Learning and International Policy
Diffusion –
The Case of Corporate Tax Policy *

Johannes Becker
University of Münster
Institute for Public Economics
Wilmergassee 6-8
48143 Münster
Germany
johannes.becker@wiwi.uni-muenster.de

Ronald B. Davies
University College Dublin
G215 Newman Building
Belfield
Dublin 4
Ireland
ronbdavies@gmail.com

November 6, 2013

Abstract

A recent empirical literature has arisen documenting the response of one nation’s policy choices, including tax, environmental, and labour policies, to those of others. This has been largely interpreted as evidence of competition, be it for mobile resources (like FDI, taxable book income, etc.) or yardstick. We present a third explanation based on learning. When countries’ tax choices reflect private information about unobserved conditions, this encourages nations to update their policies not in order to retain investment or manipulate trade flows, but because the new information conveyed by overseas tax rates allows them to fine-tune their own policies. With this “social learning”, countries converge on their optimal policies faster than in isolation. Furthermore, this convergence implies a pattern of policy convergence often attributed to competition for mobile resources. The benefits of social learning are smaller in the presence of policy adjustment costs although welfare remains higher than learning in isolation. In addition, adjustment costs result in inefficient policy adjustment because countries do not internalize the benefits conveyed by their own adjustments to other nations. We further show that these baseline results are robust to alternative network architectures, the choice of which can be used to replicate stylized facts found in the empirical tax competition literature. Finally, we discuss additional applications of social learning with adjustment costs well beyond the international tax debate.

JEL classification: H25, H32, H87

Keywords: social learning, tax competition

*We thank Ben Lockwood, Karolien De Bruyne, and seminar and conference participants at the University of Ghent, the 2013 Oxford Tax Conference, and the Trade and Networks Conference in Leuven. All errors are our own.
1 Introduction

Recently, a body of empirical literature has arisen documenting an interdependence among countries’ tax policies (see, e.g., Devereux et al. 2008, Overesch and Rincke 2009, 2011, Heinemann et al. 2010), labour policies (e.g., Davies and Valamennati, 2013, Olney, forthcoming) and environmental policies (e.g., Davies and Naughton, forthcoming, Beron, Murdoch, and Vijverberg, 2003, and Murdoch, Sandler, and Vijverberg, 2003). This work generally finds that the policies in one country are positively correlated with the policies set elsewhere. The primary conclusion is that this is evidence of competition in the tradition of Zodrow and Mieszkowski (1986) and Wilson (1986) wherein governments compete for mobile resources by implementing “business-friendly” policies. An alternative explanation is the yardstick competition of Besley and Case (1995) in which voters judge the performance of local policy makers by the policies implemented elsewhere. In both of these, the rationale behind international policy diffusion is that the policy choices undertaken in one location affect the payoffs from a given policy in other locations.

In this paper, we present and formalize a third alternative in which the preferred policy choice in one country is affected by those elsewhere without relying on changes in payoffs, i.e., without hinging on cross-border capital flows (as in tax competition) or political competition under asymmetric information (as in yardstick competition). We do so by assuming that countries only have incomplete information on the true state of nature but can learn about it by observing the policies set elsewhere. This learning results in convergence of beliefs and therefore policies, not because the taxes chosen elsewhere affect the payoffs of a given nation’s policy, but because learning results in the reduction in information asymmetries. Furthermore, our model provides predictions consistent both with the empirical evidence on overall trends in tax rates and the above interdependence literature.

Our model is one in which countries set tax rates that optimally depend on a common state of nature which is a priori unknown. In each period, each country receives a payoff that is a function of their tax policy, the state of nature, and a random variable. This payoff then acts as a private signal revealing information about the state of nature. In addition, in each period each country observes the taxes set elsewhere in the previous period. These past taxes of other countries contain information about their (lagged) beliefs on the state of nature. By incorporating this information, a given country is then able to refine its beliefs about the true state of nature and adjust its tax policy accordingly. As information diffusion results in belief convergence, so too will it lead to tax convergence. For example, this can lead to countries that will choose to reduce

---

2It should be noted, however, that there also exist a smaller group of theories in which, due to negative externalities, governments compete to drive FDI to other nations. The study by Markusen, Morey and Oleweiler (1995) is an example of this “not-in-my-backyard” literature.

3See Salmon (1987) for an initial application to taxes and Brueckner (2003) for an overview.

4Slemrod (2004), discussed further below, discusses both the overall decline in corporate tax rates and the international variation of those rates.
their tax rate in the second round if the average first round tax rate is below its own tax rate and vice versa. Thus, tax rate convergence is a welfare-improving activity and not the consequence of inefficient competition for mobile capital.

This does not mean, however, that the equilibrium is efficient. When adjustment is costly, a nation will only alter its existing tax if the expected gain from doing so outweighs the adjustment cost. An important implication of this is that the equilibrium is inefficient. This is because non-adjustment conveys less information than adjustment does because it only indicates a range of beliefs a country holds. As such, the private value of adjustment is less than the social value and adjustments are inefficiently rare. It is important to recognize that this new externality in tax setting is driven by the unintentional withholding of information but not by affecting another country’s payoff of a given policy (as occurs when taxes in one country affect capital stocks or voting in another which can result in inefficiently low taxes and the underprovision of public goods). As such, policies intended to prevent harmful tax competition such as tax harmonization would tend to exacerbate losses in our setting precisely because harmonization would inhibit the flow of information. Please note, however, that we are not attempting to suggest that tax competition for mobile capital or yardstick competition do not exist. Rather, we believe that the mechanisms in our model complement those results and provide a richer framework for the discussion of international policy diffusion. In particular, as tax convergence in our model is the result of beneficial information diffusion, our model provides a rationale for why the observed patterns in tax rates may be less damaging than what is typically supposed.

Our approach builds on the social learning literature (Gale, 1996, Gale and Kariv, 2003, Vives, 1996) which assumes that agents can observe their neighbours’ actions but not their outcomes. We translate this theory into the realm of policy-making jurisdictions and reinterpret the evidence accordingly. Moreover, we contribute to this literature by accounting for policy-specific aspects (such as ideology) that affect the informational content of another country’s actions. Finally, by introducing adjustment costs into the model, we make a further contribution to that literature.


5Brueckner (2003) offers a taxonomy of models of strategic government interaction. He proposes two categories, spill-over models and resource-shift models. In spill-over models, the objective function of an individual government directly depends on the actions taken by other governments. In resource-shift models, the objective functions depend on the quantity of resources within its borders which itself depends on the actions taken by all governments. Both models give rise to reaction functions. The learning based theory of international tax rate setting is different from both types of models.

6Other models, including Davies (2005), provide additional settings where interjurisdictional tax competition can be welfare improving.

7In contrast, the social experimentation literature (see, for instance, Bala and Goyal, 1998) assumes that agents can observe their neighbours’ actions and their outcomes.
and Mintz and Smart, 2004. Furthermore, although our discussion focuses on taxation both to focus the exposition and to demonstrate the use of a model of learning in a network with costly adjustment, it can be applied to any such setting. This includes other realms of government policy (such as labour, environmental, or social policy), firm choices (such as the location of affiliates or adoption of new technologies), and labour decisions (such as whether to relocate to a particular region because of potential work opportunities). Therefore the model we present, although framed in the context of international taxation, has general applications.

The remainder of the paper is organized as follows. Section 2 provides a review of the empirical literature on policy-making in an international setting. We do so in order to highlight key features of the data that a model of policy setting should be able to explain and to frame our overall discussion. Section 3 presents the model, derives the main results and discusses some extensions. In particular, we demonstrate how different network architectures can result in the patterns found in the empirical literature. Section 4 concludes.

2 Literature: Theory and evidence on corporate tax policy in open economies

In this section, we review the empirical literature on international tax competition (and international policy competition in general). Our goal here is two-fold. First, we seek to develop a set of empirical regularities into which a model of policy setting should be able to provide insight. Second, by reviewing the methodologies used in the empirical literature, it will ease our discussion of how our model can provide comparable patterns to what is observed in the data. With this in mind, we note three broad stylized facts found in the empirical literature on tax setting.

Fact #1:: Stickiness of corporate tax rates. A characteristic feature of international tax rate setting is the infrequent rate of adjustment. Heinemann et al. (2010) report that, on average, national governments change their tax rate every four years. When combined with the typical legislative period’s four year length, this means that the average elected government changes the tax rate once. This hints at some association with the political system, and is thus suggestive of yardstick competition. Alternatively, this could stem from an adjustment cost indicating that it is only when the current tax policy becomes sufficiently divergent from current conditions that a government makes an adjustment. As such, Heinemann et al.’s results indicate that a tax rate adjustment cost is a desirable feature of a theory of tax setting.

Fact #2: Downward trend and convergence in corporate tax rates. Despite their year-to-year persistence, as discussed in detail by Slemrod (2004), there has been a marked shift in taxes starting in the 1980s. This shift has
Table 1: Corporate taxation in selected OECD countries (1985-2005)

<table>
<thead>
<tr>
<th>Country</th>
<th>Statutory tax rate(^a)</th>
<th>1985</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td></td>
<td>61</td>
<td>25</td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Belgium</td>
<td></td>
<td>45</td>
<td>34</td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td>60</td>
<td>26</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td>50</td>
<td>34</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td>63</td>
<td>38</td>
</tr>
<tr>
<td>Greece</td>
<td></td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td>46</td>
<td>37</td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td>56</td>
<td>40</td>
</tr>
<tr>
<td>Netherlands</td>
<td></td>
<td>43</td>
<td>32</td>
</tr>
<tr>
<td>Norway</td>
<td></td>
<td>51</td>
<td>28</td>
</tr>
<tr>
<td>Portugal</td>
<td></td>
<td>55</td>
<td>28</td>
</tr>
<tr>
<td>Spain</td>
<td></td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td>60</td>
<td>28</td>
</tr>
<tr>
<td>Switzerland</td>
<td></td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td>50</td>
<td>39</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td>47.3</td>
<td>31.5</td>
</tr>
<tr>
<td><strong>Std. Dev.</strong></td>
<td></td>
<td>12.3</td>
<td>6.3</td>
</tr>
</tbody>
</table>

\(^a\) including local taxes. Sources: Devereux *et al.* (2002); http://www.ifs.org.uk/publications/3210

three key features. First, as illustrated in Figure 1, there has been a steady downward trend in corporate tax rates in nearly all OECD countries. Second, as shown in Table 1, this decline has been greatest for countries that initially had the highest tax rates. For countries with above average taxes in 1985, the average decline to 2005 was 24 percentage points. In contrast, those below the 1985 average only saw tax rates fall 6.6 percentage points. These declines were particularly great for nations with extremely high initial tax rates (above 60 percent). Third, there has been an overall convergence in taxes. This is exhibited both by the fall in the standard deviation and by the gap between the maximum and minimum taxes, which stood at 53 percentage points in 1985 compared to 27 percentage points twenty years later.

The literature provides a number of different explanations for these trends. The predominant explanations, competition for mobile capital and yardstick competition, suggest that the fall in one country’s taxes is the result of low taxes elsewhere. In the competition for FDI setting, mobile capital is driven off by relatively high tax rates (something confirmed by the bulk of the literature, with de Mooij and Ederven (2008) providing an overview). As such, these
In the empirical part, we focus on tax rate cuts. Note that our definition of a tax rate cut accounts for institutional features of the corporate tax system in countries where the observed overall tax burden on corporate income may vary even if there is no change in national tax policies. In particular, in countries with local business income taxes, we only consider countrywide reforms of the local tax, but we do not consider changes which are merely modifications at the local level. With this definition, in the 32 countries considered 148 tax rate cuts occurred between 1981 and 2007. On average, national governments have cut their corporate tax rate every 4 years. Figure 2 depicts, for all years considered, the fraction of countries that actually lowered their tax rate. Particular intensive rate-cutting activities have occurred between 1989 and 1994 and since 1998. In contrast, during the mid-1990s only very few tax rate cuts were observed. Only the transition economies in eastern Europe intensively reduced their statutory tax rates. Since 1998, our data show a general tendency toward intensified rate-cutting activities with no particular geographical pattern.

3. EMPIRICAL APPROACH AND DATA

3.1 Empirical Model

Our empirical analysis aims at identifying the key determinants of rate-cutting tax reforms among European countries. As discussed above, a sizable recent literature has analyzed the determinants of countries' business tax rates, movements can be attributed to the abolishment of capital controls beginning in the 1980s which increased the tax sensitivity of investment location and tax competition. Alternatively, yardstick competition relies on uninformed voters who compare their government's policies to those elsewhere to judge the effectiveness, and therefore re-election worthiness, of incumbent politicians. When, as formalized by Besley and Case (1995), some countries demonstrate that providing public goods is possible with lower tax rates, this pressures a high-tax government to follow suit. In conjunction with these two explanations, some suggest that these movements are attributable to a policy shift by a “leader” country such as the US. As is well-known, the US reduced its taxes between 1986 and 1988 and other countries followed suit. As first expressed by Gordon (1992), this interaction was strengthened by the US's role as the largest and most important capital exporter. Under the tax credit system, countries hosting US FDI with tax rates between the pre-reform and post-reform US rates would for the first time find that US investment was responsive to their tax. Accordingly, a drop in the US corporate tax rate could make a reduction in host countries’ tax rates necessary. In addition, there is the possibility that during the mid-1980s there was a change which gave the US leadership status in the Stackelberg sense, that is, led it to use its own tax policy to strategically manipulate those elsewhere. This change may have then resulted in a shift in US taxes as it sought to exploit this new advantage, something which then filtered through the follower countries. Alternatively, rather than a change in a “leader”, such movements could simply result in a common shift in the national characteristics that drive tax setting. For example, it could be the result of a common ideological shift across the

\[ \text{Equation} \]

\[ \text{Figure 1: Statutory tax rates in European countries, 1980-2007 (Source: Heinemann et al. 2010)} \]

\[ \text{Figure legend:} \]

\[ \text{All countries (incl. Eastern Europe)} \]

\[ \text{Western Europe} \]

\[ \text{All years considered} \]


\[ \text{Notes: Graph shows the unweighted average of statutory corporate income tax rates.} \]

\[ \text{Whalley (1990) considers the first two alternative approaches and concludes that neither of them is able to account for the similarities in the tax changes of the countries under consideration.} \]
OECD in favour of a lower tax rate regime.\footnote{Indeed, such a possibility seems in line with the fact that the political landscape during the 1980s was dominated by leaders such as Margaret Thatcher, Ronald Reagan, and Helmut Kohl; all conservatives who may have shaped their societies and policies towards a more business-oriented attitude.}

Thus, the existing literature by and large attributes the change in tax rates to either some form of competition (where the payoffs to a tax rate in one country depends on those elsewhere) or a coincidence driven by a common trend in determining factors. Our paper contributes by offering a new, learning based model consistent with these changes.

Fact #3: Taxes in one country depend on those elsewhere. Arising from the two competition theories for tax changes, there grew a collection of empirical studies seeking to estimate the sensitivity of one country’s tax to those elsewhere. Corresponding to the two theories of tax competition, this literature follows two motivating notions. On the competition for FDI side, examples include Devereux et al. (2008), Redoano (2007), Davies and Voget (2009), Overesch and Rincke (2011), among others. Those motivated by yardstick competition include Heyndels and Vuchelen (1998), Revelli (2001), and Solé-Ollé (2003). The results from this literature suggests that countries can be observed to cut their tax rates \textit{because} their neighbours did so.

To frame our discussion, we now present the primary methodology in this literature which has been a spatial econometric approach.\footnote{For a complete treatment, we defer the interested reader to Anselin (1988).} In these works, the general regression specification is:

\[ T_{i,t} = x_{i,t}\beta + \rho \sum_{j \neq i} \omega_{j,i,t}T_{j,t} + \varepsilon_{i,t} \tag{1} \]

where \( T_{i,t} \) is a policy by country \( i \) (be it tax, environmental, labor, or other policy), \( x_{i,t} \) is a vector of characteristics specific to country \( i \) (which can include lagged values of its tax rate) as well as common factors across countries, \( \varepsilon_{i,t} \) is an error term, and \( \sum_{j \neq i} \omega_{j,i,t}T_{j,t} \) is known as the spatial lag. The spatial lag is the weighted average of the policies across other countries.\footnote{Note that this variable is often considered endogenous, either due to errors correlated across countries or due to “strategic” endogeneity as suggested by the competition literature. As discussed by Anselin (1988) and many others, there are several methods available for dealing with this which are not germane to our analysis. We do wish to point out, however, that one method for dealing with this endogeneity is to use \( t - 1 \) values for the spatial lag.} It aggregates the policies of other countries into a single variable using a weighting scheme where \( \omega_{j,i,t} \) is the weight that \( i \) assigns to another country \( j \) in year \( t \).\footnote{Typically, these weights sum to one (known as row standardization).} Thus, the slope of \( i \)'s best response to \( j \)'s tax is \( \rho \omega_{j,i,t} \). Note that this is distinct from the impact of \( x_{i,t} \) which would control for common changes in factors such as ideology or the common effect driven by a leader country.

In specifying this relationship between taxes, the choice of weight is a crucial one. A number of different types weighting schemes have been used in the literature. The simplest of these is an equal weighting scheme, i.e. with \( N \) countries,
\( \omega_{i,j,t} = \frac{1}{N-1} \). More commonly, however, it is assumed that some countries matter more than others. A frequently used class of weighting schemes is one based on geographic proximity. This can be a variant of the equal weights where only neighboring countries receive a positive weight (something used by Altshuler and Goodspeed (2007) and Devereux, Lockwood, and Redoano (2008) in their studies of corporate taxes). Alternatively, weights can be a declining function of the distance between countries \( i \) and \( j \) (as was employed by Davies and Naughton, forthcoming, in their study of environmental policies and by Davies and Vadlamannati, 2013, in their study of labor policy). Klemm and van Parys (2012) bring these two together and use inverse-distance weights within a region, but assume zero weights across regions. These distance schemes are motivated the empirical regularity that FDI is deterred by distance between the parent and host (see Blonigen and Piger (2011) for a recent overview of FDI determinants). Thus, a tax rate reduction close to country \( i \) may be more successful in attracting away \( i \)'s capital, forcing it to respond more to a proximate country than a distant one.

A second class of weights presumes that some countries matter more overall. For example, Devereux, Lockwood, and Redoano (2008) use GDP-based weights, in which larger countries have higher weights.\(^{13}\) An alternative approach is to give countries with greater populations more weight, as done in Davies and Vadlamannati (2013).\(^{14}\) The rational behind such weights is again driven by the FDI determinants literature which finds that FDI is generally attracted to large countries. This suggests that country \( i \) may be forced to respond more to tax cuts by large countries than small ones. Indeed, the theory of Hauffer and Wooton (1999) shows that large countries may be expected to win over small ones when competing for FDI. More recently, some papers have begun to employ multiple spatial lags to examine whether a given country responds differently to another based on some classification such as EU membership (Davies and Voget (2009), Redoano (2007)), OECD membership (Davies and Naughton, 2006), or level of development (Davies and Vadlamannati, 2013).

An important distinction here is that the above discussion refers to large countries as “leaders” but that this does not imply leadership in the Stackelberg sense. There, leadership also conveys the ability to use one’s own tax to manipulate the tax rate of others. Altshuler and Goodspeed (2007) consider whether the US has taken the position of a Stackelberg leader in international tax competition and find some supporting evidence.\(^{15}\) Leadership within Europe by Germany or the United Kingdom cannot be confirmed, however.

Finally, just as some countries generate a greater response by others, the

\(^{13}\)Davies and Vadlamannati (2013) also use this when considering labour standards. Davies and Klasen (2013) do so in their study of the joint-determination of overseas development aid.

\(^{14}\)Others use weighting schemes based on the “attractiveness” of a given location. Examples of this include Devereux, Lockwood, and Redoano’s (2008) use of FDI as a weight, Exbrayat’s (2009) phi-ness weights, and Davies and Voget’s (2009) market potential weights.

\(^{15}\)Stackelberg leadership by the large(st) country is usually simply assumed as in Gordon (1992) or Baldwin and Krugman (2004). It should be noted that Kempf and Rota-Graziosi (2010) show that, if one endogenizes the leadership role, it is likely that the small country leads the tax competition.
literature has explored whether some countries are themselves more responsive to others’ taxes. For example, Devereux et al. (2008) report that countries with less strict capital controls are more sensitive to their neighbours’ tax policy choices.\(^{16}\)

However, it must be noted that, although this empirical literature finds that tax rates in country \(i\) depend on those elsewhere, it is unable to determine why this is so.\(^{17}\) Although the capital-control results of Devereux et al. (2008) point to tax competition as the main cause for spatial correlation in taxes, this is not to say that there are no studies which make the plausible case for yardstick competition on the community level. For instance, Bordignon et al. (2003) show for Italian municipalities that interaction in tax rates only occurs in jurisdictions with high degrees of political competition. Likewise, rather than using a spatial estimation approach, Bosch and Solè-Ollè (2007) examine the impact of taxes elsewhere on voting patterns in Spain and find evidence consistent with yardstick competition. In the next section, we add to the potential explanations for these three stylized facts by presenting a model of a learning based model of tax policy diffusion. A notable feature of this model is that it is able to generate observationally equivalent patterns of tax movements without any tax competition or yardstick competition involved. As such, it will not rely on externalities in the form of capital flows or voting behavior. Because of this, a change in one country’s policy does not impact the payoff to a given policy chosen by another country. Nevertheless, because information does move between countries, this will create an information externality. In addition, our approach differs from yardstick competition because it does not rely on an asymmetric information problem (but rather on an incomplete information problem). Therefore the information spillovers affect choices in a very different manner than in yardstick competition. Moreover, there are no political economy aspects to our story (although they could be integrated). Third, the shared tax movements in our model come about not due to changes in a common underlying parameter, but due to the arrival of new information, both from a country’s own experience and that of others.

With that said, we see our approach as complementary to the explanations discussed above.\(^{18}\) In no way do we seek to suggest that countries do not compete for mobile capital or that electorates do not base their voting decisions in part on their observations of what transpires elsewhere. While we could embed our learning-based model in a tax competition or yardstick competition model, we instead shut down those channels of policy diffusion in order to focus

\(^{16}\)It should be noted that Overesch and Rincke (2011) reexamine this linkage and show that it is not robust.

\(^{17}\)See Revelli (2005) for a discussion on this issue.

\(^{18}\)Brueckner (2003) offers a taxonomy of models of strategic government interaction. He proposes two categories, spill-over models and resource-shift models. In spill-over models, the objective function of an individual government directly depends on the actions taken by other governments. In resource-shift models, the objective functions depend on the quantity of resources within its borders which itself depends on the actions taken by all governments. Both models give rise to reaction functions. The learning based theory of international tax rate setting is different from both types of models.
on the specific contributions our approach makes.

3 The Model

We consider a four-period model with \( N \geq 2 \) countries, indexed by \( i \). Each country is characterized by a vector of country-specific, time-invariant characteristics \( x_i \). The \( N \) countries all share a common, but initially unobserved state of nature, the realization of which is denoted \( S^* \). The true state of nature is unobserved by countries. It is drawn in period 0 from a cumulated distribution function \( F_S(S) \) where \( f_S(S) = F_S'(S) \) denotes the density function. An example of the unknown state of nature could be the tax payers’ labour supply elasticity with response to taxation which is not directly observable but substantially affects welfare.

Each country’s government receives a per-period payoff, \( w_{i,t}(\cdot) \), given by

\[
 w_{i,t} = U\left(T_{i,t}; x_i, S^*_j\right) + \varepsilon_{i,t}
\]  

(2)

where \( U(\cdot) \) is a deterministic function of a policy parameter \( T_{i,t} \) to which we refer as the “tax rate”.\(^{19}\) A crucial assumption is that – in contrast to the tax competition models – the welfare of an individual country does not depend on the tax rates set elsewhere. \( \varepsilon_{i,t} \) is a random variable with mean of zero which is drawn from a cumulated distribution function \( F_\varepsilon(\varepsilon) \) where \( f_\varepsilon(\varepsilon) = F_\varepsilon'(\varepsilon) \) denotes the density function. Accordingly, the density of \( w_{i,t} \) is given by \( f_{w_i}(w_{i,t}|T_{i,t}, S^*_j) = f_\varepsilon \left(w_{i,t} - U\left(T_{i,t}; x_i, S^*_j\right)\right) \) – where we suppressed the \( x_i \) as arguments. Note that, since \( U(\cdot) \) is deterministic, it is the random variable \( \varepsilon \) that prevents governments from directly inducing \( S^* \) from the observation of \( w_{i,t} \). If \( S^* \) is the tax payers’ elasticity towards taxation, \( w_{i,t} \) could be the tax revenue. Simply by observing current tax revenue, a country’s government cannot precisely infer the taxpayers’ elasticity. Inference requires a sufficiently large number of observations to identify the true relationship between taxes and revenues. The function \( U(\cdot) \) is assumed to be twice differentiable, continuous in \( T_{i,t} \) and \( S^* \) and, for given \( x_i \) and \( T_{i,t} \), to vary monotonically in \( S^* \). Given the \( f_{x_i} \), there is a unique optimal tax rate for each level of \( S \).

The timing of the model is as follows. In period 0, each government \( i \) is exogenously endowed with an initial tax policy \( T_{i,0} \). This generates a payoff \( w_{i,0} \) for each. In period 1, government \( i \) uses \( w_{i,0} \) to refine its belief about the true state of nature and decides on a tax policy \( T_{i,1} \). If \( T_{i,1} \neq T_{i,0} \), then the

\(^{19}\)Although we could allow \( U \) to differ across countries, as long as it remains common knowledge this would add notation without affecting the results in any way. Further, we could permit \( U \) and/or \( x_i \) to vary with time. Although this would increase the complication of the discussion, as governments would condition their taxes on future functions and characteristics, again this would not affect the qualitative nature of our main results.
country incurs an adjustment cost $\alpha \geq 0$. Following this, it receives a first period payoff $w_{i,1}$. In period 2, the government observes $T_{j,1}$ for $j \in \Omega_i$, which is the set of countries which $i$ observes. Note that as all other’s period 1 taxes are observed simultaneously, there is not an information cascade in which a decision by an early tax setter would have disproportionately large impacts on the beliefs of others as its decisions are reinforced by the subsequent decisions of other countries. Define $N_i$ as the number of elements in $\Omega_i$. Following Gale and Kariv (2003), this would define the countries that $i$ is connected to, thereby defining the information network architecture. If all $i$ are connected to all the other $N - 1$ countries, this is a completely connected network; otherwise this is an incompletely connected network. Following this inflow of information, $i$ updates its beliefs regarding $S^*$ and sets its second period tax $T_{i,2}$. Again, if this differs from the tax rate used in the previous period, it incurs an adjustment cost $\alpha$. Finally, in period 3, the true state of nature is revealed and each country has the option of revising its tax one final time (where as before, adjustment costs are $\alpha$).

The functions $F_S(\cdot)$ and $F_\varepsilon(\cdot)$ are common knowledge, as are all the $x_i$’s and the shape of each country’s $U(\cdot)$. In contrast, the $w_{i,t}$ is private information of country $i$.

With the model in place, we can now derive a country’s optimal policy formation.

### 3.1 Policy Setting

Before proceeding, it will be useful to establish some notation. Define $\Pi_i(T_{i,t}; x_i, I_{i,t})$ as the present discounted stream of payoffs from period $t$ to 3 (inclusive) from choosing a tax rate $T_{i,t}$ in period $t$ based on the expectations generated by the information held in period $t$, denoted by $I_{i,t}$. Let $\hat{T}_i(x_i, I_{i,t}) = \arg \max \Pi_i(T_{i,t}; x_i; I_{i,t})$ be the tax rate that maximizes the stream of expected payoffs given the country’s characteristics and the information available in $t$. We refer to this as the information conditioned optimal (ICO) tax. Let $T^*(x_i, S^*)$ be that tax that maximizes $\Pi_i(T_{i,t}; x_i, S^*)$ when $S^*$ is known with certainty. From this point on, we will suppress the $x_i$ in both of these tax rates unless it is needed. Note that $T^*(S^*)$ also maximizes the payoff in each period as both $x_i$ and $S^*$ are constant over time.

In describing government choices, it is important to recognize that a government’s optimal policy in period $t$ depends on its expectations about what will happen in subsequent periods. With this in mind, we begin with period 3’s choice. At this point, each government knows the true state of nature and starts the period with a tax $T_{i,2}$. It can choose to retain this tax rate and receive an expected payoff of $U(T_{i,2}; x_i, S^*)$. Recall that even if $S^*$ is

---

We assume that this is equal across $i$ and $t$. However, changing this would add notation but not alter the main results. The primary difference is that low $\alpha$ countries would alter their taxes more frequently and would therefore convey more information to others.

For seminal work on information cascades, see Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992).
known, this is only in expectation due to $\epsilon_{i,3}$. Alternatively, $i$ can adjust its tax rate. If it does so, since $S^*$ is known, it will adjust to $T^*(S^*)$, leaving it with expected payoff of $U(T^*(S^*) ; x_i, S^*) - \alpha$. Whether it does so depends on whether the gain from adjustment outweighs the cost, i.e. on whether: $U(T^*(S) ; x_i, S^*) - U(T_{i,2}^* ; x_i, S^*) \geq \alpha$. As the function $U$ is continuous in taxes and $S$, this implies a range of states of nature for which $i$ will not adjust. Define this set by $\Gamma(T_{i,2}) = \{ S(T_{i,2}) \}$. Note that, since it would not adjust if $T_{i,2} = T^*(S^*)$, i.e. if it already has the optimal tax for the true state of nature, that this set is non-empty. Further, if $\alpha = 0$, this set contains a single element. Finally, as adjustment costs increase, this set expands to include more states of nature.

Moving to period 2, $i$ has an information set $I_{i,2}$ which contains its period 0 and 1 payoffs and $T_{j,1}$ for $j \in \Omega_i$. If it chooses to update its tax rate from the tax it had in period 1, $T_{i,1}$, it will do so based on its information $I_{i,2}$ and will choose it to maximize the expected stream of payoffs:

$$
\Pi^2(T_{i,2}; x_i, I_{i,2}) = \int_{-\infty}^{\infty} U(T_{i,2}; x_i, \tilde{S}) f(\tilde{S}; I_{i,2}) d\tilde{S} + \delta \int_{\tilde{S} \in \Gamma(T_{i,2})} U(T_{i,2}; x_i, \tilde{S}) f(\tilde{S}; I_{i,2}) d\tilde{S} + \delta \int_{\tilde{S} \notin \Gamma(T_{i,2})} \left( U(T^* ; x_i, \tilde{S}) - \alpha \right) f(\tilde{S}; I_{i,2}) d\tilde{S}
$$

(3)

where $f(\tilde{S}; I_{i,2})$ is the conditional probability of the state of nature being $\tilde{S}$ given $I_{i,2}$ and $\delta \leq 1$ is the discount rate. Note that this takes into account that $i$ may or may not choose to adjust its tax rate in period 3. The ICO tax is implicitly determined by the first order condition:

$$
\frac{\partial \Pi^2(T_{i,2}; x_i, I_{i,2})}{\partial T_{i,2}} \bigg|_{T_{i,2}=\hat{T}_{i,2}} = \int_{-\infty}^{\infty} UT(\hat{T}_{i,2}(I_{i,2}); x_i, \tilde{S}) f(\tilde{S}; I_{i,2}) d\tilde{S} + \delta \int_{\tilde{S} \notin \Gamma(\hat{T}_{i,2}(I_{i,2}))} U_T(\hat{T}_{i,2}(I_{i,2}); x_i, \tilde{S}) f(\tilde{S}; I_{i,2}) d\tilde{S} = 0.
$$

(4)

Note that this is not necessarily the one that maximizes the one period expected payoff. Even if expectations indicate that the state of nature is some level $\tilde{S}$, it need not be the case that the government would choose $T^*(\tilde{S})$ since it gives positive probability to keeping its period 2 tax rate in period 3 unless adjustment costs are zero. As such, depending on the sign of the second term in the above equation, it may choose to set an ICO tax above or below $T^*(\tilde{S})$ depending on whether the expected marginal utility in period 3 when not adjusting is positive or negative when evaluated at $T^*(\tilde{S})$.

As in period 3, whether or not $i$ chooses to adjust depends on how the expected gain from deviation $\Pi^2(T^*(I_{i,2}); x_i, I_{i,2}) - \Pi^2(T_{i,1}; x_i, I_{i,2})$ compares

---

22 Although $i$ also observes the period 0 tax rates of these countries, as they contain no information we ignore them here.
to the cost $\alpha$. This then implies a set of information sets $\Theta(T_{i,1})$ for which it will not adjust its tax rate. As before, there will exist an information set which indicates that the period 2 ICO tax is equal to that inherited from period 1, meaning that $\Theta(T_{i,1})$ is non-empty. Further, the size of $\Theta(T_{i,1})$ is increasing in $\alpha$ and contains a single information set for $\alpha = 0$. For future use, let the cumulative density function of these information sets be $H(I_{2,1}; I_{i,1})$ with an associated probability distribution function $h(\cdot)$.

Moving to period 1 here, the information set consists of a single piece of information which is $i$’s period 0 payoff, i.e. $I_{i,1} = w_{i,0}$. Comparable to period 2, if $i$ chooses to adjust its tax rate, it will choose the ICO tax as this maximizes:

$$
\Pi^1(T_{i,1}; x_i, I_{i,1}) = \int_{-\infty}^{\infty} U(T(\tilde{S}); x_i, \tilde{S}) f(\tilde{S}; I_{i,1}) d\tilde{S} + \delta \left( \int_{I_{i,2}\in\Theta(T_{i,1})} \Pi^2(T_{i,1}; \tilde{I}_2) h(\tilde{I}_{i,2}, I_{i,1}) d\tilde{I}_{i,2} + \int_{I_{i,2}} \Theta(T_{i,1}) \left( \Pi^2(T(\tilde{I}_2); \tilde{I}_{i,2}) - \alpha \right) h(\tilde{I}_{i,2}, I_{i,1}) d\tilde{I}_{i,2} \right)$$

(5)

where the expected period 2 payoffs account for the expected period 3 payoffs.

Lastly, as in periods 2 and 3, the decision of whether or not to adjust depends on how the gain from adjustment compares to the cost. As such, the country will adjust whenever $\Pi^1(T_{i,1}(I_{i,1}); x_i, I_{i,1}) - \Pi^2(T_{i,0}; x_i, I_{i,1}) \geq \alpha$. Again, this means that there is a non-empty set of of first period information sets $\Lambda(T_{i,0}, I_{i,1})$ for which $i$ will not adjust. As with the other such sets, if $\alpha = 0$, this set contains a single element. Finally, because the information in this period is only $i$’s period 0 payoff, $\Lambda(T_{i,0}, I_{i,1})$ also defines a set of $w_{i,0}$ for which $i$ will not adjust in period 1. Using the continuity of the welfare function, this set is compact with a lower bound $w_{i,0}(T_{i,0})$ and an upper bound of $\bar{w}_{i,0}(T_{i,0})$. Finally, note that this set is increasing in $\alpha$ and that, as adjustment costs become infinitely large, it spans the entire range of payoffs, i.e. $i$ will never adjust.

To aid intuition, these results are illustrated in Figures 2 and 3 for a specific version of the model where the government’s goal is to “match” the tax rate to the state of nature. Whenever it does so, regardless of the state, the same maximal level of per-period expected welfare is achieved, given by $U(T^*(x_t; S); x_i, \cdot)$. Figure 1 illustrates the updating of beliefs. In this, the payoff in period $t$, inclusive of the random component, is illustrated on the vertical axis and the tax rate is on the horizontal axis. If the state of nature is $S'$, then the determinant part of the one-shot payoff as a function of the tax rate is illustrated by the solid, horizontally-oriented curve $U(T^*(S); x_i, S')$. This reaches its maximum when the tax rate $T^*(S')$; deviations in the implemented tax reduce expected welfare relative to this level. Note that $T^*(S')$ yields $U(T^*(x_t; S); x_i, \cdot)$ only in expectation; there is still a distribution around this level due to $\epsilon_{t,i}$, the distribution of which is illustrated by the solid, vertically-oriented curve. If a payoff such as $U$ is obtained in $t$, it is possible that the true state of nature is $S''$. However, this is unlikely relative to some alternative state of nature such as $S''$, where the determinant part of the payoff and the distribution around that are illustrated by the horizontal and dashed lines. As such, the government
will update its beliefs, decreasing the relative probability placed on $S'$ while increasing that for $S''$.

To illustrate the decision of when to adjust the tax rate, consider Figure 3. In this, $\Pi_1$ is on the vertical axis and the tax rate implemented in period 1 is on the horizontal axis. As in Figure 2, the goal is to match the tax rate to the state of nature. When this is done, regardless of the state, the same maximal level of expected welfare is achieved denoted by $\Pi_1(T_{i,1} (x_i, I_{i,1}); x_i, I_{i,1})$. Note that this is again an expectation due to uncertainty about $\epsilon_{i,t}$, $S^*$, and whether or not there will be a future tax adjustment. The figure illustrates this relationship for three information sets (the three bell curves) where moving from left to right $I_{i,1}$ equals $w_{i,0}(T_{i,0})$, the payoff that indicates that $T_{i,0}$ is the ICO tax (with the payoff denoted by $\hat{w}_{i,0}$), and $\tilde{w}_{i,0}(T_{i,0})$. As can be seen, when the ICO tax is chosen, in each case the same level of $\Pi_1$ is achieved. When $w_{i,0} \neq \hat{w}_{i,0}$, the expected stream of payoffs from retaining $T_{i,0}$ falls. However, it is only when the difference between the maximal level of payoffs and that from maintaining $T_{i,0}$ exceeds $\alpha$ that it becomes beneficial to adjust the tax rate. This occurs below the period 0 payoff $w_{i,0}(T_{i,0})$ and above the period 0 payoff $\tilde{w}_{i,0}(T_{i,0})$. For payoffs indicating ICO taxes between those resulting from these two information sets, i.e. those close to $T_{i,0}$, $i$ will not find that the expected benefit from adjustment exceeds its cost $\alpha$.

Finally, note that in each of the decision-making periods, the presence of adjustment costs halts changes that otherwise might occur. This is then consistent
with the stickiness of tax rates highlighted in our first stylized fact.

3.2 Learning in a Network

The above describes the decision making process of a given country $i$ in each period. The one point where other countries influence $i$’s decisions is when $i$ is choosing its period 2 tax. Recall that this is driven by the expected payoff if it switches to the ICO tax, given by $\Pi^2(T_i, 2; I_i, 2)$. In this, the tax rate of other countries impacts $i$ by influencing $i$’s conditional probabilities of the various states of nature, $f(\tilde{S}; I_i, 2)$. Recall that $I_i, 2$ contains $i$’s period 0 and 1 payoffs and the period 1 tax rates for countries in $\Omega_i$. If a country $j$ switches its tax rate from periods 0 to 1, this is based on the single piece of information available to it in period 1, its period 0 payoff $w_{j, 0}$. As such, by observing $j$’s period 1 ICO tax, $i$ can perfectly infer $j$’s period 0 payoff. If $j$ does not adjust, all $i$ can infer is that $j$’s period 1 information is in the set $\Lambda(T_{j, 0}, I_{j, 1})$, which is equivalent to a range of payoffs for which $j$ would not adjust. Define the set $\bar{\Omega}_i$ as the countries in $\Omega_i$ that adjust in period 1 and $\Omega_i$ as the set of those that do not.

With this notation, we can write $i$’s condition probability of state $\tilde{S}$ given
Consider a given country $i$. If $j$ adjusts, $i$ revises its beliefs by

$$\frac{f_w(w_{j,0}|\tilde{S},T_j,0)}{f_w(w_{j,0}|T_j,0)}$$

whereas if it does not, then $i$ revises its beliefs by

$$\frac{\pi(T_j,0)}{\int w(T_j,0) f_w(w_{j,0}|T_j,0) d\tilde{w}_{j,0}}.$$

If $\tilde{S} = S^*$, i.e. it is the true state of nature, then in expectation

$$\frac{f_w(w_{j,0}|\tilde{S},T_j,0)}{f_w(w_{j,0}|T_j,0)} \geq \frac{\pi(T_j,0)}{\int w(T_j,0) f_w(w_{j,0}|T_j,0) d\tilde{w}_{j,0}}$$

and $i$’s conditional probability that $\tilde{S} = S^*$ is greater when $j$ adjusts. If $\tilde{S} \neq S^*$, the reverse is true. In each case, when $i$’s expectations become more accurate, it is able to make better choices both by making a more informed choice on whether to adjust and, in the case of adjustment, choosing an expected welfare superior period 2 ICO tax.

Furthermore, as $j$’s range of period 0 payoffs for which it does not adjust grows, less is learned from its choice to not adjust. In the limit, as $\alpha \to \infty$, for a given country $j$:

$$\frac{\pi(T_j,0)}{\int w(T_j,0) f_w(w_{j,0}|T_j,0) d\tilde{w}_{j,0}} \to 1$$
i.e. nothing is learned from the non-adjustment of \( j \).

As long as something is learned, however, the more countries that are observed, the better.

**Lemma 2** For finite adjustment costs, expected welfare of \( i \) increases as \( \Omega_i \) gets larger.

**Proof.** As \( \Omega_i \) grows, at least one of the sets \( \overline{\Omega}_i \) and \( \Omega_i \) grows as well and, as the first period adjustment decision is independent of the taxes of others, neither set gets smaller. Therefore, the second and/or third terms in (6) will increase if \( S = S^* \) or fall if \( \hat{S} \neq S^* \). This improves the accuracy of \( i \)'s conditional probability and, as in Lemma 1, \( i \)'s expected welfare improves. When \( \alpha = \infty \), however, no country adjusts and nothing is learned by observing the taxes of others.

Furthermore, when adjustment costs are finite, as the number of countries \( i \) observes approaches infinity, \( f_S \left( \hat{S} | I_{i,2} \right) \to 1 \) if \( \hat{S} \) is the true state of nature and 0 otherwise. Intuitively, this is akin to what would happen with learning in isolation but an infinite time horizon, i.e. after observing an infinite number of its own payoffs, \( i \) would be able to deduce \( S^* \) on its own.

These two lemmas thus demonstrate the benefit of learning in a network which is the ability to refine one’s own beliefs about the true state of nature and to act accordingly. It also demonstrates that each country \( i \)'s period 1 adjustment choice has a positive, but uninternalized, externality on countries for which \( i \in \Omega_j \). The impact of this is summarized in the following proposition.

**Proposition 1** When there are no adjustment costs, learning is efficient given the network architecture. As such, aggregate expected welfare is maximized. In contrast, when adjustment costs are positive but do not prohibit all adjustment, aggregate expected welfare is between that when learning occurs in isolation and when there are no adjustment costs. As such, with adjustment costs, tax adjustments are inefficiently rare.

**Proof.** When there are no adjustment costs, each country implements its ICO tax in every period. As such, the beliefs in (6) are as accurate as possible given the constraints imposed by the information network (which limit the number of countries each \( i \) can observe). This then implies that expected welfare for each \( i \) is maximized. Summing this across \( i \) implies the maximum expected aggregate expected welfare available subject to the constraints imposed by the network architecture which governs the flow of information. When \( \alpha > 0 \), there is a positive probability that some countries will not adjust. Following Lemma 1, this means that the beliefs of others are not as accurate, leading them to choose taxes which are sub-optimal to those they would make if better information were available. Thus, in expectation, aggregate welfare will be lower. The size of this aggregate welfare loss is increasing in \( \alpha \), following on from the impact this has on the various countries’ conditional probabilities. Finally, as any learning from others provides some improvement in the accuracy of \( f_S \left( \hat{S} | I_{i,2} \right) \), in expectation,
This is given by \( \Delta f \) change in information affects \( i \) differing information content of different actions. for mobile capital or yardstick competition. Instead, it is entirely driven by the not rely on the tax in \( i \) resulting from adjustment. It is important to recognize that this externality does not rely on the tax in \( i \) impacting the payoff in \( j \) as occurs with competition for mobile capital or yardstick competition. Instead, it is entirely driven by the differing information content of different actions.

To describe this externality in more detail, consider the changes for \( i \) when \( j \) adjusts versus when it does not. These changes are wrought by a change in \( i \)'s information set from \( I_{1,2} \) when \( j \) does not adjust to \( \tilde{I}_{1,2} \) when \( j \) adjusts. This change in information affects \( i \) conditional expectations on the state of nature. This is given by \( \Delta f_S = f_S \left( \tilde{S}|I_{1,2} \right) - f_S \left( S|I_{1,2} \right) \), defined by:

\[
\Delta f_S \left( \tilde{S}|I_{1,2} \right) = \frac{\prod_{s=0}^{1} f_w \left( w_{i,s}|T_{i,s} \right)}{\prod_{s=0}^{1} f_w \left( w_{i,s}|I_{i,s} \right)} \frac{\prod_{k \in \Omega_i} f_w \left( w_{k,0}|T_{k,0} \right)}{\prod_{k \in \Omega_i} f_w \left( w_{k,0}|I_{k,0} \right)} \int_{\Omega_i} \int_{\Omega_i} f_w \left( \tilde{w}_{k,0}|T_{k,0} \right) d\tilde{w}_{k,0}
\]

i.e. beliefs change by the improvement in information given by \( j \)'s adjustment. Armed with this revised belief, \( i \) would fall into one of four groups. If \( i \in G_{i,0,0} \), then \( i \) does not adjust regardless of whether or not \( j \) adjusts. This would be the case if, both before and after the change in the conditional expectation caused by \( j \)'s adjustment, \( i \) believes that the gain from adjustment in period 2 is less than \( \alpha \). In this case, the change to \( i \)'s expected welfare from \( j \)'s adjustment, denoted by \( \Delta \Pi^2_i \) equals zero, since in both cases it continues to use a tax \( T_{i,1} \).

If \( i \in G_{i,0,yes} \), then \( i \) adjusts its tax rate in both cases. However, because it holds a different set of beliefs when \( j \) adjusts, it would choose a different ICO tax. This results in an expected welfare gain of:

\[
\Delta \Pi^2_i = \Pi^2_i \left( T_{i,2} (\tilde{I}_{i,2}) ; x_i, \tilde{I}_{i,2} \right) - \Pi^2_i \left( T_{i,2} (I_{i,2}) ; x_i, \tilde{I}_{i,2} \right) > 0 \quad (10)
\]

since \( i \)'s beliefs are (in expectation) more accurate due to the additional information gleaned from \( j \). If \( i \in G_{i,0,yes} \) then \( i \) would not be adjusting were it not due to the improved information gleaned from \( j \). This results in an expected welfare gain:

\[
\Delta \Pi^2_i = \Pi^2_i \left( T_{i,2} (\tilde{I}_{i,2}) ; x_i, \tilde{I}_{i,2} \right) - \Pi^2_i \left( T_{i,1} ; x_i, \tilde{I}_{i,2} \right) - \alpha > 0. \quad (11)
\]

\[23\text{Recognize that this gain is evaluated with the knowledge of the new information from } j, \text{ i.e. it is the gain in } i \text{'s eyes following } j \text{'s adjustment.}\]
Finally, if \(i\) would reverse a decision to adjust after observing that \(j\) did, \(i \in G_{yes, no}^j\) and it’s expected welfare improves by:

\[
\Delta \Pi_i^j = \Pi_i^j \left(T_{i,1}; x_i, \tilde{I}_{i,2}\right) - \Pi_i^j \left(T_{i,2} \left(\tilde{I}_{i,2}\right); x_i, \tilde{I}_{i,2}\right) + \alpha > 0
\]

i.e. part of the gain is the savings of the adjustment cost. Summing this across the countries in each of the groups, this then gives a measure of the additional value to to \(j\)’s adjustment which \(j\) does not internalize, i.e. the size of the externality.

A secondary result is that, because cross-country beliefs converge to the correct ones via learning in the network, the tax rates chosen will tend to exhibit conditional converge as well. This convergence is conditional in two ways. First, even with the same beliefs on the state of nature, different country characteristics \((x_i)s\) will result in different ICO taxes. Second, when \(\alpha > 0\), not all countries will find it worthwhile to adjust even as their beliefs are updated. Nevertheless, in our model this conditional convergence is not the result of a welfare-reducing race to the bottom in taxes. Instead, it is the sign of a welfare-improving information exchange that leads countries to converge in ICO taxes.\(^{24}\) This convergence is consistent with that discussed under our second stylized fact. In particular, countries with initially extreme tax rates (again, extreme conditional on their characteristics) would be those most likely to adjust. Furthermore, if in reality the true state of nature called for a low tax rate, this would be consistent with the overall downward trend and the greatest movements by countries with initially high tax rates. Finally, because taxes in \(\Omega_i\) convey information, which then drives \(i\)’s tax, this would result in a positive spatial lag coefficient as high tax countries (where nearly by definition the spatial lag tends to be relatively small) cut their rates. This would therefore be a new explanation for the tax rate interdependence discussed in the third stylized fact.

3.3 Incompletely connected networks

In the above analysis, we remained agnostic about the network architecture, that is, the precise interrelation of the \(\Omega_i, s\). An obvious baseline is a network in which \(\Omega_i\) contains \(N-1\) countries for all \(i\), i.e. each country observes all others, making the information network a completely connected on. Alternatively, we can assume that each country only observes an exogenous subset of the other countries, i.e. \(\Omega_i\) is smaller than the set of countries for at least some \(i\). This is known as an incompletely connected network (Gale and Kariv, 2003). This network structure can result from either a limited flow of information (such as when a country observes only the tax rates of its neighbors) or a limited ability to process the incoming information (such as the rational inattention literature of Sims (1998, 2003) and others).\(^{25}\) In the extreme, the set of countries is simply

\(^{24}\)In the extreme, with no variation in \(x_i\)s and no adjustment costs, from period 2 on all countries would set the same tax rate which is also the one that an aggregate welfare maximizer would choose.

\(^{25}\)One example of the costs of processing information and its impact on tax policy is found in information exchange, the process by which one tax authority provides information to another.
Figure 4: Incompletely connected networks

divided into a collection of separate, non-overlapping groups in which countries observe in-group actions, but not out-of-group actions, i.e. the network is completely connected within a group but unconnected across groups. Here, the situation is simply a repeated version of the one discussed above with the prediction that what transpires within one peer group has no impact on what happens elsewhere. Indeed, Klemm and van Parys (2012), who estimate responsiveness of the tax incentives of one nation to those elsewhere find evidence of diffusion within Latin America and Africa, but not across regions. Thus, if geography inhibits the flow of information, then our model could result in such a pattern of tax policy responsiveness. These separate groups could also arise from other factors such as the level of development or ideology that inhibit the ability of a member of one group to collect or interpret the information from a member of another group.

Alternatively suppose that the $\Omega_i$s intersect with one another. A number of different network architectures can then be considered, with the nature of the network governing who learns from who, resulting in different patterns. Consider, for instance, the network displayed on the left hand side in Figure 4. Here, the $N$ countries are distributed on a circle and observe only the actions of their neighbors. More concretely, let France (A) observe Germany (B) and Spain (E), Germany (B) observe France (A) and Austria (C), and so forth. Starting in period 1, France’s tax rate is based solely on its own experience, $(w_{A,0}, T_{A,0})$, and the \textit{a priori} distribution of $S$. In period 2, France has observed the first period tax rates in Germany and Spain, but not Austria. As such, when updating its beliefs and choosing its period 2 tax, France accounts for what is learned from its neighbors but not from distant countries. This implies that the tax rate set in France in $t = 2$ will be a function of its tax rate in $t = 0, 1$ and the tax rates of its neighbors in $t = 1$. From the perspective of the empirical literature, such an incompletely connected network would lead to spatial lags that include only the policies of neighboring countries, as done by Altshuler and Goodspeed (2007).

An alternative network structure is depicted on the right hand side of Figure regarding a multinational active in both. Although intended to curb tax evasion, the cost of collecting and translating information can be a significant barrier. Vann (1996) and Krabbe (1996) provide additional discussion.
4. While the circle on the left hand side implies a symmetry across countries, the line on the right implies that there are dead ends (“peripheries”) on the left and on the right. In period 2, countries C, B and D have access to more information than do countries A and E. As such, these countries will generally have more accurate beliefs and higher expected welfare.

A twist on the incomplete network structure is one with overlapping networks. Figure 5 presents examples which build from the circular structure of Figure 4a. In Figure 5a, A is a large country surrounded by smaller countries B to F. Whereas A can observe every other country’s actions, the smaller ones only observe their direct neighbours including country A. Again, there will be different learning speeds and differently accurate beliefs. Country A has the most and best (i.e. most recent) information. Therefore, each of the small countries will base its beliefs more heavily on the large country’s past actions than on those of other small countries. In this sense, the hub nation A becomes acts as a “leader” and has a greater impact on the policies of other countries. Altshuler and Goodspeed (2007) find that other OECD countries’ corporate taxes respond more to the US tax rate than to that of neighboring countries. This would be consistent with the type of structure in Figure 5a with the US acting as a hub. An implication of this is that, all else equal, the social costs of non-adjustment are higher for hub countries because their information is used by a relatively large number of countries.

In Figure 5b, a multi-centric world is depicted with three large countries (A, B and C) each of which has its own “backyard” or subnetwork (a1, a2 and so on). Whereas the large countries observe each others’ actions and their subnetwork, the smaller backyard countries only observe each other and the large country in their subnetwork. Such a setting could occur when observation capacity is linked to trade flows. Whereas small countries only trade with the largest country in the region, the large countries engage in interregional trade. As such, the larger countries have access to more information than do the smaller nations. From an empirical point of view, spatial lags could be weighted with trade flows in order to capture connectedness.26 Alternatively, as trade flows are correlated with distance, this architecture might suggest the use of a spatial lag constructed

---

26 Although we do not know of such an estimation, Exbrayat (2009) uses a measure of overall trade openness as a weighting scheme.
with weights that are decreasing in distance, something done by Davies and Vadlamannati (2013) and others. In particular, Davies and Naughton (forthcoming) discuss the different implications of various distance-based weighting scheme in their study of environmental policy diffusion.

Regardless of the precise structure of the network, however, it is important to remember that because the model did not impose complete connectivity, the main results hold. First, because even in an incompletely connected network more information arrives to each $i$ than in isolation, learning in the group is welfare improving. However, since less information arrives each period relative to the completely connected setting, learning is less beneficial than in a completely connected network.\footnote{Gale and Kariv (2003) provide simulation results that compare the speed of convergence.}

Another important difference relative to the case of completely connected networks, however, is that the accuracy of beliefs may differ across countries. Second, in the presence of positive adjustment costs, there is underadjustment relative to the social optimum due to the uninternalized externality of conveying information. Finally, as this discussion shows, by choosing a particular information network architecture, it is possible to obtain tax rate interdependence consistent with estimates from the different weighting schemes used by the spatial econometric literature.

**Non-shared states of nature** In the above, we assumed that all countries in a network shared the same state of nature. Alternatively, assume that the realized states of nature $S^*_i$ are not identical across countries, but are correlated with one another, though imperfectly.\footnote{If the country-specific states of nature are independently distributed, then no country learns anything of use by observing the actions of others. As such, the game is equivalent to one of learning in isolation.} The $S^*_i$s are assumed to be jointly distributed according to some joint cumulated distribution function $F(S_1, \ldots, S_N)$. Observing the behavior of other countries still conveys useful information but due to the imperfect correlation of states across countries, the informational content is limited. Nevertheless, as $N_i \to \infty$ and/or $t \to \infty$, learning remains perfect and each country’s tax will converge to $T^*_i(x_i, S^*_i)$ (up to the limits created by adjustment costs). The speed of this convergence will lie between that when learning in isolation and when $S$ is common across countries. As such, there are still welfare gains relative to learning in isolation and an unrealized positive externality from adjustment.

Another implication is that countries will extract more information from countries whose states of nature are more correlated with their own. For example, states of nature may be more correlated between proximate countries than distant ones. This would provide a rationale for using a distance-based weighting matrix when estimating policy diffusion even in a completely connected network. Alternatively, the correlation between the $S^*_i$s could be greater between those with more similar $x_i$s, which could include factors such as ideology. This would suggest the use of political similarity weights, something utilized in Davies and Klasen (2013), who use a measure of political affinity based on United Nations
voting as a weighting scheme for their study of overseas development assistance donations.

Taking this idea a step farther, one can imagine a setting in which a subset of nations, for instance the EU, share a common state of nature but non-members have individual states. In this case, although all countries learn from each other, there is a distinction between what is learned from an EU and a non-EU country. This would then fit the pattern found by Redoano (2007) and Davies and Vogel (2009), who find that while non-EU members respond equally to the corporate taxes of both EU and non-EU countries, members respond less to the taxes of non-members than they do to members. In any case, the primary results hold, i.e. group learning increases welfare over the isolated learning case but happens inefficiently often in the presence of adjustment costs.

One-way information flows  In deriving our results, we did not require that if \( j \in \Omega_i \) then \( i \in \Omega_j \) as well, i.e. that information flows are bilateral. Although that assumption may fit many cases, it is instructive to consider the alternative in which for some network linkages information flows move in only one direction. For example, we might assume that countries can only observe their two most important partners’ actions which, in a multilateral world need not align across countries. For example, the US and the UK are Ireland’s primary trading partners, however, Ireland accounts for a much smaller share of trade for those two nations. If importance is associated with trade, learning would be unilateral for some links.\(^{29}\) Part c) of Figure 5 illustrates a case in which the three large countries observe each others’ actions but not those of smaller countries in their own subnetwork. The smaller countries, however, observe all of the actions in their subnetwork. This provides another avenue for asymmetries in beliefs driven by the flow of information.\(^{30}\) Nevertheless, the main results continue to hold. The primary distinction is that learning, and the benefits thereby, can vary by country. Furthermore, the potential for inefficiency is greatest for such a large, important country, i.e. if the US does not adjust in response to its updated information, this has the potential to impact a much greater number of nations (quite possibly including other large countries) than when a small country such as Ireland does not adjust.

---

\(^{29}\) Alternatively, the resources available to process information can vary across countries. If such resources are more limited for small countries than large ones, even if a large and a small country observe one another’s actions, only the large country would be able to process the information, making the information flow effectively one-way.

\(^{30}\) A variant of this model could capture the relationship between the OECD (which is observed by all) and the non-OECD (who are only observed by one another). In this case, all nations would update their beliefs in response to OECD actions but OECD nations would not respond to the policies of non-OECD countries. This is the pattern found by Davies and Naughton (forthcoming), who estimate the participation in international environmental agreements. Although they find that all countries respond to the participation of OECD countries, the same cannot always be said of non-OECD countries’ treaty participation decisions. In particular, for some specifications, although all countries respond to the OECD decisions, only non-OECD nations respond to non-OECD decisions in line with the predictions of this architecture.
3.4 Longer Time Horizons

In the above model, because $S^*$ is revealed in period 3 and tax rates are observed by others only in the period after they are set, period 2 is the only one in which learning from others occurs. If, however, the game ends in a later period $t > 3$, the model can be easily extended. The difference this causes is that it permits multiple periods of learning with each period working similar to that above – in each period $t > 1$, for countries in $\Omega_i$, $i$ is able to infer their $t - 1$ beliefs on the state of nature, perfectly so if the observed country adjusted its policy in $t - 1$ and imperfectly if it did not. Unlike the relatively tidy formulation in (6) for $i$’s conditional probability, it is necessary to account for the fact that, whenever an observed country $j$ has a period of non-adjustment starting in a period $t$ and ending in $t + k$, this implies that it is no longer to precisely intuit its stream of payoffs from $t$ to $t + k$. Once $j$ adjusts, it is possible to back out its beliefs but not the precise set of $j$’s payoffs during the non-adjustment period. Nevertheless, it remains the case that $i$ can use the information gleaned from tax rates during the periods prior to $t$ to update its beliefs and act accordingly. Therefore, the results above carry through.

In the limit as $t \to \infty$, the primary difference is that the state of nature is never revealed. Nevertheless, beliefs converge on the true state of nature. If learning occurs in isolation, as the number of own payoffs received approaches infinity, each $i$ is able to deduce the true state of nature. With learning in a network, however, this convergence is even faster. When $\alpha = 0$, this is the setting of Gale and Kariv (2003) who demonstrate that convergence occurs in a finite amount of time. With positive adjustment costs, as the flow of information is slower, so too is convergence. Nevertheless, as the number of pieces of information available to $i$ in $t$ is $t + (t - 1)N_i$, which is greater than learning in isolation (where it is simply $t$), learning in the network is faster than in isolation, i.e. it takes fewer periods to obtain the same amount of information. As such, because beliefs are on average more accurate when learning in a network as compared to in isolation, expected individual and aggregate welfare are higher when learning in a network than in isolation. Finally, as in the simple baseline model, because countries do not internalize the impact of their adjustment on others, in the presence of adjustment costs, learning is slower than without them and there is an inefficiently low number of tax rate adjustments. As a final note, if in a variant of the networks in Figure 4 distant countries are observed with a longer time delay, this would provide another rational for distance-weighted spatial lags as distant tax rates observed in period $t$ would be older and contain less information than proximate taxes observed in the same period.

31 Gale and Kariv (2003) provide simulation results that compare the speed of convergence.
4 Conclusion

The two leading traditional approaches to explain spatially correlated tax policies are competition for mobile resources and yardstick competition. The goal of this paper has been to offer a third mechanism by which the policies chosen in one country can be influenced by those elsewhere. This mechanism operates via the information that is embodied in a country’s policy choice. When the optimal policy depends on an unobserved state of nature which is correlated across nations, by observing the choices made in another location, a given country is able to refine its beliefs and base its decisions upon this. As a result, the policies of countries will converge on one another (up to the limits created by other differences across nations and potential adjustment costs). This then mirrors the patterns found in the empirical literature. A second implication, and one that mirrors the primary concern in the tax competition literature, is that the equilibrium pattern of adjustments can be inefficiently slow. This is because a given country does not internalize the benefits accruing to others from an adjustment in its own policy.

An important aspect of our approach is that the policies of one country do not influence the mapping between policy and payoffs elsewhere. This is then in direct opposition to the mechanisms resulting in correlated policies in the tax competition or yardstick competition literatures. A consequence of this difference is rather than tax-rate convergence resulting from inefficient competition, in our model it is a sign of welfare-improving learning. This then has very different implications when designing overarching policy recommendations. In particular, if efforts such as tax harmonization, minimum applicable taxes, or a common consolidated corporate tax base inhibit tax rate adjustment, this inhibits the flow of information in our setting. As a result, it slows learning and results in lower expected welfare that when countries are free(er) to adjust their policies.

Finally, we show that the model is capable of generating the same pattern in taxes as is found in the empirical literature. This then adds to the challenge for the empirical literature which has yet to distinguish between the competition for mobile capital and yardstick competition motives for tax interrelations by calling for the need to distinguish both of these from a learning motive. In light of the potentially very different policy prescriptions across motives, at the very least it suggests a note of caution in interpreting the existing empirical results as pointing towards the need for something like tax harmonization. That said, this is in no way intended to suggest that the two competition mechanisms do not occur. Indeed, it is our supposition that all three sources of international policy diffusion jointly drive the patterns observed in the data. Therefore, our goal has been to illuminate the additional interactions and the role that network architecture plays in those which result from social learning.

Finally, although our focus here has been on international policy diffusion and tax policy in particular, the model can be applied to a broad category of situations in which learning occurs in a network and adjustment costs are present. For example, one can consider a number of firms, each a monopolist in
its own product, who are considering where to locate an affiliate when relocation comes at a cost. If a given firm observes a firm re-locating to, say, China, this could lead it to revise its belief about Chinese productivity upwards. However, if it only observes that the other firm did not relocate, it can merely surmise that that other firm’s belief is such that any potential productivity improvement does not outweigh the cost of relocation. This would have similar predictions and implications as our tax model: that learning from others improves one’s location choice, that adjustment costs inhibit the flow of information, and that locations would converge, achieving agglomeration patterns without the need for cross-firm spillovers. A similar story can be told for worker re-location choices. Therefore, although the present application of our model has been rather specific, the setup it provides is useful in a number of settings including those well outside the realm of policy setting.

References


