \right] &= 0 \quad \text{and} \quad \left[\frac{\partial \lambda(I_f^{PO})}{\partial I_f} \cdot S_f^{PO} - c \right] = 0.
\end{aligned} \tag{5.4}$$

Comparing (5.3) and (5.4), it is immediately apparent that the policy limitation embodied in Assumption 2 interferes with the efficiency of politically optimal policies unless $\gamma^* = 1$ (in which case the politically optimal policy for the foreign government would be laissez faire in any event). With $\gamma^* > 1$ it follows that $t_f^{PO} > t_f^E$: even with terms-of-trade motives eliminated from the unilateral policy choices of the domestic government, the domestic sales tax on foreign service providers is too high relative to the efficient level, because the desire of the foreign government to subsidize its foreign service providers in the domestic economy has not been taken into account in the politically optimal choice of t_f (and by Assumption 2 the foreign government has no sales tax/subsidy instrument of its own with which to satisfy this desire). Hence, the inefficiency associated with Nash policy choices cannot be given a terms-of-trade interpretation under the policy constraint embodied in Assumption 2.

It is well known in the context of the terms-of-trade theory that placing constraints on the trade policy instruments available to governments can change the problem for a trade agreement to solve (see Ossa, 2011, and the discussion in Bagwell and Staiger, 2012, 2015 and forthcoming). The new point we emphasize here is that, in the context of Mode 3 services trade, there are especially good (e.g., jurisdictional) reasons to expect that governments will not possess complete sets of trade tax/subsidy instruments, and hence good reasons to expect that the problem for a trade agreement to solve will diverge from the terms-of-trade problem that arguably dominates in the context of goods trade.

Two further points now also follow. A first point is that the critical role of government market power over the terms of trade, essential for identifying a problem for a trade-in-services agreement to solve in our benchmark results, disappears under Assumption 1: that is, under Assumption 1 and provided $\gamma^* > 1$, there would be a reason for an agreement to liberalize trade in Mode 3 services even if all countries were small.

To see this first point, let us consider a small-country thought experiment, in which the domestic “country” is actually a domestic region composed of a continuum of identical domestic countries and likewise that the foreign “country” is a foreign region composed of a continuum of identical foreign countries, but everything else in the model is unchanged. Focusing on the domestic region (the only policy active region under Assumption 2), it then follows that no single country in the domestic region can impact the terms of trade \tilde{q}_w^0 with its individual policy choices, and so the Nash conditions for a representative domestic government are

$$\begin{aligned}
W_{\tilde{P}} \frac{d\tilde{P}}{dt_f} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dt_f} &= 0, \\
W_{\tilde{P}} \frac{d\tilde{P}}{dt} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dt} &= 0, \\
W_r + W_{\tilde{P}} \frac{d\tilde{P}}{dr} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dr} &= 0, \\
W_\rho + W_{\tilde{P}} \frac{d\tilde{P}}{d\rho} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{d\rho} &= 0, \\
W_{I_d} + W_{\tilde{P}} \frac{d\tilde{P}}{dI_d} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dI_d} &= 0 \text{ and} \\
W_{I_f} + W_{\tilde{P}} \frac{d\tilde{P}}{dI_f} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dI_f} &= 0.
\end{aligned} \tag{5.5}$$

And in this setting, the conditions for efficient policy intervention for a representative domestic

government are

$$\begin{aligned}
W_{\tilde{P}} \frac{d\tilde{P}}{dt_f} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dt_f} + W_{\tilde{q}_f}^* \frac{d\tilde{q}_f^0}{dt_f} &= 0, \\
W_{\tilde{P}} \frac{d\tilde{P}}{dt} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dt} + W_{\tilde{q}_f}^* \frac{d\tilde{q}_f^0}{dt} &= 0, \\
W_r + W_{\tilde{P}} \frac{d\tilde{P}}{dr} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dr} + W_{\tilde{q}_f}^* \frac{d\tilde{q}_f^0}{dr} &= 0, \\
W_\rho + W_{\tilde{P}} \frac{d\tilde{P}}{d\rho} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{d\rho} + W_{\tilde{q}_f}^* \frac{d\tilde{q}_f^0}{d\rho} &= 0, \\
W_{I_d} + W_{\tilde{P}} \frac{d\tilde{P}}{dI_d} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dI_d} + W_{\tilde{q}_f}^* \frac{d\tilde{q}_f^0}{dI_d} &= 0, \text{ and} \\
W_{I_f} + W_{\tilde{P}} \frac{d\tilde{P}}{dI_f} + W_{\tilde{q}_d} \frac{d\tilde{q}_d}{dI_f} + W_{\tilde{q}_f}^* \frac{d\tilde{q}_f^0}{dI_f} &= 0.
\end{aligned} \tag{5.6}$$

Clearly, the Nash choices in (5.5) will satisfy the conditions for efficiency in (5.6) if and only if $W_{\tilde{q}_f}^* = 0$, so that the last term on the left-hand-side of each condition in (5.6) is zero. But it is direct to show that $W_{\tilde{q}_f}^* = (\gamma^* - 1) \cdot S_f^N$ when this derivative is evaluated under Assumption 2, and hence the Nash choices will be efficient if and only if $\gamma^* = 1$.

Evidently, under the policy constraint described in Assumption 2 and provided $\gamma^* > 1$, the purpose of a trade-in-services agreement is not simply to eradicate the exercise of market power from the service-sector policy choices of governments, and even small countries who have no ability to alter their terms of services trade may have reason to engage in liberalization under the agreement. This point may take on special importance in the context of services trade, where policies toward service providers are often determined and implemented at the local/state government level. When obstacles to trade arise at the subsidiary government level, the market power wielded by the relevant government decision makers over the terms of trade of foreign service providers may be quite limited.²³ According to our Benchmark Model, this would then limit the amount of liberalization in services needed to bring countries to the international efficiency frontier, at least as compared to goods trade liberalization, for the simple reason

²³ A related point is that Mode 3 foreign service providers employ mostly domestic nationals, and this may also limit the ability of the domestic government to effectively use its power over the terms of trade to shift the costs of its policy intervention onto foreigners. In particular this could be the case if, contrary to the assumptions of our Benchmark Model, domestic workers developed firm-specific human capital from which they earned rents. In this case, the domestic nationals employed by a Mode 3 foreign service provider in the domestic market would share in the surplus earned by the foreign firm, and this would diminish the effectiveness with which the domestic government could use its power over the terms of trade to take foreign surplus and thereby shift the costs of its policy intervention onto foreigners.

that the (state and local) governments who set policies for trade in services are less likely to wield market power over world prices than are the (national) governments who set trade policy in goods markets.²⁴ But when the Benchmark Model is augmented with Assumption 2, this conclusion no longer follows. And as a result, a number of the more specific design features associated with the way market access bargaining is often organized in GATT (e.g., request-offer bilateral bargaining protocols organized along the lines of the principal supplier rule) may not work well in the context of a trade-in-services agreement.²⁵

A second point also follows from the inefficiency of the political optimum under Assumption 2: even outright prohibitions of service trade in the Nash equilibrium may be subject to liberalization under a trade-in-services agreement. This follows because the Nash equilibrium and the political optimum coincide whenever Nash policy choices prohibit trade (reflecting the fact that, as we have observed, $W_{\tilde{q}_0^w} = -\tilde{q}_f^0$ and $W_{\tilde{q}_w^*}^* = \tilde{q}_f^0$, and that prohibitive Nash policies imply $\tilde{q}_f^0 = S_f^N = 0$). When the political optimum is efficient, as in our Benchmark Model and as with the terms-of-trade theory when applied to goods trade, the Nash equilibrium will then also be efficient whenever Nash policy choices prohibit trade, and there is nothing for a trade agreement to do. But under Assumption 2, the political optimum is no longer efficient, as we have shown, and so the Nash equilibrium will then *also* be inefficient whenever Nash policy choices prohibit trade; and hence even outright prohibitions on foreign service providers may be subject to liberalization in a trade-in-services agreement. This point takes on special significance in the context of services, where as we have noted governments are often beginning their market access negotiations with prohibitions in place.

Finally, and importantly, augmenting the Benchmark Model with Assumption 2 does *not* weaken the case for shallow integration of services trade that emerges from the Benchmark Model. The only implication of Assumption 2 for our earlier discussion of shallow integration in the context of the Benchmark Model is that the particular level of market access commitments and implied import volume needed to reach the efficiency frontier would be altered. The rest of the earlier discussion would remain in tact.

²⁴There is a growing body of evidence in the context of goods trade that market power over foreign exporter (“world”) prices is fairly wide spread across countries and products, and evidence as well that the patterns of GATT/WTO market access liberalization in goods reflect the reduction or elimination of tariffs embodying these motives, as the terms-of-trade theory predicts (see Bagwell, Bown and Staiger, forthcoming, for a recent review of this literature).

²⁵The GATS bargaining protocol adopted for the now-suspended Doha Round was modeled on the request-offer bilateral bargaining practice used in many of the GATT rounds, but unlike in GATT, the principal supplier rule played no explicit role in the GATS bargaining protocol (see Feketekuty, 2008).

We summarize with:

Proposition 4. *According to the Benchmark Model augmented by Assumption 2, when the foreign government lacks a discriminatory sales tax and therefore a tariff-equivalent policy for use in the domestic service sector, and as long as the foreign government objectives include political economy motivations, the problem for a trade-in-services agreement to solve can no longer be given a terms-of-trade interpretation; as a consequence, even small countries may have reason to engage in liberalization under a trade-in-services agreement, and even outright prohibitions may be the subject of liberalizing market access negotiations. The case for shallow integration of services trade remains the same as in the Benchmark Model.*

5.3. Other Limitations on Policy Options

We now briefly consider the impact of imposing several other limitations on the policy options assumed in our Benchmark Model.

The Nondiscriminatory Domestic Sales Tax Instrument Our Benchmark Model assumes that the domestic government has access to a nondiscriminatory sales tax t for use in the domestic service industry under consideration. We have discussed evidence on the availability of such taxes, and have argued that it is reasonable to assume that governments do indeed have such policy options available to them.

Here we raise a different question: Can sales taxes be as finely tuned to individual service industries as the regulatory standards r and ρ ? Our Benchmark Model assumes that they can, a feature that is important for our finding that an NT rule channels terms-of-trade manipulation motives into the setting of the nondiscriminatory sales tax and ensures that the non-cooperative regulatory standards will be set efficiently. And it is not obvious that a consideration of transaction costs would invalidate this assumption: Why would the transaction costs associated with applying distinct sales tax rates across a collection of service industries be any higher than the transaction costs associated with applying distinct regulatory standards across those industries?²⁶

²⁶In this sense, Mode 3 services are again very different from traded goods, where as Staiger and Sykes (2011) observe the transaction costs associated with imposing distinct tariff rates at the border on imports of narrowly defined goods is quite clearly far lower than the transaction costs that would be incurred by imposing different sales tax rates across the same set of goods.

Nevertheless, it is not hard to see that the logic of shallow integration we have described above would still survive, even if our Benchmark Model were augmented to reflect an assumption that the nondiscriminatory tax had to be applied to a more aggregate set of service industries than the regulatory standards. There would have to be some modifications to this logic: in the absence of sufficiently finely tuned tax policies, governments would now have to make market access commitments on *some* regulatory measure in each industry; but once that market access commitment is made, attaining efficient choices for the remaining regulatory measures in the industry could be handled with the presence of the NT, TBT and NV rules, just as before. The one caveat to this statement is that there would need to be greater reliance on the NV rule.

The National Treatment Rule Our shallow integration analysis relies on an implicit assumption that an NT requirement for regulatory measures cannot be circumvented by legal artifice. For example, the GATS national treatment obligation (when it exists) requires a member to afford “treatment no less favourable than it accords to its own like services and service suppliers” (see GATS Art. XVII(1)). If governments can manipulate the concept of “likeness” excessively, they may be able to discriminate in regulatory standards between closely competing foreign and domestic service suppliers and an NT obligation may accomplish little (see Mattoo (2000)). Little case law exists on the concept of “likeness” in service sectors, but what little exists suggests that distinctions in the way providers are taxed or regulated may affect the determination of likeness if they affect the competitive relationship between suppliers.²⁷ If the NT obligation as to regulatory standards is easily circumvented as a legal matter, the utility of a shallow integration approach will be undermined.

The Non-Violation Rule For our results on shallow integration, we have assumed implicitly that governments have access to the data they would need to implement an effective NV rule. This requires that governments have the ability to detect when a non-contracted policy measure is used by a trading partner in a way that denies market access, a task that ultimately comes down to being able to measure and monitor changes in import volumes and prices with reasonable accuracy.

While detailed and comprehensive data on trade in goods has been around since the beginning of GATT, for Mode 3 services, where the imports do not cross international borders,

²⁷See Argentina – Measures Relating to Trade in Goods and Services, WT/DS453/AB/R, adopted May 9, 2016.

satisfying these data needs presents a special challenge. As Maurer et al (2008) discuss, detailed data on Mode 3 services trade can be found in the Foreign Affiliates Trade in Services (FATS) statistics, but very few countries currently produce these statistics. In fact, Maurer et al report that until recently the United States was the only country compiling FATS statistics. This is beginning to change, but as Maurer et al (p. 164) conclude, when it comes to data on Mode 3 service trade flows it is still the case that “the lack of data reliability and inter-country comparability is almost a general rule.”

According to our findings here, the fragmentary nature of the data on Mode 3 service trade flows currently available could represent a significant roadblock in the way of a shallow-integration approach for services trade liberalization, beyond the challenges faced in the context of goods trade, and could as a result help explain the broad differences in the approach to integration taken by GATT and GATS. However, as explained by Maurer et al, following the entry into force of GATS there has been an increasing call for detailed and comprehensive data on trade in services; and efforts are under way, sponsored by the WTO and others, to develop and collect such data. While successful efforts to compile this data would no doubt provide a valuable input into deep-integration bargaining under the current GATS structure, our findings suggest that these efforts might also be instrumental in facilitating consideration of alternative approaches to integration of services trade more along the lines of the shallow integration approach that has proven to be successful for GATT.

6. Conclusion

The terms-of-trade theory of international trade agreements focuses on tariff agreements covering trade in goods, and offers an explanation for core features of the GATT’s “shallow integration” approach to goods trade liberalization. Tariffs play almost no role in services markets, however, and the existing models cannot account for the dramatically different approach to services trade liberalization taken by GATS. We show that the key features of GATS, including its emphasis on “deep integration” – sector-by-sector negotiations on behind the border policy instruments – can be accounted for when the terms-of-trade theory is augmented with a set of restrictions, motivated by salient features of services trade, on the policy instruments and trade data available to governments. This is the main positive message of our paper. Using our augmented terms-of-trade theory, we also find that a shallow-integration approach to services trade liberalization, suitably modified to fit the needs of the services trade context, may indeed

be possible. The current, deep integration, approach of GATS raises significant challenges for negotiations seeking to expand market access, challenges that would not arise under shallow integration. With continued improvements in the data on services trade, we therefore interpret this second finding as pointing toward a possible way forward for achieving success in future efforts to liberalize trade in services. This is the main normative message of our paper.

The benchmark model on which we build our analysis abstracts from a number of complications that are likely to be important features of the world in which real trade agreements operate, such as firms that wield market power and governments that possess private information. In the context of goods trade, some of these features have been shown to introduce substantial qualifications into the case for shallow integration (see, for example, the recent surveys of this literature in Lee, forthcoming and Ederington and Ruta, forthcoming). For our purposes here, however, where we seek to account for the differences in approach between GATT and GATS, the important question is whether any of these qualifications are more likely to arise in the context of services trade.

In this regard, the most salient alternative to the account we have put forth in this paper may come from the “offshoring” model developed by Antràs and Staiger (2012a,b). Antràs and Staiger show that the case for shallow integration is undermined when international prices are determined by bilateral bargains between individual buyers and sellers, rather than by anonymous industry-wide market clearing conditions as in the standard terms-of-trade theory. And they observe that in light of its often highly specialized nature, services trade may represent a sector where this form of price-determination is especially prevalent.²⁸ These two accounts have distinct normative implications: according to the offshoring model of Antràs and Staiger, the deep-integration approach of GATS is an inevitable consequence of the nature of price determination for services trade, while according to the account we have provided here the possibility of shallow integration for services trade exists. We view the task of distinguishing between the relevance of these and other (possibly complementary) explanations of the striking differences between GATT and GATS as an important avenue for future research.

²⁸The case for shallow integration is also undermined in the commitment theory of trade agreements (see, for example, the discussion of this in the recent review of Bagwell, Bown and Staiger, forthcoming, and see Fiorini and Lebrand, 2016 for an application of the commitment theory to services trade). But again, for our purposes the key question is why the commitment theory (and its implication of deep integration) should be relevant for services trade while the terms-of-trade theory (and its support of shallow integration) should be relevant for goods trade. Without a compelling answer to this question, we do not see the commitment theory as providing a promising account of the differences between GATT and GATS. See also Mattoo (2015) for a dissenting view on market-access-centered approaches to negotiated services trade liberalization.

References

- [1] Antràs, Pol and Robert W. Staiger. 2012a. “Trade agreements and the nature of price determination.” *American Economic Review: Papers and Proceedings* 102(3): 470-6.
- [2] Antràs, Pol and Robert W. Staiger. 2012b. “Offshoring and the role of trade agreements.” *American Economic Review* 102(7): 3140-3183
- [3] Bagwell, Kyle, Bown, Chad and Robert W. Staiger. Forthcoming. “Is the WTO Passé?” *Journal of Economic Literature*.
- [4] Bagwell, Kyle and Robert W. Staiger. 1999. “An Economic Theory of GATT.” *American Economic Review*.89(1), pp. 215-48.
- [5] Bagwell, Kyle and Robert W. Staiger. 2001. “Domestic Policies, National Sovereignty and International Economic Institutions.” *Quarterly Journal of Economics* 116: 519-562.
- [6] Bagwell, Kyle and Robert W. Staiger. 2002. *The Economics of the World Trading System* (Cambridge MA: MIT Press).
- [7] Bagwell, Kyle and Robert W. Staiger. 2012. “Profit Shifting and Trade Agreements in Imperfectly Competitive Markets.” *International Economic Review* 53(4): 1067-1104.
- [8] Bagwell, Kyle and Robert W. Staiger. 2015. “Delocation and trade agreements in imperfectly competitive markets.” *Research in Economics* 69(2): 132-56.
- [9] Bagwell, Kyle and Robert W. Staiger. Forthcoming. “The Design of Trade Agreements.” in Kyle Bagwell and Robert W. Staiger (eds) *Handbook of Commercial Policy*, Elsevier.
- [10] Bagwell, Kyle and Alan O. Sykes. 2004. “Chile - Price Band System and Safeguard Measures Relating to Certain Agriculture Products.” *World Trade Review* 3(3): 507-28.
- [11] Blanchard, Emily J. 2007. “Foreign direct investment, endogenous tariffs, and preferential trade agreements.” *BE Journal of Economic Analysis & Policy* 7.
- [12] Copeland, Brian and Aaditya Mattoo. 2008. “The Basic Economics of Services Trade.” in Aaditya Mattoo, Robert M. Stern and Gianni Zanini (eds) *A Handbook of International Trade in Services*, Oxford University Press.

- [13] Due, John F., and John L. Mikesell. 1994. *Sales Taxation, State and Local Structure and Administration*. 2d ed. Washington, D.C.: Urban Institute Press.
- [14] Ederington, Josh and Michele Ruta. Forthcoming. “Non-Tariff Measures and the World Trading System.” in Kyle Bagwell and Robert W. Staiger (eds) *Handbook of Commercial Policy*, Elsevier.
- [15] Feketekuty, Geza. 2008. “Appendix: A Guide to Services Negotiations.” in Aaditya Mattoo, Robert M. Stern and Gianni Zanini (eds) *A Handbook of International Trade in Services*, Oxford University Press.
- [16] Fiorini, Matteo and Mathilde Lebrand. 2016. “The Political Economy of Services Trade Agreements.” mimeo, European University Institute.
- [17] Francois, Joseph and Bernard Hoekman. 2010. “Services trade and policy.” *Journal of Economic Literature* 48(3): 642-92.
- [18] Hendrix, Michele E. and George Zodrow. 2003. “Sales Taxation of Services: An Economic Perspective.” *Florida State University Law Review*, Vol 30, pp 411-433.
- [19] Johnson, Harry. 1953-54. “Optimum Tariffs and Retaliation.” *Review of Economic Studies* 21: 142-53.
- [20] Lee, Gea M. Forthcoming. “Subsidies and Countervailing Duties.” in Kyle Bagwell and Robert W. Staiger (eds) *Handbook of Commercial Policy*, Elsevier.
- [21] Mattoo, Aaditya. 2000. “MFN and the GATS,” in Thomas Cottier and Petros C. Mavroidis (eds.), *Regulatory Barriers and the Principle of Non-Discrimination in World Trade Law: Past, Present and Future*, The World Trade Forum Vol. II. Ann Arbor, MI: University of Michigan Press.
- [22] Mattoo, Aaditya. 2015. *Services Trade and Regulatory Cooperation*. E15 Expert Group on Services: Think Piece. July.
- [23] Maurer, Andreas, Marcus, Yann, Magdeleine, Joscelyn and Barbara d’Andrea. 2008. “Measuring Trade in Services.” in Aaditya Mattoo, Robert M. Stern and Gianni Zanini (eds) *A Handbook of International Trade in Services*, Oxford University Press.

- [24] Ossa, Ralph. 2011. "A 'new-trade' theory of GATT/WTO negotiations." *Journal of Political Economy* 119(1): 122-52.
- [25] Schwartz, Warren F. and Alan O. Sykes. 1996. Toward a Positive Theory of the Most Favored Nation Obligation and its Exceptions in the WTO/GATT System, *International Review of Law and Economics* 16: 27-51.
- [26] Schwartz, Warren F. and Alan O. Sykes. 2002. "The Economic Structure of Renegotiation and Dispute Resolution in the World Trade Organization," *The Journal of Legal Studies* 31(1): S179-S204.
- [27] Staiger, Robert W. and Alan O. Sykes. 2011. "International trade, national treatment, and domestic regulation." *Journal of Legal Studies* 40(1): 149-203.
- [28] Staiger, Robert W. and Alan O. Sykes. 2013. "Non-Violations." *Journal of International Economic Law* 16(4): 741-775.
- [29] Staiger, Robert W. and Alan O. Sykes. Forthcoming. "How important can the non-violation clause be for the GATT/WTO?." *American Economic Journal: Microeconomics*.