# Business News, Bayesian Updating, and Bilateral Trade *

Gabriel J. Felbermayr and Benjamin Jung†
University of Tübingen

February, 2007

**Abstract**

We use 15 years of Reuters-Business-Briefing (RBB) data in an empirical gravity model to analyze how content and number of business news about potential trading partners' countries affect the volume of bilateral trade. We motivate the exercise by a theoretical model, where exporters sink fixed costs before engaging into a trade relationship. On the export market, they are exposed to expropriation risk; hence, they need to form expectations about the shirking probability. We hypothesize that exporters learn in a Bayesian way from news of other agents' experiences in the respective country. Priors on the shirking behavior are especially relevant if business partners are located in the south. Our regressions in first differences robustly indicate that the shirking probability corresponds to an ad valorem tariff equivalent of 3 to 7 percent. Trade flows adjust to unanticipated information more stronger if its relative precision is high.

`JEL Codes`: F12, C23, D83

`Keywords`: Gravity Equation, Informational Frictions, Bayesian Updating.

---

*We are grateful to Gary King for sharing his data, to Peter Egger, Hartmut Egger, Robert Feenstra, Wilhelm Kohler, Josef Falkinger, Johannes Moenius, Farid Toubal, Udo Kreikemeier, Ingo Borchert, and other seminar participants at the EEA-ESEM conference 2006, at the Munich-Tubingen workshop in international economics, at ETH Zurich, and at the universities of Zurich and Nottingham for comments and suggestions, and to Walter Assenmacher and Robert Jung for help with statistical theory. We are responsible for all remaining errors. Part of this paper was written when Felbermayr was visiting Zurich university and Jung the university of Nottingham.

†Address: Nauklerstraße 47, 72074 Tübingen, Germany. Tel.: +49 7071 29 78183. Email: gabriel.felbermayr@uni-tuebingen.de, benjamin.jung@uni-tuebingen.de.

# 1 Introduction

Trade costs are of economically sensible magnitudes and matter massively for trade patterns and volumes. Measuring and decomposing them into meaningful components is a challenging task. In the gravity context trade costs are modeled as a measure of ignorance in the sense that researchers usually do not know what hides behind it. Obstfeld and Rogoff (2000) argue that trade costs are important in addressing major puzzles in international macroeconomics.

In their comprehensive survey on trade costs Anderson and van Wincoop (2004) provide a rough break down of non-tariff barriers into categories like information cost barriers, policy barriers, and security security barriers. They suggest that the information cost barrier is equivalent to a 6 percent ad valorem tariff, i.e. it accounts for approximately 14 percent of the tariff equivalent of non-tariff barriers. Anderson and van Wincoop (2004) also argue that trade barriers in developing countries are higher than those reported for industrialized countries.

A number of indirect measures for the information cost barrier has been put forward. Rauch and Trindade (2000) emphasize the role of newsletters devoted to international trade to overcome informational barriers. Recent papers stress the importance of social and ethnic networks as catalysts of information and relate variation in the existence of these networks to the pattern and volume of trade. Casella and Rauch (2002) and Rauch and Trindade (2003) develop theoretical frameworks, while Combes et al. (2005) find empirical support for the network idea. Rose (2006) and Nitsch (2007) point out that diplomatic activities ease the flow of information.

So far the literature has focused on the availability of information. However, newsletters on trading opportunities as such are worthless unless the reported news are not translated into information. An ethnic network is not formed all of a sudden by sending migrants to another country, but by the active interchange of information about the other country. But how is relevant information about the potential trading partners obtained? We provide a framework where the flow of publicly available business news is translated into a stock of decision relevant know how. For this purpose we integrate a monopolistic competition trade model with a Bayesian theory of information updating. This allows to construct a direct measure of information from observed business news that proxies bilateral business climate. We address the question, how the content of information affects bilateral trade flows in terms of ad valorem tariff equivalents. Moreover, we decompose this effect into a response to unanticipated information, i.e. the derivation of actual observed behavior from the forecast, and its relative precision, proxied by the number of underlying events. Our setup particularly accomodates trade with delevoping countries where exporters are likely to know little about potential partners and therefore bearing a high risk of being exproproiated. Moreover, we expect the business climate to have an impact only when differentiated goods are concerned, as homogeneous goods are much easier to be sold to buyers in another country.

Why does information matter at all and what is it all about? Transactions across borders often require irreversible transaction-specific investment prior to the resolution of uncertainty about the behavior of the trading partner. Exporters are therefore susceptible to ex post expropriation by their associates such that standard holdup problems arise. We argue that varieties are heterogeneous with respect to ficed export costs, while the nature of foreign behavior (un-cooperative or cooperative) is dyadic in nature and modeled as part of variable trade costs. This gives rise to a channel for the relevance of business news for bilateral trade patterns and volumes since publicly available business news – along with other sources of information – may help agents to form priors about the probability of being expropriated (shirking probability).

The measure of information is obtained by exploiting data on number and content of business news occurring between two countries. The data set has recently been compiled by The Kansas Event Data System (KEDS), funded by the U.S. National Science Foundation, and made available and documented at Gary King's homepage at Harvard. It is generated by machine-coding 10 millions of events recorded and disseminated internationally by Reuters Business Briefings (RBB). The data goes beyond conventional measures on the quality of diplomatic relations between two countries in that it includes real business news. The dyadic nature of the data and its availability over a time span of fifteen years (1990-2004) make it particularly attractive for use in a standard gravity model of bilateral international trade.

The paper makes the following contributions. It provides theoretical underpinning for the relevance of information in trade relationships and derives a mapping of business news into decision relevant information from Bayesian theory. The framework is used to perform a partial equilibrium analysis of bilateral trade flows, where we link a new data base that exploits real business news from Reuters Business Briefing and the COMTRADE data base. Aggregating individual behavior to the national level, our results suggest that the shirking probability forecasts as derived from the RBB matter for the evolution of bilateral trade volumes over time and space. In our preferred specification, an one standard-deviation increase in the shirking probability expected by exporters corresponds to an ad valorem tariff equivalent of about 3 to 7 percent on differentiated goods. Moreover, we detect robust interactions between unanticipated information and its relative precision consistent with our theoretical learning hypothesis. We further find evidence that the response to unanticipated information is especially large if the trading partner is located in the south and the relative precision is high.

The remainder of the paper is organized as follows. In Section 2 we derive the gravity equation in the presence of heterogeneous export market entry costs and uncertainty about variable trade costs, where agents update their prior beliefs according to the Bayesian learning hypothesis. A heuristic overview over the data set and a discusses the empirical strategy is provided in Section 3, while Section 4 presents our main results and discusses robustness checks. Section 5 concludes.

# 2 Theoretical framework

## 2.1 Economic environment

We base our theoretical framework on the model of monopolistic competition. Following the established literature (as survey by Feenstra, 2004), we assume a world with many countries. Each country has a sector for differentiated, homogeneous, and referenced priced goods.[1] Sectors are characterized by their elasticity of substitution. Each sector

[1]Rauch (1999) develops a classifiation that sorts industries into three categories: homogeneous goods which are traded on organized exchange, referenced priced goods which can be quoted without stating the producing firm, and differenciated goods which fall in neither of the two categories.

has a representative consumer with Dixit-Stiglitz preferences

$$U_{jt}^{k} = \sum_{i=1}^{C} \int_{N_{ijt}^{k}} x_{ijt}^{k}(z)^{\rho^{k}} dz, \; 0 < \rho^{k} < 1, \tag{1}$$

where $j$ denotes the country of consumption, $i$ the country of production, $t$ is time (which we take as discrete), $z$ describes a generic variety, and $\rho^k$ governs the degree of substitutability between varieties in sector $k = 1, 2, 3$. Sectoral utility indices $U_{jt}^k$ are nested in a Cobb-Douglas utility function with parameters $a$ and $b$ that governs the expenditure shares allocated to sector 1 and 2, respectively. In our proceeding analysis we always treat sectors separateley. Therefore we henceforth suppress the sector index $k$ and the Cobb-Douglas parameters $a$ and $b$.

The only difference to the standard specification is that not all products available in country $i$ are necessarily available in country $j$, that is $N_{ijt}$ describes the number of varieties actually exported from $i$ to $j$ at time $t$. Consumers maximize utility subject to a budget constraint that binds each period.

Under these conditions, we get the usual import demand functions

$$x_{ijt}(z) = E_{jt} P_{jt}^{\sigma-1} p_{ij}^{-\sigma}, \tag{2}$$

where $p_{ij}$ denotes the c.i.f. price of a good imported from country $i$, $E_{jt}$ is total income available in country $j$ at time $t$, and $P_{jt}$ is the price index dual to (**??**). We denote the elasticity of substitution by $\sigma = 1/(1-\rho) > 1$.

We assume that all firms in all countries share the same production technology and that labor is the only production factor. In particular, factor demand for the production of $y(z)$ units of output is given by $l(z) = \alpha + \beta y(z)$, where $\alpha$ denotes the fixed and $\beta$ the variable input requirement.

However, firms are heterogeneous with respect to the costs that they face on foreign markets. Concerning the entry process, we assume the following timing:

First, in order to enter a foreign market, firms have to *sink* fixed (beachhead) costs $c(z)$ in terms of labor, which are distributed according to some c.d.f. $G(c)$ with support on the real line. That distribution is known to all firms; moreover, $c(z)$ is a time-invariant known characteristic of the product that a firm $z$ happens to manufacture.[2]

[2]We argue that fixed export costs vary across varieties, e.g. costs of building a distribution channel

Second, after paying the entry costs, firms draw the variable distribution costs that occur when their product is sold abroad.[3] We assume that these costs take the usual iceberg form and assume that the ad valorem equivalent may then be written as $\delta_{ij} = 1 + \delta X_{ijt} \geq 1$, where $\delta$ is a constant and $X_{ij}$ is a random variable. Consistent with the structure of our data we let $X_{ij}$ be a Bernoulli variable which takes the value of unity with probability $\theta_{ij} \in \Theta = (0, 1)$ and zero with probability $1 - \theta_{ij}$. Hence, the realization of a loss in market $j$ is given by

$$\delta_{ij} \sim 1 + \delta X_{ijt}, \; X_{ijt} = \begin{cases} 1 & \text{with probability} \quad \theta_{ij} \\ 0 & \text{with probability} \quad 1 - \theta_{ij} \end{cases} . \tag{3}$$

Firms know that $X_{ij}$ is drawn from a Bernoulli distribution, but they do not know its parameter $\theta_{ij}$. We assume that $\theta_{ij}$ is constant, so that stationary learning through own experience and exposure to business news is possible. We also assume that $X_{ij}$ is a dyadic variable rather than a variable specific to the import market and is uncorrelated to the beachhead costs $c$.

Our setup implies that a share $\delta X_{ijt}$ of variety shipped from $i$ to $j$ at time $t$ is lost in the transaction. This waste relates to frictions with foreign trade partners or foreign authorities. A holdup problem arises as the exporter undertakes irreversible investment that makes him susceptible to ex post expropriation by business associates. Our preferred interpretation of *shirking* is the following. Exporting requires a partner in the foreign country. The firm knows that sinking the beachhead costs buys it a local partner. But it is uncertain about the quality of the partner. The foreign partner can cheat in numerous ways, ranging from not providing the promised marketing effort to downright theft. Another interpretation may be that it is unclear ex ante what kind of treatment an exporter receives from political authorities. The advantage of this setup is that we do not have to model in detail the contractual complexities that govern the relationship between importers and foreign producers.[4]

---

for laptops is rather different from that of, say, clothing apparel.

[3]Note that this setup has a similar structure to Schmitt and Yu (2001) who introduce firm-specific fixed costs of export to a standard monopolistic competition model. It is reverse to Melitz (2003) where all firms pay the same market entry costs but then draw different productivities.

[4]Shirking may also be possible if distribution in the foreign market is organized by a foreign affiliation. First, the ownership structure of the foreign affiliation is likely to include local partners. Second, also for a fully owned affiliation a principal-agent problem remains.

We implicitly assume that there are no reputational effects of existing trading relationships. Moreover, exporters do not have own experience regarding the reliability of the business partner. Besedes (2006) points out that it is most likely that trade relationships last just one year, especially in the case of developing countries. Hence, these assumptions may hold in particular if the trading partner is located in the south. In our empirical analysis we will put special interest on that topic.

One can easily argue that trust especially matters in cases where exporters deal with differentiated rather than with homogeneous or reference priced goods. Relationship specific investment may be higher and are surely sunk for differentiated goods since they have to address importers' very specific needs and therefore cannot be diverted to other customers. Hence, we expect shirking to play a substantial role in the case differentiated goods while the impact on trade in homogeneous goods should be much smaller if at all present.

Total trade costs $T_{ijt}(z)$ include iceberg transportation costs $\tau_{ijt} \geq 1$, which are identical over varieties and known ex ante. Hence, we have $T_{ijt}(z) = \tau_{ijt}\delta_{ijt}(z)$. We index $\delta$ by $z$ to make clear that producers are ex post heterogeneous in terms of their distribution costs.

The ex factory (f.o.b.) prices that producers set are identical across varieties, time-invariant, and equal to $p_{it}(z) = w_{it}\beta/\rho$, where $1/\rho > 1$ is the usual markup over variable costs. However, in the foreign market, the c.i.f. price depends on $z$ and is equal to $p_{ij}(z) = \tau_{ijt}\delta_{ijt}(z)\, w_{it}\beta/\rho$. Using the demand function $x_{ijt}(z)$, we can now write the additional profits that a producer $z$ with distribution costs $\delta_{ijt}(z)$ makes on the foreign market

$$\pi_{ijt}(z) = [\tau_{ijt}\delta_{ijt}(z)\, w_{it}]^{1-\sigma} B_{jt}, \tag{4}$$

where $B_{jt} \equiv E_{jt}P_{jt}^{\sigma-1}(\beta/\rho)^{1-\sigma}(1-\rho)$. Hence, extra profits from selling abroad are strictly decreasing in trade costs and the domestic wage rate.

## 2.2 Bayesian learning of the shirking probability

We now define more precisely, in which type of information we are interested in, and how to obtain it from business news. In general, information refers to knowledge, i.e. accumulated data (stock magnitude), as well as to news, i.e. the increment to acculumated data (flow magnitude).[5] In our setup, the probability $\theta_{ij}$ that an importer in country $j$ misbehaves towards the country $i$ producer is unknown. Agents do not know the true shirking probability $\theta_{ij}$, but have some prior beliefs, which allows them to evaluate which elements in the parameter space $\Theta$ are more and less likely respectively. Thus, agents need *information* on the shirking probability to reduce their uncertainty about it. Obviously, agents do not directly receive *news* that point to the true shirking probabilities as such. They rather observe a set of *business news* that reports several realizations of $X_{ij}$. We discuss below how agents can draw information about the true shirking probability and its precision from the business news they observe to update their prior beliefs according to the Bayesian rule.[6] Note that in our framework private and public information coincide, since the signal is derived from commonly observed business news.[7]

Let $f(\theta_{ij})$ denote exporters' initial homogeneous prior beliefs about $\theta_{ij}$. According to the binary structure of the underlying process $X_{ij}$ , $f(\theta_{ij})$ is a Beta distribution.[8] The initial mean forecast $\theta^F_{ij}$ and the *precision* of the mean forecast can be translated into the two parameters that governs the Beta distribution $r^A_{ij0}$,$n^A_{ij0} - r^A_{ij0} \geq 1$, and which can be interpreted as business news equivalents. Thus, mean forecast and precision are equivalent to observing $r^A_{ij}$ non-cooperative out of $n^A_{ij}$ events. In slight abuse of notation, we refer to $n^X_{ij}, X \in \{A, F\}$ as the precision, where $A$ stands for actual observed and $F$ for forecast.

Adopting a discrete view on time, we have the following sequencing. At the beginning of time ($t = 0$), agents share common prior beliefs in terms of business news equivalents $r^F_{ij0}$ and $n^F_{ij0}$, which are taken as exogenous.[9] At the beginning of period $t$, exporters form a prior $\theta^F_{ijt}$ based on business news dated $0 < s < t$ and their initial prior. The superscript $F$ indicates that this prior belief is the forecast for the next period. Since the true parameter

[5]See Hirshleifer and Riley (1992).

[6]Note, that news refers to a signal that contains valuable information on the true shirking probability and lies in the parameter space $\Theta$. It is constructed from business news, that is a vector of binary entries.

[7]See Hautsch and Hess (2005) for an application of common rather than individual learning to finance.

[8]See Clemen (1952) for the link between process and probability distribution. See Appendix A for a detailed discussion of the Beta distribution and its moments.

[9]The assumption of a common prior is usual in the learning literature, see Clemen (1952). We will discuss the construction of the prior below.

$\theta_{ij}$ is stationary, exports do not only rely on information of last period events, but also take into account all previous ones. By the law of large numbers (under the assumption of independent events), any observation of business news enlarges the sample that exporters base their priors on and makes their estimates ever more precise. Asymptotically, the estimated shirking probability gets arbitrarily close to the true parameter. Thus, at the beginning of period $t$, the set of observed business events $\mathcal{F}_{ijt}$ consists of the number of *reported* non-cooperative events (shirking) $r_{ijs}$ and the overall *reported* number of events $n_{ijs}$ for all periods $s < t$

$$\mathcal{F}_{ijt} = \left\{ \{r^A_{ijs}\}_{s<t}, \{n^A_{ijs}\}_{s<t} \right\} . \tag{5}$$

Then $n^F_{ijt} = \sum_{s<t} n^A_{ijs}$ is a measure of the precision of the set of the business news at the beginning of period $t$, which is equivalent to the precision of the mean forecast.

At the end of period $t$, $n^A_{ijt}$ news about non-cooperative $\left(r^A_{ijt}\right)$ and cooperative $\left(n^A_{ijt} - r^A_{ijt}\right)$ events arrive. We assume that news are fully informative and come from a free source. Based on this news, exporters calculate an actual observed shirking probability $\theta^A_{ijt}$. Furthermore, priors will be updated, and the posterior from period $t$ becomes the prior for period $t + 1$.[10]

At this stage, we do not model which type of events the business news agent finds optimal to report. For empirical reasons, we simply assume that the business news agent randomly reports some events from the total set of bilateral events that occur at some time $t$. Hence, we conclude that the number of reported events in some period $s$, $n_{ijs}$, is proportional to the number of firms that are active in country $j$, i.e. $n_{ijs} = \chi N_{ijs}$, where $\chi < 1$ is a constant.

We can now make different conjectures on the composition of the sample that exporters base their learning on. We have assumed above that distribution costs are bilateral in nature. However, if country $j's$ behavior towards exporters from countries different from $i$, covaries positively with its behavior towards $i$, firms in $i$ should use not only bilateral business events to, but all the data on $j$ that they can take. For the time being, we assume that the covariance between the process that governs $X_{ij}$ and the one that governs $X_{i'\neq i,j}$

[10]We assume that the true process that governs the shirking probability is time-invariant. This facilitates calculations, but is stronger an assumption than needed. We need that the true process governing $\theta_{ijt}$ has sufficient persistence, so that learning based on past news is possible and meaningful.

is zero, so that only bilateral business news is informative.

We may now state the following proposition.

**Proposition 1.** *Assume that the true data generating process is binary (non-cooperative versus cooperative behavior), strictly bilateral, and captured by an AR(1) parameter $\theta_{ij}$ . Given the information set $\mathcal{F}_{ijt}$, application of Bayes' theorem yields the following prior at the beginning of period t:*

$$\theta^F_{ijt} = \frac{\sum_{s<t} r_{ijs}}{\sum_{s<t} n_{ijs}} \tag{6}$$

**Proof.** See the appendix.■

This result is not surprising since the assumption of stationarity implies that all past business news are equally informative. One can now construct the law of motion of the mean forecast.

**Proposition 2.** *Under the assumptions stated in Proposition 1, the mean forecast evolves according to the following equation*

$$\theta^F_{ij,t+1} = \theta^P_{ijt} = \left(1 - \gamma_{ijt}\right) \theta^F_{ijt} + \gamma_{ijt}\theta^A_{ijt}, \tag{7}$$

*where $\gamma_{ijt} = n^A_{ijt}/\left(n^A_{ijt} + n^F_{ijt}\right)$ measures the relative precision of arriving information.*
**Proof.** See the appendix.■

The weight $\gamma_{ijt} \in (0,1)$ captures the relative intensity of business news of the present period, as $\gamma_{ijt} = n^A_t/\left(n^A_t + n^F_t\right)$. As intuition would suggest, the weight $\gamma_{ijt}$ of the present period decreases over time, because the number of events observed in $t$ is small as compared to business news cumulated from previous periods.

**Corollary.** *Given Proposition 2, the adjustment of the mean forecast $\Delta\theta^F_{ij,t+1}$ to the actual observed shirking probability $\theta^A_{ijt}$ can be decomposed into a response to unanticipated information $S_{ijt}$ and its relative precision $\gamma_{ijt}$.*

$$\begin{aligned} \Delta\theta^F_{ij,t+1} &= \theta^F_{ij,t+1} - \theta^F_{ijt} \\ &= \left(\theta^A_{ijt} - \theta^F_{ijt}\right) \frac{n^A_{ijt}}{n^A_{ijt} + n^F_{ijt}} \\ &= S_{ijt}\gamma_{ijt}. \end{aligned} \tag{8}$$

**Proof.** The claim follows immediately from rearranging equation (**??**).■

The first term $S_{ijt} = \theta_{ijt}^{A} - \theta_{ijt}^{F}$ evaluates the deviation of the forecast from the actual observed shirking probability. Thus, it measures unanticipated information $S_{ijt}$ (surprise) contained in the presently observed shirking probability. However, as argued above, the forecast gets arbitrarily close to the true shirking probability if the number of underlying business events is sufficiently large, while the shirking probability actually observed in period $t$ is an imprecise estimator of $\theta_{ij}$. Hence, the surprise is weighted by the relative precision $\gamma_{ijt}$, which relates the number of presently observed business events to the history of events. Analyzing the evolution of the mean forecast in differences allows to study interactions between the surprise and its relative precision, which refer to news on the shirking probability (content of information) and business news (availability of information) respectively.

Following the theory, exporters give equal weights to good $\left(\theta_{ijt}^{A} < \theta_{ijt}^{F}\right)$ and bad news $\left(\theta_{ijt}^{A} \geq \theta_{ijt}^{F}\right)$, which implies a symmetric loss function. However, in our empirical section we will go a step further and also test for asymmetric responses to good and bad news. Depending on the sign, asymmetries would point either to risk averse or to risk loving agents.

## 2.3 The producers' export market entry decision

Exports are assumed to be risk-neutral. They have to decide whether to sink the foreign market entry costs $c(z) > 0$. Note that an exporter who has decided to sink $c(z)$ never exits the foreign market, regardless what value $\delta_{ij}$ takes for her. The reason is that additional operating profits from exporting (**??**) cannot be negative, while the sunk entry costs are not relevant for her decision any more.

A firm of type $z$ decides to sink $c(z)$ and enter the foreign market $j$ if and only if expected operating profits of this alternative is equal or larger than zero, that is,

$$E_{\theta}\left\{E_{X|\theta}\left[\pi_{ijt}\left(X_{ijt}\right) - c(z)\right]\right\} \geq 0. \tag{9}$$

where $\pi_{ijt}\left(X_{ijt}\right) = \left(\tau_{ijt} w_{it}\right)^{1-\sigma}\left(1 + \delta X_{ijt}\right)^{1-\sigma} B_{jt}$. Note that the exporter first has to use available information to build an expectation about $\theta_{ij}$, the parameter that governs the distribution from which $X_{ij}$ is drawn. Given an estimate for this parameter, she computes the expected value of the objective function. This implies an expression for the cutoff entry cost $c^*_{ijt}$ which identifies that firm that is just indifferent between entering the foreign market and not doing so (For a detailed derivation see Appendix B) :

$$c^*_{ijt} = (\tau_{ijt} w_{it})^{1-\sigma} B_{jt} \Theta^F_{ijt}, \tag{10}$$

where we use the shorthand $\Theta^F_{ijt} = \left[(1+\delta)^{1-\sigma} \theta^F_{ijt} + \left(1-\theta^F_{ijt}\right)\right]$. It is clear, that $c^*_{ijt}$ monotonically increases in country $j's$ market capacity $B_{jt}$, but monotonically decreases in $\theta^F_{ijt}, \tau_{ijt}, w_{it}$, and $\delta$.

The decision rule for a producer $z$ in country $i$ is: Enter market $j$ if and only if $c(z) \leq c^*_{ijt}$. Therefore, the share of of producers from country $i$ that finds it optimal to enter $j$ is given by

$$\frac{N_{ijt}}{N_{it}} = G^i\left(c^*_{ijt}\right), \tag{11}$$

where $N_{it}$ is the number of firms active in country $i$ at time $t$. We summarize our findings in the following proposition

**Proposition 3.** *The share of firms from country $i$ active in the export market $j$ always increases in country $j's$ market size $B_{jt}$ and its competitiveness $w^{-1}_{it}$. The share decreases in the estimated shirking probability $\theta^F_{ijt}$, the damage of shirking $\delta$, transportation costs $\tau_{ijt}$.*

## 2.4 A partial equilibrium gravity equation

Using c.i.f. prices, we now express the value of *total* import demand (**??**) of country $i$ for varieties from country $i$ as $E_{jt}P^{\sigma-1}_{jt}\left[\tau_{ijt}\delta_{ij}(z) w_{it}\beta/\rho\right]^{1-\sigma}$. Note that foreign sales of any firm $z$ are independent from entry costs due to the sunk nature of the latter. Given (**??**), the sample average over all exporting firms is $E_{jt}P^{\sigma-1}_{jt}\left(\tau_{ijt}w_{it}\beta/\rho\right)^{1-\sigma}\Theta_{ij}$, where we use $\Theta_{ij} \equiv (1+\delta)^{1-\sigma}\theta_{ij} + (1-\theta_{ij})$. Note that $\Theta_{ij}$ and $\Theta^F_{ijt}$ are formally equivalent, where $\Theta_{ij}$ refers to the true $\theta_{ij}$ and $\Theta^F_{ijt}$ to its estimated counterpart. Under our assumption of $c$ and $X_{ij}$ being uncorrelated, we write the average total imports of country $j$ originating from $i$ as

$$M_{jit} = \left(\frac{\beta}{\rho}\right)^{1-\sigma} N_{it} w^{1-\sigma}_{it} P^{\sigma-1}_{jt} E_{jt} \tau^{1-\sigma}_{ijt} \Theta_{ij} G^i\left(c^*_{ijt}\right), \tag{12}$$

where $c^*_{ijt}$ is determined by (**??**) and $\theta^F_{ijt}$ is given by (**??**).

To sharpen the focus, assume that beachhead costs are distributed according to the Pareto distribution $G^i(c) = (c/\bar{c}_i)^k$, with $0 < c < \bar{c}_i$ and shape parameter $k > 0$. For $k = 1$, we retrieve the uniform distribution. Recalling the functional form of the threshold $c^*_{ijt}$ (**??**), we yield the following gravity equation

$$M_{jit} \approx K e^{-\bar{\delta}\theta_{ij}} e^{-k\bar{\delta}\theta^F_{ijt}} \left( \frac{P_{jt}}{w_{it}\tau_{ijt}} \right)^{(\sigma-1)(1+k)} N_{it} E^{1+k}_{jt}, \tag{13}$$

where $K \equiv \left(\frac{\rho}{\beta}\right)^{(\sigma-1)(1+k)} \left(\frac{1-\rho}{\bar{c}_i}\right)^k$ is a constant and $\bar{\delta} = (1+\delta)^{1-\sigma} - 1$ (See Appendix A for a detailed derivation). This gravity equation differs in some important respects from the conventional formulation: the existence of an *extensive* margin increases the import elasticities of trade costs, GDP, and the exporter's productivity (captured by $w_{it}$). Moreover, the true and the estimated shirking probabilities enter the value of imports.

As Anderson and van Wincoop (2003) forcefully argue, we do not expect that elasticities of $M_{jit}$ with respect to trade costs are constant. The reason is that variable trade costs appear directly and indirectly–through $P_{jt}$–in equation (**??**). Note, however, that $P_{jt}$ depends on the true empirical shirking probability $\theta_{ij}$ and not on the ex ante conjecture $\theta^F_{ijt}$ which lies at the center of interest in this paper.

Before moving to the empirical implementation of equation (**??**), we find it worthwhile to make a few observations. First, all firms have identical domestic sales, because they have the same technology, pay the same factor prices, and charge similar prices on their goods. However, firms that produce goods that are easier to export have lower $c$ values and pass the threshold $c^*_{ijt}$ for a larger number of countries. In each country, ex ante it expects similar sales. Hence, the more countries are served, the larger the firm is.

Second, the process exhibits path dependency. On the one hand, consider a situation where the prior about cooperative behavior is over-optimistic. Then many exporters conclude that expected operating profits are non-negative, giving rise to an increase in trade volumes and processing relative more news about non-cooperative than about cooperative behavior. Relying on a larger sample of news, exporters now adjust their export behavior in the next period. On the other hand, image a situation where the prior is over-pessimistic. Then only few potential exporters decide to sink the costs, processing news about the behavior of the importer country. Thus, adjustment of exporter behavior is very slow in the case of over-pessimistic priors.

Next, there are two sources for the cost of misinformation. Either exporters sink costs to enter the foreign market and do not generate sufficient turn-over to cover it, or they do not enter the market and forgo potential operating profits.

Fourth, the formulation features the existence of a positive international externality: Larger countries have an advantage, because the relevant amount of information about trading partners in different countries will be larger and, accordingly, the beliefs will be more accurately tracking the importer's true likelihood of shirking.

Finally, from the observers (the trade analyst's) perspective, the observed volume of exports of country $i$ to $j$ is not a sufficient statistic for the degree of cooperation of importers in country $j$. The reason is that trades take place once the sunk cost is paid, regardless of whether the exporter is cheated on. Only if the number of events in collective history grows very large, the information conveyed by the volume of exports will allow conclusions on the degree of cooperation of an importer. Hence, a larger trade volume increases the precision of our exporters' estimate of the true shirking probability, but not the expected estimator $\theta^F_{ijt}$ itself.

In this paper, we investigate very specifically the effect of news and how news is processed by firms. We need not close the model to identify average effects. In principle, it is possible to endogenize income $E_{it}$, price levels $P_{it}$, the wage rate $w_{it}$ and the number of firms $N_{it}$ in each country $i$.[11] In the next section, we show how we isolate those parameters of interest.

# 3 Data and empirical strategy

This section argues that the Reuters Business Briefing (RBB) data set contains useful information that can be brought to bear on the empirical relationship implied by our theoretical framework. First, we offer a description and provide a heuristic exploration of the data of the RBB data. Second, we discuss our empirical model and address the endogeneity problems that naturally arise in our setup.

[11]See Appendix C and Anderson and van Wincoop (2003) for conditions.

### 3.1 Reuters Business Briefing and bilateral business climate

Where do firms find a reliable source for information on the business climate between two countries? Clearly, one can turn to news archives which are usually maintained by press agencies or newspapers. But how can one make sense of this gigantic stock of news, sort out the relevant data from the irrelevant?

Business leaders rely on timely and accurate information about what is going on in the markets that they are active in. While large companies sometimes provide the required news services in house, there is also a profitable market for them. After World War II, specialized corporate news services emerged, with two products leading the field: Dow Jones Interactive and Reuters Business Briefing. In 1999, Dow Jones & Company and Reuters merged their corporate news branches and founded Factiva.

The principle of business briefings is that they contain condensed and filtered business news targeted specially to high-level executives in the private sector. The data set we use draws on the Reuters Business Briefing (RBB) data which – with some modifications and under another name – are still sold after the merger. This data is attractive for our purposes precisely because business people can be expected to draw on them when they form priors about what they have to expect when they interact with transaction partners in other countries. Moreover, in that data, the formidable task of filtering relevant business news is achieved in a market environment, since the editors of business briefings have to meet their customers' demand in order to be successful.

Still, using business briefings to construct a data set that can be used in an empirical analysis of bilateral trade flows is a formidable task. Fortunately, political scientists have produced such data. In a large research project, The Kansas Event Data System (KEDS) Project has developed a program that allows automated coding of the Reuters Business Briefing data to generate event data with a (potentially) dyadic dimension.[12] The data construction tool is discussed at length in King and Lowe (2003) who argue that machine-coding is equivalent to human coding in terms of bias and better in terms of efficiency. Their data set consists of daily news that are extracted from RBB or from a precompiled

[12]Note that the data has been collected for the purpose of forecasting military and diplomatic conflicts between countries. The data therefore contains much more information than what we can make use of in the present context. See below for more details.

database (1990-05/2003), Factiva's World News (06/2003-08/2003), and Reuters World News (09/2003-2004). The availability of information has increased throughout the period due to technological change. Especially the waves of intranet (1994) and extranet (1997) led to a rise in news considered in the data files. However, this pattern holds true for all considered countries, and is controlled for with time dummies. The data is organized according to the typology of Integrated Data for Events Analysis (IDEA) that distinguishes about 200 types of events and contains detailed information on the initiator and the addressee.[13] It contains about 2.5 millions events in 1990-1994, more than 4 millions in 1995-2000, and about 3.5 millions in 2000-2004. Data and documentation are available from Gary King's homepage at Harvard (http://gking.harvard.edu/events/).

In order to better understand how Reuters Business Briefings reflect the business climate and how the data is organized, consider the following examples which directly link Factiva business news to entries in the data that reflect a cooperative and a uncooperative business environment, respectively.

> *Former foes Turkey and Syria signed a free-trade accord and said they had agreed to put their differences behind them during a visit Wednesday by Turkish Prime Minister Recep Tayyip Erdogan. ... We are in agreement, said the Syrian premier.*
> *Factiva, 22 December 2004*

The agreement (event form AGAC), which was achieved on December $22^{nd}$, 2004 (event date) between the Turkish prime and the Syrian Premier (both government agents GAGE), is obviously made to stimulate the business climate between the two countries, even if the event is not initiated by businesses.[14] As the announcement of a free-trade accord reveals the attitudes of both countries to each other, the event shows up twice in the data. Hence, it influences both the business climate for Syrian firms in Turkey and that for Turkish firms in Syria. Coding of this example is given in the following table.

---

[13] See Bond et al. (2003) for a description of the coding scheme.

[14] In our regressions we assume that adjustment of bilateral trade flows to the enhanced business environment takes place immediately after the announcement, conditional on other bilateral business events, while the direct effect of the tariff-cut is expected to be seen after the ratification procedure. However, we control for bilateral trade policy with a FTA-dummy.

Table 1: **RBB reflects business climate. Cooperation**

| Event date | Event form | Initiator name | Initiator level | Initiator sector | Addressee name | Addressee level | Addressee sector |
|---|---|---|---|---|---|---|---|
| 22-Dec-2004 | AGAC | TUR | CTRY | GAGE | SYR | CTRY | GAGE |
| 22-Dec-2004 | AGAC | SYR | CTRY | GAGE | TUR | CTRY | GAGE |

*Argentina's new president criticized Spanish business executive sharshly and to their faces, saying it was unfair for them to moan about his country's economic woes after reaping vast profits in the 1990s...*
*Factiva, 18 July 2003*

On July $18^{th}$, 2003, Factiva records how Argentina's new president, an individual (INDI) of the national executive (NEXE), blames (BLAM) Spanish business (BUSI) organizations (ORGA). Again, this example is evidence for the interplay between politicians and business agents. Hence, in order to proxy bilateral business climate, we rely on events initiated by representatives of both politics and business.

Table 2: **RBB reflects business climate. Un-Cooperation**

| Event date | Event form | Initiator name | Initiator level | Initiator sector | Addressee name | Addressee level | Addressee sector |
|---|---|---|---|---|---|---|---|
| 18-Jul-2003 | BLAM | ARG | INDI | NEXE | SPN | ORGA | BUSI |

As pointed out above, the data spans the years 1990 to 2004, and contains a total of about 10 million events. We are interested primarily in events that involve two different countries. We suppress about 9 million data points for which initiator name and addressee name coincide. Goldstein (1992) classifies all IDEA events into cooperative, non-cooperative, and neutral. Following his classification, we drop all neutral events, which reduces the number of events roughly by another 0.9 million. Moreover, we drop all events that have to do with warfare. The reason is that for those situations we usually have no trade data. Furthermore, we focus only on events where the players's countries of residence are likely to coincide with their nationality. This leads us to drop, i.a., International Organizations such as the WTO or NATO. The final number of events considered in our study is approximately 220,000.

Un-cooperative and cooperative events respectively are summed up over all event forms, sectors, and levels to construct proxies for the annual shirking probability between the public and the business sectors of two countries. As default, we use equal weights for all events, because the weight of news is implicitly given by editor's selection of news into the RBB. Using weights according to Goldstein's classification does not change our results.

## 3.2 A heuristic exploration of the RBB data

The RBB data reproduces stylized facts of cross-countries relations worldwide. To see this, we provide some details on shirking probabilities that the U.S. displays with respect to its partner countries, and vice versa. We also try to explain bilateral shirking probabilities and the number of business events in a simple regression analysis.

Table 3: **Relations with the US**

| | U.S. is initiator | | U.S. is addressee | |
|---|---|---|---|---|
| **Partner** | **Shirking Prob.** | **Business News** | **Shirking Prob.** | **Business News** |
| Canada | 0.12 | 88 | 0.20 | 101 |
| China | 0.19 | 167 | 0.26 | 182 |
| Cuba | 0.26 | 28 | 0.50 | 35 |
| France | 0.13 | 83 | 0.21 | 99 |
| United Kingdom | 0.11 | 132 | 0.14 | 148 |
| Germany | 0.11 | 87 | 0.13 | 100 |
| Israel | 0.19 | 154 | 0.12 | 125 |
| Japan | 0.14 | 212 | 0.12 | 231 |

Shirking Probability: Sample mean of actual observed shirking probabilities (share of observed un-cooperative events)
Business News: Sample mean number of annually reported bilateral events

Table **??** provides an overview of U.S. relations with some selected partner countries. It differentiates between events in which the U.S. has been the initiator and those, where the U.S. has been the addressee. The cells in the table contain averages of the actual number of reported business news $n_{ijt}^{A}$ and the implied shirking probabilities $\theta_{ijt}^{A}$. Averages are computed over years. For example, in 16 percent of all events, the U.S. behaves cooperatively with Cuba. However, the sum of cooperative and non-cooperative events averaged over years, is 28, indicating an average number of events that is much below those of other partner countries.

Two facts stand out from Table **??**. First, the United Kingdom and Germany enjoy the lowest probabilities of being shirked by the U.S. (11 percent), while the value for Cuba (16 percent) is the highest. Vice versa, the picture is very similar. However, shirking probabilities are consistently below 50 percent, which is a well noted feature of the data (King and Lowe, 2003). Second, the total volume of bilateral news is highest for Japan, followed by China and Israel.

We go on by providing a heuristic explanation of the RBB. We run pooled OLS regressions in order to derive contitional correlations between the bilateral shirking probability and number of business news derived from RBB data and bilateral trade, income, geographical and policy controls.

Colums (1) and (2) in Table **??** summarize the results for shirking probability as dependent variable. Standardized or beta coefficients are qualitatively similar for aggregated trade flows and trade in differentiated goods.[15] It turns out that trade is not significantly correlated with the shirking probability. While GDP per capita of the addressee does not play a significant role, the shirking probability is increasing in GDP per capita of the initiator. However, this effect is mainly driven by the US. Controlling for trade and distance the shirking probability between countries with a common border and a common colonial history correlates is statistically significantly higher. Common membership in clubs like the OECD or WTO does not play a role, while a common FTA improves the relation.

When explaining the number of bilateral events we find a positive correlation with aggregated trade flows and income per capita. However, the negative coefficient of addressees' per capita income points to correlation between trade and income. A common border and common colonial history significantly increase the number of bilateral events.

## 3.3 Construction of prior beliefs

How can we create a measure that reflects exporters prior beliefs on bilateral shirking probabilities given the relatively short time span of our RBB data? One option is to use the data for back casting business news and then form a prior on the basis of these back casted events. However, we are aware of the fact that our stationarity assumption may not hold for rather long periods and that exporters discount business events that lie in the very past. The problem is, that discount factors are unknown. Hence, we divide our RBB data into two parts. We form prior beliefs on the period from 1990 to 1994 and run our regression analysis on the data for 1995 to 2004 as a default.[16]

---

[15]To obtain beta coefficients, we standardize regressor and regressands. The new coefficients are $\hat{b}_j = (\hat{\sigma}_j/\hat{\sigma}_y)\,\hat{\beta}_j$ where $j = \ldots k$ refer to covariats and $y$ to the dependent variable. A one-standard deviation increase in covariate $j$ leads to a $\hat{b}_j$ standard deviation increase in $y$

[16]This is common practice in the finance literature (e.g., see Hautsch and Hess, 2005).

Table 4: **Explaining the shirking probability and number of events by bilateral trade flows and other gravity controls (Beta coefficients)**

| **Dependent variable** | **Shirking Probability** | | **No. of Events** | |
|---|---|---|---|---|
| | **Total Trade (Total)** | **Differentiated Goods (Diff.)** | **Total** | **Diff.** |
| ln Trade | -0.007 | -0.011 | 0.201 | -0.0002 |
| | (0.01) | (0.011) | (0.023) | (0.0002) |
| ln GDPPC Addressee | -0.002 | 0.000 | -0.029 | 0.0000 |
| | (0.011) | (0.002) | (0.010) | (0.0002) |
| ln GDPPC Initiator | 0.019 | 0.043 | 0.021 | 0.0003 |
| | (0.01) | (0.021) | (0.008) | (0.0001) |
| ln Distance | 0.012 | 0.014 | 0.024 | 0.0001 |
| | (0.01) | (0.014) | (0.013) | (0.0001) |
| Contiguity | 0.037 | 0.174 | 0.079 | 0.0005 |
| | (0.01) | (0.047) | (0.016) | (0.0001) |
| Colonial history | 0.045 | 0.178 | 0.050 | 0.0007 |
| | (0.011) | (0.041) | (0.020) | (0.0002) |
| Common language | -0.005 | -0.003 | 0.042 | -0.0001 |
| | (0.009) | (0.006) | (0.012) | (0.0001) |
| OECD | -0.002 | -0.001 | 0.079 | 0.0000 |
| | (0.01) | (0.006) | (0.023) | (0.0001) |
| WTO | 0.006 | 0.006 | -0.040 | 0.0001 |
| | (0.009) | (0.008) | (0.007) | (0.0001) |
| FTA | -0.018 | -0.02 | -0.012 | -0.0003 |
| | (0.009) | (0.009) | (0.012) | (0.0001) |
| F Statistic | 6.960 | 6.916 | 19.93 | 6.916 |
| $R^2$ | 0.0110 | 0.0110 | 0.0752 | 0.0110 |
| RMSE | 0.239 | 0.239 | 5.508 | 0.239 |
| Country pairs | 5045 | 4941 | 13913 | 4941 |
| Observations | 23228 | 22915 | 130167 | 22915 |

All regressions include time fixed effects and a constant (not reported). Robust standard errors in parenthesis.

### 3.4 Trade data and other data sources

The standard monopolistic competition model assumes a world with differentiated goods, characterized by their elasticity of substitution $\sigma$. In our analysis, this $\sigma$ plays a crucial role in determining the ad valorem tariff equivalent of unanticipated information. However, Anderson and van Wincoop (2004) discuss reasonable ranges for the elasticity of substitution of different goods. To sharpen the focus, we obtain trade data from COMTRADE and employ the classification introduced by Rauch (1999). He sorts four-digit SITC industries into three categories: homogeneous goods which are traded on organized exchange, referenced priced goods which can be quoted without naming the producing firm, and differentiated goods which fall in neither of the previous categories.[17] There are 131 countries in our data set.[18] However, not all of potential 131 × 130 entries are filled.

GDPs und GDP deflators come from the World Development Indiatiors, PPPs from the Penn World Tables, and geographical controls from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) in Paris.

### 3.5 Baseline regression model and extensions

In order to derive our baseline gravity equation, we specify transportation costs $\tau_{ijt}$ in a log-linear manner according to Anderson and van Wincoop (2004)

$$\tau_{ijt} = A_t Geo_{ij}^{\gamma} e^{\pi(1-Pol_{ijt})}, \pi > 0 \tag{14}$$

where $Geo_{ij}$ and $Pol_{ijt}$ are controls for geographical variables like distance and adjacency, and bilateral trade policy, respectively.

Inserting this expression the equation (**??**), and taking natural logarithms, we get

$$\ln M_{jit} = -\alpha\theta_{ijt}^{F} + \boldsymbol{\xi}\mathbf{X_{ijt}} + \nu + \nu_t + \nu_i + \nu_j + \nu_{ij} + u_{ijt}, \tag{15}$$

where $\alpha = k\bar{\delta}$, $\beta = (\sigma - 1)(1 + k)$. $\mathbf{X}_{ijt}$ collects the remaining standard gravity controls (GDPs, populations, aggregate price levels, dummy for common FTA membership). $\nu_t$

[17]Because of ambiguities Rauch (1999) developed a liberal and a conservative classification, with the former minimizing the number of differentiated goods and the latter maximizing it. We use the liberal classification as default.

[18]See Table 12 in the Appendix.

is a set of year dummies, $\nu_i$ captures exporter specific fixed effects like $\bar{c}_i$, $\nu_j$ is a set of importer fixed effects, and $\nu_{ij}$ collects dyadic fixed effects like the true shirking probability. $u_{ijt}$ is the i.i.d. error term.

We run our regressions in first differences for the following reasons.[19] First, this allows us to analyze the effect of response to unanticipated information and its relative precision in line with our Bayesian learning hypothesis.[20] Hence, we can study the interaction between the content of information (shirking probability) and its quality (relative precision). Recall that quality of information even plays a role in a risk neutral setting.

Moreover, first differencing eliminates time-invariant exporter specific, importer specific and country-pair specific characteristics. Hence, this controls for institutional quality that has not changed over time or – together with the time fixed effects – follow a common trend for all country-pairs. Also time-invariant geographical controls like distance, contiguity, colonial ties are captured.

First differencing also drops the constant. As Baier and Bergstrand (2006) we reintroduce the constant in the differenciated gravity equation of the form

$$\Delta \ln M_{jit} = -\alpha \gamma_{ij,t-1} S_{ij,t-1} + \boldsymbol{\xi} \boldsymbol{\Delta} \mathbf{X_{ijt}} + \nu + \nu_t + \Delta u_{ijt}. \tag{16}$$

However, including or excluding the constant does not change the results. In our regression analysis $\alpha$ is the parameter of interest. But what is a reasonable measure for the effect of the estimated shirking probability on bilateral trade flows? Depending on the underlying elasticity of substitution $\sigma$ and the shape parameter $k$ we can compute ad valorem tariff equivalents of no information about the shirking probability

$$\hat{\delta} = (1 - \hat{\alpha}/k)^{\frac{1}{1-\sigma}} - 1, \tag{17}$$

where we assume reasonable values for $\sigma$ and $k$. Following Anderson and van Wincoop (2004), say $\sigma$ lies between 2 and 8 for differentiated goods. $k$ captures the heterogeneity of beachhead costs. We will consider cases where the distribution is (close to) uniform.

[19]First differenced gravity equations recently have been advanced by Baier and Bergstrand (2006).

[20]We roughly follow Hautsch and Hess (2005) who study the effect of information on price responses in stock markets.

**The role of relative precision.** In line with our theory, the response to unanticipated information $S_{ijt}$ on bilateral trade flows is the stronger, the higher the precision of the actual observed shirking probability. In order to test this hypothesis, we interact the variable surprise with a dummy variable $D^{\gamma_{\text{high}}}$ which takes the value unity if the relative precision $\gamma_{ijt}$ is larger than the sample mean and zero otherwise. Hence, we construct a subsample where the relative precision of the unanticipated information is high. Correspondingly, $D^{\gamma_{\text{low}}}$ is unity if the relative precision is lower that the sample mean.

**The role of south-trade.** We have suggested above, that our theory especially applies to trade where the importer is in the south, i.e. a poor country. Hence, the interaction of surprise $S_{ijt}$ with a dummy variable $D^{\gamma_{\text{poor}}}$ which takes the value unity if the importer belongs to the 50 percent poorest countries in terms of per capita income should affect bilateral trade flows stronger than the interaction with the corresponding dummy for rich countries $D^{\gamma_{\text{rich}}}$.

However, from the comparison of precision and importer effect we cannot conclude whether asymmetries in the response are due to the former or the latter or both effects. Therefore we interact our poor and rich dummies with the precision dummies. In particular we expect that if poor/rich plays a role at all, responses should be strongest to news with relatively high precision from poor countries.

**The role of sign.** An alternative candidate for asymmetric responses to unanticipated information is its sign. As pointed out above, a stronger response to bad than to good news would indicate asymmetric loss functions and thus reveal risk aversion. We interact the variable surprise $S_{ijt}$ with a dummy for good news $D^{good} = 1$ which equals unity if $S < 0$ and zero else, and a corresponding dummy for bad news.

## 3.6 Econometric issues

In this subsection we discuss potential endogeneity of the shirking probability and dependence of business events.

**Endogeneity bias.** The time structure of the model is as follows. The current prior $\theta^F_{ijt}$, which is based on business news from previous periods $s < t$, may affect the total bilateral trade volume positively. It affects bilateral trade entirely on the extensive margin, as a higher $\theta^F_{ijt}$ increases $c^*_{ijt}$, making it optimal for more firms to be active in country $j$.

However, it does not affect the size of sales in the foreign market. The marginal firm has the same ex ante and ex post likelihood to be shirked than any other firm. In other words, given the structure of our theoretical model, the ex ante size of the trading relationship of any firm $z$ with a foreign partner is identical and independent of the aggregate volume of trade. Therefore a higher volume of bilateral trade should not affect the shirking probabilities.

Moreover, even if on the aggregate level, country $j$ has an incentive to signal a lower shirking probability to exporters in country $i$, on the disaggregate level there is a massive free-riding problem. Individual importers are too small to influence the aggregate shirking probability. This argument also implies that individual importers have no incentive to build reputation.

**Dependence of business events.** Another problem arises due to the fact that single transactions (of whatever type they may be) often are not independent from each other. One business event triggers the next so that it is difficult to exactly identify what an event is. Most event studies share this feature. We are not particularly worried by this fact, since the reappearance of the same event may indicate that this event is more important relative to the others and should therefore obtain a higher weight. If an un-cooperative event triggers a cooperative event later in the same year, there is also no problem, since the effect on the shirking probability chancels out, while the precision of information is increased.

# 4 Results and robustness checks

This section summarizes results and robustness checks of estimating the gravity equation in the presence of unanticipated information about the shirking behavior of the importer. We find robust evidence that the ad valorem tariff equivalent of non-information about the shirking probability is between 3 and 7 percent for differenciated goods.

## 4.1 Results

Table **??** contrasts the results of estimating the gravity equation with different estimators and different sets of fixed effects on our COMTRADE data set. In line with theory, coefficients for exporters and importers GDP is close to unity in the first column. Coefficients for geographical controls show expected signs and sizes. However, as well known in the literature, GDP coefficients deviate slightly from unity if we include exporter and importer fixed effects in the Pooled OLS regression.[21]

Table **??** in Appendix D shows the results for trade in differentiated goods that look qualitatively and quantitatively similar.

We now analyze how prior beliefs about bilateral shirking probabilities affect bilateral trade flows. Table **??** distinguishes between aggregated trade flows and three categories according to Rauch's (1999) classification: differentiated, homogeneous, and referenced priced goods. The effect of shirking on aggregated trade and trade in differenciated goods is statistical significant and bears the right sign, while coefficients for homogeneous and referenced priced goods are insignificant.

Assuming an elasticity of substitution $\sigma = 4$ and a uniform distribution of beachhead costs ($k = 1$), we find an ad valorem tariff equivalent of $\left((1 - 0.189)^{-1/3} - 1\times\right) 100 = 7.2$ percent for differentiated goods. However, the ad valorem tariff equivalent is decreasing in $k$ and $\sigma$, e.g. the corresponding ad valorem tariff equivalent is 3 percent if $\sigma = 8$ or 2.2 if $k = 2$. Figure shows a plausible range of the effect for a set of reasonable parameters.

According to the Bayesian learning hypothesis, the effect of the shirking probability can be dissected into a response to unanticipated information and its relative precision. The role of the surprise is statistically significant and bears the right sign in both samples (see Table **??**). As predicted, we find that the effect is stronger if relative precision is high (p-values of 0.053 and 0.076 for total trade and trade in differentiated goods, respectively) as compared to a low relative precision (p-values: 0.07 and 0.207, respectively). Moreover, we show that the theory especially applies to cases where the importer is located in the

[21] See Baier and Bergstrand (2006) for a recent discussion of panel gravity equations.

Table 5: **Baseline gravity equations: Total bilateral trade flows**

| | **Dependent variable: Total bilateral trade flows** | | | |
|---|---|---|---|---|
| | **Pooled OLS** | **Pooled OLS** | **Fixed-Effects** | **First Diff.** |
| ln GDP Exp. | 1.532 | 0.570 | 0.691 | 0.573 |
| | (0.019) | (0.098) | (0.092) | (0.13) |
| ln GDP Imp. | 1.002 | 1.367 | 1.671 | 2.239 |
| | (0.018) | (0.10) | (0.096) | (0.13) |
| ln Pop. Exp. | -0.453 | -1.202 | -0.881 | -0.707 |
| | (0.021) | (0.20) | (0.19) | (0.27) |
| ln Pop. Imp. | -0.0836 | -0.243 | -0.262 | -0.311 |
| | (0.019) | (0.18) | (0.18) | (0.25) |
| ln Price level Exp. | -1.010 | -0.675 | -0.837 | -0.619 |
| | (0.049) | (0.11) | (0.10) | (0.14) |
| ln Price level Imp. | -0.0723 | -0.710 | -0.994 | -1.453 |
| | (0.046) | (0.12) | (0.11) | (0.14) |
| ln Distance | -1.141 | -1.357 | | |
| | (0.019) | (0.024) | | |
| Contiguity | 0.660 | 0.472 | | |
| | (0.11) | (0.12) | | |
| Colonial history | 0.887 | 0.910 | | |
| | (0.097) | (0.10) | | |
| Common language | 0.776 | 0.769 | | |
| | (0.046) | (0.050) | | |
| FTA | 0.300 | 0.172 | 0.0847 | -0.0752 |
| | (0.043) | (0.045) | (0.030) | (0.030) |
| Country FE | NO | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| F | 2974 | 313.5 | 167.5 | 64.47 |
| $R^2$ | 0.715 | 0.774 | 0.0514 | 0.0126 |
| RMSE | 1.973 | 1.759 | 0.997 | 1.139 |
| Country-pairs | 13881 | 13881 | 13881 | 11377 |
| Observations | 100477 | 100477 | 100477 | 80289 |

All regressions include a constant (not reported). Robust standard errors in parenthesis.

Table 6: **Shirking probability and bilateral trade**

| | **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|---|
| | **Total Trade** | **Differentiated Goods** | **Homogeneous Goods** | **Ref. priced Goods** |
| Shirking Prob. | -0.199 | -0.189 | -0.156 | 0.196 |
| | (0.095) | (0.074) | (0.14) | (0.12) |
| Ln GDP Exp. | 0.358 | 0.599 | 0.333 | 0.372 |
| | (0.12) | (0.15) | (0.19) | (0.15) |
| Ln GDP Imp. | 1.635 | 2.027 | 1.798 | 1.679 |
| | (0.13) | (0.14) | (0.22) | (0.17) |
| Ln Pop. Exp. | 0.462 | -0.210 | -0.591 | 2.766 |
| | (0.36) | (0.31) | (0.51) | (0.52) |
| Ln Pop. Imp. | -0.0726 | -0.0240 | 1.264 | -1.095 |
| | (0.26) | (0.27) | (0.47) | (0.32) |
| Ln Price level Exp. | -0.394 | -0.624 | -0.402 | -0.499 |
| | (0.14) | (0.17) | (0.22) | (0.17) |
| Ln Price Level Imp. | -0.961 | -1.265 | -1.179 | -1.077 |
| | (0.15) | (0.15) | (0.25) | (0.19) |
| FTA | -0.0633 | -0.0561 | 0.00948 | 0.0418 |
| | (0.027) | (0.026) | (0.055) | (0.035) |
| F Statistic | 45.68 | 46.75 | 29.68 | 37.67 |
| $R^2$ | 0.0214 | 0.0220 | 0.0139 | 0.0176 |
| RMSE | 0.805 | 0.830 | 1.205 | 0.914 |
| Country-pairs | 5490 | 5404 | 4835 | 5163 |
| Observations | 42332 | 41274 | 35097 | 38584 |

All regressions are estimated in a first-differenced panel and include year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis

south, as unanticipated information induces a response of aggregated trade flows only if the importer is poor (p-values: 0.033 and 0.111, respectively). Interaction of precision and poor/rich dummies reveal that this effect is significant only in the case where the relative precision is high (p-values 0.068 and 0.111, respectively).

Our theory has no prediction on asymmetry of responses to good and bad news. However, exporters do respond asymmetricly to bad and good news (see Table **??**: In the case of an over-optimistic prior, i.e. bad news, adjustment of trade flows is statistically significant and stronger than in the case of over-pessimistic priors for both total trade and trade in differentiated goods (p-values 0.052 and 0.028, respectively). This behavior

Table 7: **Relative precision and relevance for trade with the south**

| **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|
| | | **Total Trade** | **Trade in Diff. Goods** |
| **Surprise** | $S$ | -0.0491 | -0.0370 |
| | | (0.019) | (0.017) |
| **Precision** | $S \times D^{\gamma_{\text{high}}}$ | -0.0658 | -0.0500 |
| | | (0.034) | (0.028) |
| | $S \times D^{\gamma_{\text{low}}}$ | -0.0350 | -0.0260 |
| | | (0.019) | (0.021) |
| **Poor/Rich** | $S \times D^{\text{poor}}$ | -0.0976 | -0.0650 |
| | | (0.046) | (0.041) |
| | $S \times D^{\text{rich}}$ | -0.0285 | -0.0251 |
| | | (0.018) | (0.017) |
| **Interaction** | $S \times D^{\text{poor}} \times D^{\gamma_{\text{high}}}$ | -0.159 | -0.106 |
| | | (0.087) | (0.066) |
| | $S \times D^{\text{poor}} \times D^{\gamma_{\text{low}}}$ | -0.0477 | -0.0317 |
| | | (0.043) | (0.049) |
| | $S \times D^{\text{rich}} \times D^{\gamma_{\text{high}}}$ | -0.0274 | -0.0269 |
| | | (0.032) | (0.029) |
| | $S \times D^{\text{rich}} \times D^{\gamma_{\text{low}}}$ | -0.0295 | -0.0236 |
| | | (0.021) | (0.021) |
| $R^2$ | | 0.0213 | 0.0220 |
| RMSE | | 0.805 | 0.830 |
| Country-pairs | | 5490 | 5404 |
| Observations | | 42332 | 41274 |

All regressions are estimated in a first-differenced panel and include GDPs, populations, price levels, as well as year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis mitigates the path dependence of prior beliefs.

Table 8: **Asymmetric responses to good and bad news**

| **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|
| | | **Total Trade** | **Trade in Diff. Goods** |
| **Good/Bad** | $S \times D^{\text{bad}}$ | -0.0368 | -0.0416 |
| | | (0.019) | (0.019) |
| | $S \times D^{\text{good}}$ | -0.0800 | -0.0254 |
| | | (0.050) | (0.039) |
| $R^2$ | | 0.0213 | 0.0220 |
| RMSE | | 0.805 | 0.830 |
| Country-pairs | | 5490 | 5404 |
| Observations | | 42332 | 41274 |

Regression is estimated in a first-differenced panel and includes GDPs, populations, price levels, as well as year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis.

**Multilateral learning.** One can easily extend our framework to a case, where exporters take all importer-specific business news into account and thus form an importer-specific multilateral rather than a bileral shirking forecast. Obviously prior beliefs on multilateral behavior come with a higher precision as much more business events are involved. However, consider that the business climate between the UK and the US is different from that between France and the US. A multilateral forecasts neglects the bilateral dimension and therefore may be biased.

Of course there are a lot of country-pairs where bilateral information is not available. Then the best exporters can do is to form priors is on the basis of business events initiated by the country of interest and addressed to other exporters. However, it is questionable whether exporters rely on a pure multilateral measure or rather take only a reasonable subset into account.[22] Table **??** shows the effect of multilateral priors on bilateral trade

[22]We focus on two extreme scenarios, bilateral and multilateral learning. Formation of mixed priors requires similarity measures of exporters and importers to imputate shirking probabilities in cases where no information is available and/or similarity measures of the underlying processes to broaden the basis of business events. Furthermore, it involves the joint binomial distribution which to our knowledge has

flows. While the number of included exporting and importing countries is unaltered, much more country-pairs enter the picture when we allow for multilateral learning. Exporters of differentiated goods do not rely on a multilateral measure but seem to prefer fines ones.

The multilateral prior has a significant impact only on trade in homogeneous goods. One can think of shirking also in the context of homogeneous goods. As goods are traded on organized exchange, the bilateral dimension of shirking is less important, while the multilateral shirking behavior of the importer is of interest. The size of the estimated coefficient ($-0.568$) is much larger than in the default case of differentiated goods. However, the corresponding ad valorem tariff equivalent is decreasing on the elasticity of substitution $\sigma$ and the distribution of beachhead costs $k$. For $\sigma = 15$ and $k = 1$ we obtain approximately 6 percent.

Table 9: ***Multilateral*** **shirking and bilateral trade**

| | **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|---|
| | **Total Trade** | **Differentiated Goods** | **Homogeneous Goods** | **Ref. priced Goods** |
| Shirking Prob. | -0.0759 | 0.237 | -0.568 | -0.0165 |
| | (0.20) | (0.19) | (0.34) | (0.23) |
| F | 55.78 | 56.53 | 34.91 | 40.96 |
| $R^2$ | 0.0113 | 0.0122 | 0.0102 | 0.0103 |
| RMSE | 1.145 | 1.126 | 1.328 | 1.158 |
| Country-pairs | 11958 | 11266 | 8300 | 9574 |
| Observations | 83513 | 77188 | 54053 | 64161 |

All regressions are estimated in a first-differenced panel and include GDPs, populations, price levels, as well as year fixed effects and a constant (all not reported). Trade classified according to Rauch’s liberal industry classification. Robust standard errors in parenthesis

## 4.2 Robustness checks

This subsection provides several robustness checks. First, we weight un-cooperative and cooperative events according to the Goldstein classification. Second, we relax our learning hypothesis and assume that events in the past are discounted when forming prior beliefs. Finally, we combine these two extensions. In all cases we find evidence for our hypothesis

not been described in the literature yet.

that a high shirking probability reduces bilateral trade flows.

**Goldstein weights.** Goldstein (1992) not only classifies events in un-cooperative and cooperative, but also gives different weights to different types of events. Adapting this weighting scheme, we find a qualitatively similar effect of the shirking probability on total trade flows and trade in differentiated goods (see Table **??**, p-values: 0.04). However, the corresponding ad valorem tariff equivalent is slightly smaller, e.g. $\left((1-0.128)^{-1/4}\right) \times 100 = 4.7$ percent for $k=1$ and $\sigma=5$.

Table 10: **Using Goldstein weights to calculate shirking probabilities**

| | **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|---|
| | **Total Trade** | **Differentiated Goods** | **Homogeneous Goods** | **Ref. priced Goods** |
| Shirking Prob. | -0.144 | -0.128 | -0.147 | 0.132 |
| | (0.073) | (0.062) | (0.13) | (0.096) |
| Ln GDP Exp. | 0.359 | 0.600 | 0.333 | 0.370 |
| | (0.12) | (0.15) | (0.19) | (0.15) |
| Ln GDP Imp. | 1.635 | 2.027 | 1.798 | 1.679 |
| | (0.13) | (0.14) | (0.22) | (0.17) |
| Ln Pop. Exp. | 0.461 | -0.211 | -0.594 | 2.768 |
| | (0.36) | (0.31) | (0.51) | (0.52) |
| Ln Pop. Imp. | -0.0728 | -0.0238 | 1.264 | -1.095 |
| | (0.26) | (0.27) | (0.47) | (0.32) |
| Ln Price level Exp. | -0.395 | -0.625 | -0.402 | -0.497 |
| | (0.14) | (0.17) | (0.22) | (0.17) |
| Ln Price Level Imp. | -0.961 | -1.265 | -1.179 | -1.077 |
| | (0.15) | (0.15) | (0.25) | (0.19) |
| FTA | -0.0635 | -0.0563 | 0.00924 | 0.0420 |
| | (0.027) | (0.026) | (0.055) | (0.035) |
| F Statistic | 45.69 | 46.50 | 29.70 | 37.57 |
| $R^2$ | 0.0213 | 0.0220 | 0.0139 | 0.0176 |
| RMSE | 0.805 | 0.830 | 1.205 | 0.914 |
| Country-pairs | 5490 | 5404 | 4835 | 5163 |
| Observations | 42332 | 41274 | 35097 | 38584 |

All regressions are estimated in a first-differenced panel and include year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis

**Discounting.** In a next step we relax our Bayesian learning hypothesis by discounting events in the past. If the true process that drives the shirking behavior is stationary, it is theoretically not optimal to discount. However, if many events take place, one can easily argue that business news that lie in the past should have a lower impact on the actual business climate between two countries. In order to allow for a decreasing impact on the shirking forecast, we discount business events of previous periods with a discount rate $r = 3/7$. This implies implies that the discount factor that is applied to a business event that took place five years ago is given by $\frac{1}{(1+3/7)^5} \approx 0.17$, which is approximately one sixth of the weight applied to the present period.[23]

We find that the shirking forecast has an statistically significant impact on total bilateral trade flows and trade in differentiated goods (see Table **??** in Appendix D, p-values: 0.022 and 0.033, respectively). The ad valorem tariff equivalent for differentiated goods is about 2.2 percent ($\sigma = 5$ and $k = 1$).

Turning to the role of the relative precision of the priors, importers in the south, and interactions between relative precision and poor/rich dummies, we find the same pattern as above, namely that response to unexpection information about the shirking probabiliy is larger if the relative precision is high or the importer is in the south (see Table **??**).

We also find an asymmetric response to bad and good news in the presence of discounting (see Table **??** in the Appendix D). However, the pattern differ for total trade and trade in differentiated goods.

**Goldstein weights and discounting simultanously.** One can also apply Goldstein weights and discounting simultanously. Table **??** in Appendix D presents the results. We find that the shirking probability robustly affects significantly total trade and trade in differentiated goods (p-values: 0.04). The corresponding ad valorem tariff equivalent for differentiated goods is of 3.5 percent.

---

[23] We have also successfully experimented with discount rates different from 3/7.

Table 11: **Relative precison and trade with the south in the presence of discounting**

| Dependent variable: Bilateral trade flows | | | |
|---|---|---|---|
| | | **Total Trade** | **Trade in Diff.** |
| **Surprise** | $S$ | -0.0532 | -0.0380 |
| | | (0.019) | (0.017) |
| **Precision** | $S \times D^{\gamma_{\text{high}}}$ | -0.0695 | -0.0484 |
| | | (0.034) | (0.027) |
| | $S \times D^{\gamma_{\text{low}}}$ | -0.0393 | -0.0292 |
| | | (0.019) | (0.020) |
| **Poor/Rich** | $S \times D^{\text{poor}}$ | -0.108 | -0.0678 |
| | | (0.046) | (0.041) |
| | $S \times D^{\text{rich}}$ | -0.0300 | -0.0255 |
| | | (0.018) | (0.017) |
| **Interaction** | $S \times D^{\text{poor}} \times D^{\gamma_{\text{high}}}$ | -0.165 | -0.105 |
| | | (0.087) | (0.064) |
| | $S \times D^{\text{poor}} \times D^{\gamma_{\text{low}}}$ | -0.0614 | -0.0373 |
| | | (0.044) | (0.050) |
| | $S \times D^{\text{rich}} \times D^{\gamma_{\text{high}}}$ | -0.0303 | -0.0252 |
| | | (0.031) | (0.028) |
| | $S \times D^{\text{rich}} \times D^{\gamma_{\text{low}}}$ | -0.0298 | -0.0258 |
| | | (0.021) | (0.020) |
| $R^2$ | | 0.0214 | 0.0220 |
| RMSE | | 0.805 | 0.830 |
| Country-pairs | | 5490 | 5404 |
| Observations | | 42332 | 41274 |

All regressions are estimated in a first-differenced panel and include year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis

# 5 Conclusions

In this paper, we argue that successful cross-border transactions in international trade require ex ante information on the trade partner's likelihood of un-cooperative behavior. The innovation of this paper is to use an original data base extracted from Reuters Business Briefing (RBB) and to apply it in a gravity-type framework. We assume that the probability of encountering un-cooperative behavior by some importer is stationary, which gives rise to our Bayesian learning hypothesis.

In our regressions in first differences we detect a robust effect of bilateral priors on shirking probabilities derived from RBB on total trade flows and trade in differentiated goods The corresponding ad valorem tariff equivalent of unexpected shirking is about 3 to 7 percent. We find evidence for our Bayesian learning hypothesis, namely that the response to unexpected information is especially high if the precision of the new arriving information is high. Moreiver, we show that our theory applies particular to trade scenarios where the importer is located in the south. The findings are robust to weighting of events according to the Goldstein (1992) scheme and to discounting of past events.

There are a couple of extensions that one might consider in future work. Reuters Business Briefing data assigns the day and even the hour of each event it records. Since bilateral trade data is available in monthly frequency at least for the last years, it seems promising to work with higher frequencies to better identify the time pattern of the news-trade nexus. Second, it would be highly desirable to extend our theoretical to a world with repeated games. Then waiting for more precise informations becomes an option.

# References


[1] Anderson, J.E. and E. van Wincoop (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review* **93**(1): 170-92.

[2] Anderson, J.E. and E. van Wincoop (2004). Trade Costs. *ournal of Economic Literature* **42**: 691-751.

[3] Baier, S., and J.H. Bergstrand (2006). Do Free Trade Agreements actually Increase Member's International Trade? *Journal of International Economics*. forthcoming.

[4] Besedes, T. (2006). A Search Perspective On the Duration of Trade. Mimeo: Louisiana State University.

[5] Bond, D., J. Bond, O. Churl O., J.C. Jenkins, and C.L. Taylor (2003). Integrated Data for Events Analysis (IDEA): An Event Typology for Automated Events Data Development. *Journal of Peace Research* **40**(6): 733–745.

[6] Casella, A., and J.E. Rauch (2002). Anonymous Markets and Group Ties in International Trade. *Journal of International Economics* **58**:19-48.

[7] Combes, P-P., M. Lafourcade and T. Mayer (2005). The Trade Creating Effects of Business and Social Networks: Evidence from France, *Journal of International Economics* **66**: 1-29.

[8] Clemen, R.T. (1952). Making Hard Decisions. An Introduction to Decision Analysis. Boston: PWS-KENT Publishing.

[9] Feenstra, R. (2004). *Advanced International Trade. Theory and Evidence.* Princeton: Princeton University Press.

[10] Goldstein, Joshua S. (1992). A Conflict-Cooperation Scale for WEIS Events Data. *Journal of Conflict Resolution*, **36**: 369-385.

[11] Hautsch, N., and D. Hess (2005). Bayesian Learning in Financial Markets - Testing for the Relevance of Information Precision in Price Discovery . *Journal of Financial and Quantitative Analysis* forthcoming.

[12] Hirshleifer, J., and J.G. Riley (1992). *The Analytics of Uncertainty and Information.* Cambridge Surveys of Economic Literature.

[13] King, G., and W. Lowe (2003). An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design. *International Organization* **57**: 617-642.

[14] Lence, S.H. and D.J. Hayes (1994). Parameter-Based Decision Making Under Estimation Risk: An Application to Futures Trade. *The Journal of Finance* **49**: 345-357.

[15] Nitsch, V. (2007). State Visits and International Trade. Mimeo: Free University Berlin.

[16] Obstfeld, M. and K. Rogoff (2000). The Six Major Puzzles in International Macroeconomics: Is There a Common Cause? *NBER Working Paper* 7777.

[17] Rauch, J.E., and V. Trindade (2000). Information and Globalization: Wage Co-Movements, Labor Demand Elasticity, and Conventional Trade Liberalization. *NBER Working Paper* 7671.

[18] Rauch, J.E., and V. Trindade (2003). Information, International Substitutability, and Globalization. *American Economic Review* **93**: 775-91

[19] Rose, A.K. (2006). *The Foreign Service and Foreign Trade: Embassies as Export Promotion.* Mimeo: University of Berkeley.

[20] Schmitt, N., and Z. Yu (2001). Economies of scale and the volume of intra-industry trade *Economics Letters* **74**: 127-132.

# A Beta distribution

Exporters do not know the true shirking probability $\theta_{ij}$ itself, but have a homogeneous prior on the probability distribution of $\theta_{ij}$, which is denoted by $f(\theta_{ij})$. As the underlying process that govers $X_{ij}$ is binomial, it is reasonabe to assume $f(\theta_{ij})$ to be a the Beta distribution with parameters $a_{ij}, b_{ij} \geq 1$. For convenience we suppress the subscript $ij$.

$$f(\theta; a, b) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}, \tag{18}$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \Gamma(x) = (x-1)!, \tag{19}$$

where the scaling parameter $B(a,b)$ is the Beta function, and $\Gamma(x)$ is the Gamma function. The mean $\theta^F = E[\theta]$ is given by $a/(a+b)$ and the variance by $Var[\theta] = ab/\left[(a+b)^2(a+b+1)\right]$. These two moments are sufficient to derive the parameters $a, b$ from a mean forecast and a notion about the precision. In line with the theory, precision $\rho^F$ is defined as the inverse of the variance. As $a$ reflects the equivalent number of non-cooperative business news and $b$ the corresponding number of cooperative business news, we use an additional condition $a + b = n^F$ and solve the following system for $a, b$

$$\theta^F = \frac{a}{a+b} = \frac{a}{n^F} \Leftrightarrow a = \theta^F n^F \tag{20}$$

$$\begin{aligned} \rho^F &= \left(\frac{ab}{(a+b)^2(a+b+1)}\right)^{-1} \\ &= \frac{n^F+1}{\theta^F(1-\theta^F)} \\ \Leftrightarrow\ & n^F = \rho^F\theta^F(1-\theta^F) - 1 \end{aligned} \tag{21}$$

Hence, the Beta distribution is parameterized with $a = \theta^F n^F$ and $b = n^F - a = (1-\theta^F)n^F$. The Beta distribution features some reasonable characteristics. First, when absolutly no information about the parameter is available, a common prior used for the parameters is $\theta^F n^F = n_F(1-\theta^F) = 1$. In this case, the Beta distribution equals the standard uniform distribution, which gives equal mass to all potential parameters $\theta$ between 0 and 1. Second, if $n^F\theta^F$ equals $n^F(1-\theta^F)$, the Beta distribution is symmetric around 1/2.

Consider again the precision $\rho^F$. It is increasing in the number of underlying events $n^F$. Furthermore, it depends on the variance of the outcome $X$, which can either be

1 (shirking) with a probability $\theta$ or 0 (non-shirking) with a probability $(1-\theta)$. Then $Var(X) = \theta^F - (\theta^F)^2$. The higher this variance, the lower the precision. Consider a case where the true $\theta$ equals 1/2. Then $\rho^F$ is given by $2(n^F+1)$. Consider now a case where the true $\theta \neq 1/2$, and $\rho^F$ will always be larger than $2(n^F+1)$. In a slight abuse of terms we refer to $n^F$ as the precision of the forecast in order to suppress dependence on the forecast itself.

# B Proofs and Derivations

**Proof of Proposition 1.** We suppress the subscripts $ij$ for convenience. The conditional probability density funtion of $\theta_t^A$ given $\theta$ is Binomial, i.e.

$$g\left(\theta_t^A|\theta\right)=\binom{n_t^A}{r_t^A}\left(\theta_t^A\right)^{r_t^A}\left(1-\theta_t^A\right)^{n_t^A-r_t^A}. \tag{22}$$

Let $f\left(\theta|\mathcal{F}_t\right)$ denote exporters' posterior beliefs after observing events in $t$, where the business news set $\mathcal{F}_t$ combinds the prior parametrization and the currently observed events.

From Bayes theorem, the posterior distribution of $\theta$ is calculated from the prior $g\left(\theta;\mathcal{F}_t\right)$ and the likelihood function function $\Pr\left(x|\theta\right)$ as

$$f(\theta|x)=\frac{\Pr\left(x|\theta\right)g\left(\theta\right)}{\int\Pr\left(x|\theta\right)g(\theta)\mathrm{d}\theta}, \tag{23}$$

where $x$ denotes realiziations of the underlying random variable $X$, say, $r_t^A$ un-cooperative out of $n_t^A$ events . The likelihood function is then given by

$$\Pr(r_t^A,n_t^A|\theta)=\binom{n_t^A}{r_t^A}\theta^{r_t^A}\left(1-\theta\right)^{n_t^A-r_t^A}. \tag{24}$$

Working out this expression we find that the posterior beliefs $g\left(\theta|\mathcal{F}_t\right)$ are also beta distributed with parameters $\theta^F n^F+\theta_t^A n_t^A$ and $\left(1-\theta_t^F\right)n_t^F+\left(1-\theta_t^A\right)n_t^A$

$$\begin{aligned}
f\left(\theta|\mathcal{F}_t\right) &= \frac{\binom{n_t}{r_t}\theta^{\theta_t^F n_t^F-1+r_t^A}\left(1-\theta\right)^{n_t^F-\theta_t^F n_t^F+n_t^A-r_t^A-1}/B(\theta_t^F n_t^F,\left(1-\theta_t^F\right)n_t^F)}{\int\left(\binom{n_t}{r_t}\theta^{\theta_t^F n_t^F-1+r_t}\left(1-\theta\right)^{n_t^F-\theta_t^F n_t^F+n_t^A-r_t^A-1}/B(\theta_t^F n_t^F,\left(1-\theta_t^F\right)n_t^F)\right)\mathrm{d}\theta} \\
&= \frac{\theta^{\theta_t^F n_t^F-1+r_t}\left(1-\theta\right)^{n_t^F-\theta_t^F n_t^F+n_t^A-r_t^A-1}}{B(\theta_t^F n_t^F,\left(1-\theta_t^F\right)n_t^F)} \\
&= \mathit{Beta}\left[\theta;\theta_t^F n_t^F+\theta_t^A n_t^A,\left(1-\theta_t^F\right)n_t^F+\left(1-\theta_t^A\right)n_t^A\right].
\end{aligned} \tag{25}$$

If a posterior distribution is of the same algebraic form as the prior distribution, the literature talks about *conjugate* priors. It follows by recursive iteration that the posterior distribution at time $t-1$ has the parameters $r_t^F=\sum_{s<t}^t r_s^A$ and $\sum_{s<t}\left(n_s^A-r_s^A\right)$ . Let $n_t^F$ denote the sum over all events $\sum_{s<t}^t n_s^A$.

We proceed with calculating the posterior predictive distribution of the random variable $X$ at time $t$. In particular, the producer is interested in the probability of encountering shirking behavior, i.e. the probability of the next draw ($n=1$) being "shirking" ($x=1$)

or not. Formally, she has to calculate

$$\begin{aligned}
\Pr(1|x) &= \int \left( \binom{1}{1} \theta^{1-1+r_t^F} (1-\theta)^{n_t^F - r_t^F + 1 - 1 - 1} / B(r^F, n^F - r^F) \right) \mathrm{d}\theta \\
&= 1/B(r^F, n^F - r^F) \int \left( \theta^{r^F} (1-\theta)^{n^F - r^F - 1} \right) \mathrm{d}\theta \qquad (26) \\
&= B\left(r_t^F + 1, n_t^F - r_t^F\right) / B(r_t^F, n_t^F - r_t^F) \\
&= \frac{\Gamma\left(r_t^F + 1\right) \Gamma\left(n_t^F - r_t^F\right)}{\Gamma\left(n_t^F + 1\right)} \frac{\Gamma\left(n_t^F\right)}{\Gamma\left(r_t^F\right) \Gamma\left(n_t^F - r_t^F\right)} \\
&= \frac{\left(n_t^F - 1\right)! r_t^F!}{(r_t^F - 1)! n_t^F!} = \frac{r_t^F}{n_t^F} = \frac{\sum_{s<t}^t r_s^A}{\sum_{s<t}^t n_s^A} \qquad (27)
\end{aligned}$$

**Proof of proposition 2.** The mean $\theta_t^P$ of the posterior distribution is given by

$$\begin{aligned}
\theta_{t+1}^F &= \theta_t^P = E\left[\theta | \mathcal{F}_t \cup \left\{r_t^A, n_t^A\right\}\right] \\
&= \frac{\theta_t^F n_t^F + \theta_t^A n_t^A}{\theta_t^F n_t^F + \theta_t^A n_t^A + \left(1 - \theta_t^F\right) n_t^F + \left(1 - \theta_t^A\right) n_t^A} \\
&= \frac{\theta_t^F n_t^F + \theta_t^A n_t^A}{\theta_t^F n_t^F + \theta_t^A n_t^A - \theta_t^F n_t^F - \theta_t^A n_t^A + n_t^F + n_t^A} \\
&= \frac{\theta_t^F n_t^F + \theta_t^A n_t^A}{n_t^F + n_t^A} \\
&= \theta_t^F \frac{n_t^F}{n_t^F + n_t^A} + \theta_t^A \frac{n_t^A}{n_t^F + n_t^A} \\
&= \left(1 - \gamma_t\right) \theta_t^F + \gamma_t \theta_t^A \qquad (28)
\end{aligned}$$

which corresponds to the expressions in Proposition 2.

**Derivation of the cutoff level $c_{ijt}^*$.** We denote by $x$ a concrete realization of the random variable $X_{ijt}$. Let $\Pr\left(X_{ijt} = x | \theta_{ij}\right)$ be the likelihood function, and $\Pr(\theta_{ij} | \mathcal{F}_{ijt})$ the (posterior) predictive distribution of $X_{ijt}$, given the forecast $\theta_{ijt}^F$. Recall that the forecast contains all relevant information of the previous periods. Following Bayes' decision criterion, the optimal decision is made after integrating out the unknown parameter $\theta_{ij}$.[24]

Hence, we can write expected operational profit as

[24] For an application to finance, see Lence and Hayes (1994).

$$
\begin{aligned}
& E_{\theta_{ij}}\left\{E_{X_{ij}|\theta_{ij}}\left[\pi_{ijt}\left(X_{ijt}\right)-c\left(z\right)\right]\right\} \\
= & \int_{\theta\in[0,1]}\left(\sum_{X_{ij}\in\{0,1\}}\left[\pi_{ijt}\left(X_{ijt}\right)-c\left(z\right)\right]\Pr\left(X_{ijt}|\theta\right)\right)\ \Pr(\theta|\mathcal{F}_{ijt})\ \mathrm{d}\theta \\
= & \sum_{X_{itj}\in\{0,1\}}\left[\pi_{ijt}\left(X_{ijt}\right)-c\left(z\right)\right]\left(\int_{\theta\in[0,1]}\Pr\left(X_{ijt}=x|\theta\right)\ \Pr(\theta|\mathcal{F}_{ijt})\ \mathrm{d}\theta\right) \\
= & \sum_{X_{ijt}\in\{0,1\}}\left[\pi_{ijt}\left(X_{ijt}\right)-c\left(z\right)\right]\Pr\left(X_{ijt}=x|\mathcal{F}_{ijt}\right)\geq 0 \qquad (29)
\end{aligned}
$$

The shirking probability $\Pr\left(X_{ijt}=1|\mathcal{F}_{ijt}\right)$ is given by the estimator $\theta^F_{ijt}$. We can insert (**??**) into (**??**) to find the following expression

$$\left(\left(\tau_{ijt}w_{it}\right)^{1-\sigma}B_{jt}-c^*_{ijt}\right)\left(1-\theta^F_{ijt}\right)+\left(\left(\tau_{ijt}\left(1+\delta\right)w_{it}\right)^{1-\sigma}B_{jt}-c^*_{ijt}\right)\theta^F_{ijt}=0, \qquad (30)$$

from which follows that

$$c^*_{ijt}=\left(\tau_{ijt}w_{it}\right)^{1-\sigma}B_{jt}\left[\left(1+\delta\right)^{1-\sigma}\theta^F_{ijt}+\left(1-\theta^F_{ijt}\right)\right]. \qquad (31)$$

**Derivation of the gravity equation.** Given the functional form of the threshold $c^*_{ijt}$ (**??**), we get

$$G^i\left(c^*_{ijt}\right)=\left\{\left(\tau_{ijt}w_{it}\right)^{1-\sigma}B_{jt}\Theta^F_{ij}/\bar{c}_i\right\}^k. \qquad (32)$$

Then, we have

$$
\begin{aligned}
M_{jit} &= K\Theta_{ij}\left(\Theta^F_{ij}\right)^k\left(\frac{P_{jt}}{w_{it}\tau_{ij}}\right)^{(\sigma-1)(1+k)}N_{it}E^{1+k}_{jt} \qquad (33)\\
&\approx Ke^{-\bar{\delta}\theta_{ij}}e^{-k\bar{\delta}\theta^F_{ijt}}\left(\frac{P_{jt}}{w_{it}\tau_{ijt}}\right)^{(\sigma-1)(1+k)}N_{it}E^{1+k}_{jt}, \qquad (34)
\end{aligned}
$$

where $K\equiv\left(\frac{\rho}{\beta}\right)^{(\sigma-1)(1+k)}\left(\frac{1-\rho}{\bar{c}_i}\right)^k$ is a constant and $\bar{\delta}=\left(1+\delta\right)^{1-\sigma}-1$.[25]

[25]Note that we have used the Taylor series expansion $\ln\left[1+x\right]=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}\pm...$ for $x=\theta_{ij}\left[\left(1+\delta\right)^{1-\sigma}-1\right]\in[0,-1]$ and $x=\theta^F_{ijt}\left[\left(1+\delta\right)^{1-\sigma}-1\right]\in[0,-1]$ respectively.

# C General equilibrium conditions

Here we characterize the equilibrium conditions for risk neutral firms. The timing is the following: First, firms pay the fixed cost $\alpha$ to enter the market. Then, they draw a beachhead cost $c(z)$, identical for all foreign markets. Finally, they decide whether or not to enter the foreign market. Firms die every period, so that there is no issue of strategic intertemporal behavior.

**Free entry condition.** All firms that pay the fixed cost $\alpha$ will be active and produce for the domestic market. The reason is that all firms have identical variable costs. Profits from domestic sales are given by $\pi_{iit}(w_{it}, B_{it}) = w_{it}^{1-\sigma} B_{it}$. All firms $z$ make the same profits since they face the same technological and preference parameters. Let $J$ denote the set of foreign markets, and denote $j \neq i$ a generic foreign market. Profits in the foreign market are $\pi_{ijt}(z) = [\tau_{ij}\delta_{ij}(z) w_{it}]^{1-\sigma} B_{jt}$. All domestic firms face the same distribution from which $\delta_{ij}$ is drawn and have the same prior $\hat{\theta}_{ijt}$ on the parmater of this distribution. Hence, upon entry into the foreign market $j$ they expect to make average profits

$$\bar{\pi}_{ijt}\left(\hat{\theta}_{ijt}, B_{jt}, w_{it}\right) = [\tau_{ij} w_{it}]^{1-\sigma} B_{jt} E_{\hat{\theta}_{ijt}}[\delta_{ij}]. \tag{35}$$

The *ex ante* probability of a firm to find it optimal to enter the foreign market $j$ is given by $G\left(c_{ijt}^*\right)$. Summing over all potential export markets, total expected foreign profits net of beachhead entry costs are $\sum_{j\in J} G\left(c_{ijt}^*\right)\left[\bar{\pi}_{ijt} - w_{it}\int_0^{c_{ijt}^*} cdG(c)\right]$.The free entry condition requires that the equilibrium number of firms is such that expected total profits are zero, i.e.

$$\pi_{iit} + \sum_{j\in J} G\left(c_{ijt}^*\right)\left[\bar{\pi}_{ijt} - w_{it}\int_0^{c_{ijt}^*} cdG(c)\right] = 0. \tag{36}$$

Note that the number of firms enters the above expression through the terms $B_{it}$. There are $C$ free entry conditions.

**Labor market equilibrium condition.** A firm characterized by beachhead costs $c(z)$ enters into the export market $j$ if $c(z) \leq c_{ijt}^*$ and stays out else. However, any firm that does enter the foreign market expects the same sales. Total shipments of firm $z$, given some conjecture $\hat{\theta}_{ijt}$, can be written as

$$X_{it}(z) = \left(\frac{w_{it}\beta}{\rho}\right)^{-\sigma} \sum_{j=1}^{C} \mathbf{1}_{c(z)\leq c^*_{ijt}} E_{jt} P_{jt}^{\sigma-1} \tau_{ij}^{1-\sigma} E_{\hat{\theta}_{ijt}}\left[\delta_{ijt}^{1-\sigma}\right], \tag{37}$$

where $\mathbf{1}_{c(z)\leq c^*_{ijt}}$ is an indicator function that takes value 1 if the criterion is met and zero else. Firm $z$ labor demand is

$$l_{it}(z) = \alpha + c(z)\left(\sum_{j=1}^{C} \mathbf{1}_{c(z)\leq c^*_{ijt}}\right) + \beta X_{it}(z).$$

Next, integrating over all $z$, we find aggregate labor demand

$$\begin{aligned} L_{it} &= N_{it}\left[\alpha + \int_0^{c^*_{ijt}} c dG(c)\right] + \left(\frac{w_{it}}{\rho}\right)^{-\sigma} \beta^{1-\sigma} \int_{n_{it}} \sum_{j=1}^{C} \mathbf{1}_{c(z)\leq c^*_{ijt}} E_{jt} P_{jt}^{\sigma-1} \tau_{ij}^{1-\sigma} E_{\hat{\theta}_{ijt}}\left[\delta_{ijt}^{1-\sigma}\right] dz \\ &= N_{it}\left[\alpha + \int_0^{c^*_{ijt}} c dG(c)\right] + N_{it}\left(\frac{w_{it}}{\rho}\right)^{-\sigma} \beta^{1-\sigma} \times \\ &\quad \left\{E_{it} P_{it}^{\sigma-1} + \sum_{j\in J} G\left(c^*_{ijt}\right) E_{jt} P_{jt}^{\sigma-1} \tau_{ij}^{1-\sigma} E_{\hat{\theta}_{ijt}}\left[\delta_{ijt}^{1-\sigma}\right]\right\}. \end{aligned} \tag{38}$$

The first term in the above expression refers to labor requiremtn induced by fixed production and beachhead costs. The second term refers to variable labor inputs. The first part in the curly brackets relates to domestic shipments and the summation is shipments to all foreign markets. We have $C$ free entry conditions that pin down the wage rate in each country. Note that if labor markets clear in every country, we need not also check whether goods markets are in equilibrium.

**General equilibrium.** We may normalize the wage rate in one country. Wage rates and goods prices are linked by the monopolistic pricing rule. GDP in each country $i$ is given by $E_{it} = w_{it} L_{it}$. The cut-off cost level $c^*_{ijt}$ is given by (**??**) and depends on $\hat{\theta}_{ijt}, w_{it}, E_{jt}$, and $P_{jt}$. The price index dual to (**??**) depends on $p_{it}$ and $N_{ijt}$, where $N_{ijt} = G\left(c^*_{ijt}\right) N_{it}$. Hence, once $w_{it}$ and $N_{it}$ are known, all other variables follow. We have $C - 1$ labor market clearing conditions and $C$ free entry conditions to solve for the endogenous wages and number of firms.

# D Additional Tables

Table 12: **List of countries**

| | | | |
|---|---|---|---|
| Angola | El Salvador | Kyrgyzstan | Portugal |
| Argentina | Equatorial guinea | Laos | Russian Federation |
| Armenia | Estonia | Latvia | Saint Lucia |
| Australia | Ethiopia | Lebanon | Saint Vincent |
| Austria | Fiji | Lithuania | Sao Tome and Principe |
| Bahamas | Finland | Luxembourg | Saudi Arabia |
| Bahrain | France | Macedonia | Senegal |
| Bangladesh | Gabon | Madagascar | Sierra Leone |
| Belgium | Gambia | Malawi | Singapore |
| Belize | Georgia | Malaysia | Slovakia |
| Bolivia | Germany | Mali | Slovenia |
| Brazil | Ghana | Malta | South Africa |
| Bulgaria | Greece | Mauritania | Spain |
| Burkina faso | Grenada | Mauritius | Sri Lanka |
| Burundi | Guinea | Mexico | Sudan |
| Cameroon | Guinea-Bissau | Moldova | Sweden |
| Canada | Guyana | Mongolia | Switzerland |
| Cape verde | Haiti | Morocco | Syria |
| Centr. Afr. Republic | Honduras | Nepal | Tanzania |
| Chad | Hungary | Netherlands | Thailand |
| Chile | Iceland | New Zealand | Togo |
| China | Indonesia | Nicaragua | Trinidad and Tobago |
| Colombia | Iran | Niger | Tunisia |
| Comoros | Ireland | Nigeria | Turkey |
| Congo, Rep. of | Israel | Norway | Uganda |
| Costa Rica | Italy | Oman | Ukraine |
| Cote d'Ivoire | Jamaica | Pakistan | United Kingdom |
| Croatia | Japan | Panama | United States |
| Cyprus | Jordan | Papua New Guinea | Uruguay |
| Czech Republic | Kazakhstan | Paraguay | Viet nam |
| Denmark | Kenya | Peru | Zambia |
| Ecuador | Korea, Rep. of | Philippines | Zimbabwe |
| Egypt | Kuwait | Poland | |

Table 13: **Baseline gravity equations: Trade in differentiated goods**

| | **Dependent variable: Bilateral trade in differentiated goods** | | | |
|---|---|---|---|---|
| | **Pooled OLS** | **Pooled OLS** | **Fixed-Effects** | **First Diff.** |
| ln GDP Exp. | 1.599 | 0.721 | 1.017 | 0.573 |
| | (0.020) | (0.094) | (0.090) | (0.13) |
| ln GDP Imp. | 0.920 | 1.317 | 1.718 | 2.239 |
| | (0.019) | (0.10) | (0.097) | (0.13) |
| ln Pop. Exp. | -0.496 | -1.131 | -0.646 | -0.707 |
| | (0.022) | (0.18) | (0.17) | (0.27) |
| ln Pop. Imp. | -0.114 | -0.849 | -0.898 | -0.311 |
| | (0.020) | (0.18) | (0.16) | (0.25) |
| ln Price level Exp. | -0.918 | -0.747 | -1.096 | -0.619 |
| | (0.051) | (0.11) | (0.10) | (0.14) |
| ln Price level Imp. | 0.00311 | -0.599 | -0.964 | -1.453 |
| | (0.049) | (0.12) | (0.11) | (0.14) |
| ln Distance | -1.227 | -1.464 | | |
| | (0.019) | (0.024) | | |
| Contiguity | 0.564 | 0.441 | | |
| | (0.11) | (0.11) | | |
| Colonial history | 0.931 | 0.870 | | |
| | (0.11) | (0.11) | | |
| Common language | 0.756 | 0.777 | | |
| | (0.048) | (0.050) | | |
| FTA | 0.324 | 0.179 | 0.0644 | -0.0752 |
| | (0.044) | (0.044) | (0.033) | (0.030) |
| Country FE | NO | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| F | 2485 | 310.0 | 218.9 | 64.47 |
| $R^2$ | 0.709 | 0.794 | 0.0791 | 0.0126 |
| RMSE | 1.991 | 1.678 | 0.966 | 1.139 |
| Country-pairs | 13478 | 13478 | 13478 | 11377 |
| Observations | 101111 | 101111 | 101111 | 80289 |

All regressions include a constant (not reported). Robust standard errors in parenthesis.

Table 14: **Discounting past events**

| | **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|---|
| | **Total Trade** | **Differentiated Goods** | **Homogeneous Goods** | **Ref. priced Goods** |
| Shirking Prob. | -0.113 | -0.0820 | -0.0697 | 0.0740 |
| | (0.049) | (0.038) | (0.075) | (0.061) |
| Ln GDP Exp. | 0.359 | 0.601 | 0.335 | 0.369 |
| | (0.12) | (0.15) | (0.19) | (0.15) |
| Ln GDP Imp. | 1.636 | 2.027 | 1.798 | 1.679 |
| | (0.13) | (0.14) | (0.22) | (0.17) |
| Ln Pop. Exp. | 0.461 | -0.210 | -0.592 | 2.767 |
| | (0.36) | (0.31) | (0.51) | (0.52) |
| Ln Pop. Imp. | -0.0748 | -0.0249 | 1.263 | -1.094 |
| | (0.26) | (0.27) | (0.47) | (0.32) |
| Ln Price level Exp. | -0.395 | -0.626 | -0.404 | -0.496 |
| | (0.14) | (0.17) | (0.22) | (0.17) |
| Ln Price Level Imp. | -0.962 | -1.265 | -1.179 | -1.077 |
| | (0.15) | (0.15) | (0.25) | (0.19) |
| FTA | -0.0634 | -0.0563 | 0.00934 | 0.0419 |
| | (0.027) | (0.026) | (0.055) | (0.035) |
| F Statistic | 45.76 | 46.58 | 29.64 | 37.56 |
| $R^2$ | 0.0214 | 0.0220 | 0.0139 | 0.0176 |
| RMSE | 0.805 | 0.830 | 1.205 | 0.914 |
| Country-pairs | 5490 | 5404 | 4835 | 5163 |
| Observations | 42332 | 41274 | 35097 | 38584 |

All regressions are estimated in a first-differenced panel and include year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis

Table 15: **Using Goldstein weights and discounting past events**

| | **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|---|
| | **Total Trade** | **Differentiated Goods** | **Homogeneous Goods** | **Ref. priced Goods** |
| Shirking Prob. | -0.144 | -0.128 | -0.147 | 0.132 |
| | (0.073) | (0.062) | (0.13) | (0.096) |
| Ln GDP Exp. | 0.359 | 0.600 | 0.333 | 0.370 |
| | (0.12) | (0.15) | (0.19) | (0.15) |
| Ln GDP Imp. | 1.635 | 2.027 | 1.798 | 1.679 |
| | (0.13) | (0.14) | (0.22) | (0.17) |
| Ln Pop. Exp. | 0.461 | -0.211 | -0.594 | 2.768 |
| | (0.36) | (0.31) | (0.51) | (0.52) |
| Ln Pop. Imp. | -0.0728 | -0.0238 | 1.264 | -1.095 |
| | (0.26) | (0.27) | (0.47) | (0.32) |
| Ln Price level Exp. | -0.395 | -0.625 | -0.402 | -0.497 |
| | (0.14) | (0.17) | (0.22) | (0.17) |
| Ln Price Level Imp. | -0.961 | -1.265 | -1.179 | -1.077 |
| | (0.15) | (0.15) | (0.25) | (0.19) |
| FTA | -0.0635 | -0.0563 | 0.00924 | 0.0420 |
| | (0.027) | (0.026) | (0.055) | (0.035) |
| F Statistic | 45.69 | 46.50 | 29.70 | 37.57 |
| $R^2$ | 0.0213 | 0.0220 | 0.0139 | 0.0176 |
| RMSE | 0.805 | 0.830 | 1.205 | 0.914 |
| Country-pairs | 5490 | 5404 | 4835 | 5163 |
| Observations | 42332 | 41274 | 35097 | 38584 |

All regressions are estimated in a first-differenced panel and include year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis

Table 16: **Asymmetric responses in the presence of discounting**

| **Dependent variable: Bilateral trade flows** | | | |
|---|---|---|---|
| | | **Total Trade** | **Trade in Diff. Goods** |
| **Good/Bad** | $S \times D^{\text{bad}}$ | -0.0366 | -0.0421 |
| | | (0.019) | (0.019) |
| | $S \times D^{\text{good}}$ | -0.0913 | -0.0286 |
| | | (0.047) | (0.036) |
| $R^2$ | | 0.0214 | 0.0220 |
| RMSE | | 0.805 | 0.830 |
| Country-pairs | | 5490 | 5404 |
| Observations | | 42332 | 41274 |

Regression is estimated in a first-differenced panel and includes year fixed effects and a constant (all not reported). Trade classified according to Rauch's liberal industry classification. Robust standard errors in parenthesis