How Robust is the Skill-Dispersion-Complementarity Hypothesis?

Matilde Bombardini*, Giovanni Gallipoli† and Germán Pupato‡

August 2013

Abstract

We explore the robustness of the hypothesis, first put forward by Grossman and Maggi (2000) (GM), that countries with higher skill dispersion specialize in the sector characterized by a submodular production function, i.e. the industry that cross-matches workers of different skills (henceforth referred to as SDC hypothesis). We relax the assumption of constant returns to skill, breaking the link between submodularity and the concavity of isoquants, a key feature in GM. We show that when a submodular sector displays convex isoquants, it no longer benefits from higher skill dispersion and higher skill dispersion countries may specialize in the supermodular sector. We investigate this theoretical possibility by performing a variety of simulations, based on empirical skill distributions, and find that in the vast majority of cases the SDC hypothesis is not violated.

JEL Classification codes: F12, F16, J82.

We would like to thank seminar participants at the West Coast Trade Workshop and the Minneapolis Fed, Patrick Francois, Vadim Marmer, Francesco Trebbi, and especially Hiro Kasahara and Kevin Song for very useful comments. Evan Calford and Tom Cornwall provided excellent research assistance. Bombardini thanks CIFAR for financial support.

*University of British Columbia, CIFAR and NBER

†University of British Columbia

‡Ryerson University

1 Introduction

In recent years the trade literature has seen an increasing interest in how worker skill heterogeneity affects a variety of outcomes, such as the pattern of specialization and the impact of trade liberalization on inequality.[1]

In this paper we revisit the theory, pioneered by Grossman and Maggi (2000) (henceforth GM), linking cross-country differences in observable skill dispersion and comparative advantage and show that its predictions are not always robust under relatively common generalizations of the production technology. Such theoretical difficulties are hard to overcome without imposing further restrictions. This fact motivates us to use a different approach to verify whether these theoretical ambiguities result in actual reversals of the predicted patterns of comparative advantage. In the last part of the paper we perform several direct checks in which empirical skill distributions are mapped into predicted patterns of trade.

GM first showed how different degrees of skill dispersion across countries can generate comparative advantage in a competitive model in which sectors vary in the degree of complementarity of skills in the production process. In one sector the production function, $F_1(t_A, t_B)$, is supermodular in the skills of the two workers, t_A and t_B, employed by a firm. In this sector skills are complementary and positive assortative matching is optimal. In the other sector the production function, $F_2(t_A, t_B)$, is submodular in skills and negative assortative matching is optimal. Importantly, GM assume that the production technologies exhibit constant returns to skill, so that the supermodular sector also exhibits convex isoquants. Analogously, under the assumption of constant returns to skill, the submodular sector exhibits concave isoquants[2] and a taste for employing workers of

[1] Relevant work in this area includes, among others, Grossman (2004), Ohnsorge and Trefler (2007), Costinot and Vogel (2010), Bougheas and Riezman (2007), Costinot (2009), Helpman et al. (2012), Grossman et al. (2013), Huang and Chang (2012), Bombardini et al. (2012b) and Bombardini et al. (2012a). For an exhaustive review of this literature see Davidson and Sly (2010).

[2] Throughout the paper we refer to an isoquant as concave if the lower contour set is convex.

different skills.[3]

While the curvature of the isoquants has no consequence for the optimal matching pattern, it is crucial to assess the impact of skill dispersion on production and trade patterns. In GM this issue is simplified by the coincidence of these properties so that, for example, the submodular sector necessarily experiences an increase in output when skill dispersion increases, everything else equal. This is because the submodular sector has concave isoquants. The result is a clear-cut pattern of trade: the country with higher skill dispersion has a comparative advantage in the submodular good because that sector benefits from a higher dispersion. This pattern of trade is what we henceforth refer to as the Skill-Dispersion-Complementarity (SDC) hypothesis.

In this paper we analyze, among others, the case of submodular production functions that, despite featuring negative assortative matching, do not benefit from those cross-matched workers being very different from one another. The possibility of dispersion-averse submodular functions and dispersion-loving supermodular functions generate new insights on the impact of skill dispersion on comparative advantage. In particular, in the context of a two-sector two-country model we show that, under these more general configurations of production technologies, the pattern of trade may be such that the country with lower skill dispersion specializes in the submodular sector, violating the SDC hypothesis.

These violations occur when we relax the not innocuous assumption of constant returns to skill in the production functions $F_i(t_A, t_B)$.[4] It is well known that, in general, quasi-concavity and supermodularity are distinct properties, coinciding in the special case of homogeneity of degree one

[3] As mentioned by GM, an alternative way of generating different matching patterns in different sectors is to assume production functions that are asymmetric in the two tasks as in Kremer and Maskin (1996). Although potentially very interesting, Legros and Newman (2002) point out that this production function is neither supermodular nor submodular. This implies that the optimal matching pattern depends on the skill distribution, unlike in the setup that we and GM study, and complicates the problem greatly.

[4] GM do note that the supermodular production function has convex isoquants and the submodular production function has concave isoquants under the stated assumption of linear homogeneity of $F(\cdot)$, but they do not further investigate the case in which these properties do not coincide.

of the production function.[5] At the outset, however, it is not obvious whether constant, decreasing or increasing returns to skill are more accurate descriptions of real-world production functions, so it is important to investigate whether the SDC hypothesis applies in all of these instances. The matching literature, for example Shimer and Smith (2000),[6] disregards whether returns to skill are increasing or decreasing and often chooses production functions that have different behaviors over different skill ranges. A common assumption is, for example, that output is determined by the product of the skills of the two workers, a production function which entails increasing returns to skill. We take this as an indication that focusing on constant returns to skill is at least uncommon, if not restrictive. It is worth emphasizing that the concept of returns to skill is unrelated to the degree of returns to scale, which does have implications for the existence of a competitive equilibrium. In this respect, we maintain the assumption of constant returns to scale and a competitive setup.

Our theoretical analysis suggests that, in general, the SDC hypothesis cannot be expected to hold for arbitrary skill distributions that can be ordered according to a standard definition of dispersion (e.g. mean-preserving spreads), but are otherwise fully unrestricted. In order to assess whether such potential violations are of concern from an empirical point of view, we perform two kinds of exercises with a set of empirically relevant skill distributions. These exercises use distributions of literacy scores in 19 countries that participated in the International Adult Literacy Survey (IALS). First, we fit parametric density functions to the score distribution of each country, experimenting with truncated normals and beta distributions. We then feed these estimated distributions into our model under the assumption that countries differ only in their skill distributions. Finally we compute the relative output produced by each country at different relative prices and for a wide range of technology configurations, and compare the resulting patterns of comparative advantage

[5]Marinacci and Montrucchio (2008) generalize the study of the Choquet property (Choquet, 1953), i.e. that supermodularity implies concavity for positively homogeneous functionals.

[6]Other examples include Atakan (2006), Becker (1973), Eeckhout and Kircher (2011), Lentz (2010), Lopes de Melo (2011), Postel-Vinay and Robin (2002) and Teulings and Gautier (2004).

for each country pair against the SDC hypothesis. The second exercise is similar in spirit, but consists of numerical simulations that directly rely on the IALS score distributions, without the need of parametric assumptions. We find that patterns of comparative advantage consistent with the SDC hypothesis prevail in the vast majority of the cases considered.

It is essential to clarify that this result is not a test of the SDC hypothesis, as this would require one to posit clear-cut theoretical restrictions and test them with trade flow data. Instead, we start from the premise that theoretical ambiguities make it hard to set out these very predictions and, then, we investigate the extent to which they hold under many different technology parameterizations. Our objective is to check the theoretical robustness of the model's predictions, rather than testing them using trade data.

In the next section we lay out the model and describe the features of the production function we analyze. Section 3 characterizes the equilibrium. Section 4 provides a simple example in which, counterintuitively, the country with higher skill dispersion specializes in the supermodular sector. Section 5 presents a series of simulations aimed at quantifying the empirical relevance of such counterintuitive theoretical scenarios. Section 6 concludes.

2 Model setup

Consider an economy with two sectors. In each sector i, with $i = \{1, 2\}$, workers are paired in teams of two to produce output according to the following production function:

$$F_i\left(t_A, t_B\right) = \left(t_A^{\lambda_i} + t_B^{\lambda_i}\right)^{\frac{\gamma}{\lambda_i}}, \quad \gamma > 0,\ \lambda_i > 0 \tag{1}$$

where subindices A and B denote the two tasks required in the production process, t_A and t_B are the skills of the workers allocated to each task, the parameter λ_i measures the degree of concavity or convexity of the isoquants associated with the production function in sector i and γ governs the degree of returns to skill. Total output in the sector, denoted y_i, is equal to the number or mass of pairs of workers allocated to the sector times the average output of each pair. Therefore, regardless of the returns to skill, the property of constant returns to scale in the number or mass of workers is preserved. This condition, along with the assumption of price-taking behavior, allows us to focus on competitive equilibria.

There are two countries, Home and Foreign, each populated by a mass of workers of measure one.[7] In Home, skills are distributed in the working population according to the cumulative distribution function $\Phi(t)$, while in Foreign such distribution is denoted by $\Phi^*(t)$. We maintain the assumptions of symmetry and continuity of the associated density functions $\phi(t)$ and $\phi^*(t)$, bounded and positive support (i.e. $t \in [t_{\min}, t_{\max}]$ and $0 < t_{\min} < t_{\max} < \infty$ for Home), and a common skill mean μ across countries. The latter allows us to focus squarely on the effect of skill dispersion on comparative advantage. Assuming continuous (i.e. atomless) density functions simplifies the general proofs that follow, but is not essential and will be relaxed in a simple example in Section 4.

While GM's analysis features a general production function with constant returns to skill, in this paper we focus on the specific parametric form in (1). There are two reasons for this choice. First, we want to be able to separately control the degree of convexity/concavity of isoquants and the degree of returns to scale. Second, we want to be able to hold the degree of returns to skill constant across sectors. This is important in view of the discussion by GM and the results in Bougheas and Riezman (2007), which show how different returns to skill across sectors will also generate comparative advantage. Similarly to GM, we want to shut down this mechanism and focus on the matching channel. Both these objectives can be achieved using the relatively common CES specification in (1).

[7] Since the production function features constant returns to scale, this assumption is made without loss of generality.

The hypothesis we focus on is the one put forward by GM and that we refer to as Skill-Dispersion-Complementarity (SDC) hypothesis.

Definition 1 *Skill-Dispersion-Complementarity (SDC) Hypothesis: countries characterized by a more dispersed skill distribution have a comparative advantage in the sector featuring a submodular production function (and thus relatively lower complementarity of skills).*

We start by examining the properties of the production function in (1). Contrary to the case of constant returns to skill, whether the production function is supermodular or submodular does not depend exclusively on the degree of concavity λ_i, but also on the parameter γ. In particular, because the function is continuous and twice differentiable, it is supermodular if and only if $\frac{\partial F_i}{\partial t_A \partial t_B} \geq 0$, a condition that is satisfied for $\gamma \geq \lambda_i$.[8] Conversely, the production function is submodular if and only if $\frac{\partial F_i}{\partial t_A \partial t_B} \leq 0$, i.e. when $\gamma \leq \lambda_i$. The special case $\lambda_1 < \gamma = 1 < \lambda_2$ is covered in GM, where sector 1 is supermodular and quasi-concave and sector 2 is submodular and quasi-convex. Hence we focus on the following parameter configurations: $\lambda_1 < \gamma < \lambda_2 < 1$ and $1 < \lambda_1 < \gamma < \lambda_2$.[9] While the proofs are similar in both cases, it is worth discussing their features separately.

Case I: dispersion-averse submodular function

Under the first parameter configuration, $\lambda_1 < \gamma < \lambda_2 < 1$, both sectors are characterized by convex isoquants and decreasing returns to skill, but sector 1's production function is supermodular, while sector 2's is submodular. In a frictionless economy like the one under consideration, the results in Becker (1973) apply and thus positive assortative matching (PAM) is optimal in the supermodular sector, while negative assortative matching (NAM) is optimal in the submodular sector. It is interesting to explore how this otherwise standard result emerges in this context.

[8] See Milgrom and Roberts (1990) and Shimer and Smith (2000).

[9] Other parameter configurations, such as $\lambda_1 < \gamma < 1 < \lambda_2$, are more similar to GM's case. We believe those cases involve less intuitive tradeoffs and do not deserve particular attention. Although we do not discuss them directly, they are covered in all simulations in Section 5.

There are two forces at play. The convexity of the isoquants creates a tendency towards PAM, or matching workers of similar skills. Conversely, decreasing returns to skill will tend to make matching two high-skill workers (and two low-skill workers) less desirable than cross-matching them. Which force prevails depends on whether $\gamma \gtrless \lambda_i$. In sector 1 the high degree of convexity of isoquants prevails over the strength of decreasing returns to skill, while in sector 2 the opposite happens.

As discussed in the introduction, it is interesting to note that despite being submodular, sector 2 always loses from an increase in skill dispersion: even though cross-matching workers is optimal, output decreases when those cross-matched workers become more different. As Figure 1 illustrates, this is a direct consequence of the convexity of the isoquants. The figure shows that the output of a pair of workers of skill $\mu + \varepsilon$ and $\mu - \varepsilon$ decreases as dispersion ε increases (keeping the average skill constant along the line $t_1 + t_2 = 2\mu$).

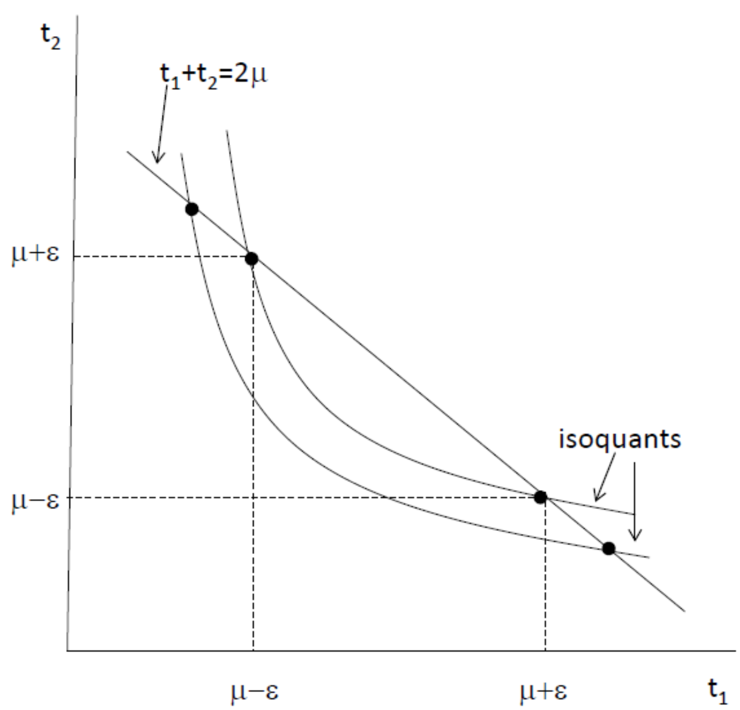

Figure 1: Isoquants of a dispersion-averse submodular production function

To the contrary, in GM the submodular sector always benefits from an increase in the skill dispersion of the workers employed. This different response to an increase in dispersion could in principle lead to a reversal of the SDC hypothesis, according to which countries with higher skill dispersion specialize in the submodular sector. Nevertheless, to establish the direction of comparative advantage, one also needs to assess the impact of skill dispersion on the supermodular sector. Under constant returns to skill, the supermodular sector's output is unaffected by an increase in skill dispersion, because output depends only on the aggregate allocation of skills in the sector. In our framework, however, an increase in dispersion reduces output in the supermodular sector, because of the decreasing returns to skill. Therefore whether higher dispersion raises or lowers *relative* output in the two sectors depends on which effect dominates. Section 4 presents an example in which the SDC hypothesis is violated.

Case II: dispersion-loving supermodular function

The analysis in the previous section carries through to the second parameter configuration we consider here: $1 < \lambda_1 < \gamma < \lambda_2$. Under these conditions both sectors exhibit increasing returns to skill and concave isoquants. This is again a case in which two forces shape the optimal matching of workers. On the one hand, increasing returns to skill create a tendency towards PAM because matching two high skill workers leads to very high output. On the other hand, concave isoquants indicate a preference for workers of different skills, and therefore NAM. Which force will prevail depends once again on whether $\gamma \gtrless \lambda_i$. Under the assumed parameter configuration, sector 2 is submodular and sector 1 is supermodular. Therefore, while it is optimal to assign workers of similar skills to each firm in sector 1, it is also true that increasing the dispersion of skills will increase aggregate output in this sector because of decreasing returns to skills. Since sector 2 is submodular and has concave isoquants, it is not different from the case described by GM. High skill dispersion in this case benefits both sectors, but we need to show whether it benefits sector 2 relatively more, i.e. check whether the SDC hypothesis holds for this parameter configuration. Although we do not discuss this case in detail, the analysis in sections 4 and 5 applies.

The next two sections have distinct roles. Section 3 characterizes the competitive equilibrium and shows that, for a given relative price of output, the optimal allocation of workers across sectors within a country is identical to the one obtained by GM, i.e. more extreme skills are assigned to sector 2. Section 4 introduces a second country in the analysis and shows how violations of the SDC hypothesis can emerge.

3 Equilibrium characterization

In this section we characterize the equilibrium allocation of workers in the Home economy with parameter configuration $\lambda_1 < \gamma < \lambda_2$, spanning cases I and II discussed above. The equilibrium features are that workers of extreme skills, i.e. those with skills below and above given thresholds are cross-matched and assigned to sector 2, while workers in the middle of the distribution are self-matched and produce good 1. These features are formally stated in the following two lemmas.[10]

Lemma 1 *Since F_1 is supermodular, then for any allocation of workers to sector 1, positive assortative matching is optimal. Since F_2 is submodular, then for any allocation of workers to sector 2, maximal cross-matching is optimal.*

These results linking whether the production function is supermodular or submodular to the optimal matching pattern in a frictionless economy are standard in the matching literature, e.g. Becker (1973) and Shimer and Smith (2000), so we refer the reader to those proofs. The next result shows that the relatively more productive pairs in sector 2 are the ones with extreme skills, i.e. the

[10] These lemmas have a parallel with Lemmas 2 and 3 in GM.

ones with the highest and lowest skills.

Lemma 2 *The efficient allocation of workers to the two sectors is characterized by a threshold $\widehat{t} \leq \mu$ such that all workers with skills below $\widehat{t}$ and above $2\mu - \widehat{t}$ produce good 2, while workers with skills between $\widehat{t}$ and $2\mu - \widehat{t}$ produce good 1.*

Proof. Consider an initial allocation in which all workers are assigned to sector 1. Then we ask what workers should first be reallocated from sector 1 to sector 2 in order to efficiently produce a small positive amount of good 2. When equal measures of workers of types t_A and t_B are reallocated from sector 1 to sector 2, the relative increase in output of good 2 relative to the loss of output of good 1 is given by:

$$s(t_A, t_B) \equiv 2^{1-\frac{\gamma}{\lambda_1}} \frac{\left(t_A^{\lambda_2} + t_B^{\lambda_2}\right)^{\frac{\gamma}{\lambda_2}}}{\left(t_A^{\gamma} + t_B^{\gamma}\right)}$$

The skill levels t_A and t_B that maximize s are the first workers to be optimally reallocated from sector 1 to sector 2. Any solutions to this maximization problem that are different from $(t_A, t_B) = (t_{\min}, t_{\max})$ or $(t_{\max}, t_{\min})$ must satisfy at least one of two conditions: $\frac{\partial s}{\partial t_A} = 0$ and $\frac{\partial s}{\partial t_B} = 0$, which simplify to:

$$\frac{t_A^{\gamma} + t_B^{\gamma}}{t_A^{\lambda_2} + t_B^{\lambda_2}} = t_j^{\gamma-\lambda_2}, \; j = A, B$$

and are only individually satisfied for $t_A = t_B$. At $t_A = t_B$, the second derivative with respect to t_j is equal to $2^{\frac{\gamma}{\lambda_2}-3} t_j^{-2} \gamma (\lambda_2 - \gamma)$ and always positive under the assumption $\lambda_2 > \gamma$, which implies that these critical points are not maxima. Therefore $s(t_A, t_B)$ is maximized at the extremes of the support $t_A = t_{\min}$ and $t_B = t_{\max}$ (or vice versa, given symmetry). Once those extreme types are matched and assigned to sector 2, the same logic applies to the remaining workers. Furthermore, by Lemma 1, a worker of skill $t \leq \widehat{t}$ is always matched with a worker of skill $2\mu - t$ in sector 2, while a worker of skill $\widehat{t} < t < 2\mu - \widehat{t}$ is self-matched in sector 1. ■

In light of the analysis in Section 5.3, it is important to note that the proofs in Lemmas 1 and 2 do not rely on the symmetry of the skill distribution $\Phi(t)$. These results can be readily extended with minor modifications. For example, in the asymmetric case, a worker of skill $t \leq \widehat{t}$ that is assigned to sector 2 is matched to a worker of skill t', such that $\Phi(t') = 1 - \Phi(t)$.

In order to complete the characterization of the equilibrium in the case of symmetric skill distributions, we have to describe how the equilibrium threshold is determined and related to the relative price of good 2, which we denote by p. In a competitive equilibrium, and assuming a positive production of both goods, the relative price p is equal to the absolute value of the slope of the production possibilities frontier (PPF). The economy produces the output levels of goods 1 and 2 determined by this condition. This determines the equilibrium allocation of workers in each sector, which is characterized by an equilibrium threshold, denoted $\widehat{t}_p$. The proof of Lemma 2 implies that the slope of the PPF is given by $s(t_A, t_B)$ evaluated at $t_A = \widehat{t}_p$ and $t_B = 2\mu - \widehat{t}_p$. It is straightforward to verify that $s\left(\widehat{t}_p, 2\mu - \widehat{t}_p\right)$ is strictly decreasing in $\widehat{t}_p$ given $\gamma < \lambda_2$ and $\widehat{t}_p \leq \mu$, which implies a one-to-one relationship between p and $\widehat{t}_p$.

This result is important in light of the next step in the analysis which involves introducing a second country and assessing comparative advantage. We assume countries can trade without frictions, so that under free trade they face the same relative price p. This implies that the two countries will feature the same threshold $\widehat{t}_p$. The key difference between the two countries is therefore the type and number of individuals with skills lower than $\widehat{t}_p$ and higher than $2\mu - \widehat{t}_p$, which in turn depends on the shape of the skill distribution and, in particular, the degree of skill dispersion. This is where the analysis by GM and this more general case differ substantially. The definition of skill dispersion adopted in GM (p. 1264) is a restricted version of mean-preserving spread, which dictates that Foreign has a more dispersed skill distribution than Home if its skill distribution can be generated by a series of elementary increases in risk. This means that Foreign's skill distribution can be generated by shifting mass from "intermediate" skill levels to more extreme ones. Under constant returns to skill these shifts of probability mass generate a comparative advantage in the submodular sector 2 for Foreign because, regardless of where $\widehat{t_p}$ lies, a shift of mass either increases y_2 or it leaves it unchanged and it either reduces or leaves the amount y_1 unchanged. After discussing the more general case, it should be clear that this logic will not always apply when we relax constant returns to skill since now y_2 could also decrease. In order to provide more intuition for this result, the next section presents an example where the SDC hypothesis is violated.

Finally, we would like to remark that, similarly to GM, under parameter configurations in which both sectors have supermodular technologies, we find a general 'no-trade' result, i.e. that differences in skill dispersion across countries do no generate comparative advantage.

Proposition 3 *If $\lambda_1 < \lambda_2 < \gamma$, i.e. both sectors have supermodular production functions, the competitive equilibrium has no trade.*

Proof. It is immediate to verify that the slope of the PPF and thus the autarky relative price are given by $2^{\frac{\gamma}{\lambda_2}-\frac{\gamma}{\lambda_1}}$, which is independent of the distribution of skills. ■

The intuition for this result relies on the fact that, independently of whether sectors benefit or lose from higher skill dispersion, PAM in both sectors implies that the relative productivity of any pair of self-matched workers is constant across sectors. This implies that two countries that only differ in terms of skill dispersion will have identically shaped PPFs.

4 Engineering a violation of the SDC hypothesis

To understand why violations of the SDC hypothesis might happen, it is useful to provide a simple example. To this end we assume a discrete support for the skill distribution and allow workers to have one of six skill levels: $\{\mu-\varepsilon'', \mu-\varepsilon', \mu-\varepsilon, \mu+\varepsilon, \mu+\varepsilon', \mu+\varepsilon''\}$ with $0 < \varepsilon < \varepsilon' < \varepsilon'' < \mu$. We refer to workers of skill $\mu-\varepsilon$ and $\mu+\varepsilon$ as workers of skill type ε.

Figure 2 displays the distribution of skills in both Home and Foreign country. Using the original definition by GM, Foreign has a more dispersed skill distribution. In fact, Foreign's skill distribution can be generated from the Home skill distribution through a series of elementary mean preserving spreads.

To study the pattern of trade, we construct the PPF for both countries in Figure 3 using expression (1) for parameter values $\lambda_1 < \gamma < \lambda_2 < 1$, corresponding to Case I discussed earlier. The PPF is built by shifting the threshold $\widehat{t}$ over the range $[\mu-\varepsilon'', \mu-\varepsilon]$: for any $\widehat{t}$, workers of extreme types are cross-matched and employed in sector 2, while the rest are self-matched and employed in sector 1. Therefore the PPF consists of three segments. Points along segment AB of

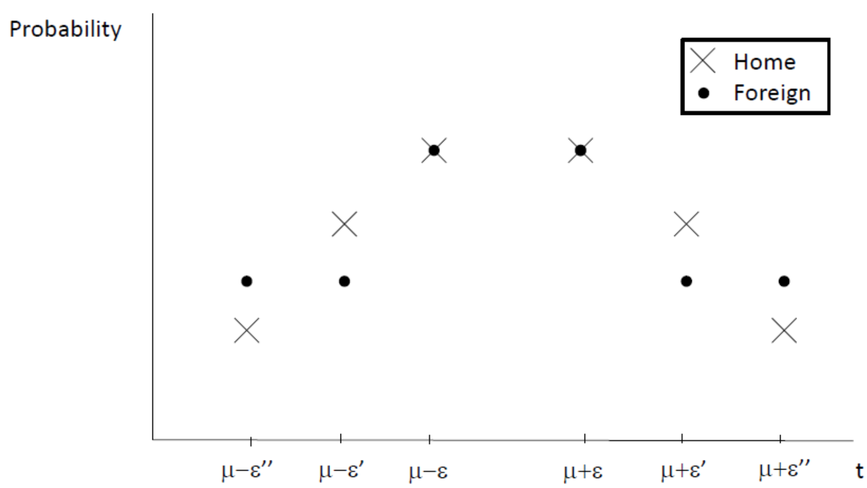

Figure 2: Skill distributions in Home and Foreign

the PPF in Home correspond to an allocation of workers such that skill types ε' and ε'' are fully employed in sector 2, while skill types ε are employed in both sectors (albeit matched differently). As we move down the PPF and reach point B, every worker of type ε is employed in sector 1.

Between points B and C, workers of skill type ε' are split between sectors 1 and 2, and so on. The slope of each segment of the PPF is determined by the relative productivity of the marginal pairs of workers in each sector, which in turn depends on the skill type that is being employed in both sectors. For example, for points along segment AB, the slope is given by $s(\mu + \varepsilon, \mu - \varepsilon)$. Note that the same argument is valid for the PPF in Foreign. However, the lengths of the three segments vary by country, reflecting unequal proportions of workers with different skills.

Figure 3 shows that PPF's segments AB and A*B* are identical in both Home and Foreign (in bold). Foreign's PPF lies inside Home's PPF because Foreign has higher skill dispersion, which reduces absolute productivity in both sectors. It is easy to verify that, in this example, if the common free trade price is the slope of the dashed line in Figure 3, the relative production y_2/y_1 may be higher in Home, which is the country with lower skill dispersion. This violates the SDC hypothesis.

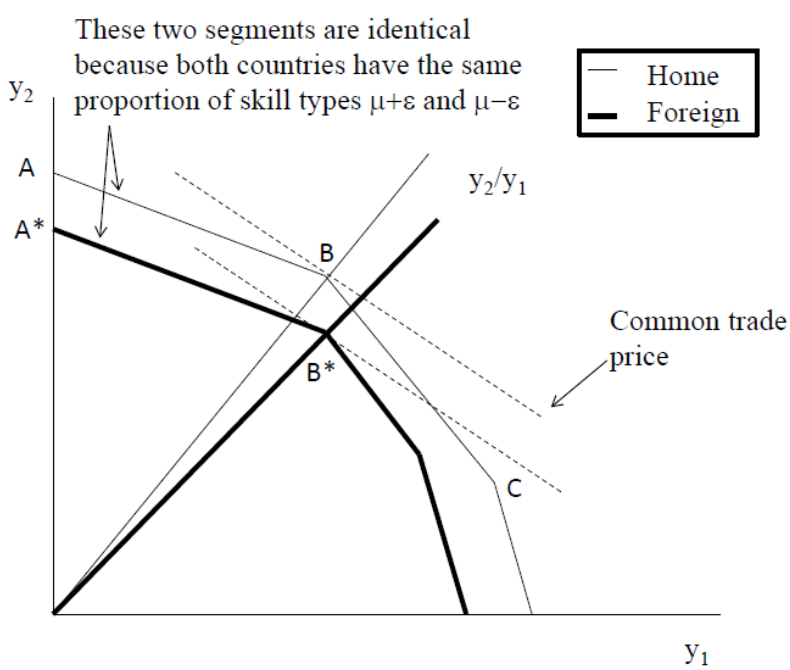

Figure 3: Production Possibility Frontiers of Home and Foreign

This happens because at kinks like B and B*, Home and Foreign produce the same amount of good 1 (as they have the same number of workers of skill type ε), but Foreign is producing *less* y_2, due to higher skill dispersion. This follows from the fact that, despite being submodular, sector 2 does not benefit from having a relatively higher share of workers of skill type ε''. This is in stark contrast to the assumption made by GM that sector 2 always benefits from dispersion, so that Foreign would produce more, not less, y_2.[11]

One may wonder whether SDC violations could generally be avoided by resorting to some other standard definition of dispersion. The example above rules out mean-preserving spreads (and thus second-order stochastic dominance) and the stronger criterion based on the monotone likelihood ratio condition used in Costinot and Vogel (2010). This suggests that although it might be possible to devise other ad hoc definitions of dispersion to rule out SDC violations, they would likely be restrictive. For this reason we do not further explore this avenue and, in what follows, we take the alternative approach of assessing the empirical relevance of SDC violations. To do so, we employ empirical skill distributions to evaluate how often the SDC hypothesis can be violated for a wide set of parametrizations of the production technologies.

5 Numerical exercises using empirical skill distributions

The previous section engineered an example in which a country with a more dispersed skill distribution may choose not to specialize in the submodular sector. The example questions the general validity of the SDC hypothesis. However, it is unclear to what extent such potential violations should be a concern from an empirical point of view. In what follows, we investigate how frequently the SDC hypothesis is violated, under an extensive range of technology parameters, by

[11] Notice that this analysis does not depend on the two countries being of the same size, since country size only affects the location of the PPF, but not the shape of the PPF.

restricting the analysis to the distributions of literacy scores observed in 19 countries.

We perform two kinds of exercises. First, we fit parametric density functions to the empirical IALS score distributions of each country. Such fitted distributions are symmetric by construction, have identical means across countries, but differ in their dispersion. Then, we feed the estimated distributions into our two-sector model, compute the relative output produced by each country for a range of technology configurations and compare the resulting patterns of comparative advantage for each country pair against the SDC hypothesis. Second, we perform a numerical simulation exercise that is conceptually similar, yet directly relies on the IALS score distributions without the need for parametric assumptions. We discuss both these approaches and report results after describing the data.

5.1 Data

We use test scores from the 1994-1998 International Adult Literacy Survey (IALS) to approximate the skill distributions in 19 countries. The IALS consists of an internationally comparable test of work-related literacy skills, administered to a representative sample of adults between the ages of 16 and 65 in each of the participant countries. It focuses on measuring skills that are needed for everyday tasks across three different dimensions of literacy: quantitative, prose and document literacy (see OECD and StatsCan (2000) for further details). As in Bombardini et al. (2012a), we combine the results of these three tests into a single average score for each individual and restrict the sample to adults participating in the labor market at the time of the survey. The scores are originally measured on a scale from 0 to 500. However, we normalize them by the corresponding country mean, as a way of eliminating cross-country variation in skill abundance. Figures 4 and 5 present the histograms of the score distributions for every country in the sample.

5.2 Simulations based on fitted score distributions

First, we generate a grid of different degrees of returns to skill and convexity (technology parameters γ, λ_1 and λ_2) as well as a grid for the threshold $\widehat{t}$, which sorts workers by sector. In particular, maintaining the assumption that $\lambda_1 < \gamma < \lambda_2$, we let λ_1, γ and λ_2 vary between 0.05 to 5 with increments of 0.3.[12] The threshold $\widehat{t}$ varies between 0.01 and 1 with increments of 0.1. Second, for each pair of countries in the sample, we use the estimated skill distributions together with the production technologies in (1) and Lemmas 1 and 2 to evaluate relative output y_2/y_1 for each country in the pair, under the assumption that they face the same relative price and therefore the same threshold $\widehat{t}$:

$$\frac{y_2}{y_1} = \frac{\int_{t_{\min}}^{\widehat{t}} \left[t^{\lambda_2} + (2-t)^{\lambda_2} \right]^{\frac{\gamma}{\lambda_2}} \phi(t)\, dt}{2^{\frac{\gamma}{\lambda_1}} \int_{\widehat{t}}^{1} t^{\lambda_1} \phi(t)\, dt}$$

where $\phi(t)$ is the estimated density. By evaluating the relative production for each pair of countries, we are able to compute the share of violations of the SDC hypothesis. We define a violation as a case in which the more dispersed country in the pair produces a higher ratio y_2/y_1.

Each country's skill distribution is approximated by a parametric function. The fitted distributions satisfy, by construction, the theoretical requirements of bounded support, identical means, and are constrained to be symmetric, following the restrictions imposed in the model. This approach inevitably means that some features of the empirical distributions are lost or misrepresented. However, it also allows us to focus exclusively on the effects of cross-country differences that are relevant in the theory (e.g. differences in skill dispersion) in a tractable way, using parametric distributions which approximate the empirical ones.

We use truncated Normals to fit the empirical score distributions because of their simplicity (they can be summarized by two moments) and their widespread application. However we are

[12] Note that we allow for decreasing and increasing returns to skill, encompassing cases I and II discussed in section 2, among others.

careful to allow for symmetric truncation points, accounting for the fact that IALS scores only assume values on a bounded support. As a robustness check, we also repeat the analysis using beta distributions, which are quite flexible.[13] These additional results are presented in the Appendix and strongly confirm our findings for the truncated Normal approximations.

Fitting a truncated Normal distribution to the skill distribution of each country i requires the estimation of two parameters: the standard deviation, σ_i, and the lower and upper truncation points, $t_{\min,i}$ and $t_{\max,i}$, under the constraint $t_{\min,i} = 2 - t_{\max,i}$. We employ maximum likelihood estimation (MLE) to obtain point estimates for these country-specific parameters. The estimation results are reported in Table 1 and the fitted distributions are displayed in Figure 4 and 5.

5.2.1 Rankings and violations

Inspection of the fitted parameters immediately reveals a difficulty in ranking countries according to their skill dispersion. This difficulty follows from the fact that two features of the distribution, the σ and the support, both determine the dispersion. These features sometimes conflict with each other. More specifically we notice that within a pair, a country may have higher σ, but a narrower support. Looking at Table 1 this is the case for example of the US and Chile, with σ higher in Chile and the range wider in the US. In such a situation it is difficult to clearly state which of the two countries in question exhibits more skill dispersion. We refer to these cases as 'ambiguous'.

To keep things simple, we start by sidestepping the problem due to ambiguous rankings, and we perform a simulation exercise in which we vary each parameter one at a time. We vary σ between 0.1 and 0.3 in steps of 0.03, a range that encompasses all the estimates in the sample. We also let the range vary between 1 and 1.9, once again covering much more than the variation in the data. The results of this exercise show that the SDC hypothesis is never violated when comparing two

[13] In the subset of symmetric functions, the beta encompasses such common cases as uniform, U-shaped densities and various types of unimodals.

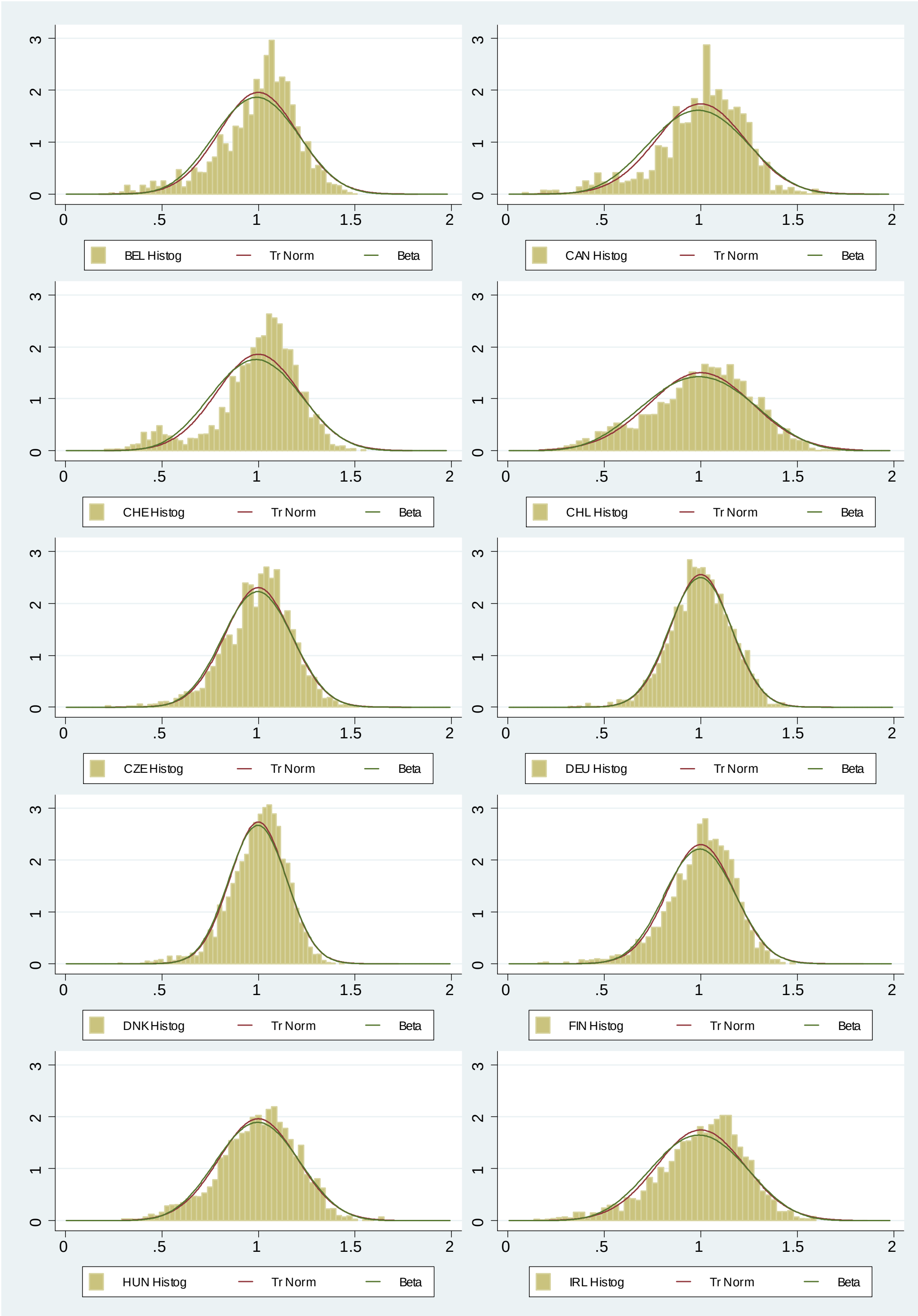

Figure 4: IALS scores histograms, fitted truncated normal and beta distributions for each country.

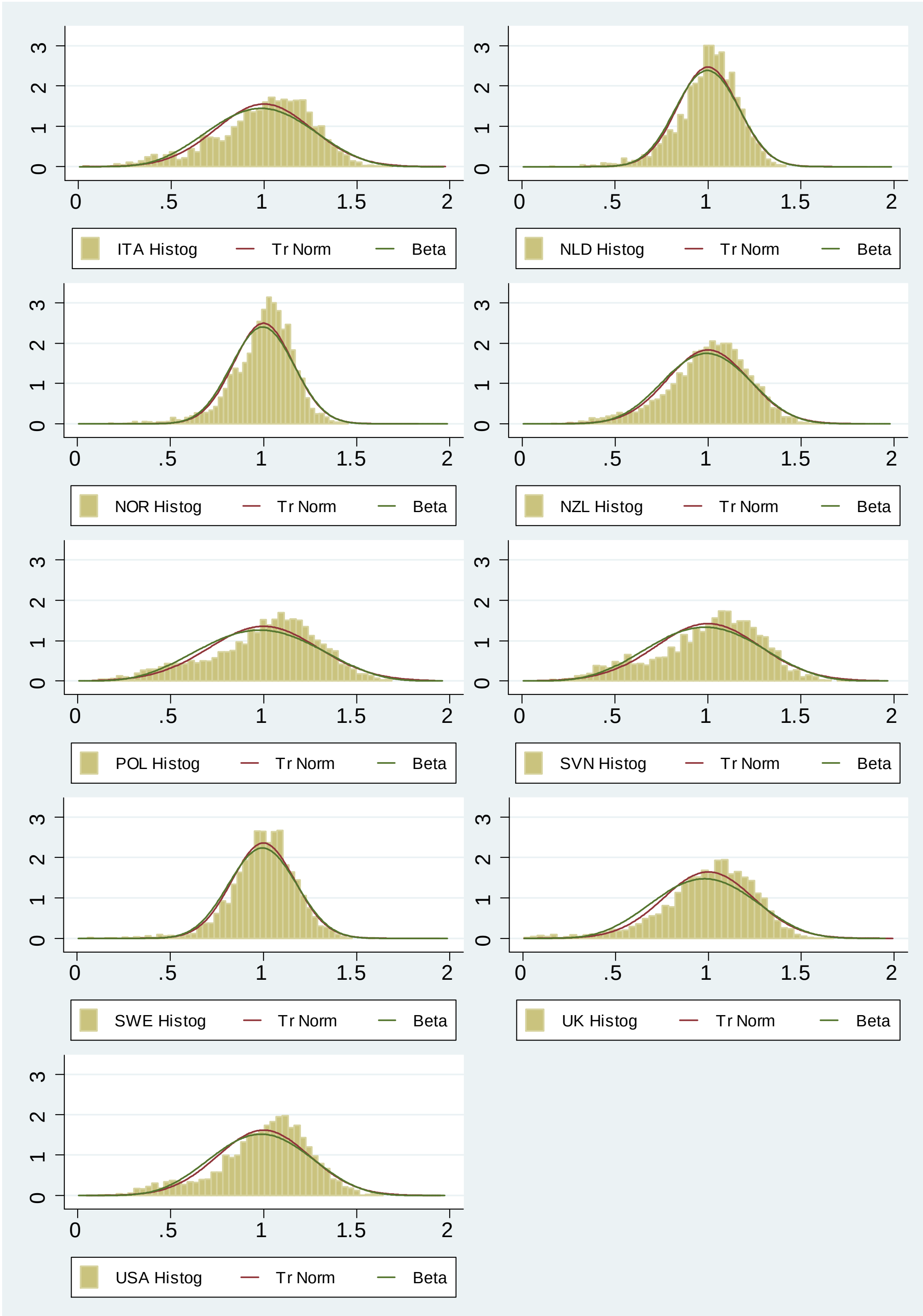

Figure 5: IALS scores histograms, fitted truncated normal and beta distributions for each country.

countries exclusively ranked in terms of σ or in terms of the range.

Next, we consider the fitted distributions, allowing for possibly conflicting rankings over σ and range. We choose to rank countries according to the standard deviation of the skill dispersion[14] Out of 171 pairs of countries we observe violations in 26 cases. Overall we observe a violation in 1.82% of the grid of technology parameters and prices. It is important to note that the only two factors that drive violations are the values of the threshold $\widehat{t}$ (which corresponds to the equilibrium price) and the shape of the skill distribution of each country in the pair. Table 2 reports whether for each pair (i) the dispersion ranking is ambiguous, and (ii) a violation of the SDC hypothesis occurs. A clear and important result in Table 2 is that all of the violations occur for country pairs in which the ranking is ambiguous. We should stress that what we denote as violations are patterns of trade which contradict the SDC when we use a ranking based on standard deviations: in other words, these patterns of trade may still be consistent with the SDC hypothesis if one were to use a dispersion ranking different from the one we adopt.

We also find that violations occur mostly for countries that have similar standard deviation of skills. In other words, the SDC hypothesis is violated in country pairs where the intensity of comparative advantage due to differences in the standard deviation of skills is weak enough that the quantitative impact of a higher σ is roughly offset by the narrower range. In fact, while ambiguities exist also for pairs far from the main diagonal in the table, they do not lead to violations.

5.2.2 Testing for a common range

So far we have shown two main results: (1) no violation of the SDC hypothesis occurs when we only vary σ across countries, holding the skill range constant; (2) violations only occur when there is ambiguity in dispersion ranking and countries have relatively similar standard deviation of skills.

[14] It turns out that this ranking coincides with the σ ranking.

In this section we show that cross-country differences in the skill range are both negligible and statistically insignificant. When we assume that the skill range is common across countries we revert to the no violation scenario, providing empirical support for the SDC hypothesis.

To test for common skill support, we consider the estimated truncation points of the IALS distributions, as reported in table 1. In principle one could use bootstrap methods to approximate standard errors for all fitted distribution parameters. However the bootstrap cannot be used to produce confidence intervals for the truncation points' estimates.[15] MLE yields estimates of the lower and upper bounds of the support that coincide with the sample minimum and maximum scores, respectively. Bootstrapping is inconsistent for this type of order statistics (see Andrews, 2000). For this reason, we use asymptotic standard errors to construct confidence intervals for the extremes of the truncated Normal distributions.[16] The estimated standard errors for the truncation points of the Normals are very large, suggesting that: (i) the truncation points lie at the far tails of the support for all countries, in areas with very low mass; (ii) there is no significant difference in the extremal skills in our sample of countries; (iii) truncating the distributions adds little information and using untruncated Normal densities would be an equally satisfactory way to approximate the empirical distributions. A detailed discussion of issues related to testing for differences in truncation points is in Appendix A.

In summary, these findings suggest that, based on IALS data, it is reasonable to assume common skill ranges for the countries in our sample, something that leads to the conclusion that the SDC hypothesis is not violated for the wide range of technology parameters, and prices, that we consider.

[15] We do bootstrap standard errors for the estimated σ_i. They are available from the authors.

[16] We use the fact that the estimate $X_{(1)}$ of the minimum of a bounded, continuous distribution $F(X)$ is asymptotically distributed as an extreme value random variable, that is $P\left\{nX_{(1)} > x\right\} = e^{-x}$. Ignoring the country index, we construct confidence intervals for the lower bound using the approximation $P\left\{n\left(X_{(1)} - t_{\min}\right) > x\right\} \approx \exp\left(-\left\{\frac{\phi(t_{\min})x}{\Phi(t_{\max})-\Phi(t_{\min})}\right\}\right)$. By symmetry, the upper bounds of the distributions are estimated as $X_{(N)} = 2\mu - X_{(1)}$.

5.3 Simulations based on empirical score distributions

In this section we take a different approach to evaluating the frequency of violations of the SDC hypothesis. Rather than using symmetric fitted distributions, we use the empirical distributions of scores in each country. As before, we simulate the model using the same grid of technology parameters. However, given the fact that empirical distributions are discrete and asymmetric, the exercise entails constructing numerical PPFs and checking that, at every possible price, the output ratio y_2/y_1 is larger in the country characterized by a higher skill dispersion, according to the SDC hypothesis.

The procedure can be summarized as follows: the numerical PPF is constructed starting from the actual IALS score distributions for the 19 countries. Without loss of generality, we normalize the size of each sample to 10,000, by scaling the sampling weight of each individual score.[17] Then, for each country, the maximum output in sector 2 is calculated using the production technologies in (1) and Lemma 1, by cross-matching all workers and allocating them to sector 2. For a country to start producing a positive y_1, the first workers to be reallocated to sector 1 are the four around the median $(t_{4999}, t_{5000}, t_{5001}, t_{5002})$, as dictated by Lemma 2. When these 4 workers are reallocated to sector 1, they are positively assortatively matched. Although the PPF is not continuous, we can calculate the slope of the line connecting the first and the second point on the PPF. Such slope is given by:

$$s_1 = \frac{\left(t_{4999}^{\lambda_2} + t_{5002}^{\lambda_2}\right)^{\frac{\gamma}{\lambda_2}} + \left(t_{5000}^{\lambda_2} + t_{5001}^{\lambda_2}\right)^{\frac{\gamma}{\lambda_2}}}{\left(t_{4999}^{\lambda_1} + t_{5000}^{\lambda_1}\right)^{\frac{\gamma}{\lambda_1}} + \left(t_{5001}^{\lambda_1} + t_{5002}^{\lambda_1}\right)^{\frac{\gamma}{\lambda_1}}}$$

The remaining 2500 slopes are calculated analogously, by reallocating 4 more "extreme" workers at a time from sector 2 to sector 1. After constructing this vector of 2500 slopes for all 19 countries, we compute the relative production y_2/y_1 of all countries at different relative prices and record,

[17] The individual IALS scores are adjusted with a sampling weight, which we treat as a frequency.

for each pair of countries in the sample, whether the country that presents a more dispersed skill distribution (as defined by the standard deviations) produces a higher y_2/y_1 ratio. We experimented with the range of prices, since at the extremes some countries do not produce one of the two goods; therefore, varying the price over that region does not produce informative behavior. In the results reported, we adopt a range of prices between the highest of the slopes s_1 among our 19 countries, and the lowest of the slopes s_{2500} among the same countries. We adopt the same range for remaining technology parameters.

Overall the violation rate is 4.4%, i.e. for 4.4% of the points in the grid of prices and technology parameters we observe a violation of the SDC hypothesis. As one would expect the violation rate in this exercise is slightly higher than what we find in the parametric exercise. This follows from the fact that the empirical distributions differ in many characteristics other than the variance. Unlike in the parametric exercise, in this case we cannot hold these other features constant across countries and they may generate some of the observed violations. We further explore bilateral violations in Table 3 and we investigate whether the patterns align with the analysis based on parametric distributions. Table 3 reports whether for each pair (i) the variances of the two countries are statistically different at the 1% level based on an F-test[18] and (ii) a violation of the SDC hypothesis occurs. This exercise broadly confirms the findings from the parametric analysis. Most of the violations happen near the main diagonal, where countries are difficult to rank in terms of dispersion, i.e. when countries have variances that are non-significantly different. In other words, the SDC hypothesis is violated in country pairs where the intensity of comparative advantage due to skill dispersion is low and perhaps dominated by other features of the distribution.

[18] We adopt a critical value of 1.177 based on a two-sided test with 1000 degrees of freedom for both numerator and denominator.

6 Concluding remarks

This paper has shown that, in the context of a trade model with two countries and two sectors, decoupling supermodularity and quasi-concavity (or submodularity and quasi-convexity) may produce situations in which submodular sectors do not benefit from skill dispersion in an absolute sense. We show that this decoupling makes it possible to engineer mean-preserving spreads in skills which contradict the main prediction of the SDC hypothesis, namely that the country with higher skill dispersion specializes in the submodular sector.

The theoretical violations we engineer are driven by peculiar changes in skill dispersion, and this makes it difficult to obtain more robust theoretical predictions without imposing more restrictions on the problem. To assess the relevance of these theoretical difficulties we pursue a different route, which does not require us to impose more structure on the model, using empirical distributions of literacy scores for 19 countries. Our results suggest that the SDC predictions are reverted exclusively when countries in a pair are not easy to rank in terms of dispersion parameters. For example, in the parametric exercise, these situations can be easily portrayed by the cases in which one country has higher σ, but narrower skill range than the other. We see this as a positive result, suggesting that the predictions of the SDC hypothesis remain robust when we generalize the production technology to cases in which concavity and supermodularity are decoupled and over a very wide range of technology parameters.

Lastly, we have not discussed the case of unobservable skills, studied by GM and more recently by Bombardini et al. (2012b) in a multi-country, multi-sector setting. However, their results are robust to our analysis. It is indeed the case that adding unobservable skill dispersion reinforces the pattern of comparative advantage. The intuition relies, similarly to those papers, on the fact that the supermodular sector always features 'more convex' isoquants than the submodular sector (under the stated assumption that the two sectors have the same degree of returns to skill). When random matching prevails, then a more dispersed skill distribution causes the largest productivity drops in the sector with relatively more convex isoquants, even if both sectors feature concave isoquants.

References

Andrews, D., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68 (2), 399–405.

Atakan, A., 2006. Assortative matching with explicit search costs. Econometrica 74 (3), 667–680.

Becker, G., 1973. A theory of marriage: Part i. The Journal of Political Economy 81 (4), 813–846.

Bombardini, M., Gallipoli, G., Pupato, G., 2012a. Skill dispersion and trade flows. American Economic Review 102 (5), 2327–48.

Bombardini, M., Gallipoli, G., Pupato, G., 2012b. Unobservable Skill Dispersion and Comparative Advantage. Mimeo, The University of British Columbia.

Bougheas, S., Riezman, R., 2007. Trade and the distribution of human capital. Journal of International Economics 73 (2), 421–433.

Choquet, G., 1953. Theory of capacities. In: Annales de l'institut Fourier. Vol. 5. p. 87.

Costinot, A., 2009. An Elementary Theory of Comparative Advantage. Forthcoming Econometrica.

Costinot, A., Vogel, J., 2010. Matching and Inequality in the World Economy. Journal of Political Economy 118 (4), 747–786.

Davidson, C., Sly, N., 2010. Forthcoming.Ştrade and the labor market: Recent developments and new frontiers.Ť.

Eeckhout, J., Kircher, P., 2011. Identifying sorting: In theory. The Review of Economic Studies 78 (3), 872–906.

Grossman, G. M., 2004. The Distribution of Talent and the Pattern and Consequences of International Trade. Journal of Political Economy 112 (1), 209–239.

Grossman, G. M., Helpman, E., Kircher, P., 2013. Matching and sorting in a global economy. Working paper.

Grossman, G. M., Maggi, G., 2000. Diversity and trade. American Economic Review, 1255–1275.

Helpman, E., Itskhoki, O., Muendler, M.-A., Redding, S. J., 2012. Trade and inequality: From theory to estimation. National Bureau of Economic Research.

Huang, F., Chang, P.-L., 2012. Trade and Divergence in Education Systems. Forthcoming International Economic Review.

Kremer, M., Maskin, E., 1996. Wage inequality and segregation by skill. National Bureau of Economic Research.

Legros, P., Newman, A. F., 2002. Monotone matching in perfect and imperfect worlds. The Review of Economic Studies 69 (4), 925–942.

Lentz, R., 2010. Sorting by search intensity. Journal of Economic Theory 145 (4), 1436–1452.

Lopes de Melo, R., 2011. Sorting in the labor market. Mimeo, Chicago University.

Marinacci, M., Montrucchio, L., 2008. On concavity and supermodularity. Journal of Mathematical Analysis and Applications 344 (2), 642–654.

Milgrom, P., Roberts, J., 1990. The economics of modern manufacturing: Technology, strategy, and organization. The American Economic Review, 511–528.

OECD, StatsCan (Eds.), 2000. Literacy in the Information Age: Final Report of the International Adult Literacy Survey. OECD, Paris, France.

Ohnsorge, F., Trefler, D., 2007. Sorting It Out: International Trade with Heterogeneous Workers. Journal of Political Economy 115 (5), 868–892.

Postel-Vinay, F., Robin, J., 2002. Equilibrium wage dispersion with worker and employer heterogeneity. Econometrica 70 (6), 2295–2350.

Shimer, R., Smith, L., 2000. Assortative matching and search. Econometrica 68 (2), 343–369.

Teulings, C., Gautier, P., 2004. The right man for the job. Review of Economic Studies 71 (2), 553–580.

Table 1: Estimated parameters - Truncated Normal distribution

Country	Sigma (1)	Lower truncation point (2)	Upper truncation point (3)
DNK	0.15	0.27	1.73
DEU	0.16	0.31	1.69
NOR	0.16	0.17	1.83
NLD	0.16	0.15	1.85
SWE	0.17	0.06	1.94
CZE	0.17	0.21	1.79
FIN	0.17	0.16	1.84
HUN	0.20	0.29	1.71
BEL	0.20	0.17	1.83
NZL	0.22	0.16	1.84
CHE	0.21	0.20	1.80
IRL	0.23	0.14	1.86
CAN	0.23	0.08	1.92
USA	0.25	0.05	1.95
UK	0.24	0.01	1.99
ITA	0.26	0.02	1.98
CHL	0.27	0.16	1.84
SVN	0.28	0.08	1.92
POL	0.29	0.08	1.92

Notes: This table reports ML estimation results for truncated Normal distributions. Columns 1, 2 and 3 report sigma, lower truncation point and upper truncation point for a truncated Normal with mean 1. Countries are ordered by increasing skill variance

Table 2: Ranking ambiguities and SDC violations

	DNK	DEU	NOR	NLD	SWE	CZE	FIN	HUN	BEL	CHE	NZL	IRL	CAN	UK	USA	ITA	CHL	SVN	POL
DNK		A,V	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-
DEU			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NOR				-	-	A,V	-	A,V	A	A	-	-	-	-	-	-	-	-	-
NLD					-	A,V	A,V	A,V	A	A	A	-	-	-	-	-	A	A	-
SWE						A,V	A,V	A,V	A,V	A	A	A	A	-	-	-	A	A	A
CZE							-	A,V	-	-	-	-	-	-	-	-	-	-	-
FIN								A,V	A	A	-	-	-	-	-	-	A	-	-
HUN									-	-	-	-	-	-	-	-	-	-	-
BEL										A,V	-	-	-	-	-	-	-	-	-
CHE											-	-	-	-	-	-	-	-	-
NZL												-	-	-	-	-	A	-	-
IRL													-	-	-	-	A,V	-	-
CAN														-	-	-	A,V	A	A
UK															A,V	A,V	A,V	A,V	A,V
USA																-	A,V	A,V	A,V
ITA																	A,V	A,V	A,V
CHL																		-	-
SVN																			-
POL																			

Note: The letter A indicates ambiguous dispersion ranking between the countries in the pair, i.e. higher range and lower σ or the opposite. The letter V indicates a 'violation', i.e. cases in which the %more dispersed country according to standard deviation produces relatively more of good 2.

Table 3: Statistically different variances and SDC violations

	DNK	DEU	NOR	NLD	SWE	CZE	FIN	HUN	BEL	CHE	NZL	IRL	CAN	UK	USA	ITA	CHL	SVN	POL
DNK		N,V	V*	V*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DEU		-	N,V	N,V	N,V	V*	V*	-	-	-	-	-	-	-	-	-	-	-	-
NOR				N,V	N,V*	N,V	V	V*	-	-	-	-	-	-	-	-	-	-	-
NLD					N,V*	N	N	-	-	-	-	-	-	-	-	-	-	-	-
SWE						N,V	N,V	V	V*	-	-	-	-	-	-	-	-	-	-
CZE							N,V	-	V*	V*	-	-	-	-	-	-	-	-	-
FIN								-	-	-	-	-	-	-	-	-	-	-	-
HUN									N,V	N,V	N,V	-	-	-	-	-	-	-	-
BEL										N,V	N,V	-	-	-	-	-	-	-	-
CHE											N,V	N,V	N,V	-	-	-	-	-	-
NZL												N	N,V	-	-	-	-	-	-
IRL													N,V	N	N,V	-	-	-	-
CAN														N,V	N,V*	-	-	-	-
UK															N,V	N,V*	V	V*	-
USA																N,V	N,V*	-	-
ITA																	N,V	N,V*	-
CHL																		N,V	-
SVN																			N
POL																			

Note: The letter N indicates that one cannot reject the null that the variances of the two countries are the same at the 1% significance level. The letter V indicates a ‘violation’, i.e. cases in which the more dispersed country according to standard deviation produces relatively more of good 2. V* indicates that the violation incidence is less than 5% over the search grid defined in Section 5.2.

A Appendix

A.1 Estimation and inference about boundary values

Consider a sample of observations of a random variable (say, test scores), drawn from a population which assumes values on a finite range $[\theta_L, \theta_H]$, with $0 < \theta_L < \theta_H < \infty$. Using this observed sample, one may estimate the shape of the distribution of test scores by making some parametric assumptions and using a maximum likelihood estimator.

The choice of parametric function should take into account the fact that the random variable assumes values on a compact set. Therefore it is good practice to start by using functions which are defined on a bounded and closed domain. However, this means that the boundaries, or truncation points, must also be estimated. For our specific case, we also restrict the probability distribution to be symmetric, which implies that the boundaries are equidistant from the mean. In this case, one can concentrate on estimating only one of them (e.g. the minimum).

Denote the estimator of the lowest point in the range as $\hat{\theta}_L$ (by symmetry, we assume that $\hat{\theta}_H = 2\mu - \hat{\theta}_L$). Building a confidence interval around $\hat{\theta}_L$ might be desirable in order to compare the lower bounds estimated from samples drawn from two different countries and assess whether they are significantly different from each other.

A.2 Distribution of the estimator

The bootstrap method cannot be used to construct confidence intervals in this case, as it is not consistent if the parameter of interest is on a boundary of the parameter space as described in Andrews (2000) .[19] Given the problem with bootstrapping confidence intervals for a minimum, one

[19] Bootstrap methods remain a valid way to approximate the distribution of other estimators, e.g. the estimator of the mean of the distribution of interest.

can derive the asymptotic distribution of the estimator.

Begin by considering a set $\{U_1, ..., U_n\}$ of i.i.d. draws from a uniform distribution $[0,1]$. Let $U_{(1)} = \min\{U_i\}$. One can show that

$$P\left\{nU_{(1)} > x\right\} \to e^{-x}$$

where e^{-x} is an extreme value density of type III (see "Asymptotic Statistics", van der Vaart, p.312).[20] Notice also that $P\left(nU_{(1)} < x\right) = 1 - e^{-x}$ and that $f(x) = e^{-x}$.

Next, suppose that X_1,...,X_n are continuous random variables that are i.i.d. draws from a non-degenerate cumulative distribution F with a bounded support $[\theta_L, \theta_H]$. Finally, let $U_i = F(X_i)$. Then the asymptotic distribution of the order statistic $X_{(1)}$ is approximately

$$P\left\{n\left(X_{(1)} - \theta_L\right) > n\left(F^{-1}\left(\frac{x}{n}\right) - \theta_L\right)\right\} \simeq e^{-x}$$

This probability can also be rewritten as,

$$P\left\{n\left(X_{(1)} - \theta_L\right) > x\right\} \simeq \exp\left\{-n\left(F\left(\frac{x}{n} + \theta_L\right)\right)\right\}$$

If we choose the estimator $\hat{\theta}_L = X_{(1)}$ for the lower bound of the support, the last expression provides a density function for the asymptotic distribution of this estimator. This asymptotic distribution will depend on the specific function F that is being fit to the data. Below we discuss the case in which F is a truncated Normal, with symmetric truncation points given by θ_L and θ_U.

[20] Simple proof: $P\left(k_n U_{(1)} > x\right) = P\left(U_{(1)} > \frac{x}{k_n}\right)$ where k_n is an arbitrary constant. As long as $0 < \frac{x}{k_n} < 1$ we have that $P\left(k_n U_{(1)} > x\right) = \left(1 - \frac{x}{k_n}\right)^n$, where n denotes the sample size. Setting $k_n = n$, we write $P\left(nU_{(1)} > x\right) = \left(1 - \frac{x}{n}\right)^n$ and, if $x > 0$, we have that as n increases onae can show (lemma 21.12, page 312, van der Vaart) that $P\left(nU_{(1)} > x\right) \longrightarrow e^{-x}$.

A.3 The case of a truncated Normal distribution

Suppose that one fits a truncated Normal distribution to a sample of test scores. Then we can write

$$F(x) = \frac{\Phi(x) - \Phi(\theta_L)}{\Phi(\theta_U) - \Phi(\theta_L)}$$

Hence

$$\begin{aligned} nF\left(\frac{x}{n} + \theta_L\right) &= \frac{n\left\{\Phi\left(\frac{x}{n} + \theta_L\right) - \Phi(\theta_L)\right\}}{\Phi(\theta_U) - \Phi(\theta_L)} \\ &\simeq \frac{\phi(\theta_L)\, x}{\Phi(\theta_U) - \Phi(\theta_L)} \end{aligned}$$

We can then write the density of the estimator as

$$P\left\{n\left(X_{(1)} - \theta_L\right) > x\right\} \simeq \exp\left(-\left\{\frac{\phi(\theta_L)\, x}{\Phi(\theta_U) - \Phi(\theta_L)}\right\}\right)$$

A.4 Constructing confidence intervals

Given the distribution derived above, we are now ready to construct confidence intervals for our estimator. In general, we know that

$$n\left(X_{(1)} - \theta_L\right) \sim Z$$

where $Z = 1 - \exp\left(-\left\{\frac{\phi(\theta_L)x}{\Phi(\theta_U) - \Phi(\theta_L)}\right\}\right)$ is the exponential function we have derived before. We want to derive upper and lower bounds for a confidence interval, denoted as $\hat{C}_L$ and $\hat{C}_U$.

Say that we want to build a 95% confidence interval, that is

$$P\left(\hat{C}_L \leq \theta_L \leq \hat{C}_U\right) \simeq 0.95$$

This can be rewritten as

$$\begin{aligned} P\left(\hat{C}_L - \hat{\theta}_L \leq \theta_L - \hat{\theta}_L \leq \hat{C}_U - \hat{\theta}_L\right) &\simeq 0.95 \\ P\left(\hat{\theta}_L - \hat{C}_U \leq \hat{\theta}_L - \theta_L \leq \hat{\theta}_L - \hat{C}_L\right) &\simeq 0.95 \end{aligned}$$

This means that

$$P\left\{n\left(\hat{\theta}_L - \hat{C}_U\right) \leq n\left(\hat{\theta}_L - \theta_L\right) \leq n\left(\hat{\theta}_L - \hat{C}_L\right)\right\} \simeq 0.95$$

Now, denote $C_{\alpha,L} = n\left(\hat{\theta}_L - \hat{C}_U\right)$ and $C_{\alpha,U} = n\left(\hat{\theta}_L - \hat{C}_L\right)$. Note that by construction $n\left(\hat{\theta}_L - \theta_L\right) \in [0, \infty)$. Recall that $n\left(\hat{\theta}_L - \theta_L\right) = z \sim Z$. Therefore,

$$P\left\{n\left(\hat{\theta}_L - \hat{C}_U\right) \leq z \leq n\left(\hat{\theta}_L - \hat{C}_L\right)\right\} \simeq 0.95$$

or

$$P\left\{C_{\alpha,L} \leq z \leq C_{\alpha,U}\right\} \simeq 0.95$$

The last step is to express $\hat{C}_L$ and $\hat{C}_U$ as a function of $C_{\alpha,L}$ and $C_{\alpha,U}$, that is

$$\begin{aligned} \hat{C}_L &= \hat{\theta}_L - \frac{C_{\alpha,U}}{n} \\ \hat{C}_U &= \hat{\theta}_L - \frac{C_{\alpha,L}}{n} \end{aligned}$$

In what follows we derive a 95% confidence interval for the estimate of the lowest value of a truncated Normal:

1. $P\left(z \leq C_{\alpha,L}\right) = 1 - \exp\left(-\left\{\beta C_{\alpha,L}\right\}\right)$ where $\beta = \frac{\phi(\theta_L)}{\Phi(\theta_U) - \Phi(\theta_L)}$. Solving for a 2.5% probability that the random variable lies to the left of $C_{\alpha,L}$ we get $1 - \exp\left(-\left\{\beta C_{\alpha,L}\right\}\right) = 0.025$, which

gives $C_{\alpha,L} = -\frac{\ln 0.975}{\beta} > 0$;

2. $P(z \geq C_{\alpha,U}) = \exp(-\{\beta C_{\alpha,U}\})$ and setting this probability to 2.5% we get $C_{\alpha,U} = -\frac{\ln 0.025}{\beta} > 0$;

3. Finally, we derive $\hat{C}_L$ and $\hat{C}_U$ as follows:

$$\begin{aligned} \hat{C}_L &= \hat{\theta}_L + \frac{\ln 0.025}{n\beta} \\ \hat{C}_U &= \hat{\theta}_L + \frac{\ln 0.975}{n\beta} \end{aligned}$$

Notice that $\hat{C}_L < \hat{C}_U < \hat{\theta}_L$, as the estimator does not lie in the confidence interval around the true parameter value θ_L. The intuition is the following: every sample from the true population is bound to have a lowest element that is, at most, just as low as the true population minimum. Most samples will have a lowest element which is larger than the true population minimum: this explains why a 95% confidence interval lies to the left of the estimated minimum. This simply means that almost certainly the true population minimum is smaller than the lowest element in a random sample.

In order to convey this intuition, we plot the distribution of the estimator $\hat{\theta}_L$ in Figure A1.

In order to compute the actual confidence interval, it is necessary to assign numerical values to the following parameters: $\beta = \frac{\phi(\theta_L)}{\Phi(\theta_U)-\Phi(\theta_L)}$ and n. The first can be computed by using estimated values for θ_L and θ_U, as well as for the standard deviation and mean of the Normal $\Phi(\cdot)$. For example, using point estimates for Denmark we get $\sigma = 0.15$, $\theta_L = 0.27$, $\theta_U = 1.73$.[21] The sample size is $n = 3000$.

Then we obtain $\beta = 0.0000191341$ and $\hat{C}_L = 0.272 + \frac{\ln 0.025}{3000 \times 0.0000191341} = -63.99$ and $\hat{C}_U =$

[21] The mean μ is normalized to 1 for all countries.

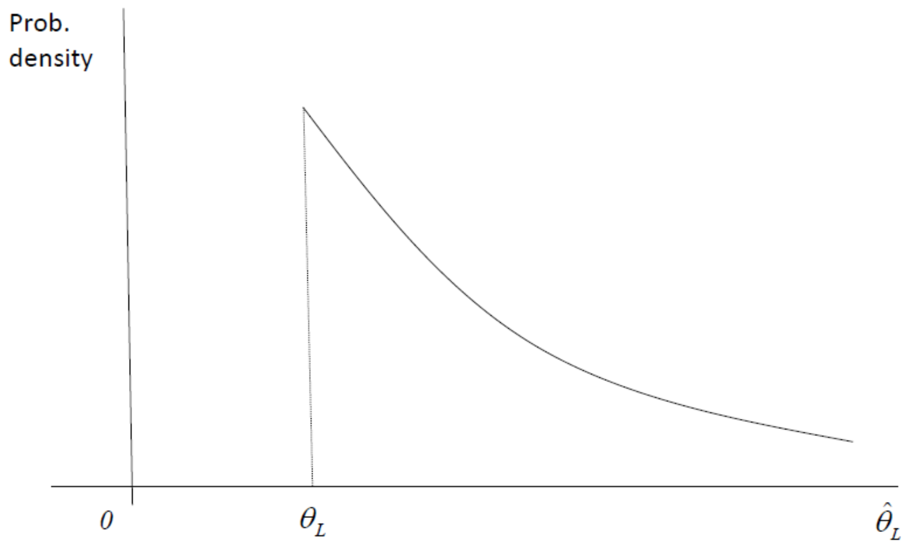

Figure A.1: Asymptotic distribution of the lower truncation point estimator $\widehat{\theta}_L$.

$0.272 + \frac{\ln 0.975}{3000 \times 0.0000191341} = -0.17$. Since by assumption the true population minimum is $0 < \theta_L < \theta_H < \infty$, we conclude that with more than a 95% probability the true minimum is between zero and the point estimate $\hat{\theta}_L = 0.27$. Table A1 reports the asymptotic confidence intervals for lower and upper truncation points for all countries in the sample.

A.5 Beta densities

We also replicate the analysis using symmetric Beta distributions. A symmetric Beta is characterized by a shape parameter (α_i for each country i) as well as upper and lower bounds of the support, $t_{\min,i}$ and $t_{\max,i}$, respectively. The density function is given by:

$$\phi(t) = \frac{\left[(t - t_{\min})(t_{\max} - t)\right]^{\alpha-1}}{Beta(\alpha,\alpha)(t_{\max} - t_{\min})^{2\alpha-1}}.$$

where $Beta(\cdot,\cdot)$ is the Beta function. We estimate these parameters for each country by MLE and replicate the analysis performed under the assumption of normality. The estimates and fitted

Table A-1: Truncation points: Confidence Intervals

Country	Lower Truncation Point	95% Confidence Interval		Upper Truncation Point	95% Confidence Interval	
DNK	0.27	-63.99	-0.17	1.73	2.17	65.99
DEU	0.31	-5.08	0.27	1.69	1.73	7.08
NOR	0.17	-343.74	-2.19	1.83	4.19	345.74
NLD	0.15	-662.76	-4.40	1.85	6.40	664.76
SWE	0.06	-2282.69	-15.61	1.94	17.61	2284.69
CZE	0.21	-25.41	0.03	1.79	1.97	27.41
FIN	0.16	-104.80	-0.56	1.84	2.56	106.80
HUN	0.29	-0.05	0.29	1.71	1.71	2.05
BEL	0.17	-3.22	0.15	1.83	1.85	2.83
CHE	0.20	-0.72	0.19	1.80	1.81	2.72
NZL	0.16	-0.83	0.15	1.84	1.85	2.83
IRL	0.14	-0.63	0.13	1.86	1.87	2.63
CAN	0.08	-2.03	0.07	1.92	1.93	4.03
UK	0.01	-3.65	-0.02	1.99	2.02	5.65
USA	0.05	-1.00	0.04	1.95	1.96	3.00
ITA	0.02	-0.95	0.01	1.98	1.99	2.95
CHL	0.16	0.06	0.16	1.84	1.84	1.95
SVN	0.08	-0.11	0.08	1.92	1.92	2.11
POL	0.08	-0.06	0.08	1.92	1.92	2.06

Notes: Table reports the asymptotic 95 percent confidence intervals for the ML estimates of the lower and upper truncation points of a Normal distribution, as reported in Table 1.

Table A-2: Estimated parameters - Beta distribution with variable support

Country	Alpha (1)	Lower bound (2)	Upper bound (3)
DNK	22.5	0.00	2.00
DEU	19.8	0.00	2.00
NOR	18.3	0.00	1.99
NLD	18.0	0.00	1.99
SWE	15.8	0.00	1.99
CZE	15.7	0.00	1.99
FIN	15.5	0.00	1.99
HUN	11.4	0.00	1.99
BEL	11.0	0.00	1.98
NZL	9.7	0.00	1.99
CHE	9.7	0.00	1.98
IRL	8.6	0.00	1.98
CAN	8.2	0.00	1.97
USA	7.3	0.00	1.97
UK	6.8	0.00	1.96
ITA	6.6	0.00	1.97
CHL	6.5	0.00	1.98
SVN	5.7	0.00	1.97
POL	5.1	0.00	1.96

Notes: This table reports ML estimation results for Beta distributions. Columns 1, 2 and 3 report alpha, lower bound and upper bound for a Beta distribution with mean 1. Countries are ordered by increasing skill variance

distributions are displayed in Table A-2 and Figure 4, respectively.

The results are very similar to the case in which we fit truncated Normals: the share of violations of the SDC hypothesis is 1.05% when comparing every country with all other countries.

Just like it was the case for the Normal approximations, when we restrict the range of the distribution to be constant across countries (leaving the share parameters different) the violation rate goes to zero.