Estimating Cross-Industry Cross-Country Interaction Models Using Benchmark Industry Characteristics

Antonio Ciccone
University of Mannheim, Barcelona GSE, and CEPR

Elias Papaioannou*
London Business School, CEPR, and NBER

April 6, 2017

Abstract

Empirical cross-industry cross-country models are applied widely in economics, for example to investigate the determinants of economic growth or international trade. Estimation generally relies on US proxies for unobservable technological industry characteristics, for example industries' dependence on external finance or relationship-specific inputs. We examine the properties of the estimator and find that estimates can be biased towards zero (attenuated) or away from zero (amplified), depending on how technological similarity with the US covaries with other country characteristics. While the possibility of attenuation bias due to the use of US technology proxies is often acknowledged, the cross-industry cross-country literature appears unaware that the same technology proxies may actually result in the contrary, an amplification bias. We also develop an alternative estimator that yields a lower bound on the true effect in cross-industry cross-country models of comparative advantage under two assumptions (one widely used and one straightforward to check). We illustrate the newly-developed estimator by reassessing the impact of contract enforcement on comparative advantage in sectors that depend more on relationship-specific intermediate inputs.

Keywords: International Specialization and Trade, Industry Growth, Measurement Error, Financial Development, Institutions, Regulation, Factor Endowments

JEL Classification Numbers: G30, F10, O40

*Emails: antonio.ciccone@gmail.com and eliaspapaioannou@london.edu. We would like to thank Sebastian Hohmann for comments. Any errors are ours.

Abstract

Empirical cross-industry cross-country models are applied widely in economics, for example to investigate the determinants of economic growth or international trade. Estimation generally relies on US proxies for unobservable technological industry characteristics, for example industries' dependence on external finance or relationship-specific inputs. We examine the properties of the estimator and find that estimates can be biased towards zero (attenuated) or away from zero (amplified), depending on how technological similarity with the US covaries with other country characteristics. While the possibility of attenuation bias due to the use of US technology proxies is often acknowledged, the cross-industry cross-country literature appears unaware that the same technology proxies may actually result in the contrary, an amplification bias. We also develop an alternative estimator that yields a lower bound on the true effect in cross-industry cross-country models of comparative advantage under two assumptions (one widely used and one straightforward to check). We illustrate the newly-developed estimator by reassessing the impact of contract enforcement on comparative advantage in sectors that depend more on relationship-specific intermediate inputs.

Keywords: International Specialization and Trade, Industry Growth, Measurement Error, Financial Development, Institutions, Regulation, Factor Endowments

JEL Classification Numbers: G30, F10, O40

1 Introduction

Recent empirical work in macroeconomics and international trade has relied extensively on cross-industry cross-country models that relate cross-country differences in industry performance – industry growth or industry exports for example – to an interaction between (i) country characteristics like financial development, institutional quality, or human capital endowments and (ii) industry characteristics like external-finance dependence, the complexity of production, or skill intensity. The approach has proven useful for examining a surprisingly wide variety of interesting economic questions, briefly reviewed below. Two strands of research stand out. First, following Rajan and Zingales (1998), cross-industry cross-country models have been used to examine how economic growth and development is affected by financial development, property rights protection, contract enforcement, and human capital endowments. Second, building on Romalis (2004) and subsequent theoretical contributions in international trade, cross-industry cross-country models have served as the basis for empirical studies of the effect of factor endowments and institutions on comparative advantage (for a review, see Nunn and Trefler, 2014). For example, Nunn (2007) uses the approach to show that better contract enforcement is a source of comparative advantage in industries that use relationship-specific inputs more intensively.

Because there is little industry data for most countries, the cross-industry cross-country literature generally treats the relevant technological industry characteristics – for example, external-finance dependence in Rajan and Zingales (1998) or relationship-specific input intensity in Nunn (2007) – as unobservable and employs proxies from a benchmark country, typically the United States (US). Another reason for using US industry data to obtain proxies for the relevant industry characteristics is that technological industry characteristics must be inferred from industry behavior, which is likely to yield more reliable results in countries where markets are not too distorted. Our goal here is to understand the widely used cross-industry cross-country estimator and formally analyze the implications of using data from a benchmark country to proxy unobservable technological industry characteristics.

Our starting point is an empirical framework that encompasses the cross-industry cross-country models in the literature. A basic feature of the framework is that the technological characteristics of industries may be more similar for some pairs of countries than others (e.g., Bernard and Jones, 1996; Acemoglu and Zilibotti, 2001; Schott, 2004; Caselli, 2005). We then show that the benchmarking estimator used in the cross-industry cross-country literature is subject to a bias shaped by two countervailing forces. Unsurprisingly, proxying the technological industry characteristics of countries using data from a benchmark country may result in a bias toward zero (an attenuation bias). The reasoning is similar to that of the classical measurement error bias. But benchmarking may also result in a bias away from zero, which we refer to as amplification bias. The amplification bias can be very strong if technologically similar countries are similar in other dimensions.[1] The literature using cross-industry cross-country models appears to be unaware of the possibility that using technology proxies from a benchmark country may result in estimates being amplified rather than attenuated, as we were unable to find any mention of this issue (by contrast, it is widely acknowledged in the literature that such technology proxies may result in an attenuation bias). In fact, as far as we know, the cross-industry cross-country model appears to be the only standard model in applied economics where the bias is either an attenuation bias – where estimates are biased towards zero whether the true effect is positive or negative– or an amplification bias –where estimates are biased away from zero whether the true effect is positive or negative.[2]

A main area of application of cross-industry cross-country models is international trade, where these models have been used to examine the effect of factor endowments and institutions on comparative advantage (e.g. Romalis, 2004; Levchenko, 2007; Nunn, 2007; Cuñat and Melitz, 2012; Krishna and Levchenko, 2013; Manova, 2013). We show that in this context there is a benchmarking estimator that is biased towards zero and therefore yields a lower bound on the true effect, as long as some countries differ in the direction of their comparative advantage and industry fetaures in the benchmark country proxy for unobserved global sectoral characteristics. We illustrate this estimator by applying it to Nunn's (2007) study of the effect of contract enforcement on comparative advantage in industries that depend more on relationship-specific intermediate inputs.

The rest of the paper is structured as follows. Next we briefly review some of the applications of the cross-industry cross-country approach. Section 2 examines the estimator used in the cross-industry cross-country literature. Section 3 develops the alternative estimator that yields a lower bound on the true effect in models of comparative advantage and illustrates the estimator in the context of Nunn (2007). Section 4 concludes.

[1] It is tempting to think of the amplification bias as an omitted variable bias, but there are differences that make this analogy less useful. For example, the two forces determining the bias of the benchmarking estimator result in either amplification or attenuation. In contrast, the simple omitted variable bias is either upwards or downwards. Nevertheless, the bias of the benchmarking estimator can – just like the classical measurement error bias – be modelled as a nonstandard omitted variable bias (although, just like in the case of classical measurement error bias, this does not help much in understanding it).

[2] Note that an amplification bias as defined here is not the same as so-called "bias amplification". In the economics literature, bias amplification is used to describe instances where the supposed solution for a bias actually increases the bias.

Some Applications of the Cross-Industry Cross-Country Approach The cross-industry cross-country approach is widely used in economics and our brief review here is only meant to illustrate the range of empirical applications. See Appendix Table 1 for a summary of the variety of applications.

Many applications of the cross-industry cross-country approach investigate the effects of financial markets on economic growth, firm entry and exit, investment, and innovation (e.g., Rajan and Zingales, 1998; Fisman and Love, 2003, 2007; Braun and Larrain, 2005; Aghion, Fally, and Scarpetta, 2007; Beck, Demirgüc-Kunt, Laeven, and Levine, 2008; Brown, Martinson, and Petersen, 2013; Hsu, Tian, and Xu, 2014; Calomiris, Larrain, Liberti, and Sturgess, 2016).

The cross-industry cross-country approach has been widely used to examine the determinants of international trade and industrial specialization. Nunn (2007), Levchenko (2007), and subsequent works show that institutionally advanced countries tend to specialize in sectors that rely on differentiated intermediate inputs (see also Ranjan and Lee, 2007; Ferguson and Formai, 2013; Nunn and Trefler, 2014). Manova (2008, 2013) links financial development to the patterns of international trade (see also Chan and Manova, 2015; Manova, Wei, and Zhang, 2015). Building on Romalis (2004), Ciccone and Papaioannou (2009) show that countries with an educated workforce tend to specialize in human capital intensive sectors. The cross-industry cross-country approach has also been used to investigate the effect of product and labor market institutions on comparative advantage, productivity, entrepreneurship, and innovation (e.g., Ciccone and Papaioannou, 2007; Cingano, Leonardi, Messina, and Pica, 2010; Cuñat and Melitz, 2012; Tang, 2012; Griffith and Macartney, 2014). And recent works have employed the cross-industry cross-country approach to study the effects of environmental protection laws and water supply on comparative advantage (Broner, Bustos, and Carvalho, 2015; Debaere, 2015).

Other applications of the cross-industry cross-country approach investigate a variety of different economic issues. For example, the driving forces of outsourcing, foreign direct investment, and the fragmentation of production (e.g., Alfaro and Charlton, 2009; Carluccio and Fally, 2012; Basco, 2013; Blyde and Danielken, 2015; Paunov, 2016). The cross-industry cross-country approach has also been used to examine the economic consequences of cross-country differences in firm size distributions, entry regulation, transaction costs, risk sharing possibilities, skill dispersion, and foreign aid inflows (e.g. Pagano and Schivardi, 2003; Klapper, Laeven, and Rajan, 2006; Acemoglu, Johnson, and Mitton, 2009; Rajan and Subramanian, 2010; Aizenman and Sushko, 2011; Bombardini, Gallipoli, and Pupato, 2012; Michelacci and Schivardi, 2013; Larrain, 2014; Aghion, Howitt, and Prantl, 2014). Recent applications use the cross-industry cross-country setup to assess the effects of financial crises on macroeconomic performance and international trade (e.g. Dell'Ariccia, Detragiache, and Rajan, 2008; Iacovone and Zavacka, 2009; Duchin, Ozbas, and Sensoy, 2010; Claessens, Tong, and Wei, 2012; Laeven and Valencia, 2013) and to examine the effects of fiscal and monetary policy over the business cycle (e.g. Aghion, Farhi, and Kharroubi, 2013; Aghion, Hemous, and Kharroubi, 2014).

Variations of the cross-industry cross-country approach have been employed to examine the economic effects of differences in financial development, institutional quality and trust across regions and over time (e.g. Cetorelli and Strahan, 2006; Bertrand, Schoar, and Tesmar, 2007; Hsieh and Parker, 2007; Aghion, Askenazy, Berman, Cette, and Eymar, 2012; Fafchamps and Schündeln, 2013; Feenstra, Hong, Ma, and Spencer, 2013; Duygan-Bump, Levkov, and Montoriol-Garriga, 2015; Jacobson and von Schedvin, 2015; Cingano and Pinotti, 2016).

2 The Benchmarking Bias

2.1 Empirical Framework

The basis of cross-industry cross-country models are theories linking outcomes for industries in different countries to an interaction between country characteristics and technological industry characteristics. For example, in Rajan and Zingales (1998), the outcome variable is industry growth and the interaction is between country-level financial development and the external-finance dependence of industries. In Nunn (2007), the outcome variable is industry exports and the interaction is between country-level contract enforcement and the intensity with which industries use relationship-specific inputs. As the main theoretical prediction concerns the effect of the interaction between country and industry characteristics, cross-industry cross-country models allow controlling for country and industry fixed effects. An empirical framework that encompasses the models used in the cross-industry cross-country literature is

$$y_{in} = \alpha_n + \alpha_i + \beta x_n z_{in} + v_{in} \tag{1}$$

where y_{in} is the outcome in industry $i = 1, .., I$ and country $n = 1, .., N$; x_n the relevant country characteristic; and z_{in} the relevant industry characteristic. The α_n and α_i denote country and industry fixed effects and v_{in} unobservable determinants of the outcome. The parameter of interest is the coefficient on the industry-country interaction, β. We take v_{in} to be distributed independently of z_{in} and x_n to abstract from omitted variable and reverse causation issues. We also assume that v_{in} has a finite variance and $E(v_{in}|n) = E(v_{in}|i) = 0$, and take x_n to be given with $\sum_{n=1}^{N}(x_n - \overline{x})^2 > 0$ where $\overline{x}$ is the average of x_n.

Estimation of β in (1) would be straightforward if there were data on the technological industry characteristics z_{in} for a broad set of countries. But detailed industry data are unavailable for most countries. Moreover, the cross-industry cross-country literature often focuses on technological industry characteristics that are not directly observable and must therefore be inferred from industry behavior. Such inference is likely to be more reliable in countries where markets are not too distorted; where markets are distorted, actual industry characteristics may only be very loosely linked to the underlying technological characteristics. In practice, z_{in} is generally proxied using industry data from a single, relatively undistorted benchmark country, almost always the US. In a few cases, z_{in} is proxied using average industry data from a group of relatively undistorted countries.

It is therefore important to understand whether β in (1) can be estimated using industry characteristics from a benchmark country as a proxy for z_{in}. For such a benchmarking estimator to stand a chance, there must be some global element to an industry's technological characteristics. At the same time, it seems unreasonable to presume that industries use the same technology in all countries, as the optimal technology choice depends on many factors that vary across countries (e.g., Bernard and Jones, 1996; Acemoglu and Zilibotti, 2001; Schott, 2004; Caselli, 2005). We therefore model the industry characteristics z_{in} in (1) as the sum of a global industry characteristic (z_i^*) and a country-specific industry characteristic (ε_{in})

$$z_{in} = z_i^* + \varepsilon_{in} \tag{2}$$

where z_i^* is i.i.d. with variance $Var(z^*)$ and independent of other elements of the model.[3] The country-specific industry characteristics ε_{in} allow us to capture that industry characteristics may be more similar for some country-pairs than others in a simple way. We assume $E(\varepsilon_{in}|n) = E(\varepsilon_{in}|i) = 0$; $E(\varepsilon_{in}^2|n) = \sigma^2$; $E(\varepsilon_{in}\varepsilon_{jm}|n, m) = 0$ for all industries $i \neq j$; and the following correlation of idiosyncratic industry characteristics for country pairs $n \neq m$

$$Corr(\varepsilon_{in}\varepsilon_{im}|n, m) = \rho_{mn} \ . \tag{3}$$

Hence, the correlation of industry characteristics z_{in} for country pairs $n \neq m$ is

$$Corr(z_{in}, z_{im}|n, m) = \frac{Var(z^*)}{Var(z^*) + \sigma^2} + \frac{\sigma^2}{Var(z^*) + \sigma^2}\rho_{mn}. \tag{4}$$

[3] While it is reasonable to think of industry characteristics as also reflecting a country-specific component, we can omit such components in (2) without any loss of generality as they can be absorbed into the country fixed effects in (1).

$Corr(z_{in}, z_{im}|n, m)$ can be interpreted as an index of technological similarity and country pairs with greater ρ_{mn} are therefore more similar technologically.

2.2 The Bias

As data on the technological industry characteristics z_{in} are unavailable for a broad set of countries, the cross-industry cross-country literature generally proceeds using a proxy from a relatively undistorted benchmark country. We refer to this proxy as z_{iUS} as the benchmark country is almost always the US.[4] Hence, the equation estimated in the cross-industry cross-country literature is

$$y_{in} = a_n + a_i + bx_n z_{iUS} + residual_{in} \tag{5}$$

where a_n and a_i stand for country and industry fixed effects. The main coefficient of interest in the literature is b and the method of estimation is least squares. (Our analysis abstracts from all estimation issues other than those leading to an amplification bias. For example, some applications where the exogeneity of x_n is an issue – in our analysis no such issue arises as x_n is assumed to be exogenous – try instrumenting x_n. Amplification bias, however, carries through to the case where an instrument for x_n is used; in this case our analysis applies to the reduced-form equation where the instrument replaces x_n.

To understand the relationship between the least-squares estimator of b in (5) and β in (1), which is the parameter of interest, it is useful to rewrite the least-squares estimator in terms of demeaned data (e.g., Baltagi, 2008)

$$\widehat{b} = \frac{\frac{1}{N}\frac{1}{I}\sum_n \sum_i (z_{iUS} - \overline{z}_{US})(x_n - \overline{x})\,(y_{in} - \overline{y}_n - \overline{y}_i + \overline{y})}{\frac{1}{N}\frac{1}{I}\sum_i \sum_n (z_{iUS} - \overline{z}_{US})^2 (x_n - \overline{x})^2} \tag{6}$$

where $\overline{y}$ is the average of y_{in} across industries and countries; $\overline{y}_i$ the average of y_{in} for industry i; $\overline{y}_n$ the average of y_{in} for country n; $\overline{z}_{US}$ the average of z_{iUS}; and $\overline{x}$ the average of x_n. The

[4] The analysis and conclusions are very similar if z_{in} is proxied using average industry characteristics from a group of countries, as long as countries in the group are not more distorted than other countries in the sample.

probability limit of $\widehat{b}$ is[5]

$$\widehat{b}^a = \operatorname*{plim}_{I\to\infty} \widehat{b} = \beta\left[1+\lambda\right]. \tag{7}$$

with

$$\lambda = \frac{\sum_n (x_n - \overline{x})\left[Corr(z_{US}, z_n)x_n\right]}{\sum_n (x_n - \overline{x})^2} - 1 \tag{8}$$

where $Corr(z_{US}, z_n) \equiv Corr(z_{iUS}, z_{in}|n)$.

2.2.1 The Case of Attenuation Bias

It follows from (7) that the benchmarking estimator $\widehat{b}$ used in cross-industry cross-country empirics will be attenuated (biased towards zero) if and only if $0 < 1+\lambda \leq 1$; equivalently using (8)

$$0 < \frac{\sum_n (x_n - \overline{x})\left[Corr(z_{US}, z_n)x_n\right]}{\sum_n (x_n - \overline{x})^2} \leq 1. \tag{9}$$

For example, this will be the case if the index of technological similarity with the US is the same for all countries and technological industry characteristics in the US therefore proxy equally well for technological industry characteristics in all other countries, $Corr(z_{US}, z_n) = \pi > 0$. In this case, $\widehat{b}^a = \pi\beta$ where π plays the role of the reliability ratio in the classical measurement error model (e.g. Wooldridge, 2002).

A somewhat more general sufficient condition for $\widehat{b}$ to be biased towards zero is that the index of technological similarity with the US, $Corr(z_{US}, z_n)$, is decreasing in the country characteristic x_n, but that $Corr(z_{US}, z_n)x_n$ is increasing in x_n (if the latter condition is not satisfied, the benchmarking estimator may have the wrong sign).

2.2.2 The Case of Amplification Bias

But the benchmarking estimator $\widehat{b}$ can yield estimates of β that are biased away from zero (amplified). From (7) and (8) it follows that this will be the case if and only if $\lambda > 0$ or

[5]Substituting (1) into (6) and taking the probability limit as $I \to \infty$ of the numerator yields $\operatorname*{plim}_{I\to\infty} \beta\left(\frac{1}{N}\sum_n\left((x_n-\overline{x})x_n\frac{1}{I}\sum_i(\varepsilon_{in}-\varepsilon_{iUS})\varepsilon_{iUS}\right)\right) + \operatorname*{plim}_{I\to\infty}\beta\left(\frac{1}{N}\frac{1}{I}\sum_i\sum_n(z_{iUS}-\overline{z}_{US})^2(x_n-\overline{x})^2\right)$. Using (3) this simplifies to $\beta\left(\frac{1}{N}\sum_n(x_n-\overline{x})x_n(\rho_{USn}-1)\sigma^2\right) + \beta Var(z_{US})\left(\frac{1}{N}\sum_n(x_n-\overline{x})^2\right)$. The probability limit of the denominator when substituting (1) into (6) is $Var(z_{US})\left(\frac{1}{N}\sum_n(x_n-\overline{x})^2\right)$. Hence, the probability limit of (6) is $\beta + \beta\left(\frac{1}{N}\sum_n(x_n-\overline{x})x_n(\rho_{USn}-1)\sigma^2\right) / \left(Var(z_{US})\left(\frac{1}{N}\sum_n(x_n-\overline{x})^2\right)\right)$ which defining $Corr(z_{US}, z_n) \equiv Corr(z_{iUS}, z_{in}|n)$ and making use of (4) and $Var(z_{US}) = Var(z^*) + \sigma^2$ can be written as in (7) and (8).

equivalently

$$\frac{\sum_n (x_n - \overline{x}) \left[Corr(z_{US}, z_n) x_n\right]}{\sum_n (x_n - \overline{x})^2} > 1. \tag{10}$$

The left-hand side of the inequality in (10) turns out to be the standard formula for the least-squares slope of a regression of $Corr(z_{US}, z_n)x_n$ on x_n. Hence, the condition for an amplification bias in (10) is equivalent to a least-squares slope greater unity when regressing $Corr(z_{US}, z_n)x_n$ on x_n. For this to be the case, the index of technological similarity of country n with the US, $Corr(z_{US}, z_n)$, must be strictly increasing in the country characteristic x_n over some range.

To develop some intuition for the amplification bias, it is useful to rewrite the model in (1) in terms of two equations

$$y_{in} = \alpha_n + \alpha_i + \gamma_n z_{in} + v_{in} \tag{11}$$

where

$$\gamma_n = \beta x_n. \tag{12}$$

The country-specific slope parameters γ_n capture cross-country differences in how industry outcomes covary with industry characteristics. For example, in Rajan and Zingales (1998) these slope parameters would capture cross-country differences in the covariation between industry growth and the external-finance dependence of industries. In Nunn (2007), the slope parameters would capture cross-country differences in the covariation between industry exports and the relationship-specific input intensity of industries.

Now imagine estimating the country-specific slopes γ_n in (11) with least squares using US industry characteristics z_{iUS} as a proxy of industry characteristics z_{in}. The resulting least-squares slopes $\widehat{\gamma}_n$ reflect the covariation between industry outcomes in country n and US industry characteristics z_{iUS}. Substituting the least-squares slopes $\widehat{\gamma}_n$ in (6) yields that the benchmarking estimator can be expressed as the least-squares slope of a regression of the country-specific slope estimates $\widehat{\gamma}_n$ on the country characteristics x_n

$$\widehat{b} = \frac{\sum_n \widehat{\gamma}_n (x_n - \overline{x})}{\sum_n (x_n - \overline{x})^2}. \tag{13}$$ [6]

[6] To see this, note that the least-squares estimates of the country-specific slopes expressed in terms of demeaned data (e.g., Baltagi, 2008) are $\widehat{\gamma}_n = \sum_i z_{iUS} \left(y_{in} - \overline{y}_n - \overline{y}_i + \overline{y}\right) \big/ \sum_i z_{iUS}^2$.

Similarly, the probability limit of the benchmarking estimator $\widehat{b}^a$ can be written as the least-squares slope when regressing $\widehat{\gamma}_n^a$ on x_n

$$\widehat{b}^a = \frac{\sum_n \widehat{\gamma}_n^a (x_n - \overline{x})}{\sum_n (x_n - \overline{x})^2} \tag{14}$$

where

$$\widehat{\gamma}_n^a = \gamma_n Corr(z_{US}, z_n). \tag{15}$$

Equation (14) shows that the bias of the benchmarking estimator will reflect how the bias of the country-specific least-squares slopes covaries with the country characteristic x_n. As a result, the amplification bias can arise even if all country-specific slope estimates are attenuated because of classical measurement error, as long as the attenuation bias is weaker for countries with greater x_n.

A setting where countries fall into two groups The amplification bias emerges most clearly in a setting where countries except the US fall into two groups, A and B, and countries in the same group are identical. In this two-group setting, (13) simplifies to

$$\widehat{b} = \frac{\widehat{\gamma}_A - \widehat{\gamma}_B}{x_A - x_B}. \tag{16}$$

That is, the benchmarking estimator is simply the slope of the line connecting the two points $(x_A, \widehat{\gamma}_A)$ and $(x_B, \widehat{\gamma}_B)$. Making use of (15), the probability limit of (16) is

$$\widehat{b}^a = \frac{\widehat{\gamma}_A^a - \widehat{\gamma}_B^a}{x_A - x_B} = \beta \left(\frac{Corr(z_{US}, z_A) x_A - Corr(z_{US}, z_B) x_B}{x_A - x_B} \right) \tag{17}$$

where $Corr(z_{US}, z_A) \equiv Corr(z_{iUS}, z_{in}|n)$ for all countries n in group A and $Corr(z_{US}, z_B)$ is defined analogously. There is an amplification bias if and only if the term in parenthesis is greater than unity. The simplest way to see that the amplification bias can be very large is to consider the case where where (i) countries in group A have the same technological characteristics as the US and US industry characteristics therefore proxy perfectly for industry characteristics of these countries, $Corr(z_{US}, z_A) = 1$, but (ii) countries in group B have technological characteristics that differ from the US to the point where US industry characteristics are uncorrelated with industry characteristics of these countries, $Corr(z_{US}, z_B) = 0$. In this case, (17) simplifies to

$$\widehat{b}^a = \beta \left(\frac{x_A}{x_A - x_B} \right). \tag{18}$$

Hence, there will be an amplification bias if $x_A > x_B > 0$ and the bias will be very large if the two groups of countries have very similar characteristics x. This is because in this case there is a strong positive association between the country characteristic x_n and technological similarity with the US.

Figure 1 illustrates the true model and the estimated model in the two-group setting for $\beta > 0$. In figure $1A$, we graph the true country-specific slopes γ_A and γ_B against x_A and x_B. As $\gamma_n = \beta x_n$, the true parameter of interest β is simply the slope of the line connecting the two points (x_A, γ_A) and (x_B, γ_B). In figure $1B$, we also graph the probability limits of the country-specific slope estimates $\widehat{\gamma}_A^a$ and $\widehat{\gamma}_B^a$ against x_A and x_B. Equation (17) implies that the probability limit of the benchmarking estimator $\widehat{b}^a$ is simply the slope of the line connecting the two points $(x_A, \widehat{\gamma}_A^a)$ and $(x_B, \widehat{\gamma}_B^a)$. The amplification bias $\widehat{b}^a > \beta > 0$ follows because US industry characteristics are a perfect proxy for industry characteristics of countries in group A, which implies $\widehat{\gamma}_A^a = \gamma_A$, but do not proxy for industry characteristics of countries in group B, which implies $\widehat{\gamma}_B^a = 0 < \gamma_B$. More generally, the amplification bias of the benchmarking estimator arises when the attenuation bias of the country-specific slope estimates (which reflects technological dissimilarity with the US) is sufficiently stronger for countries that are less similar to the US in the country characteristic x.

Figure 1A: True Model

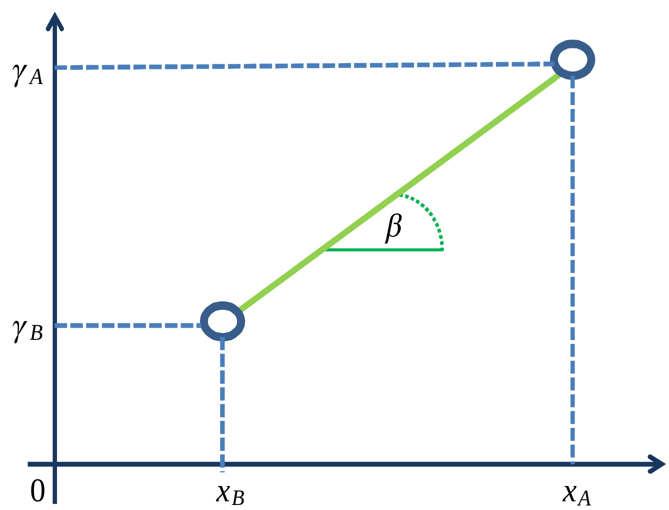
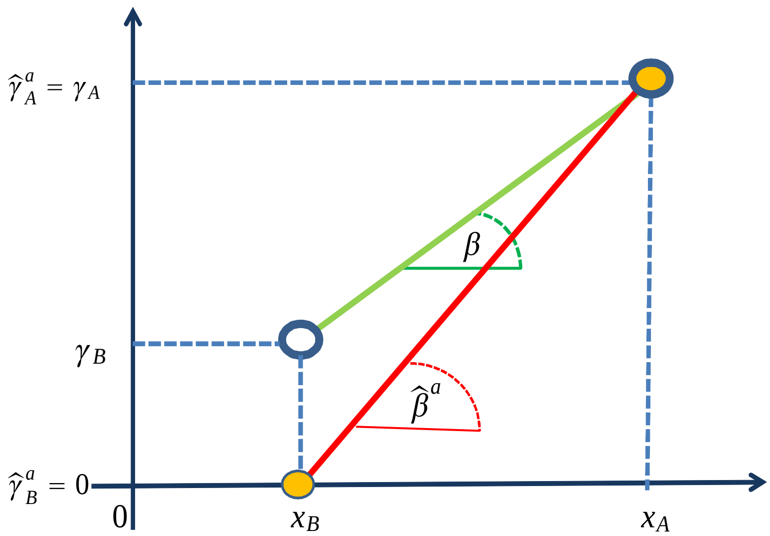


Figure 1B: Amplification Bias



Notes: True country-specific slopes (bold circles) and estimated country-specific slopes (filled circles) in the two-group model where the benchmarking estimator is biased away from zero (amplification bias). There is amplification bias although the country-specific slope estimates are weakly biased towards zero (attenuated).

3 Estimating Comparative Advantage Models with a Benchmarking Estimator

The (standard) benchmarking estimator of the empirical cross-industry cross-country literature has been used widely to investigate the determinants of comparative advantage in international trade (e.g. Romalis, 2004; Levchenko, 2007; Nunn, 2007; Manova, 2008, 2013; Cuñat and Melitz, 2012). In this context, there turns out to be a new benchmarking estimator that yields a lower bound on the strength of comparative advantage under the assumption that at least one pair of countries differs in the direction of comparative advantage. We first illustrate the argument in a model of comparative advantage where all countries except the US fall into two groups and countries in the same group are identical. A special feature of this setting is that the new benchmarking estimator turns out to be identical to the (standard) benchmarking estimator used in the literature. Then we discuss the new benchmarking estimator in a more general setting (where the new benchmarking estimator is no longer identical to the benchmarking estimator used in the cross-industry cross-country literature).

3.1 Model and Assumptions

It is useful to rewrite (without loss of generality) the country characteristic x_n in (11) and (12) as $x_n = q_n - q^*$. This yields

$$y_{in} = \alpha_n + \alpha_i + \gamma_n z_{in} + v_{in} \tag{19}$$

$$\gamma_n = \beta(q_n - q^*). \tag{20}$$

q_n is the country characteristic that may determine a country's comparative advantage and q^* the value of q_n where comparative advantage switches from high-z industries to low-z industries as long as $\beta \neq 0$. We can obtain a lower bound on the strength of comparative advantage β under two assumptions. The first assumption, which is standard in the comparative advantage literature using the cross-industry cross-country approach, is that high-z industries in the US also tend to be high-z industries elsewhere. The second assumption – which will turn out to be testable – is that there is at least one country on either side of the threshold q^*. Formally:

(A1) High-z industries in the US tend to be high-z industries elsewhere, $Corr(z_{US}, z_n) > 0$.

(A2) There is at least one country on either side of the threshold q^*, that is $(q_n - q^*)(q_m - q^*) < 0$ for at least one pair of countries n, m. Or equivalently, as long as $\beta \neq 0$, at least one country has a comparative advantage in high-z industries and at least one country has a comparative advantage in low-z industries.

For example, both assumptions are satisfied in the theoretical model of Romalis (2004), which is often seen as the conceptual foundation of empirical cross-industry cross-country models of international trade and industrial specialization.

A setting where countries fall into two groups To illustrate why these two assumptions allow for a benchmarking estimator that yields a lower bound on the true strength of comparative advantage, we return to the setting where countries except the US fall into two groups and countries in the same group are identical. A special feature of this setting is that the new benchmarking estimator turns out to be identical to the (standard) benchmarking estimator used in the literature. We can therefore illustrate the argument using the standard benchmarking estimator and postpone the introduction of the new benchmarking estimator.

As shown above, in the setting where countries except the US fall into two groups and countries in the same group are identical, the key formulas for the (standard) benchmarking estimator $\widehat{b}$ used in the cross-industry cross-country literature simplify to (16) and (17). Substituting $x_n = q_n - q^*$ yields

$$\widehat{b} = \frac{\widehat{\gamma}_A - \widehat{\gamma}_B}{q_A - q_B} \tag{21}$$

and

$$\widehat{b}^a = \frac{\widehat{\gamma}_A^a - \widehat{\gamma}_B^a}{q_A - q_B} = \beta\left(\frac{Corr(z_{US}, z_A)(q_A - q^*) - Corr(z_{US}, z_B)(q_B - q^*)}{q_A - q_B}\right). \tag{22}$$

The benchmarking estimator $\widehat{b}$ will be attenuated and therefore yield a lower bound on the true effect β, if and only if the term in parenthesis on the right-hand side of (22) is strictly greater than zero but smaller than unity. This is equivalent to[7]

$$\begin{aligned} & [Corr(z_{US}, z_B) + Corr(z_{US}, z_A)]\,(q_A - q^*)(q_B - q^*) \\ < \; & Corr(z_{US}, z_A)(q_A - q^*)^2 + Corr(z_{US}, z_B)(q_B - q^*)^2 \end{aligned} \tag{23}$$

and

$$\begin{aligned} & [2 - Corr(z_{US}, z_A) - Corr(z_{US}, z_B)](q_A - q^*)(q_B - q^*) \\ \leq \; & [1 - Corr(z_{US}, z_B)]\,(q_B - q^*)^2 + [1 - Corr(z_{US}, z_A)]\,(q_A - q^*)^2. \end{aligned} \tag{24}$$

Both conditions will be satisfied if assumptions (A1) and (A2) hold. To see this, notice that because countries in the same group are identical, assumption (A2) is equivalent to $(q_A - q^*)(q_B - q^*) < 0$. Combined with assumption (A1), this implies that the left-hand side of (23) is strictly negative while the right-hand side is positive. Assumptions (A1) and (A2) also imply that the left-hand side of (24) is negative while the right-hand side is positive. Hence, assumptions (A1) and (A2) imply that the term in parenthesis on the right-hand-side of (22) is strictly greater than zero but smaller than unity and that the benchmarking estimator $\widehat{b}$ will be biased towards zero. When US industry characteristics are an imperfect proxy for industry characteristics of countries in group A or group B, the inequality in (24) will be strict and the benchmarking estimator $\widehat{b}$ will be strictly biased towards zero.

Figure 2 illustrates the true model and the estimated model for $\beta > 0$. In figure 2A, we graph the true country-specific slopes γ_A and γ_B against $q_A - q^* > 0$ and $q_B - q^* < 0$. As $\gamma_n = \beta(q_n - q^*)$, the true parameter of interest β is simply the slope of the line connecting the two points. In figure 2B, we also graph the probability limits of the country-specific slope

[7] To derive the conditions in (23) and (24), it is convenient to write the term in parenthesis in (22) as θ_1/θ_2. As long as $q_A \neq q_B$, the condition $0 < \theta_1/\theta_2 \leq 1$ is equivalent to $\theta_1\theta_2 > 0$, which is the condition in (23), and $(\theta_1/\theta_2)(q_A - q_B)^2 \leq (q_A - q_B)^2$, which making use of $q_A - q_B = (q_A - q^*) - (q_B - q^*)$ is the condition in (24).

estimates $\widehat{\gamma}_A^a$ and $\widehat{\gamma}_B^a$ against $q_A - q^*$ and $q_B - q^*$. According to (22), the slope of the line connecting these two new points yields the probability limit of the benchmarking estimator $\widehat{b}^a$. As $\widehat{\gamma}_A^a > 0$ and $\widehat{\gamma}_B^a < 0$ are biased towards zero, it follows that the line connecting the country-specific slope estimates must be less steep than the line connecting the true country-specific slopes. Hence, $\widehat{b}^a$ is biased towards zero (attenuated). For $\beta < 0$ the argument is analogous.

To better understand this result, it is useful to compare figure $2B$ where the benchmarking estimator $\widehat{b}$ is biased towards zero, with figure $1B$ where $\widehat{b}$ is biased away from zero. In both figures, all country-specific slope estimates are biased towards zero. But in figure $1B$ this results in the benchmarking estimator $\widehat{b}$ being biased away from zero, while in figure $2B$ $\widehat{b}$ is biased towards zero. From the figures it becomes clear that this is because in figure $2B$, there are countries on both sides of q^* and these countries differ in the direction of their comparative advantage. As a result, the line connecting the country-specific slope estimates in figure $2B$ is a clockwise rotation of the line connecting the true country-specific slopes. Hence, $\widehat{b}^a$ is necessarily biased towards zero.

Figure 2A: True Model

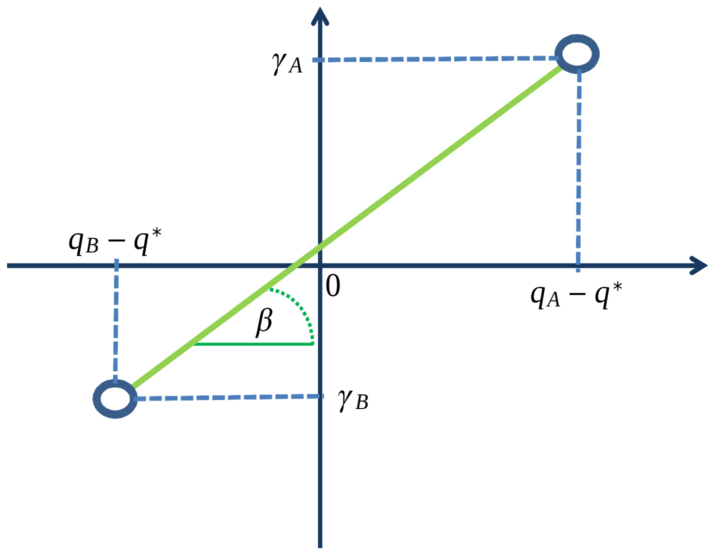
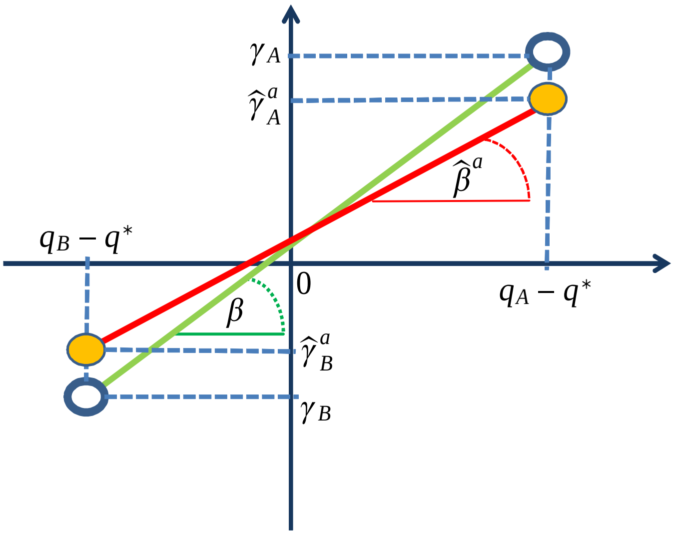


Figure 2B: Estimated Model



Notes: True country-specific slopes (bold circles) and estimated country-specific slopes (filled circles) in the two-group model of comparative advantage where the benchmarking estimator of the cross-industry cross-country approach is necessarily biased towards zero (attenuation bias).

.

3.2 A New Benchmarking Estimator

There continues to be a benchmarking estimator yielding a lower bound on the true strength of comparative advantage associated with country characteristic q_n in (19) and (20) when there are many different countries (but this estimator is no longer the benchmarking estimator used in the cross-industry cross-country literature). To show this, we start with the case where it is known which countries are on the same side of q^*. Or equivalently as long as $\beta \neq 0$, the case where it is known which countries have a comparative advantage going in the same direction. Then we turn to the case where the grouping of countries by the direction of their comparative advantage is unknown.

3.2.1 Known Country Grouping

If it were known which countries are on the same side of q^*, we could put countries on one side of q^* into group A and countries on the other side of q^* into group B. Then we could estimate the strength of comparative advantage associated with country characteristic q_n

using the following new benchmarking estimator

$$\widehat{b}_G = \frac{\overline{\widehat{\gamma}}_{nA} - \overline{\widehat{\gamma}}_{nB}}{\overline{q}_A - \overline{q}_B} \tag{25}$$

where $\overline{\widehat{\gamma}}_A$ and $\overline{\widehat{\gamma}}_B$ denote the average country-specific slope estimate for countries in group A and group B; $\overline{q}_A$ and $\overline{q}_B$ are the average country characteristic in groups A and B (it does not matter which group countries with q_n at the threshold q^* are assigned to; we denote the new benchmarking estimator with a subscript G because the estimator can be seen as a grouped-data estimator).[8] It is immediate that the new benchmarking estimator in (25) is identical to the standard benchmarking estimator in (21) when countries in the same group are identical. Hence, the argument in section 3.1 that the standard benchmarking estimator is biased towards zero under assumptions (A1) and (A2) when countries in the same group are identical, implies that the new benchmarking estimator is also biased towards zero in this special case. In the general case where the strength of comparative advantage differs among countries with the same direction of comparative advantage, (i) the new benchmarking estimator and the standard benchmarking estimator are no longer identical, and (ii) the standard benchmarking estimator may be biased upward or downward even if assumptions (A1) and (A2) hold. However, assumptions (A1) and (A2) imply that the new benchmarking estimator in (25) is biased towards zero this general case also, and the new benchmarking estimator therefore continues to yield a lower bound on the strength of the true effect. To see this, we first obtain the probability limit of (25) using (15), which yields

$$\widehat{b}^a_G = \frac{\overline{\widehat{\gamma}^a_{nA}} - \overline{\widehat{\gamma}^a_{mB}}}{\overline{q}_A - \overline{q}_B} = \beta \left(\frac{\overline{Corr(z_{US}, z_{nA})(q_{nA} - q^*)} - \overline{Corr(z_{US}, z_{mB})(q_{mB} - q^*)}}{\overline{q}_A - \overline{q}_B} \right) \tag{26}$$

where $\overline{Corr(z_{US}, z_{nA})(q_{nA} - q^*)}$ is the average of $Corr(z_{US}, z_{nA})(q_{nA} - q^*)$ across countries n in group A and $\overline{Corr(z_{US}, z_{mB})(q_{mB} - q^*)}$ is defined analogously for countries m in group B. (26) implies that the new benchmarking estimator will be attenuated and therefore yield a lower bound on the true effect, if and only if the term in parenthesis is strictly greater than zero but smaller than unity. This turns out to be equivalent to[9]

$$\begin{aligned} & \overline{Corr(z_{US}, z_{nA})(q_{nA} - q^*)\,(\overline{q}_B - q^*)} + \overline{Corr(z_{US}, z_{mB})(q_{mB} - q^*)\,(\overline{q}_A - q^*)} \\ < \; & \overline{Corr(z_{US}, z_{nA})(q_{nA} - q^*)\,(\overline{q}_A - q^*)} + \overline{Corr(z_{US}, z_{mB})(q_{mB} - q^*)\,(\overline{q}_B - q^*)} \end{aligned} \tag{27}$$

and

$$\begin{aligned} & \overline{[1 - Corr(z_{US}, z_{nA})](q_{nA} - q^*)\,(\overline{q}_B - q^*)} + \overline{[1 - Corr(z_{US}, z_{mB})](q_{mB} - q^*)\,(\overline{q}_A - q^*)} \\ \leq \; & \overline{[1 - Corr(z_{US}, z_{nA})]\,(q_{nA} - q^*)\,(\overline{q}_A - q^*)} + \overline{[1 - Corr(z_{US}, z_{mB})]\,(q_{mB} - q^*)\,(\overline{q}_B - q^*)}. \end{aligned} \tag{28}$$

[8] See Angrist (1991) for an application and a brief historical review of grouped-data estimation.

[9] The argument is analogous to that in footnote 6.

Both conditions will be satisfied if assumptions (A1) and (A2) hold. To see this, notice that the left-hand side of (27) is strictly negative as $Corr(z_{US}, z_{nA}) > 0$, $Corr(z_{US}, z_{mB}) > 0$, and at least one pair of countries differs in the direction of their comparative advantage, $(q_{nA} - q^*)(q_{mB} - q^*) < 0$; and the right-hand side of (27) is positive as countries in the same group have the same direction of comparative advantage, $(q_{nA} - q^*)(\overline{q}_A - q^*) \geq 0$ and $(q_{mB} - q^*)(\overline{q}_B - q^*) \geq 0$. A similar argument yields that the left-hand side of (28) is negative while the right-hand side is positive. Hence, the term in parenthesis in (26) is strictly greater than zero but smaller than unity and the grouping estimator $\widehat{b}_G$ is biased towards zero. When US industry characteristics are an imperfect proxy for industry characteristics in at least one country where $q_n \neq q^*$, the inequality in (28) will be strict and the new benchmarking estimator $\widehat{b}_G$ will be strictly biased towards zero.

Figure 3 illustrates the true model and the estimated model for $\beta > 0$. In figure $3A$, we graph the true country-specific slopes γ_n against $q_n - q^*$ with each circle representing a country. As $\gamma_n = \beta(q_n - q^*)$, the true parameter of interest β is the slope of the line through the circles. In figure $3B$, we also graph the probability limits of the country-specific slope estimates $\widehat{\gamma}_n^a$ against $q_n - q^*$. All country-specific slope estimates are biased towards zero. This means that we underestimate the country-specific slopes for countries with comparative advantage in high-z industries and we overestimate the country-specific slopes for countries with comparative advantage in low-z industries. As a result, the new benchmarking estimator $\widehat{b}_G$ – which according to (26) is the slope of the line connecting the average country-specific slope estimate for countries with comparative advantage in high-z industries with the average country-specific slope estimate for countries with comparative advantage in low-z industries – will necessarily be biased towards zero.

Figure 3A: True Model

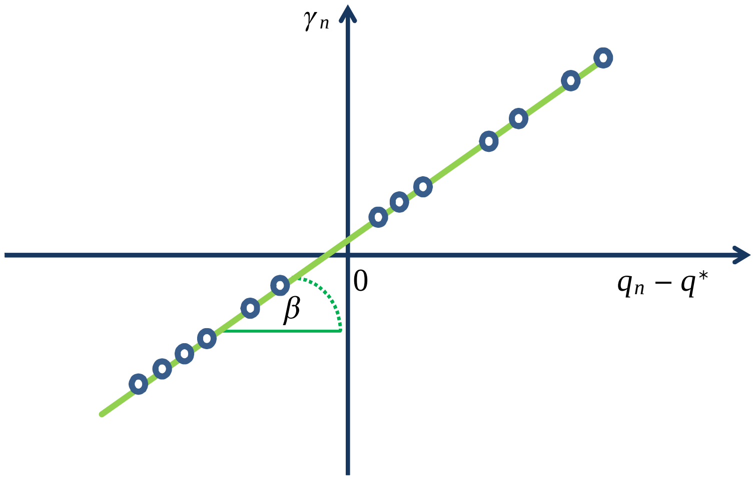
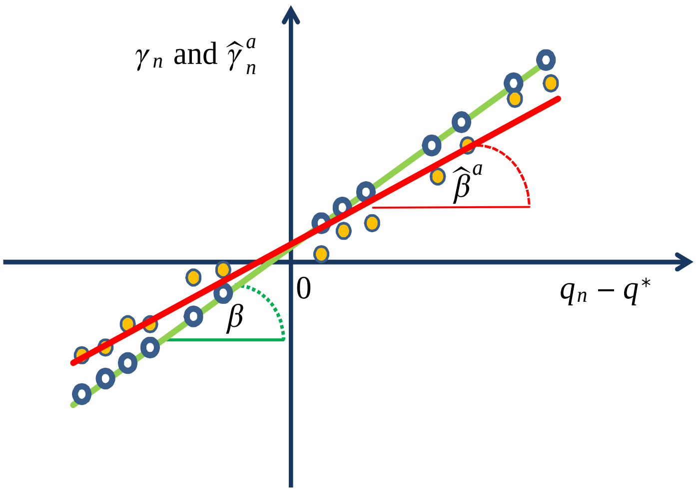


Figure 3B: Estimated Model



Notes: True country-specific slopes (bold circles) and estimated country-specific slopes (filled circles) in a model of comparative advantage with many different countries (each circle represents a country). In this case the new benchmarking estimator is necessarily biased towards zero (attenuation bias).

Summarizing, the new benchmarking estimator is downward biased in figure 3 – and hence a lower bound on the true effect – because (A1) and (A2) combined with (25) imply

$$\frac{\overbrace{\widehat{\overline{\gamma}}_{nA}}^{\downarrow bias} - \overbrace{\widehat{\overline{\gamma}}_{nB}}^{\uparrow bias}}{\overline{q}_A - \overline{q}_B} \quad \Rightarrow \quad \overbrace{\widehat{b}_G}^{\downarrow bias}. \tag{29}$$

For $\beta < 0$, the argument is analogous.

A 2SLS Interpretation The new benchmarking estimator in (25) turns out to have an interpretation as a 2SLS estimator applied to the cross-industry cross-country model

$$y_{in} = \alpha_n + \alpha_i + bz_{iUS}q_n + residual_{in}. \tag{30}$$

To see this, define an indicator function 1_n^* that assigns a value of 1 to countries in group A and a value of 0 to all other countries (or the other way round). Now we can estimate (30) using 2SLS with the product of the indicator function and the US industry characteristics $z_{iUS}1_n^*$ as an instrument for the interaction term $z_{iUS}q_n$. This 2SLS estimator can be expressed in terms of demeaned data as

$$\widehat{b}_{G,2SLS} = \frac{\sum_n \widehat{\gamma}_n \left(1_n^* - \overline{1}^*\right)}{\sum_n q_n \left(1_n^* - \overline{1}^*\right)} \tag{31}$$

where the $\widehat{\gamma}_n$ are the country-specific least-squares slopes estimated using US industry characteristics as a proxy for industry characteristics in all other countries and $\overline{1}^* = \frac{1}{N}\sum_n 1_n^*$. It is now straightforward to show that the right-hand side of (31) is the same as the right-hand side of (25) and hence $\widehat{b}_{G,2SLS} = \widehat{b}_G$.

3.2.2 Estimated Country Grouping

The 2SLS estimator in (31) cannot be implemented directly because we generally do not know whether countries have a comparative advantage in high-z or low-z industries. As a result, we cannot generate the necessary indicator function 1_n^*. But it turns out that we can estimate 1_n^* consistently under the (testable) assumption $\beta \neq 0$. As shown in Wooldridge (2002, Section 6.1.2), the 2SLS estimator using a consistently estimated instrument is not only consistent but has the same asymptotic distribution as the 2SLS estimator using the actual instrument under weak conditions. Hence, we can obtain an estimate with the same asymptotic distribution as $\widehat{b}_{G,2SLS}$ by estimating the cross-industry cross-country model in (30) with 2SLS and instrumenting $z_{iUS}q_n$ with $z_{iUS}\widehat{1}_n^*$ where $\widehat{1}_n^*$ is a consistent estimator of 1_n^*. We now discuss two approaches to obtain such a consistent estimator. A simple approach that only relies on the sign of the country-specific slope estimates and a second, somewhat more complex, approach that also considers the country characteristic shaping the direction of comparative advantage.

Simple Approach The first estimator, which we refer to as $\widehat{1}^*_{1n}$, is an indicator that takes the value of 1 for countries n with $\widehat{\gamma}_n \geq 0$ and the value of 0 for all other countries. Recall that $\widehat{\gamma}_n$ converges to $\widehat{\gamma}^a_n = \gamma_n Corr(z_{US}, z_n) = \beta(q_n - q^*)Corr(z_{US}, z_n)$, where we made use of (15) and (20). Hence, as long as $\beta \neq 0$, assumption (A1) implies that $\widehat{\gamma}^a_n$ has the same sign for countries on the same side of q^* and $\widehat{1}^*_{1n}$ is a consistent estimator of 1^*_n . The hypothesis $\beta = 0$ can be tested, as it implies that $\gamma_n = 0$ for all countries n. We can therefore proceed in three steps. First, estimate the least-squares slopes $\widehat{\gamma}_n$ and test the hypothesis $\gamma_n = 0$ for all n. Second, if this hypothesis can be rejected, obtain the estimate of the indicator function for each country $\widehat{1}^*_{1n}$. Third, estimate the model in (30) with 2SLS using $z_{iUS}\widehat{1}^*_{1n}$ as an instrument for $z_{iUS}q_n$.

Alternative Approach There is a second, somewhat more complex, approach to obtain a consistent estimator of the indicator function 1^*_n. This approach differs from the first approach in that it also uses information on the characteristics q_n that may be driving countries' comparative advantage. To see the basic idea, suppose that $\beta > 0$ and that (20) holds. In this case, countries with $q_n \geq q^*$ have a comparative advantage in high-z industries and countries with $q_n < q^*$ have a comparative advantage in low-z industries. The idea of the approach is to estimate q^* and then group countries according to whether q_n is above or below q^*. Estimating q^* would be simple if we observed γ_n. We could chose a value $\widehat{q}^*$ that maximizes the share of countries with $q_n \geq \widehat{q}^*$ and $\gamma_n \geq 0$ plus the share of countries with $q_n < \widehat{q}^*$ and $\gamma_n < 0$. This can be thought of estimating q^* so as to maximize the share of countries whose direction of comparative advantage conforms to (20). Once we have obtained the threshold $\widehat{q}^*$ we could generate the indicator function 1^*_n by assigning a value of 1 to countries n with $q_n \geq \widehat{q}^*$ and a value of 0 to all other countries. This approach would yield a unique indicator function, although the threshold $\widehat{q}^*$ would not be unique, as the data for the country characteristic q_n are discrete. If one wants to ensure a unique threshold also, this can be easily done by choosing $\widehat{q}^*$ from the set of values taken by the country characteristic q_n. An analogous approach can be used to obtain $\widehat{q}^*$ when $\beta < 0$. If we observed γ_n, we could chose a threshold $\widehat{q}^*$ from the set of values taken by the country characteristic q_n that maximizes the share of countries with $q_n \geq \widehat{q}^*$ and $\gamma_n \leq 0$ plus the share of countries with $q_n < \widehat{q}^*$ and $\gamma_n > 0$. This can again be thought of as estimating q^* to maximize the share of countries whose direction of comparative advantage conforms to (20).

In practice, we generally neither observe the γ_n nor do we know whether $\beta > 0$ or $\beta < 0$. But instead of the γ_n we can use the least-squares estimates $\widehat{\gamma}_n$, as their sign is a consistent estimate of the direction of countries' comparative advantage under assumption

(A1). That we do not observe whether $\beta > 0$ or $\beta < 0$ can be taken care of by choosing either the threshold estimated under the assumption $\beta > 0$ or the threshold estimated under the assumption $\beta < 0$, depending on which yields a greater share of countries whose direction of comparative advantage conforms to (20). Summarizing, the alternative approach generates a consistent estimate of the indicator function 1_n^* by splitting countries into two groups based on an estimate of the threshold q^*. This estimate is obtained by maximizing the share of countries whose estimated direction of comparative advantage conforms to (20).

To explain the second approach more formally, we need to introduce a considerable amount of notation. Let Q be the set that collects the values of q_n for all countries n. Define $p(q|\, q \in Q)$ as the share of countries with $q_n \geq q$ and a comparative advantage in high-z industries plus the share of countries with $q_n < q$ and a strict comparative advantage in low-z industries,

$$(32) \qquad p(q|\, q \in Q) = \text{share of countries with } \begin{cases} q_n < q \text{ and } \gamma_n < 0 \\ q_n \geq q \text{ and } \gamma_n \geq 0 \end{cases}.$$

Also define $m(q|\, q \in Q)$ as the share of countries with $q_n \geq q$ and a comparative advantage in low-z industries plus the share countries with $q_n < q$ and a strict comparative advantage in high-z industries

$$(33) \qquad s(q|\, q \in Q) = \text{share of countries with } \begin{cases} q_n < q \text{ and } \gamma_n > 0 \\ q_n \geq q \text{ and } \gamma_n \leq 0 \end{cases}.$$

Let q_Q^* be the value $q \in Q$ such that countries with $q_n \geq q_Q^*$ have a comparative advantage going in the same direction and countries with $q_n < q_Q^*$ also have a comparative advantage going in the same direction. If $\beta > 0$, q_Q^* is straightforward to determine as it is the unique value maximizing $p(q|\, q \in Q)$. Similarly, q_Q^* is also straightforward to determined if $\beta < 0$, as it is the unique value maximizing $s(q|\, q \in Q)$. Collecting the cases $\beta > 0$ and $\beta < 0$ it follows that as long as $\beta \neq 0$, we can determine q_Q^* as

$$(34) \qquad q_Q^* = \begin{cases} \text{argmax } p(q|\, q \in Q) \text{ if max } p(q|\, q \in Q) \geq \text{max } s(q|\, q \in Q) \\ \text{argmax } s(q|\, q \in Q) \text{ if max } s(q|\, q \in Q) > \text{max } p(q|\, q \in Q) \end{cases}.$$

To see this, notice that if $\beta > 0$, max $p(q|\, q \in Q) = 1$ while max $s(q|\, q \in Q) = 0$ except if there are countries that happen to have a value of q_n exactly equal to q^*; in this case, max $s(q|\, q \in Q) = M/N$ with N the number of countries and M the number of countries with $q_n = q^*$. On the other hand, if $\beta < 0$, max $s(q|\, q \in Q) = 1$ while max $p(q|\, q \in Q) = 0$ except if there are countries that happen to have a value of q_n exactly equal to q^*; in this case, max $p(q|\, q \in Q) = M/N$.

Using (34), we can obtain a consistent estimator of q_Q^* once we have consistent estimators of $p(q)$ and $s(q)$. Moreover, consistent estimators of $p(q)$ and $s(q)$ are straightforward to find under assumption (A1). In this case, $\widehat{\gamma}_n^a \geq 0$ if and only if $\gamma_n \geq 0$, see (15). Hence, we can obtain consistent estimators $\widehat{p}(q)$ and $\widehat{s}(q)$ of $p(q)$ and $s(q)$ by replacing γ_n by $\widehat{\gamma}_n$ in (32) and (33). Then we can replace $p(q)$ and $s(q)$ by $\widehat{p}(q)$ and $\widehat{s}(q)$ in (34) to obtain a consistent estimator $\widehat{q}_Q^*$ of q_Q^*. Finally, we can obtain our alternative consistent estimator of 1_n^* as the indicator function $\widehat{1}_{2n}^*$ that assigns a value of 1 to all countries n with $q_n \geq \widehat{q}_Q^*$ and the value of 0 to all other countries (or the other way around).

3.2.3 Applying the 2SLS Grouping Estimator

We now illustrate the alternative benchmarking estimator in the context of Nunn's (2007) empirical analysis of the effect of contract enforcement on comparative advantage in industries that depend more on relationship-specific intermediate inputs (see also Levchenko, 2007, and Costinot, 2009, for related empirical and theoretical findings). Nunn's analysis is based on the cross-industry cross-country model

$$\ln e_{in} = a_n + a_i + b z_{iUS} q_n + residual_{in} \tag{35}$$

where $\ln e_{in}$ is the log value of exports of country n in industry i; q_n the quality of contract enforcement in country n; and z_{iUS} a measure of industry i's dependence on relationship-specific intermediate inputs obtained using US data. Nunn's key finding is that b is positive and statistically significant, indicating that countries with better contract enforcement export relatively more in industries that depend more on relationship-specific intermediate inputs.

To apply our 2SLS benchmarking estimator, we first need to estimate the country-specific slopes $\gamma_n = bq_n$ in

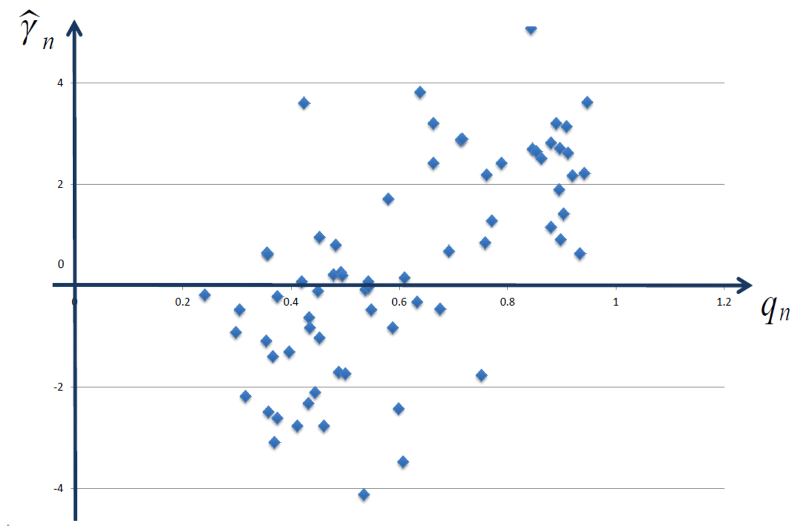
$$\ln e_{in} = \alpha_n + \alpha_i + \gamma_n z_{iUS} + residual_{in}. \tag{36}$$

The least-squares slope estimates $\widehat{\gamma}_n$ tell us how much more country n exports in industries that depend more on relationship-specific intermediate inputs. We plot these estimates against the quality of contract enforcement q_n in figure 4. The second step is to use the least-squares slope estimates $\widehat{\gamma}_n$ to test the hypothesis that $\beta = 0$ (by testing whether $\gamma_n = 0$ for all n). This hypothesis is rejected at any conventional confidence level. The third step is to use the least-squares slope estimates to obtain the two indicators $\widehat{1}_{1n}^*$ and $\widehat{1}_{2n}^*$ that group countries by the direction of their comparative advantage.[10] We can then obtain

[10] The estimator $\widehat{1}_{1n}^*$ assigns countries with a positive least-squares slopes $\widehat{\gamma}_n$ a value of 1 and all other countries a 0. The estimator $\widehat{1}_{2n}^*$ assigns countries with a value for the quality of contract enforcement q_n above 0.588 a value of 1 and all other countries a 0 (that is $\widehat{q}_Q^*$ is estimated to be 0.588).

an estimate of the effect of better contract enforcement on exports in relationship-specific input industries by applying 2SLS to (35) and instrumenting the interaction $z_{iUS}q_n$ with $z_{iUS}\widehat{1}^*_{1n}$, $z_{iUS}\widehat{1}^*_{2n}$, or both. We proceed using both instruments simultaneously as this is the most efficient approach. Using Nunn's baseline specification (in his Table IV), this yields a standardized beta coefficient of 0.361 with a standard error of 0.015. Nunn's estimate using the standard cross-industry cross-country benchmarking estimator is 0.289 with a standard error of 0.013. Hence, our new benchmarking estimator – which provides a lower bound on the strength of the true effect under assumptions (A1) and (A2) – yields that better contract enforcement is even more important for exports in relationship-specific input industries than the estimator of the cross-industry cross-country literature.

Figure 4: Country-Specific Slopes, Nunn (2007)



Notes: Country-specific least-squares slope estimates for the cross-industry cross-country model of Nunn (2007) plotted against the quality of contract enforcement.

4 Conclusion

Cross-industry cross-country models are used extensively in economics. The approach has attractive features, like its focus on theoretical mechanisms and the possibility to control for country-level determinants of economic activity. But there are also drawbacks. Implementation requires specifying technological industry characteristics that are generally unobservable and must therefore be proxied with industry characteristics in a benchmark country. That this can lead to an attenuation bias is both well understood and unsurprising. What appears to not be understood, as we failed to find any mention of it, is that the exact same approach generally thought to give rise to an attenuation bias can actually lead to a (large) amplification bias.

A main area of application of cross-industry cross-country models is international trade, where these models have been used to examine the effects of factor endowments and institutions on comparative advantage. We show that in this context there is an estimator that yields a lower bound on the true effect, as long as the data indicate that some countries differ in the direction of their comparative advantage.

References

Acemoglu, Daron and Fabrizio Zilibotti (2001). Productivity Differences. *Quarterly Journal of Economics*, *116*(2), pp. 563-606

Acemoglu, Daron, Simon Johnson, and Todd Mitton (2009). Determinants of Vertical Integration: Financial Development and Contracting Costs. *Journal of Finance*, *64*(3), pp. 1251-1290.

Aghion, Philippe, Thibault Fally, and Stefano Scarpetta (2007). Credit Constraints as a Barrier to the Entry and Post-entry Growth of Firms. *Economic Policy*, *52*(6), pp. 731-779.

Aghion, Philippe, Philippe Askenazy, Nicolas Berman, Gilbert Cette, and Laurent Eymard (2012). Credt Constraints and the Cyclicality of R&D Investment: Evidence from France. *Journal of the European Economic Association. 10*(5), pp. 1001-1024.

Aghion, Philippe, Emmanuel Farhi, and Enisse Kharroubi (2013). Countercyclical Monetary Policy, Liquidity Dependence, and Economic Growth. Mimeo, Harvard University.

Aghion, Philippe, David Hemous, and Enisse Kharroubi (2014). Cyclical Fiscal Policy, Credit Constraints, and Industry Growth. *Journal of Monetary Economics, 62*(1), pp. 41-58.

Aghion, Philippe, Peter Howitt, and Susanne Prantl (2014). Patent Rights, Product Market Reforms, and Innovation. Harvard University, updated version of NBER WP 18854.

Aizenman, Joshua and Vladyslav Sushko (2011). Capital Flow Types, External Financing Needs, and Industrial Growth: 99 countries, 1991-2007. NBER Working Paper 17228.

Alfaro, Laura, and Andrew Charlton (2009). Intra-Industry Foreign Direct Investment. *American Economic Review*, *99*(5), pp. 2096-2119.

Angrist, Joshua D. (1991). Grouped-Data Estimation and Testing in Simple Labor-Supply Models. *Journal of Econometrics*, *47*(2-3), pp. 243-266.

Basco, Sergi (2013). Financial Development and the Product Cycle. *Journal of Economic Behavior and Organization*, *94*(1)*,* pp. 295-313.

Baltagi, Badi (2008). *Econometric Analysis of Panel Data.* John Wiley and Sons, USA.

Beck, Thorsten (2003). Financial Dependence and International Trade. *Review of International Economics*, *11*(1), pp. 296-316.

Beck, Thorsten, Asli Demirgüc-Kunt, Luc Laeven, and Ross Levine (2008). Finance, Firm Size, and Growth. *Journal of Money, Banking, and Finance*, *40*(8), pp. 1371-1405.

Bernard, Andrew and Charles I. Jones (1996). Comparing Apples to Oranges, pp. Productivity Convergence and Measurement Across Industries and Countries, *American Economic Review*, *86*(6), pp. 1216-1238.

Bertrand, Marianne, Antoinette Schoar, and David Tesmar (2007). Banking Deregulation and Industry Structure, pp. Evidence from the 1985 French Banking Act. *Journal of Finance*, *62*(2), pp. 597-628.

Blyde, Jun, and Danielken Molina (2015). Logistic Infrastructure and the International Location of Fragmented Production. *Journal of International Economics, 95*(2), pp. 319-332

Bombardini, Matilde, Giovanni Gallipoli and German Pupato (2012). Skill Dispersion and Trade Flows. *American Economic Review, 102*(5), pp. 2327-2348.

Braun, Matias, and Borja Larrain (2005). Finance and the Business Cycle: International Inter-Industry Evidence. *Journal of Finance*, *60*(3), pp. 1097-1128.

Broner, Fernando, Paula Bustos, and Vasco Carvalho (2015). Sources of Comparative Advantage in Polluting Industries. Working Paper, CREI.

Brown, James R., Gustav Martisson, and Bruce C. Petersen (2013). Law, Stock Markets, and Innovation. *Journal of Finance*, *68*(4), pp. 1517-1549.

Calomiris, Charles, Mauricio Larrain, José Liberti, and Jason Sturgess (2016). How Collateral Laws Shape Lending and Sectoral Activity. NBER Working Paper 21911.

Carluccio, Juan and Thibault Fally (2012). Global Sourcing under Imperfect Capital Market. *Review of Economics and Statistics*, *94*(3), pp. 764–788.

Caselli, Francesco (2005). Development Accounting. In *The Handbook of Economic Growth* edited by Philippe Aghion and Steven N. Durlauf. North Holland, Amsterdam, Netherlands.

Cetorelli, Nicola, and Philip Strahan (2006). Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local US Markets. *Journal of Finance*, *61*(1), pp. 437-461.

Ciccone, Antonio, and Elias Papaioannou (2007). Red Tape and Delayed Entry. *Journal of the European Economic Association*, *Papers and Proceedings*, *5*(2-3), pp. 444-458.

Ciccone, Antonio, and Elias Papaioannou (2009). Human Capital, the Structure of Production, and Growth. *Review of Economics and Statistics*, *91*(2), pp. 66-82.

Cingano, Federico, Marco Leonardi, Julian Messina, and Giovanni Pica (2010).

The Effect of Employment Protection Legislation and Financial Market Imperfections on Investment: Evidence from a Firm-Level Panel of EU Countries. *Economic Policy*, *25*(1), pp. 117-163.

Cingano, Federico, and Paolo Pinotti (2016). Trust, Firm Organization, and the Pattern of Comparative Advantage. *Journal of International Economics*, *100*(1), pp. 1-13.

Chan, Jackie M.L. and Kalina Manova (2015). Financial Development and the Choice of Trade Partners. *Journal of Development Economics*, *116*(1), pp. 122-145.

Chor, David (2010). Unpacking Sources of Comparative Advantage, pp. A Quantitative Approach. *Journal of International Economics*, *82*(2), pp. 152-167.

Claessens, Stijn, and Luc Laeven (2003). Financial Development, Property Rights, and Growth. *Journal of Finance*, *58*(7), pp. 2401-2436.

Claessens, Stijn, Hui Tong, and Shang-Jin Wei (2012). From the Financial Crisis to the Real Economy: Using Firm-level Data to Identify Transmission Channels. *Journal of International Economics*, *88*(2), pp. 375-387.

Costinot, Arnaud (2009). On the Origins of Comparative Advantage. *Journal of International Economics*, *77*(2), pp. 255-264.

Cuñat, Alejandro, and Marc Melitz (2012). Volatility, Labor Market Flexibility, and Comparative Advantage. *Journal of the European Economics Association*, *10*(2), pp. 225-254.

Debaere, Peter (2014). The Global Economics of Water: Is Water a Source of Comparative Advantage? *American Economic Journal: Applied Economics*, *6*(2), pp. 32-48.

Dell'Ariccia, Giovanni, Enrica Detragiache, and Raghuram G. Rajan (2008). The Real Effects of Banking Crises. *Journal of Financial Intermediation*, *17*(1), pp. 89-112.

Duchin, Ran, Ogguzhan Ozbas, and Berk A. Sensoy (2010). Costly External Finance, Corporate Investment, and the Subprime Mortgage Credit Crisis. *Journal of Financial Economics*, *97*(3), pp. 418-435.

Duygan-Bump, Burcu, Alexey Levkov and Judith Montoriol-Garriga (2015). Financing Constraints and Unemployment, pp. Evidence from the Great Recession. *Journal of Monetary Economics*. *75*(1), pp. 89-105.

Fafchamps, Marcel and Matthias Schündeln (2013). Local Financial Development and Firm Performance, pp. Evidence from Morocco. *Journal of Development Economics, 103*(1), pp. 15-28.

Feenstra, Robert (2004). *Advanced International Trade: Theory and Evidence*. Princeton University Press, USA.

Feenstra, Robert, Chang Hong, Hong Ma, and Barbara J. Spencer (2013).

Contractual Versus Non-Contractual Trade: The Role of Institutions in China. *Journal of Economic Behavior and Organization*, *94*(1), pp. 281-294.

Ferguson, Shon and Sara Formai (2013). Institution-Driven Comparative Advantage and Organizational Choice. *Journal of International Economics*, *90*(1), pp. 193-200.

Fisman, Raymond, and Inessa Love (2003). Trade Credit, Financial Intermediary Development, and Industry Growth. *Journal of Finance*, *58*(1), pp. 353-374.

Fisman, Raymond, and Inessa Love (2007). Financial Development and Growth in the Short and Long Run. *Journal of the European Economic Association*, *5*(2-3), pp. 470-479.

Griffith, Rachel and Gareth Macartney (2014). Employment Protection Legislation, Multinational Firms, and Innovation. *Review of Economics and Statistics, 96*(1), pp. 135-150.

Hsieh, Chang-Tai and Jonathan Parker (2007). Taxes and Growth in a Financially Underdeveloped Country. Evidence from Chile's Investment Boom. *Economia*, *8*(1), pp. 41-53.

Hsu, Po-Hsuan, Xuan Tian and Yan Xu (2014). Financial Development and Innovation: Cross-Country Evidence. *Journal of Financial Economics, 112*(1), pp. 116-135.

Hummels, David and Peter Klenow (2005). The Variety and Quality of a Nation's Exports. *American Economic Review*, *95*(6), pp. 704-723.

Iacovone, Leonardo and Veronica Zavacka (2009). Banking Crises and Exports: Lessons from the Past. *World Bank Research Working Paper* 5016.

Jacobson, Tor, and Erik von Schedvin (2015). Trade Credit and the Propagation of Corporate Failure. An Empirical Analysis. *Econometrica*, *83*(4), pp. 1351-1371.

Klapper, Laura, Luc Laeven, and Raghuram G. Rajan (2006). Entry Regulation as a Barrier to Entrepreneurship. *Journal of Financial Economics*, *82*(3), pp. 591-629.

Klingebiel, Daniela, Randall S. Kroszner, and Luc Laeven (2007). Financial Crises, Financial Dependence, and Industry Growth. *Journal of Financial Economics*, *84*(1), pp. 187-228.

Krishna, Pravin, and Andrei Levchenko (2013). Comparative Advantage, Complexity, and Volatility. *Journal of Economic Behavior and Organization, 94*(2), pp. 314-329.

Laeven, Luc and Fabian Valencia (2013). The Real Effects of Financial Sector Interventions during Crises. *Journal of Money, Credit and Banking*, *45*(1), pp. 147-177.

Larrain, Mauricio (2014). Capital Account Opening and Wage Inequality. *Review of Financial Studies, 28*(6), pp. 1555-1587.

Levchenko, Andrei (2007). Institutional Quality and International Trade. *Review of*

Economic Studies, *74*(3), pp. 791-819.

Manova, Kalina (2008). Credit Constraints, Equity Market Liberalizations and International Trade. *Journal of International Economics*, *76*(1), pp. 33-47.

Manova, Kalina (2013). Credit Constraints, Heterogeneous Firms, and International Trade. *Review of Economic Studies*, *80*(2), pp. 711-744.

Manova, Kalina, Shang-Jin Wei, and Zhiwei Zhang (2015). Firm Exports and Multinational Activity under Credit Constraints. *Review of Economics and Statistics, 97*(3), pp. 574-588.

Michelacci, Claudio, and Fabiano Schivardi (2013). Does Idiosyncratic Business Risk Matter for Growth? *Journal of the European Economic Association, 11*(2), pp. 343-368.

Nunn, Nathan (2007). Relationship-Specificity, Incomplete Contracts and the Pattern of Trade. *Quarterly Journal of Economics*, *122*(2), pp. 569-600.

Nunn, Nathan and David Trefler (2014). Domestic Institutions as a Source of Comparative Advantage. In *Handbook of International Economics* edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. North Holland, Netherlands.

Pagano, Patrizio, and Fabiano Schivardi (2003). Firm Size Distribution and Growth, *Scandinavian Journal of Economics*, *105*(1), pp. 255-274.

Paunov, Caroline (2016). Corruption's Asymmetric Impacts on Firm Innovation. *Journal of Development Economics*, *118*(1), pp. 216-231.

Rajan, Raghuram G., and Arvind Subramanian (2011). Aid, Dutch Disease, and Manufacturing Growth. *Journal of Development Economics*, *94*(1), pp. 106-118.

Rajan, Raghuram G., and Luigi Zingales (1998). Financial Dependence and Growth. *American Economic Review*, *88*(3), pp. 559-586.

Ranjan, Priya and Jae Young Lee. (2007). Contract Enforcement and International Trade. *Economics and Politics*, *19*(2): 191-218.

Rauch, James E (1999). Networks versus Markets in International Trade. *Journal of International Economics*, *48*(1), pp. 7-35.

Romalis, John (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, *94*(1), pp. 67-97.

Schott, Peter (2004). Across-Product versus Within-Product Specialization in International Trade. *Quarterly Journal of Economics*, *119*(2), pp. 647-678.

Tang, Heiwai (2012). Labor Market Institutions, Firm-specific Skills, and Trade Patterns. *Journal of International Economics*, *87*(2), pp. 337-351.

Wooldridge, Jeffrey M. (2002). *Econometric Analysis of Cross-Section and Panel Data*. MIT Press, USA.