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

Antonio Ciccone
Elias Papaioannou

May 2019

Barcelona GSE Working Paper Series

Working Paper nº 1100

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 and CEPR

May 2019

Abstract

Cross-industry cross-country models are applied widely in economics. For example, to investigate the effect of financial development on economic growth or the effect of institutional quality on international trade. The literature estimates the effect of interest by examining the interaction between country characteristics—for example, financial development or institutional quality—and theoretically relevant technological industry characteristics—for example, dependence on external finance or relationship-specific inputs. As the relevant industry characteristics are unobservable for most countries, they are proxied by industry characteristics of a benchmark country. We analyze this approach when there is cross-country heterogeneity in technological industry characteristics. First, we show that the estimation approach in the literature is biased and that the bias cannot be signed if technologically more similar countries are also more similar in terms of other characteristics. Second, we derive necessary and sufficient conditions for identification of the effect of interest. Third, we use the new identification approach to reestimate the impact of institutional quality on comparative advantage in industries that rely on relationship-specific inputs.

Keywords: Economic Growth, International Specialization and Trade, Country-Specific Technology, Financial Development, Institutions

JEL classification numbers: G30, F10, O40

*Emails: antonio.ciccone@uni-mannheim.de and eliaspapaioannou@london.edu. We thank two anonymous referees and Joseba Martinez for very useful comments. We also thank Sebastian Hohmann, Claudio Kretz, and Tomasz Olma for their great help in editing the paper, and Yogam Tchokni for outstanding research assistance. Financial support by the German Research Foundation through CRC TR 224 and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Program for Centers of Excellence in R&D SEV-2015-0563 is gratefully acknowledged.

1 Introduction

Empirical work in economics has relied extensively on cross-industry cross-country models over the past 20 years. These models relate cross-country differences in industry performance—for example, industry growth or industry exports—to an interaction between (i) country characteristics like financial development, institutional quality, or human capital and (ii) theoretically relevant technological industry characteristics like external-finance dependence, reliance on certain inputs, or skill intensity. The approach has proven useful for examining a wide variety of questions, reviewed below. Two strands of research stand out. First, following Rajan and Zingales (1998), cross-industry cross-country models are used to study how financial development, property rights protection, contract enforcement, and human capital affect industry employment, output, and value added growth. Second, cross-industry cross-country models serve as the basis for empirical studies of the institutional determinants of comparative advantage; see Nunn and Trefler (2014) for a review. For example, Nunn (2007) uses the approach to show that institutional quality is a source of comparative advantage in industries that rely on relationship-specific inputs.

The theoretically relevant technological industry characteristics in the cross-industry cross-country literature depend on the economic question asked, but a common feature is that they are unavailable for most countries. This is for two main reasons. First, there is little industry data for most countries. Second, technological industry characteristics must be inferred from observed, endogenous industry characteristics (e.g., Rajan and Zingales, 1998). Such inference is challenging for countries where firms have adapted to, for example, low financial development, institutional quality, or human capital. As a result, the cross-industry cross-country literature generally proceeds by treating technological industry characteristics as unobservable for all countries except a highly-developed benchmark country with relatively undistorted markets, usually the USA. The effect of interest is then estimated using the technological industry characteristics of the benchmark country as proxies for the technological industry characteristics of all other countries.

To better understand this estimation approach, we study cross-industry cross-country models with two main features. First, following the literature, the relevant technological industry characteristics are unobservable for all but a benchmark country. Second, there is some cross-country heterogeneity in technological industry characteristics (e.g., Bernard and Jones, 1996; Acemoglu and Zilibotti, 2001; Schott, 2004; Caselli, 2005).

We show that the estimation approach used in the cross-industry cross-country literature results in a bias shaped by two countervailing forces. A key determinant of the relative strength of these forces—and hence the ultimate bias of the estimation approach—is whether technologically more similar countries are also more similar in terms of other characteristics, for example, their levels of financial development, institutional quality, or human capital. If there is no relationship between how similar countries are technologically and how similar they are in terms of other characteristics, the approach in the literature yields estimates of the effect of interest that are biased towards zero (attenuated). That estimates may be attenuated is generally understood in the cross-industry cross-country literature and explained in terms of a classical measurement error bias due to the industry characteristics of the benchmark country measuring the industry characteristics of other countries with some error.[1] On the other hand, if technologically more similar countries are also more similar in terms of other characteristics, the approach in the cross-industry cross-country literature may yield amplified or entirely spurious estimates. There appears to be no discussion of this issue in the literature.[2]

The size of the amplification bias and of spurious effects generated by the estimation approach in the cross-industry cross-country literature depends on how much more similar (dissimilar) countries are technologically to the benchmark country as they become more similar (dissimilar) in other characteristics. As a result, these biases can be sizable if there is a drop-off in technological similarity with the benchmark country as countries become more dissimilar in, for example, their levels of financial development, institutional quality, or human capital—even if countries are on average quite similar to the benchmark country.

As the estimation approach used in the cross-industry cross-country literature does not identify the effect of interest when there is cross-country heterogeneity in technological industry characteristics, it is important to develop alternatives. To do so, we first provide an analysis of identification. Identification is substantially more challenging with cross-country heterogeneity in technological industry characteristics than without, and sometimes exact identification is impossible. Our analysis develops necessary and sufficient conditions for exact identification and provides bounds when exact identification is impossible. We use the new identification approach to reestimate the effect of institutional quality on comparative advantage in relationship-specific-input intensive industries in Nunn (2007).

The rest of the paper is structured as follows. Section 2 reviews applications of cross-industry cross-country models. Section 3 examines the estimation approach in the literature. Section 4 discusses identification of the effect of interest. Section 5 uses our identification results to reestimate Nunn (2007). Section 6 provides the conclusion.

[1]For example, see Rajan and Zingales (1998), p. 567. Attenuated estimates are sometimes seen as a relatively minor drawback as they can be interpreted as lower bounds on the strength of true effects.

[2]The literature does discuss the possibility of endogeneity bias (endogenous country characteristics or industry characteristics of the benchmark country) or omitted variable bias (relevant industry-country interactions are omitted). We abstract from these well-understood sources of estimation bias by assuming that country as well as industry characteristics of the benchmark country are exogenous and that there is a single relevant industry-country interaction.

2 Applications in the Literature

Cross-industry cross-country models have been applied extensively in many areas of economics. Our review is only meant to illustrate the range of applications. Appendix Table 1 provides brief summaries of more papers in the literature.

The Economic Effects of Financial Markets. Starting with the influential work of Rajan and Zingales (1998), who showed that financial development exerts a disproportionately large impact on sales growth in industries that depend more on external sources of finance, cross-industry cross-country models have been applied extensively to investigate the effects of financial markets on economic growth, firm entry and exit, investment, and innovation. For example, Fisman and Love (2003) find that in countries with less developed financial markets, industries that rely more on trade credit grow faster, and Fisman and Love (2007) show that better developed financial markets spur growth in industries facing better global growth opportunities. Claessens and Laeven (2003), Braun and Larrain (2005), and Lei, Qiu, and Wan (2018) extend the cross-industry cross-country model of Rajan and Zingales (1998) to account for the role of intangible assets. Brown, Martinson, and Petersen (2013), Hsu, Tian, and Xu (2014), and Acharya and Zu (2017) use cross-industry cross-country models to examine the impact of financial markets on innovation. Pagano and Shivardi (2003), Aghion, Fally, and Scarpetta (2007), and Beck, Demirguüc-Kunt, Laeven, and Levine (2008) analyze how financial markets affect firm entry and exit and the growth of smaller versus larger firms.

Cross-industry cross-country models are also used to examine the economic effects of specific financial market policies or institutions, such as bank recapitalizations (Laeven and Valencia, 2013), insider trading legislation (Edmans, Jayaraman, and Schneemeir, 2017), and collateral laws (Calomiris, Larrain, Liberti, and Sturgess, 2017). A more recent strand of research employs cross-industry cross-country models to assess the effects of financial crises and capital account liberalization on macroeconomic performance (Dell'Ariccia, Detragiache, and Rajan, 2008; Iacovone and Zavacka, 2009; Duchin, Ozbas, and Sensoy, 2010; Claessens, Tong, and Wei, 2012; Larrain and Stumpner, 2018).

International Specialization and Trade. Cross-industry cross-country models are widely used to examine the determinants of international trade and international specialization. Levchenko (2007) and Nunn (2007) use cross-industry cross-country models to examine the effect of institutional quality on international specialization (see also Ferguson and Formai, 2013; Nunn and Trefler, 2014). Manova (2008, 2013) uses cross-industry cross-country models to link financial development to the patterns of international trade (see also Chan and Manova, 2015; Manova, Wei, and Zhang, 2015; Claessens, Hassib, and van Horen, 2017; Crin and Oglirari, 2017). Ciccone and Papaioannou (2009) and Debaere (2015) use cross-industry cross-country models to examine the effects of human capital and natural resources on international specialization. Cingano, Leonardi, Messina and Pica (2010), Mueller and Phillippon (2011), Cuñat and Melitz (2012), Tang (2012), Griffith and Macartney (2014), and Broner, Bustos, and Carvalho, (2016) use cross-industry cross-country models to examine role of labor-market and environmental regulation for international trade.

Other Applications. Cross-industry cross-country models have proven useful for examining a surprisingly wide variety of additional economic questions. For example, Alfaro and Charlton (2009), Carluccio and Fally (2012), Basco (2013), Blyde and Danielken (2015), Paunov (2016), and Fort (2017) use cross-industry cross-country models to analyze the determinants of outsourcing, foreign direct investment, and the fragmentation of production. Pagano and Schivardi (2003), Klapper, Laeven, and Rajan (2006), Acemoglu, Johnson, and Mitton (2009), Aizenman and Sushko (2011), Bombardini, Gallipoli, and Pupato (2012), Michelacci and Schivardi (2013), Larrain (2014), and Aghion, Howitt, and Prantl (2014) use cross-industry cross-country models to analyze the economic consequences of cross-country differences in firm size distributions, entry regulation, transaction costs, risk sharing possibilities, and skill dispersion. Rajan and Subramanian (2010) and Chauvet and Ehrhart (2018) use cross-industry cross-country models to understand the economic effects of foreign aid, while Pierce and Snyder (2017) and Levine, Lin, and Xie (2018) use them to study the legacy of slave trade. Aghion, Farhi, and Kharroubi (2015), Aghion, Hemous, and Kharroubi (2014), and Cecchetti and Kharoubi (2018) use cross-industry cross-country models to analyze the economic effects of fiscal and monetary policy over the business cycle and Avdjiev, Bruno, Koch, and Shin (2018) to analyze the economic impact of exchange rates.

3 The Standard Benchmarking Estimator

3.1 The Model

The basis of cross-industry cross-country models are theories linking industry outcomes 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 financial development and the external-finance dependence of industries. In Nunn (2007), the outcome variable is industry exports and the interaction is between institutional quality 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 in the literature is

$$y_{in} = (\alpha + \beta x_n) z_{in} + \nu_{in} \tag{1}$$

where y_{in} is the outcome in I industries indexed by i and N countries indexed by n; x_n is the relevant country characteristic; z_{in} denotes the relevant industry characteristic in different countries; and ν_{in} captures country and industry fixed effects as well as any unobserved determinants of industry outcomes that are independent of z_{in}. The main parameter of interest is β, the coefficient on the industry-country interaction term. The parameter α captures any direct effects of industry characteristics on outcomes.[3] We take the relevant country characteristic x_n to be non-stochastic.

Estimation of β in (1) would be straightforward if there was data on the relevant industry characteristics z_{in} for a broad set of countries. But the necessary data is unavailable for most countries. Moreover, the cross-industry cross-country literature often focuses on technological industry characteristics that must be inferred from observed, endogenous industry characteristics. Such inference is challenging in countries where firms have adapted to, for example, low financial development, institutional quality, or human capital. As a result, the cross-industry cross-country literature generally proceeds by proxying the relevant technological industry characteristics of all countries with industry characteristics from a highly-developed benchmark country with relatively undistorted markets, usually the USA.

To better understand this benchmarking approach, it is useful to distinguish between the relevant technological industry characteristics z_{in} in (1) and observed industry characteristics $\tilde{z}_{in}$. Observed industry characteristics are endogenous and may therefore depend on the country characteristic x_n the model in (1) focuses on, as well as other country characteristics h_n, $\tilde{z}_{in} = g(i, x_n, h_n)$. The objective of the model in (1) is to determine the economic effects of cross-country differences in x_n through the industry-specific channel captured by z_{in}. Because $\tilde{z}_{in}$ is endogenous to the cross-country differences in x_n, using $\tilde{z}_{in}$ as right-hand-side industry characteristics in (1) would generally produce misleading conclusions. A better choice for the right-hand-side industry characteristics would be the hypothetical industry characteristics of countries n if they all had the same country characteristic x^*, $\tilde{z}^*_{in} = g(i, x^*, h_n)$ for some function $g(\cdot)$. The industry data to infer these hypothetical industry characteristics is unavailable for most countries. As a result, the cross-industry

[3] For example, Rajan and Zingales use the external-finance dependence of industries to capture the extent to which technological shocks raise an industry's investment opportunities beyond what internal funds can support. In this application, the parameter β in (1) allows testing RZ's hypothesis that financial development fosters growth disproportionally in industries with greater demand for external finance. The parameter α allows to capture direct effects of the technological shocks raising an industry's investment opportunities on industry growth. Technological shocks may affect industry growth directly in several ways, for example by changing the marginal productivity of labor, and hence equilibrium employment, across industries.

cross-country literature generally proceeds by proxing the industry characteristics z_{in} in (1) with the industry characteristics of a highly-developed benchmark country with relatively undistorted markets.[4]

It is important to understand if the benchmarking approach used in the cross-industry cross-country literature can identify the effect of interest β. Clearly, the approach works if the technological industry characteristics z_{in} of all countries were identical, i.e. $z_{in} = z_i$. In this case, using the industry characteristics of a benchmark country as a proxy for the industry characteristics of all other countries would not involve any measurement error.

But as countries generally differ in a range of characteristics that could be relevant for industry structure and technology adoption, it seems extremely implausible that the relevant technological industry characteristics of all countries are identical (e.g., Bernard and Jones, 1996; Acemoglu and Zilibotti, 2001; Schott, 2004; Caselli, 2005).[5] The technological industry characteristics of any benchmark country will therefore most likely be a noisy proxy for the technological industry characteristics of other countries. How good a proxy, can be expected to be country specific.

This point can be illustrated with the study of Nunn (2007). The key industry-level variable in Nunn is the technological relationship-specific-input intensity of industries and the country characteristic of interest is institutional quality. Clearly, the observed relationship-specific-input intensity of industries in a country may depend on its institutional quality, as firms might make fewer relationship-specific investments when they operate in an environment with worse institutional quality. For this reason, and because there is little industry data for most countries, Nunn proxies the technological relationship-specific-input intensity of industries of all countries by the observed relationship-specific-input intensity of US industries. However, even if all countries had the US level of institutional quality, the technological relationship-specific-input intensity of industries might still be different across countries, as industry structure and technology may not depend solely on institutional quality. Put differently, the relationship-specific-input intensity of US industries may be a noisy proxy for the technological relationship-specific-input intensity of other countries even if these countries had the institutional quality of the US. How good a proxy, depends on country characteristic other than institutional quality that affect industry structure and the technological relationship-specific-input intensity of industries.

For example, Nunn documents that industries that rely more on relationship-specific inputs also use human capital more intensively. Hence, the level of human capital of a

[4]In a few cases, z_{in} is proxied using industry data from several highly-developed countries. This does not affect our analysis below at all, except that the place of US industry characteristics would be taken by the industry characteristics of the alternative benchmark country/countries.

[5]In fact, the cross-industry cross-country literature has regarded this assumption as unreasonable since its beginnings, see Rajan and Zingales (1998), p. 563.

country may affect which of the many industries with different technological relationship-specific-input intensities produce in the country. That is, the technological relationship-specific-input intensity of industries may depend on the country's human capital. As a result, the relationship-specific-input intensity of industries in a high human capital country like the US could be similar to the technological relationship-specific-input intensity of countries with similar human capital but substantially different from the technological relationship-specific-input intensity of countries with low human capital.

We want a framework that allows us to capture in a flexible way that technological industry characteristics may be more similar for some country pairs than others. The first step is to take the relevant technological industry characteristics z_{in} in (1) to be the sum of a country-specific component z_n; a global industry-specific component z_i; and a country-specific industry component ε_{in}

$$z_{in} = z_n + z_i + \varepsilon_{in}. \tag{2}$$

The country-specific component z_n allows us to capture all country-specific factors that shift the entire distribution of technological industry characteristics. We treat this component as non-stochastic. The global industry component z_i allows us to capture factors that make two industries i and j different from each other independently of the country where they are located.[6] We treat this component as an independent and identically distributed random variable with $Var(z_i) > 0$. For the ε_{in} we chose a model that allows us to capture that:

(i) How different any two industries are technologically may be country specific.

(ii) Some countries may be more similar technologically than others.

To capture (i) and (ii) we assume that the ε_{in} in (2) are jointly normally distributed for all i and n. For any pair of countries $n \neq m$, the correlation of the ε_{in} across industries is allowed to be an arbitrary function of country characteristics

$$Corr(\varepsilon_{in}, \varepsilon_{im}) = \rho_{nm}. \tag{3}$$

As ρ_{nm} can be different for each country pair, (3) will yield a flexible model of the relationship between the characteristics of any pair of countries and their technological similarity. Our analysis of the bias of the estimation approach in the cross-industry cross-country literature will show that, whether the bias is upwards or downwards is partly determined by how ρ_{nm} changes as country n and country m become more dissimilar in terms of their x-characteristics. Across industries, the ε_{in} are taken to be independent, and

$$E(\varepsilon_{in}) = 0 \text{ and } E(\varepsilon_{in}^2) = \sigma^2. \tag{4}$$

[6]That there must be such a global component for the estimation approach in the cross-industry cross-country literature to make sense was already pointed out in Rajan and Zingales (1998), p. 563.

The variance across industries of the technological industry characteristics z_{in} in (2) is $Var(z_{in}) = Var(z_i) + \sigma^2$ for all countries n. Hence, larger values of σ^2 imply that more of the heterogeneity in technological industry characteristics is country specific. The assumption that $Var(z_i) > 0$ implies that σ^2 is strictly smaller than the variance across industries of the technological industry characteristics in each country, $\sigma^2 < Var(z_{in})$ for all n. This is because $Var(z_i) > 0$ and (2) imply that at least some of the variance across industries of technological characteristics in each country reflects a global component.

If $\sigma^2 = 0$, the variance across industries of the technological industry characteristics is entirely driven by the global component in each country. As a result, there is no cross-country heterogeneity in technological differences between industries. This is because in this case, $\varepsilon_{in} = 0$ for all i, n and hence (2) implies that $z_{in} - z_{jn} = z_i - z_j$. Hence, technological differences between industries $z_{in} - z_{jn}$ do not vary at all across countries n. Because our model for z_{in} in (2) allows for a country-specific component z_n, the levels of technological industry characteristics could still vary across countries. But such cross-country heterogeneity does not play an important role in our analysis, as it is absorbed by the country fixed effects always present in cross-industry cross-country models.

If $\sigma^2 > 0$, there is cross-country heterogeneity in technological differences between industries. To see this, note that (2) implies $z_{in} - z_{jn} = (z_i - z_j) + (\varepsilon_{in} - \varepsilon_{jn})$ and generically $\varepsilon_{in} - \varepsilon_{jn} \neq \varepsilon_{im} - \varepsilon_{jm}$ for all $n \neq m$. To see the implications of this heterogeneity, it is useful to relate $z_{in} - z_{jn}$ for country n and any pair of industries i and j to differences in US industry characteristics, $z_{iUS} - z_{jUS}$, and differences in the global industry component, $z_i - z_j$. This yields

$$z_{in} - z_{jn} = \rho_{nUS}(z_{iUS} - z_{jUS}) + (1 - \rho_{nUS})(z_i - z_j) + u_{ijnUS} \tag{5}$$

where ρ_{nUS} refers to the correlation in (3) between the specific industry characteristics of country n and the US, and u_{ijnUS} is a random variable with $E(u_{ijnUS}) = 0$ that is independent of the z_i and the z_{iUS}.[7] Hence, in expectation:

(i) The difference between the technological characteristics of any two industries in country n can be thought of as a weighted average of industry differences in the US and industry differences in the global component.

(ii) The weight on the technological industry characteristics of the US is the correlation coefficient ρ_{nUS} between the specific industry characteristics of country n and the US.

As the coefficients ρ_{nUS} in (5) can be arbitrary functions of country characteristics, our model of technological industry characteristics allows for a flexible relationship between the

[7]This holds for any pair of countries n and m. It follows from (2)–(4) and joint normality of the distribution of ε_{in} for all i and n.

x-characteristics of countries and their technological similarity with the US.

It is useful to see what the model in (5) allows us to capture in the context of Rajan and Zingales (1998) and of Nunn (2007) for example. In Nunn, (5) allows us to capture that—even if all countries had the US level of institutional quality—the technological relationship-specific-input intensity of a high human capital country like the US could be different from countries with low human capital. As the coefficients ρ_{nUS} in (5) can be arbitrary functions of country characteristics, they allow us to capture in a flexible way that how similar countries are in terms of the technological relationship-specific-input intensity of their industries may depend on human capital—that is, $\rho_{nUS} = g(h_n, h_{US})$ for some function $g(\cdot)$.[8] The estimation approach in the cross-industry cross-country literature fails to take this into account and estimates of the effect of institutional quality on industry outcomes could therefore be biased upwards or downwards. Our analysis of the bias of the estimation approach in the cross-industry cross-country literature will show that when the US is used as a benchmark country, the bias depends on how ρ_{nUS} changes as country n and the US become more dissimilar in terms of their x-characteristics. In the context of Nunn's study, where the x-characteristic is institutional quality, the bias would therefore depend how ρ_{nUS} changes as country n and the US become more dissimilar in terms of their institutional quality. If $\rho_{nUS} = g(h_n, h_{US})$, this depends on both the effect of countries' human capital on the technological relationship-specific-input intensity of their industries and on whether countries with more dissimilar institutional quality are also more dissimilar in human capital.

In Rajan and Zingales (1998), the key industry-level variable is the external-finance intensity of industries. Rajan and Zingales use this variable to capture technological shocks that raise an industry's investment opportunities beyond what internal funds could support. As the benchmark country in Rajan and Zingales is the US, the external-finance intensity of industries used in their empirical analysis is that of US industries. The technological shocks affecting US industries could be similar to shocks affecting industries in countries with similar human capital, for example, but quite different from shocks affecting industries in countries with low human capital. As the coefficients ρ_{nUS} in (5) can be arbitrary functions of country characteristics, our framework allows us to capture flexibly that technological shocks in countries with low human capital could be quite different from technological shocks in high human capital countries like the US. The estimation approach in the cross-industry cross-country literature does not take this into account and can therefore yield upwards or downwards biased estimates of the effect of financial development on industry outcomes. If the technological shocks affecting industries vary with the human capital of countries, the bias depends on both the relationship between human capital and technological shocks and

[8]As an aside, a country's human capital could also affect industry outcomes through the technological human-capital intensity of industries of course.

on whether countries with more dissimilar financial development are also more dissimilar in human capital.

It is interesting to note that (5) does not determine whether the difference between the technological characteristics of any two industries in country n increases or decreases relative to the US as ρ_{nUS} increases. The answer depends on whether the difference between the global component of technological industry characteristics, $z_i - z_j$, is greater or smaller than the difference between the technological industry characteristics of the US, $z_{iUS} - z_{jUS}$. This gives the model additional flexibility that seems desirable. For example, consider the effect of a country's human capital on the relationship-specific-input intensity of its industries discussed above in the context of Nunn's (2007) study. Compared to the US, industries might be less relation-specific-input intensive in countries with low human capital. However, there seems no reason to suppose that this effect is stronger in some industries than others. Hence, the difference in the use of relation-specific inputs between industries i and j in countries with low human capital may be greater or smaller than in the US.

We could also model the endogeneity of US industry characteristics by extending our conceptual framework to include a fictional frictionless (ff) country. This would allow us to write US industry characteristics in terms of industry characteristics in the fictional frictionless country, $z_{iUS} - z_{jUS} = \rho_{USff}(z_{iff} - z_{jff}) + (1 - \rho_{USff})(z_i - z_j) + u_{ijUSff}$ where the derivation and variable definitions are analogous to (5). This results in a flexible model of how US industry characteristics—and, using (5), the characteristics of any other country—differ compared to a frictionless baseline as ρ_{USff} is allowed to be an arbitrary function of US characteristics.

3.2 Characterizing the Standard Benchmarking Estimator

We now apply the estimation approach used in the cross-industry cross-country literature to the model in (1) and (2). This yields what we refer to as the *standard benchmarking estimator*. We then discuss the forces shaping the bias of this estimator.

3.2.1 Deriving the Standard Benchmarking Estimator

The estimating equation in the cross-industry cross-country literature is

$$y_{in} = a_i + a_n + bx_n z_{iUS} + residual_{in} \tag{6}$$

where a_i and a_n are industry and country fixed effects, and z_{iUS} denotes the industry characteristics of the benchmark country (we use the subscript *US* as the benchmark country is usually the US). The effect of interest is captured by the coefficient b on the industry-country interaction, and the method of estimation is least squares.[9]

It is useful to write the least-squares estimator of b in (6) in terms of demeaned variables

$$\widehat{b} = \frac{\frac{1}{N}\frac{1}{I}\sum_{n=1}^{N}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})(x_n - \overline{x})\left(y_{in} - \overline{y}_n - \overline{y}_i + \overline{y}\right)}{\frac{1}{N}\frac{1}{I}\sum_{n=1}^{N}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})^2(x_n - \overline{x})^2} \tag{7}$$

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

To see when the standard benchmarking estimator identifies the main parameter of interest β, we consider the probability limit of $\widehat{b}$ as the number of industries goes to infinity. Substituting (1) in (7) and taking the probability limit—see the Appendix for details—yields

$$b = \operatorname*{plim}_{I\to\infty}\widehat{b} = \left(1 - \frac{\sigma^2}{\sigma^2_{US}}\right)\beta + \left(\frac{\sigma^2}{\sigma^2_{US}}\right)(\alpha A + \beta B) \tag{8}$$

where σ^2 is the variance of ε_{in} and σ^2_{US} is the variance of the US industry characteristic z_{iUS}, with $\sigma^2/\sigma^2_{US} < 1$; α captures direct effects of industry characteristics on industry outcomes; and A and B capture the relationship between the characteristic x_n of country n and how similar the country is technologically to the US (as measured by ρ_{nUS})

$$A = \frac{Cov(x_n, \rho_{nUS})}{Var(x_n)} = \frac{\sum_{n=1}^{N}(x_n - \overline{x})\rho_{nUS}}{\sum_{n=1}^{N}(x_n - \overline{x})^2} \tag{9}$$

and

$$B = \frac{Cov(x_n, \rho_{nUS}x_n)}{Var(x_n)} = \frac{\sum_{n=1}^{N}(x_n - \overline{x})x_n\rho_{nUS}}{\sum_{n=1}^{N}(x_n - \overline{x})^2}. \tag{10}$$

For example, suppose that the US is a high-x country, i.e. the US has a high level of financial development, institutional quality, or human capital. Then A is positive if countries that are

[9]We assume x_n to be exogenous. In some applications in the literature, exogeneity is an issue and x_n is therefore instrumented. In these applications, our analysis applies to the reduced-form equation. We always include the US (benchmark country) as one of the countries in our analysis. The literature sometimes drops the benchmark country but, given the relatively large number of countries included, this generally makes very little difference for the estimates.

similar technologically to the US are also similar to the US in terms of the x-characteristic. In the typical application of cross-industry cross-country models in the literature, B would also tend to be positive in this case.[10]

An immediate implication of (8) is that the standard benchmarking estimator identifies β when there is no cross-country heterogeneity in technological industry characteristics, $\sigma^2 = 0$. In this case, the technological differences between US (benchmark country) industries are identical to the technological differences between industries of all other countries. Using US industry characteristics as a proxy for the technological industry characteristics of all other countries does therefore not involve any measurement error.[11]

When there is cross-country heterogeneity in technological industry characteristics, $\sigma^2 > 0$, the standard benchmarking estimator in (8) is biased and the bias is shaped by two main forces. First, how much country-specific heterogeneity there is in technological industry characteristics (captured by σ^2/σ_{US}^2). Second, how the technological similarity of countries with the US (captured by ρ_{nUS}) covaries with their characteristics x_n (captured by A and B). We now discuss the forces shaping the bias in some interesting special cases and show that the standard benchmarking estimator may be biased towards zero (attenuated); biased away from zero (amplified); or entirely spurious.

3.2.2 The Bias of the Standard Benchmarking Estimator: a First Approach

The expressions in (8)–(10) allow us to discuss the forces shaping the bias of the standard benchmarking estimator and illustrate three main types of biases.

Attenuation Bias. We start with the case that we see as corresponding to the implicit assumption in the cross-industry cross-country literature. In this case, differences between the technological industry characteristics of a country and global technological industry characteristics are assumed to be completely idiosyncratic to the country. Put differently, the technological industry characteristics of different countries are related through global industry characteristics only, i.e. $\rho_{nm} = 0$ for all country pairs $n \neq m$.

In this case, (9) and (10) imply $A = B = 0$ and the expression for the standard benchmarking estimator in (8) simplifies to $b = \beta(1 - \sigma^2/\sigma_{US}^2)$. As already mentioned, the assumption $Var(z_i) > 0$ implies $\sigma^2/\sigma_{US}^2 < 1$ as at least some of the variation in technological industry characteristics in each country, including the US, is due to the global component.

[10] Theoretically, the sign of B could depend on the distribution of the x-characteristics across countries even if A is positive.

[11] As already mentioned, our model for z_{in} in (2) allows for a country-specific component z_n and the levels of technological industry characteristics could therefore vary across countries even if $\sigma^2 = 0$. But such cross-country heterogeneity does not play an important role in our analysis, as it is absorbed by the country fixed effects always present in cross-industry cross-country models.

Hence, the standard benchmarking estimator b has the same sign as the parameter of interest β but is biased towards zero. This possibility is generally understood in the cross-industry cross-country literature and explained in terms of a classical measurement error bias due to US (benchmark country) industry characteristics measuring the technological industry characteristics of other countries with some error (e.g. Rajan and Zingales, 1998, p. 567). Intuitively, $\rho_{nm} = 0$ for all $n \neq m$ implies that US industry characteristics are an equally imperfect proxy for the technological industry characteristics of all other countries. US industry characteristics become a uniformly worse proxy for the technological industry characteristics of other countries for larger values of σ^2/σ^2_{US}. As a result, the attenuation bias is stronger the greater the country-specific component of technological industry characteristics.

Spurious Interaction Effect. When there is cross-country heterogeneity in technological industry characteristics, the standard benchmarking estimator can indicate a positive effect of the country characteristic x_n on industry outcomes even though x_n does not actually enter the true model at all. To see this, suppose that $\beta = 0$, which implies that the country characteristic x_n drops out from the true model in (1). Suppose also that there is cross-country heterogeneity in technological industry characteristics, $\sigma^2 > 0$. In this case, the standard benchmarking estimator in (8) is $b = \alpha A \sigma^2/\sigma^2_{US}$. Hence, if $\alpha A > 0$, the standard benchmarking estimator indicates a positive effect of the industry-country interaction $x_n z_{iUS}$ on industry outcomes, although the country characteristic is in fact irrelevant for industry outcomes. This is because $\alpha A > 0$ implies that cross-country heterogeneity in technology is such that industry outcomes in high-x countries are more closely correlated with US industry characteristics than industry outcomes in low-x countries.[12] The standard benchmarking estimator misinterprets this as a positive effect of the industry-country interaction $x_n z_{iUS}$ on industry outcomes, and therefore leads to the erroneous conclusion that the country characteristic x_n has an effect on industry outcomes.[13]

The size of the spurious effect generated by the standard benchmarking estimator depends on A in (9). A is the slope of a least-squares regression of ρ_{nUS}, which measures technological similarity of country n with the US, on the x-characteristic of countries. As a result, the bias

[12]This could be because the technological industry characteristics of high-x countries are more similar to US industry characteristics and there is a positive direct effect of technological industry characteristics on industry outcomes ($A > 0$ and $\alpha > 0$). Alternatively, technological industry characteristics of high-x countries could be less similar to US industry characteristics and there could be a negative direct effect of technological industry characteristics on industry outcomes ($A < 0$ and $\alpha < 0$).

[13]More formally, when $\beta = 0$, the benchmarking estimator solely reflects the covariation between the direct effect of country-specific industry characteristics on industry outcomes $\alpha\varepsilon_{in}$ and the interaction $x_n z_{iUS}$. This covariation is $\alpha \frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})E\varepsilon_{in}(z_{iUS} - z_i) = \alpha \frac{1}{N}\sum_{i=1}^{N}(x_n - \overline{x})\sigma^2\rho_{nUS} = \alpha\sigma^2 Cov(x_n, \rho_{nUS})$ where we made use of the definition of ρ_{nUS}. Hence, as long as there is cross-country heterogeneity in technological industry characteristics, the covariation is positive if and only if $\alpha Cov(x_n, \rho_{nUS}) > 0$. Using the definition of A, this is equivalent to $\alpha A > 0$.

of the standard benchmarking estimator could be sizable although countries that are similar to the US in the x-characteristic are also similar technologically, if there is a drop-off in technological similarity with the US as countries become less similar in the x-characteristic. In fact, if (i) countries similar to the US in the x-characteristic are also similar technologically; (ii) countries are on average similar to the US in the x-characteristic; and (iii) there is a drop-off in technological similarity as countries become less similar to the US in the x-characteristic, then the bias of the standard benchmarking estimator can be sizable although the average country is technologically quite similar to the US.

Amplification Bias. The standard benchmarking estimator can also result in an amplification bias. To see this in the simplest case, assume there is no direct effect of industry characteristics on outcomes, $\alpha = 0$. In this case, (8) simplifies to $b = \beta\,[1 + (B-1)\sigma^2/\sigma^2_{US}]$. Hence, if $B > 1$ and there is cross-country heterogeneity in technological industry characteristics ($\sigma^2 > 0$), the standard benchmarking estimator b will be an amplified version of β, $|b| > |\beta|$ and $sign(b) = sign(\beta)$.

The amplification bias of the standard benchmarking estimator is the most difficult bias to understand intuitively. At the most general level, for there to be an amplification bias, US industry characteristics must be a better proxy for the technological industry characteristics of some countries than others. Specifically, US industry characteristics must be a better proxy for the technological industry characteristics of countries that have x-characteristics similar to the US. In our framework, this is the case if countries that are more similar technologically to the US are also more similar in terms of their x-characteristics.

To see the sources of the amplification bias of the standard benchmarking estimator formally, it is useful to rewrite the model in (1) as

$$y_{in} = \gamma_n z_{in} + \nu_{in} \tag{11}$$

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

where we continue to assume $\alpha = 0$. We simplify further by treating the disturbance ν_{in} as an independent and identically distributed random variable. The parameters γ_n in (11) capture the effect of industry characteristics on outcomes in different countries. We refer to these parameters as country-specific slopes. The parameter β in (12) captures how these country-specific slopes covary with the country characteristic x_n.

Now imagine estimating the country-specific slopes γ_n in (11) separately for each country. As we only observe the technological industry characteristics of the US, we use US industry characteristics z_{iUS} as a proxy for the technological industry characteristics z_{in} of each country. We denote the least-squares slope estimates of γ_n by $\widehat{g}_n$. Clearly, $\widehat{g}_n$ will generally be biased. To see the factors shaping the bias we take the probability limit of $\widehat{g}_n$ as the number of industries I goes to infinity. This yields

$$g_n = \operatorname*{plim}_{I\to\infty} \widehat{g}_n = \gamma_n \left[\left(1 - \frac{\sigma^2}{\sigma^2_{US}} \right) + \left(\frac{\sigma^2}{\sigma^2_{US}} \right) \rho_{nUS} \right] \tag{13}$$

where $\sigma^2/\sigma^2_{US} < 1$. The term in square brackets turns out to be the correlation coefficient between the technological industry characteristics of country n and the US, $corr(z_{in}, z_{iUS})$. Hence, the bias of the least-squares slopes, $g_n - \gamma_n$, reflects the technological similarity between country n and the US as captured by $corr(z_{in}, z_{iUS})$. This yields two insights: (i) the more similar a country is technologically to the US (the closer $corr(z_{in}, z_{iUS})$ to 1), the smaller the bias of the least-squares slopes in (13); and (ii) the least-squares slopes in (13) are biased towards zero (attenuated) for all countries n, as long as the technological industry characteristics of all countries are positively correlated with those of the US ($corr(z_{in}, z_{iUS}) \geq 0$ for all n). Hence, as long as $corr(z_{in}, z_{iUS}) \geq 0$ for all countries n, the term in square brackets in (13) can be thought of as the so-called attenuation factor in the classical measurement error literature. This attenuation factor is larger—and hence the attenuation bias is smaller—for countries that are more similar technologically to the US.

That the country-specific least-squares slope estimates in (13) might be attenuated for all countries is not difficult to understand from the perspective of the classical measurement error literature, as US industry characteristics will generally proxy for industry characteristics of other countries with error. It is harder to see why, if all the slope estimates in (13) are attenuated, the standard benchmarking estimator may be subject to an amplification bias. This is possible because the attenuation bias of the least-squares slope estimates is heterogeneous across countries, with a smaller attenuation bias for countries that are more similar technologically to the US.

To see this point, it is useful to express the standard benchmarking estimator in (8) as a slope of slopes. We start from the least-squares slopes g_n in (13) obtained by regressing outcomes across industries on US industry characteristics separately for each country n. These country-specific slopes g_n are then regressed on the country characteristics x_n across countries. The least-squares slope of the second, cross-country regression is the standard benchmarking estimator in (8). To see this, note that

$$\frac{\sum_{n=1}^{N} g_n(x_n-\overline{x})}{\sum_{n=1}^{N}(x_n-\overline{x})^2} = \beta\left(\frac{\sum_{n=1}^{N}\left[\left(1-\frac{\sigma^2}{\sigma^2_{US}}\right)+\left(\frac{\sigma^2}{\sigma^2_{US}}\right)\rho_{nUS}\right]\gamma_n(x_n-\overline{x})}{\sum_{n=1}^{N}(x_n-\overline{x})^2}\right) \tag{14}$$

$$= \beta\left[\left(1-\frac{\sigma^2}{\sigma^2_{US}}\right)+\left(\frac{\sigma^2}{\sigma^2_{US}}\right)B\right] = b.$$

The left-most expression in (14) is the standard expression for the slope of a least-squares regression, in this case of g_n on x_n. The first equality follows from substituting the least-squares slopes in (13) for g_n. The second equality uses (12) and the definition of B in (10), and the last equality uses the expression for b in (8) for the case $\alpha = 0$. The key message of the slope-of-slopes expression for the standard benchmarking estimator in (14) is that the bias of the estimator reflects how the attenuation factor of the country-specific least-squares slopes in (13) covaries with the country characteristics x_n. The amplification bias can emerge when the attenuation factor (bias) is larger (smaller) for countries with greater x_n.

We now illustrate the amplification bias in the simplest version of our framework.

The Amplification Bias in the Simplest Setting. The source of the amplification bias emerges most clearly when there are two different groups of countries and countries in the same group are identical. In this two-group setting, the formula for the benchmarking estimator in (14) simplifies to

$$b = \frac{g_S - g_D}{x_S - x_D} \tag{15}$$

where g_S and g_D are the country-specific slope estimates in (13) for countries in group S and group D, and x_S and x_D are the x-characteristics of countries in the two groups.

Now suppose that the US is part of group S. As countries in the same group are identical, this implies that all countries n in group S are identical technologically to the US, $\rho_{nUS} = 1$. As a result, (13) implies that the estimated country slopes and the true country slopes are the same for all countries in group S: $g_S = \gamma_S$. This is unsurprising as using US technological industry characteristics as a proxy for the technological industry characteristics of other countries in group S does not involve any measurement error.

On the other hand, suppose that countries in group D are technologically somewhat different from the US. The simplest approach is to think of these countries as having specific industry characteristics that are uncorrelated with US-specific industry characteristics. That is, $\rho_{nUS} = 0$ for all countries n in group D. If there is some cross-country heterogeneity in technological industry characteristics ($\sigma^2 > 0$), (13) implies that the estimated country slopes for all countries in group D are an attenuated version of the true slopes: $g_D = (1-\sigma^2/\sigma_{US}^2)\gamma_D$. This is because US industry characteristics are a noisy proxy for the technological industry characteristics of countries in group D.

Substituting the expressions for g_S and g_D we just obtained into (15) and using (12) yields

$$b = \beta \left[1 + \left(\frac{\sigma^2}{\sigma_{US}^2}\right) \frac{x_D}{x_S - x_D}\right]. \tag{16}$$

Hence, there will be an amplification bias, $|b| > |\beta|$ and $sign(b) = sign(\beta)$, if $x_S > x_D > 0$. The bias can be large if the two groups of countries have very similar x-characteristics. 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: The amplification bias in the simplest possible case.

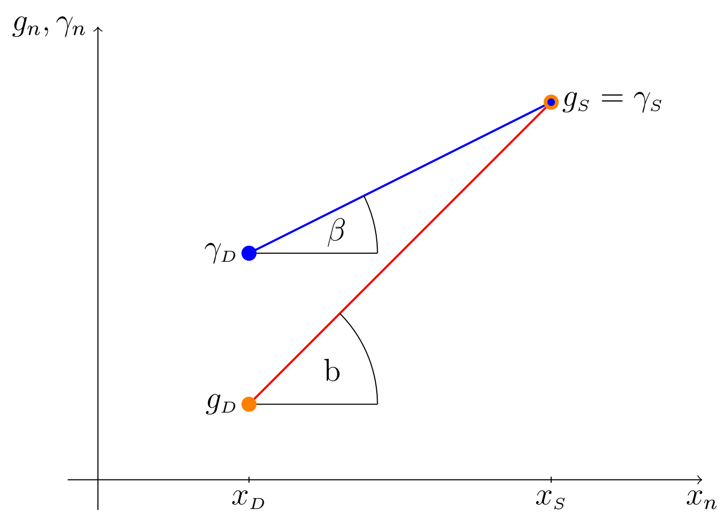


Figure 1 is a graphical illustration of the amplification bias in the two-group setting for $\beta > 0$. The two blue dots plot the true country-specific slopes γ_S and γ_D against x_S and x_D. The parameter β we want to estimate is the slope of the blue line connecting the two blue dots as (12) implies $\beta = (\gamma_S - \gamma_D)/(x_S - x_D)$. The two red dots plot the least-squares slope estimates g_S and g_D against x_S and x_D. Equation (15) implies that the benchmarking estimator b is the slope of the red line connecting the two red dots, $b = (g_S - g_D)/(x_S - x_D)$. The amplification bias $b > \beta > 0$ emerges because:

(i) Countries in group S with high x-values have the same technological industry characteristics as the US, and US industry characteristics are therefore a perfect proxy for the industry characteristics of all high-x countries. Hence, there is no measurement error when the US is used to proxy for the industry characteristics of these countries. This implies that the least-squares slope estimates for these countries are equal to the true slopes, $g_S = \gamma_S$. That is, the blue and the red dot lie on top of each other.

(ii) Countries in group D with low x-values have technological industry characteristics that are somewhat different from those of the US, and US industry characteristics therefore proxy for the technological industry characteristics of all low-x countries with some error. Hence, the least-squares slopes estimates g_D for these countries underestimates the true slopes, $g_D < \gamma_D$. That is, the red dot lies below the blue dot.

Hence, cross-country heterogeneity in technological industry characteristics implies that using the US industry proxy yields a consistent estimate of γ_S for high-x countries that are technologically identical to the US, but a downwards biased estimate of γ_D for low-x countries that are technologically different from the US. Because the standard benchmarking estimator b is the slope of the red line connecting the red dots while the parameter of interest β is the slope of the blue line connecting the blue dots, this leads to an amplification bias, $0 < \beta < b$. More generally, the amplification bias of the standard benchmarking estimator arises when greater technological similarity between high-x countries and the US leads to a sufficiently smaller attenuation bias for the country-specific slope estimates of high-x countries.

It is interesting to note that the size of the amplification bias in the two-group example does not depend on the relative number of countries in the two groups. But the more countries there are in group S with high x-values relative to group D with low x-values, the more similar the average country becomes technologically to the US (the benchmark country). Hence, the amplification bias could be sizable although the average country is quite similar technologically to the US.

3.2.3 The Bias of the Standard Benchmarking Estimator: the General Case

To characterize the bias of the standard benchmarking estimator more generally, it is useful to distinguish the case $\beta = 0$ and the case $\beta \neq 0$.

If $\beta = 0$, (8) simplifies to $b = \alpha A \sigma^2/\sigma^2_{US}$ with $\sigma^2/\sigma^2_{US} < 1$. Hence, with cross-country heterogeneity in technological industry characteristics, $\sigma^2 > 0$, the standard benchmarking estimator is biased upwards if $\alpha A > 0$ and is biased downwards if $\alpha A < 0$.

If $\beta \neq 0$, the standard benchmarking estimator in (8) can be written as

$$b = \beta \left[\left(1 - \frac{\sigma^2}{\sigma^2_{US}}\right) + \left(\frac{\sigma^2}{\sigma^2_{US}}\right) \delta \right] \tag{17}$$

where $\sigma^2/\sigma^2_{US} < 1$ and δ is a function of A and B in (9)–(10)

$$\delta = \theta A + B \tag{18}$$

with

$$\theta = \frac{\alpha}{\beta}. \tag{19}$$

Hence, when there is cross-country heterogeneity in technological industry characteristics, $\sigma^2 > 0$, the bias of the standard benchmarking estimator depends on δ. If $\delta = 0$, the standard benchmarking estimator is attenuated. For example, our framework yields $\delta = 0$ when country-specific industry characteristics are uncorrelated across countries. If $\delta > 0$, there is a countervailing force that can weaken the attenuation bias or result in an amplification bias. If $\delta < 0$, the standard benchmarking estimates may have the wrong sign.

We now summarize how the bias of the standard benchmarking estimator depends on δ.

Proposition 1. [Bias of standard benchmarking estimator when $\boldsymbol{\beta} \neq \mathbf{0}$]

1. If $0 \leq \delta \leq 1$, the standard benchmarking estimator is subject to an attenuation bias: b has the same sign as β but is biased towards zero, $sign(b) = sign(\beta)$ and $|b| \leq |\beta|$.

2. If $\delta > 1$, the standard benchmarking estimator is subject to an amplification bias: b has the same sign as β but is biased away from zero, $sign(b) = sign(\beta)$ and $|b| > |\beta|$.

3. If $\delta < 0$, the standard benchmarking estimator may be subject to an attenuation bias, an amplification bias, or may have a different sign than β, depending on σ^2/σ^2_{US}.

4 Identification of β

We have seen that the standard benchmarking estimator used in the cross-industry cross-country literature does not identify the effect of interest when there is cross-country heterogeneity in technological industry characteristics ($\sigma^2 > 0$). Moreover, the bias cannot be signed if technologically similar countries are similar in terms of other characteristics ($A \neq 0$ or $B \neq 0$). We now examine how the effect of interest can be identified when there is cross-country heterogeneity in technological industry characteristics.

To get a first idea how the effect of interest might be identified and where the challenges lie, we return to the expression for the benchmarking estimator in (17). Inverting it yields $\beta = b/[1 + (\delta - 1)\sigma^2/\sigma^2_{US}]$. The right-hand-side parameter b can be identified using the standard benchmarking approach in the literature, and the variance of the US industry characteristics σ^2_{US} is observable. If we can identify δ and σ^2, we can therefore identify β. As we will show, δ can be identified from the variances and covariances of industry outcomes for different country pairs. If these variances and covariances would also identify the variance of country-specific industry characteristics σ^2, identification of β would be straightforward. But the variances and covariances of industry outcomes do not identify σ^2.

To see how the variances and covariances of industry outcomes for different country pairs help to identify β, we rewrite the model in (1) as

$$y_{in} = v_i + v_n + \gamma_i x_n + u_{in} \tag{20}$$

where

$$\gamma_i = \beta z_i \tag{21}$$

and

$$u_{in} = (\alpha + \beta x_n)\varepsilon_{in} \tag{22}$$

and v_i and v_n denote industry and country fixed effects.[14] The industry-specific slopes γ_i capture the effect of the country characteristic on outcomes in different industries.

The effect of (unobservable) country-specific technological industry characteristics ε_{in} on industry outcomes is captured by u_{in} in (22). $E(u_{in}u_{im})$, the variances and covariances of u_{in} for industry i and countries n, m, reflect the effect of cross-country heterogeneity in technological industry characteristics on the variances and covariances of industry outcomes in countries n, m. As a result, they play a central role for the identification of δ and β.

To see this, note that (3) and (22) imply that the variances and covariances $E(u_{in}u_{im})$ are

$$E(u_{in}u_{im}) = (\alpha\sigma + \beta\sigma x_n)(\alpha\sigma + \beta\sigma x_m)\rho_{nm} = \omega_{nm}. \tag{23}$$

That the ω_{nm} may allow us to identify δ is quite straightforward. From (18) and A and B in (9)–(10), it can be seen that δ depends on the ρ_{nm}, which capture how similar any two countries are technologically, and on α/β, which captures the direct effect of technological industry characteristics on industry outcomes relative to the industry-country-interaction effect. We should be able to infer these parameters entering δ from the ω_{nm} under some conditions as according to (23), the ω_{nm} depend on the ρ_{nm}; on $\alpha\sigma$, which captures the direct effect of country-specific heterogeneity in technological industry characteristics on industry outcomes; and on $\beta\sigma$, which captures the industry-country-interaction effect of country-specific heterogeneity in technological industry characteristics on industry outcomes.

However, it can also be seen that the ω_{nm} in (23) will not allow us to identify the variance of country-specific industry characteristics σ^2. This is because the ω_{nm} solely reflect σ^2 through its effects on outcomes, which is why σ only appears multiplied by either α or β. This is what makes the identification of β challenging.

To see when and how β can be identified, we now proceed in two steps. We first examine the identification of β for known ω_{nm}. Then we discuss how the ω_{nm} can be identified.

[14]These industry and country fixed effects capture the industry and country fixed effects in v_{in} and absorb αz_i in the industry fixed effect and z_n in the country fixed effect.

4.1 Identification of β for Known $\mathbf{\Omega}$

It is convenient to collect the variances and covariances ω_{nm} in (23) for all countries n, m in the $N \times N$ variance-covariance matrix $\mathbf{\Omega}$. The straightforward part of identification of β for known $\mathbf{\Omega}$ is determining whether or not $\beta = 0$. The elements on the diagonal of $\mathbf{\Omega}$ are equal to $\omega_{nn} = (\alpha\sigma + \beta\sigma x_n)^2$ for all countries n. As long as there is some cross-country heterogeneity in technological industry characteristics, $\sigma^2 > 0$, the ω_{nn} are independent of country characteristics if and only if $\beta = 0$. Hence, we obtain that $\beta = 0$ if the ω_{nn} are independent of x_n. On the other hand, $\beta \neq 0$ if the ω_{nn} depend on x_n.

The next question is how to identify β if the ω_{nn} depend on the country characteristics x_n. We first explain how $\mathbf{\Omega}$ can be used to obtain two key parameters for the identification of β, namely δ and $(\beta\sigma)^2$. Then we show how δ and $(\beta\sigma)^2$ can be used to identify β.

Obtaining δ and $(\beta\sigma)^2$ from $\mathbf{\Omega}$ is simple. We start by determining $\alpha\sigma$ and $\beta\sigma$—and hence $(\beta\sigma)^2$—from the variances $\omega_{nn} = (\alpha\sigma + \beta\sigma x_n)^2$. This is possible if there are at least two countries with different x-values, so that we have at least two equations in the two unknowns $\alpha\sigma$ and $\beta\sigma$.[15] Then we invert the expression for the covariances ω_{nm} for $n \neq m$ in (23) to get $\rho_{nm} = \omega_{nm}/[(\alpha\sigma + \beta\sigma x_n)(\alpha\sigma + \beta\sigma x_m)]$. This allows us to obtain the ρ_{nm} by combining the ω_{nm} with $\alpha\sigma$ and $\beta\sigma$. Once we have obtained $\alpha\sigma$, $\beta\sigma$, and the ρ_{nm}, it is straightforward to obtain A and B in (9)–(10), θ in (19), and hence $\delta = \theta A + B$ in (18).

To see when and how δ and $(\beta\sigma)^2$ obtained from $\mathbf{\Omega}$ allow us to identify β, we start from the expression for the bias of the standard benchmarking estimator $b - \beta = \beta(\delta - 1)\sigma^2/\sigma^2_{US}$ obtained by rearranging (17). Multiplying both sides of this equation by β yields $(b - \beta)\beta = (\delta - 1)(\beta\sigma)^2/\sigma^2_{US}$. The right-hand side parameters δ and $(\beta\sigma)^2$ can be obtained from $\mathbf{\Omega}$, and σ^2_{US} is the observable variance of US industry characteristics. The parameter b is identified by the standard benchmarking approach in the literature. Hence, β is the only unknown of the quadratic equation

$$(b - \beta)\beta = \eta(\delta - 1) \tag{24}$$

where we defined

$$\eta = \frac{(\beta\sigma)^2}{\sigma^2_{US}}. \tag{25}$$

This establishes a key result: β is one of the solutions for q of the quadratic equation

$$(b - q)q = \eta(\delta - 1). \tag{26}$$

Generally, the quadratic equation in (26) has two solutions. In addition to the solution $q_1 = \beta$, there is a second solution $q_2 = \beta(\delta - 1)\sigma^2/\sigma^2_{US}$. We therefore need to analyze when

[15] There is no gain of using more than two ω_{nn} equations as additional equations leave results unchanged. When we use our identification results for estimation, we use all ω_{nn} equations of course.

we can determine which of the two solutions for q in (26) identifies β.

We start with the simplest case, which is when δ is positive and smaller than 2. In this case $\sigma^2/\sigma^2_{US} < 1$ implies $(\delta-1)\sigma^2/\sigma^2_{US} \in (-1, 1)$. As the two solutions for q in (26) are $q_1 = \beta$ and $q_2 = \beta(\delta-1)\sigma^2/\sigma^2_{US}$, this yields that β can be identified as the solution for q in (26) with the larger absolute value, $\beta = \max(|q_1|, |q_2|)$.

This is the simplest expression for β when δ is positive and smaller than 2. But the expression does not generalize to other cases where β is exactly identified. An alternative expression that holds for all cases where β is exactly identified is $\beta = \kappa b$, where b is the standard benchmarking estimator and κ is a function of the two solutions for q in (26)

$$\kappa = \max\left(\frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2}\right). \tag{27}$$

The next proposition, which is proven in the Appendix, summarizes this result.

Proposition 2. [Identifying β: sufficient condition in terms of identifiable δ] If $\delta \in [0, 2]$, β can be identified as $\beta = \kappa b$ where b is the probability limit of the standard benchmarking estimator and κ is defined in (27).

The next proposition gives a necessary and sufficient condition for the exact identification of β for known $\mathbf{\Omega}$.

Proposition 3. [Identifying β: necessary and sufficient condition in terms of identifiable δ and κ]
The effect of interest β can be exactly identified if and only if

$$\begin{array}{llll} \text{either} & \delta \geq 0 & \text{and} & \kappa \geq \frac{\delta-1}{\delta} \\ \text{or} & \delta < 0 & \text{and} & \kappa \leq \frac{\delta-1}{\delta} \end{array} \tag{28}$$

where δ is defined in (18) and κ is defined in (27). If this condition is not satisfied, β is equal to one of the two solutions for q in (26), but it cannot be determined which.

When β is exactly identified, it can be obtained as

$$\beta = \kappa b \tag{29}$$

where b is the probability limit of the standard benchmarking estimator.

The proposition is proven in the Appendix. The idea is the following. The two solutions for q of the quadratic equation in (26) yield two candidate solutions for β. Each of these two candidate solutions can be combined with the variance of US industry characteristics and the identifiable parameter η in (25) to yield two candidate solutions for the country-specific technological heterogeneity parameter σ^2. As at least some of the variation in technological industry characteristics reflects a global component, it must be that $0 \leq \sigma^2 < \sigma^2_{US}$. It turns out that this restriction is only satisfied by one of the two candidate solutions for σ^2 if the condition in (28) holds. Hence, only one of the two candidate solutions for β is consistent with the model and this solution is $\beta = \kappa b$. On the other hand, if the condition in (28) fails, both candidate solutions for β imply candidate solutions for σ^2 that are positive and smaller than σ^2_{US}. As a result, both candidate solutions are consistent with the model and it is impossible to say which of the two solutions of (26) identifies β.

The necessary and sufficient condition in Proposition 3 is not easily interpreted in terms of the parameters of the underlying model. The next proposition gives the necessary and sufficient condition for identification in terms of σ^2 and δ.

Proposition 4. [Identifying β: necessary and sufficient condition in terms of model parameters] β can be exactly identified if and only if

$$(\delta - 1)^2 \left(\frac{\sigma^2}{\sigma^2_{US}}\right) \leq 1. \tag{30}$$

If this condition is not satisfied, β is one of the two solutions for q in (26), but it cannot be determined which.

Intuitively, Proposition 4 implies that β can be identified exactly if cross-country heterogeneity in technological industry characteristics is not too large (σ^2/σ^2_{US} not too large) and/or if the association between countries' technological similarity with the US and their x-characteristics is not too strong (δ not too large in absolute value). On the other hand, if there is substantial cross-country heterogeneity in technological industry characteristics and/or countries' technological similarity with the US is strongly associated with their x-characteristics, it cannot be established which of the two solutions of (26) identifies β.

When exact identification of β is impossible, one could report both solutions for q in (26) as possible values for β. An alternative is to establish bounds on β in terms of the standard benchmarking estimator b. For $\delta > 2$, we have already established upper and lower bounds in Proposition 1. The next proposition establishes somewhat tighter bounds under the condition that $\delta > 2$ and that exact identification of β is impossible. For completeness, the proposition also gives bounds for the case $\delta < 0$ even though these are less useful. The proof of the proposition is in the Appendix.

Proposition 5. [Bounds on β] If the condition in (28) does not hold and exact identification of β is impossible, then

$$\begin{aligned}
&\text{if } \delta > 2 \text{ then } \frac{\beta}{b} \in \left(\frac{1}{\delta}, \frac{\delta - 1}{\delta}\right) \\
&\text{if } \delta < 0 \text{ then } \frac{\beta}{b} \notin \left[\frac{1}{\delta}, \frac{\delta - 1}{\delta}\right].
\end{aligned} \tag{31}$$

For example, suppose that $\delta = 2.5$, b is positive, and (28) does not hold. In this case Proposition 5 implies that β is between $0.4b$ and $0.6b$. Hence, we can infer the range and the sign of the parameter of interest β from the standard benchmark estimator b. As another example, suppose that $\delta = -2.5$, b is positive, and (28) does not hold. Proposition 5 then implies that β is smaller than $-0.4b$ or larger than $0.6b$. Hence, we cannot establish an upper or lower bound for β, nor can we infer the sign of β from the sign of b.

4.2 Identification of $\mathbf{\Omega}$

Now that we have shown when and how β can be identified for known variance-covariance matrix $\mathbf{\Omega}$, we turn to the identification of $\mathbf{\Omega}$. Our approach is closely related to the identification of variance-covariance matrices in general least squares theory. In particular, the first step to identify $\mathbf{\Omega}$ consists of least-squares estimation and the second step involves understanding when and how the least-squares residuals can be used to identify $\mathbf{\Omega}$.[16]

The starting point to identify $\mathbf{\Omega}$ is least-squares estimation of the model in (20). The least-squares residuals $\widehat{u}_{in} = y_{in} - \widehat{v}_i - \widehat{v}_n - \widehat{\gamma}_i x_n$, with hats denoting least-squares estimates, allow us to estimate $\frac{1}{I}\sum_{i=1}^{I} \widehat{u}_{in}\widehat{u}_{im}$ for all country pairs n, m. These estimated variances and covariances depend on the ω_{nm} we collected in the variance-covariance matrix $\mathbf{\Omega}$ and can therefore be used to identify $\mathbf{\Omega}$ under some conditions.

Relating $\mathbf{\Omega}$ to the variances and covariances of the residuals across industries. We now derive the relationship between the variances and covariances across industries of the residuals $\widehat{u}_{in}$ for all pairs of countries n, m, $\frac{1}{I}\sum_{i=1}^{I} \widehat{u}_{in}\widehat{u}_{im}$, and the elements ω_{nm} of $\mathbf{\Omega}$. The first step is to express the least-squares residuals $\widehat{u}_{in}$ in terms of the underlying disturbances u_{in} in (20)

$$\widehat{u}_{in} = \upsilon_{in} - (x_n - \overline{x}) \sum_{k=1}^{N} \psi_k \upsilon_{ik} \tag{32}$$

[16]The main difference with GLS analysis is that we are interested in the variance-covariance matrix of the disturbances in (20) but not other model parameters, like the industry slopes for example. GLS analysis is generally interested in variance-covariance matrices because of their role in the efficient estimation of other model parameters. The reason we are not interested in the other model parameters in (20) is that they not help to identify β. For example, while the industry slopes depend on β, they also depend on the unobservable variation in the global component of technological industry characteristics.

where the v_{in} are the demeaned versions of u_{in}

$$v_{in} = u_{in} - \frac{1}{N}\sum_{m=1}^{N} u_{im} - \frac{1}{I}\sum_{j=1}^{I} u_{jn} + \frac{1}{N}\frac{1}{I}\sum_{m=1}^{N}\sum_{j=1}^{I} u_{jm} \tag{33}$$

and the ψ_k are the least-squares regression weights

$$\psi_k = \frac{x_k - \overline{x}}{\sum_{p=1}^{N}(x_p - \overline{x})^2}. \tag{34}$$

The second step is to calculate the probability limit as the number of industries goes to infinity of the variances and covariances of the residuals across industries for all country pairs, which we refer to as π_{nm}

$$\pi_{nm} = \operatorname*{plim}_{I\to\infty} \frac{1}{I}\sum_{i=1}^{I} \widehat{u}_{in}\widehat{u}_{im}. \tag{35}$$

We show in the Appendix that using (32)-(33) in (35) yields the following equations linking π_{nm} and the elements ω_{nm} of $\mathbf{\Omega}$

$$\pi_{nm} = \omega_{nm} - \mu_n - \mu_m - (x_n - \overline{x})\lambda_m - (x_m - \overline{x})\lambda_n \tag{36}$$

where μ_n and λ_n are functions of the ω_{nm} detailed in the Appendix and

$$0 = \sum_{n=1}^{N} \lambda_n. \tag{37}$$

These equations are the basis for the identification of the variance-covariance matrix $\mathbf{\Omega}$ from the least-squares residuals of (20)-(22).

A structure for $\mathbf{\Omega}$. It is well understood that the identification of the variance-covariance matrix $\mathbf{\Omega}$ is impossible for an arbitrary matrix as (36) and (37) has more unknowns than linearly independent equations (e.g., Amemiya, 1985).[17] For identification to be possible, the empirical framework must put some structure on $\mathbf{\Omega}$. The structures used in the literature depend on the application (e.g., Amemiya, 1985; Wooldridge, 2002; Conley, 2010).

We chose a structure for $\mathbf{\Omega}$ that has the implicit structure in the cross-industry cross-country literature as a special case but allows for substantial deviations from this baseline. The implicit structure for $\mathbf{\Omega}$ in the cross-industry cross-country literature is that differences

[17] We show this in the Appendix.

between the technological industry characteristics of a country and global technological industry characteristics are completely idionsyncratic to each country. This implies that the technological industry characteristics of different countries are related through the global component only. Or put differently, the country-specific technological industry characteristics ε_{in} in (2) for any pair of countries $n \neq m$ are uncorrelated, i.e. $\rho_{nm} = 0$. As we have seen above, and as acknowledged in the cross-industry cross-country literature, the standard benchmarking estimator is attenuated in this case.

We chose a structure for $\mathbf{\Omega}$ that follows the cross-industry cross-country literature in that the technological industry characteristics of *some country pairs* are related through the global component only. I.e. for some country pairs $n \neq m$, $\rho_{nm} = 0$. But for all other country pairs, we allow for an entirely arbitrary correlation ρ_{nm} between the country-specific technological industry characteristics.

Specifically, our structure for $\mathbf{\Omega}$:

(i) Allows for an arbitrary correlation ρ_{nm} with $n \neq m$ between the country-specific technological industry characteristics of two countries if they are sufficiently similar. Two countries are taken to be sufficiently similar if the distance between their x-characteristics is below a threshold τ. When we set large values for the threshold τ, many country pairs satisfy $|x_n - x_m| \leq \tau$, and our structure for $\mathbf{\Omega}$ therefore allows for arbitrary correlations ρ_{nm} between the country-specific technological industry characteristics of many country pairs. Formally, for these country pairs, technological similarity as measured by $corr(z_{in}, z_{im})$ is $[Var(z_i) + \sigma^2 \rho_{nm}]/(Var(z_i) + \sigma^2)$. Hence, the technological industry characteristics of these country pairs are *not* assumed to be related through the global component only and can be related in arbitrary ways to all country characteristics.

(ii) When the distance between the x-characteristics of a country pair exceeds the threshold τ, their country-specific industry characteristics are taken to be uncorrelated, $\rho_{nm} = 0$.[18] $\rho_{nm} = 0$ implies that the technological industry characteristics of these country pairs *are* related through the global technological component only, as implicitly assumed for all country pairs in the cross-industry cross-country literature. Formally, technological similarity as measured by $corr(z_{in}, z_{im})$ for country pairs with $\rho_{nm} = 0$ is $Var(z_i)/(Var(z_i) + \sigma^2)$. By increasing τ, we can reduce the number of country pairs with $\rho_{nm} = 0$ and therefore deviate substantially from the implicit assumption in the cross-industry cross-country literature that $\rho_{nm} = 0$ for all country pairs $n \neq m$.

We refer to this structure for the variance-covariance matrix $\mathbf{\Omega}$ as $\mathbf{\Omega}^{\tau}$ to capture that it depends on the threshold τ. What makes this structure for $\mathbf{\Omega}$ interesting in our context is that

[18] The approach can be thought of as a cross-country analogue of so-called K-dependence in time-series econometrics, which allows for any correlation between random variables at t and T if $|t - T| \leq \tau$ but assumes independence if $|t - T| > \tau$ (e.g., Amemiya, 1985).

it corresponds to the implicit structure in the cross-industry cross-country literature for $\tau = 0$. We can move away from this baseline quite continuously and substantially by increasing τ. Moreover, the structure does not impose any functional form on how the technological similarity of country pairs with $|x_n - x_m| \leq \tau$ depends on country characteristics.

The size of the threshold τ must be interpreted relative to the distribution of the x-characteristic across countries. It is therefore often easier to think about the fraction of unrestricted ρ_{nm} with $n \neq m$ implied by a threshold τ. For example, when τ is chosen very small, the fraction of unrestricted ρ_{nm} will be small as few country pairs will satisfy $|x_n - x_m| \leq \tau$. As a result, the assumed structure for $\mathbf{\Omega}$ will be similar to the implicit structure in the cross-industry cross-country literature. On the other hand, when the threshold τ is chosen large, the fraction of unrestricted ρ_{nm} will be large as many country pairs will satisfy $|x_n - x_m| \leq \tau$. As a result, the structure for $\mathbf{\Omega}$ can deviate quite substantially from the implicit structure in the cross-industry cross-country literature. (If the threshold τ is chosen so large that all country pairs with $n \neq m$ can have different ρ_{nm}, we are not imposing any structure on the variance-covariance matrix $\mathbf{\Omega}$ and identification is impossible.)

As the choice is difficult in practice, we vary the threshold τ over the whole range that permits identification of $\mathbf{\Omega}$. Put differently, we allow the fraction of unrestricted ρ_{nm} with $n \neq m$ to vary between zero and the maximum that still permits identification of $\mathbf{\Omega}$. As this maximum can be surprisingly large, our structure for $\mathbf{\Omega}$ can deviate substantially from the implicit structure in the cross-industry cross-country. In some cases, $\mathbf{\Omega}$ can be identified for values of τ that leave 90% of the ρ_{nm} unrestricted. This amounts to little structure being put on the cross-country heterogeneity in technological industry characteristics. By varying the fraction of the unrestricted ρ_{nm} between zero and the maximum that permits identification, we can examine how sensitive the results for β are to the restrictions put on $\mathbf{\Omega}$.

Of course, other, more parsimonious structures for $\mathbf{\Omega}$ could be chosen (and would generally be simpler to deal with). For example, the structures used in spatial econometrics for spatial dependence could be adapted to capture the technological similarity of countries as a function of their x-characteristics and other country characteristics (e.g., Conley, 2010).

Summarizing, we assume that if countries have sufficiently similar x-characteristics $|x_n - x_m| < \tau$, ρ_{nm} with $n \neq m$ is unrestricted. On the other hand, $\rho_{nm} = 0$ if $|x_n - x_m| \geq \tau$. The threshold τ is set by us and we present results for the largest possible range allowing for the identification of $\mathbf{\Omega}$. Larger τ translate into a greater fraction of ρ_{nm} that are unrestricted.

A condition for identification of $\mathbf{\Omega}$. The structure $\mathbf{\Omega}^\tau$ for the variance-covariance matrix $\mathbf{\Omega}$ assumes $\rho_{nm} = 0$ and hence $\omega_{nm} = 0$ in (23) for all country pairs with relatively different x-characteristics, $|x_n - x_m| \geq \tau$. We denote the number of such country pairs by

Q. For these Q country pairs, (36) simplifies to

$$\pi_{nm} = -\mu_n^\tau - \mu_m^\tau - (x_n - \overline{x})\lambda_m^\tau - (x_m - \overline{x})\lambda_n^\tau. \tag{38}$$

These equations are the starting point for the identification of $\mathbf{\Omega}^\tau$ from the π_{nm}. In particular, we use these equations to try and determine the μ_n^τ and λ_n^τ for all n. Then we use (36) to determine the ω_{nm}^τ for all other country pairs.

To take the first step and determine μ_n^τ and λ_n^τ, it is useful write the Q equations in (38) and the restriction in (37) in normal form

$$\boldsymbol{\pi} = \mathbf{G}^\tau \begin{pmatrix} \boldsymbol{\mu}^\tau \\ \boldsymbol{\lambda}^\tau \end{pmatrix} \tag{39}$$

where $\boldsymbol{\mu}^\tau = (\mu_1^\tau, \ldots, \mu_N^\tau)'$ and $\boldsymbol{\lambda}^\tau = (\lambda_1^\tau, \ldots, \lambda_N^\tau)'$ collect the $2N$ unknowns; $\boldsymbol{\pi}$ is a column vector of length $Q+1$ that collects the values on the left-hand side of equations (37) and (38); and $\mathbf{G}^\tau$ is a $(Q+1) \times 2N$ matrix of coefficients implied by the right-hand side of equations (37) and (38). By writing the equations in (37) and (38) in normal form, it becomes clear that $\boldsymbol{\mu}^\tau$ and $\boldsymbol{\lambda}^\tau$ can be determined if the matrix $\mathbf{G}^\tau$ has full rank.

An illustration of the identification condition. We can identify the variance-covariance matrix $\mathbf{\Omega}^\tau$ if the matrix $\mathbf{G}^\tau$ has full rank. This depends on the distance threshold τ and the distribution of the x-values across countries.

Table 1 illustrates this for three types of distributions for the x-values across countries. For each distribution, we draw x-values for 150 countries.[19] We repeat this 300 times. For each draw we calculate the value for the maximum threshold τ such that $\mathbf{G}^\tau$ has full rank for all smaller τ. We refer to this value as $\tau_{\max}$. As this value is somewhat difficult to interpret, we do two things to put it into perspective:

(i) We calculate the average distance $|x_n - x_m|$ across all possible country pairs for each draw. This allows comparing $\tau_{\max}$ with the average distance in the x-characteristics across all country pairs and get a sense whether $\tau_{\max}$ is relatively large or small.

(ii) We calculate the number of countries with unrestricted ρ_{nm} with $n \neq m$ that are implied by $\tau_{\max}$. We then report this number relative to the total number of country pairs. For example, if this ratio is 0.8, the ρ_{nm} are unrestricted for 80% of all country pairs.

Table 1 reports these statistics averaged across the 300 draws we take. We first present results for the case where the country characteristics are uniformly distributed between 0

[19] This is approximately the number of countries in our application of the new benchmarking estimator below. We obtain very similar results for 75, 250, and 500 countries.

Table 1: Identification of the variance-covariance matrix.

Distribution	Average distance between x_n across all country pairs	Maximum threshold τ allowing identification ($\tau_{\max}$)	Country pairs $n \neq m$ with unrestricted ρ_{nm} relative to total number of country pairs at $\tau_{\max}$
Uniform on $[0, 1]$	0.33	0.49	0.74
Standard normal	1.13	2.58	0.93
Exp. with $\lambda = 1$	1.01	2.55	0.91

and 1. The distance $|x_n - x_m|$ averaged across all country pairs is 0.33. The maximum value of the distance threshold τ that still permits identification ($\tau_{\max}$) is 0.49. The number of country pairs with unrestricted ρ_{nm} with $n \neq m$ relative to the total number of country pairs at $\tau_{\max}$ is 74%. The statistics in the last two columns remain nearly unchanged when we vary the support of the uniform distribution (not in the table).

As a second illustration, Table 1 shows results for the case where the country characteristics are drawn from a normal distribution with mean 0 and a standard deviation of 1. The distance $|x_n - x_m|$ averaged across all country pairs is 1.13. $\tau_{\max}$ is 2.42. The number of country pairs with unrestricted ρ_{nm} with $n \neq m$ relative to the total number of country pairs at $\tau_{\max}$ is 93%. The statistics in the last two columns do not vary with the mean of the normal distribution and remain nearly unchanged when we vary the standard deviation (not in the table). The third illustration in Table 1 is for the exponential distribution and yields results similar to the normal distribution.

5 An Application

We now apply our identification results. We start by explaining how to go from identification to estimation. Then we use the approach to reestimate Nunn (2007).

5.1 From Identification to Estimation

We first explain how our identification results can be used to obtain consistent estimates of q in (26) in five steps. Once we have estimated the solutions for q, we estimate β using Proposition 2 or Proposition 3, or obtain bounds for β using Proposition 5.

Step 1: Estimate (20) with least squares and then use the residuals to estimate the variances and covariances across industries of the residuals for all country pairs

$$\widehat{\pi}_{nm} = \frac{1}{I}\sum_{i=1}^{I} \widehat{u}_{in}\widehat{u}_{im}. \tag{40}$$

These variances and covariances are consistent estimators of the π_{nm} in (35) as the number of industries I goes to infinity.

Step 2: Estimate $\boldsymbol{\mu}^{\tau}$ and $\boldsymbol{\lambda}^{\tau}$ on the basis of (39). We start by obtaining the matrix $\mathbf{G}^{\tau}$ for different distance cutoffs τ. We begin with very small values of τ. If all countries have different x-characteristics (as in our application below), this implies that the $\rho_{nm} = 0$ condition is imposed for all country pairs $n \neq m$ and that $\boldsymbol{\Omega}$ is a diagonal matrix (as implicitly assumed in the cross-industry cross-country literature). The implied matrix $\mathbf{G}^{\tau}$ is of full rank. We then increase τ up to the maximum value still yielding a matrix $\mathbf{G}^{\tau}$ of full rank. To estimate $\boldsymbol{\mu}^{\tau}$ and $\boldsymbol{\lambda}^{\tau}$ on the basis of (39), we also need an estimator of the column vector $\boldsymbol{\pi}$. We obtain this estimator by replacing the π_{nm} collected in the vector $\boldsymbol{\pi}$ with the estimates $\widehat{\boldsymbol{\pi}}_{nm}$ in (40). Of course, we cannot estimate $\boldsymbol{\mu}^{\tau}$ and $\boldsymbol{\lambda}^{\tau}$ by simply replacing $\boldsymbol{\pi}$ with $\widehat{\boldsymbol{\pi}}$ in (39). This is because generally $\widehat{\boldsymbol{\pi}} \neq \boldsymbol{\pi}$ due to sampling error and the equation system in (39) would therefore be overdetermined. Instead, $\boldsymbol{\mu}^{\tau}$ and $\boldsymbol{\lambda}^{\tau}$ are estimated by applying least squares to

$$\widehat{\boldsymbol{\pi}} = \mathbf{G}^{\tau}\begin{pmatrix} \boldsymbol{\mu}^{\tau} \\ \boldsymbol{\lambda}^{\tau} \end{pmatrix} + \mathbf{v} \tag{41}$$

where $\mathbf{v}$ is a column vector of length $Q+1$ that captures the sampling error $\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}$. Because $\widehat{\boldsymbol{\pi}}$ is a consistent estimator of $\boldsymbol{\pi}$ as the number of industries I goes to infinity, the least-squares estimators $\widehat{\boldsymbol{\mu}}^{\tau}$ and $\widehat{\boldsymbol{\lambda}}^{\tau}$ are consistent estimators of $\boldsymbol{\mu}^{\tau}$ and $\boldsymbol{\lambda}^{\tau}$.

Step 3: Estimate the non-zero elements ω_{nm}^{τ} of $\boldsymbol{\Omega}^{\tau}$ by combining (36) with $\widehat{\boldsymbol{\mu}}^{\tau}$, $\widehat{\boldsymbol{\lambda}}^{\tau}$, and $\widehat{\boldsymbol{\pi}}$. This yields

$$\widehat{\omega}_{nm}^{\tau} = \widehat{\mu}_{n}^{\tau} + \widehat{\mu}_{m}^{\tau} + (x_n - \overline{x})\widehat{\lambda}_{m}^{\tau} + (x_m - \overline{x})\widehat{\lambda}_{n}^{\tau} + \widehat{\pi}_{nm}. \tag{42}$$

Consistency of the $\widehat{\omega}_{nm}^{\tau}$ follows from the consistency of $\widehat{\boldsymbol{\mu}}^{\tau}$, $\widehat{\boldsymbol{\lambda}}^{\tau}$, and $\widehat{\boldsymbol{\pi}}$. The estimates of ω_{nm}^{τ} allow us to estimate θ, $\beta\sigma$, and the nonzero ρ_{nm}. The estimates of θ and $\beta\sigma$ are obtained by combining the expressions for the variances $\omega_{nn} = (\theta + x_n)^2(\beta\sigma)^2$ in (23) with our estimates $\widehat{\omega}_{nn}^{\tau}$. This yields

$$\widehat{\omega}_{nn}^{\tau} = (\theta + x_n)^2(\beta\sigma)^2 + \upsilon_{nn} \tag{43}$$

where υ_{nn} captures sampling error. The nonlinear least-squares estimates of θ and $\beta\sigma$ are then combined with our estimates of the nonzero ω_{nm}^{τ} with $n \neq m$ and the expression for the covariances in (23) to estimate the nonzero ρ^τ_{nm} using that $\rho^\tau_{nm} = \omega^\tau_{nm}/[(\theta + x_n)(\theta + x_m)(\beta\sigma)^2]$. Moreover, our estimate of $\beta\sigma$ can be combined with the variance of the industry characteristics in the benchmark country σ^2_{US} to estimate $\widehat{\eta}$ using (25). Consistency follows from the consistency of the $\widehat{\omega}^{\,\tau}_{nm}$.

Step 4: Use the estimates $\widehat{\rho}^{\,\tau}_{nm}$ to estimate $\widehat{A}^{\,\tau}$ using (9) and $\widehat{B}^{\,\tau}$ using (10). Hence, we have all the elements to estimate $\widehat{\delta}^{\,\tau}$ using (18)

$$\widehat{\delta}^{\,\tau} = \widehat{\theta}\widehat{A}^{\,\tau} + \widehat{B}^{\,\tau}. \tag{44}$$

Step 5: Replace δ and η in (26) by the consistent estimates $\widehat{\delta}^{\,\tau}$ and $\widehat{\eta}$. This allows us to obtain consistent estimates of q by solving

$$(\widehat{b} - q)q = \widehat{\eta}^{\,\tau}(\widehat{\delta}^{\,\tau} - 1) \tag{45}$$

where $\widehat{b}$ is the standard benchmarking estimator.

The estimates of q based on (45) can be used to estimate β as explained in Proposition 2 and Proposition 3 or to obtain bounds on β as explained in Proposition 5. Confidence bands of all our estimates are obtained by bootstrapping.[20]

5.2 Reestimating Nunn (2007)

Nunn employs data on exports for up to 222 industries in up to 146 countries to show that institutional quality has a positive effect on comparative advantage in industries that depend more on relationship-specific intermediate inputs.[21] In terms of the model in (1), the institutional quality of countries takes the place of x_n and their log exports in industry i takes the place of y_{in}. The theoretically relevant industry characteristic z_{in} is the relationship-specific intermediate-input intensity of production. The benchmark country used to obtain proxies for how intensively industries use relationship-specific inputs is the US.

We apply the approach in the previous section to reestimate Nunn's baseline specification without controls and his specification with controls for human and physical capital. We report our estimates of δ and β as a function of the threshold distance τ and the share of unrestricted, and hence estimated, correlation coefficients ρ_{nm}.

[20]Bootstrapping the confidence intervals of our estimates of δ and β involves reshuffling the u_{in} in (20) across industries for each country 300 times and each time reestimating $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$, ω_{nm}, θ, $\beta\sigma$, A, B, ρ_{nm}, δ, η, q_1, q_2, and β. The confidence intervals for our point estimates of δ and β are obtained from the distributions of the estimated δ and β.

[21]See Levchenko (2007) and Costinot (2009) for related empirical and theoretical findings on the effect of institutional quality on comparative advantage.

5.2.1 Results for the Baseline Specification

Figure 2: Share of unrestricted ρ_{nm} with $n \neq m$ as a function of τ for Nunn's baseline specification.

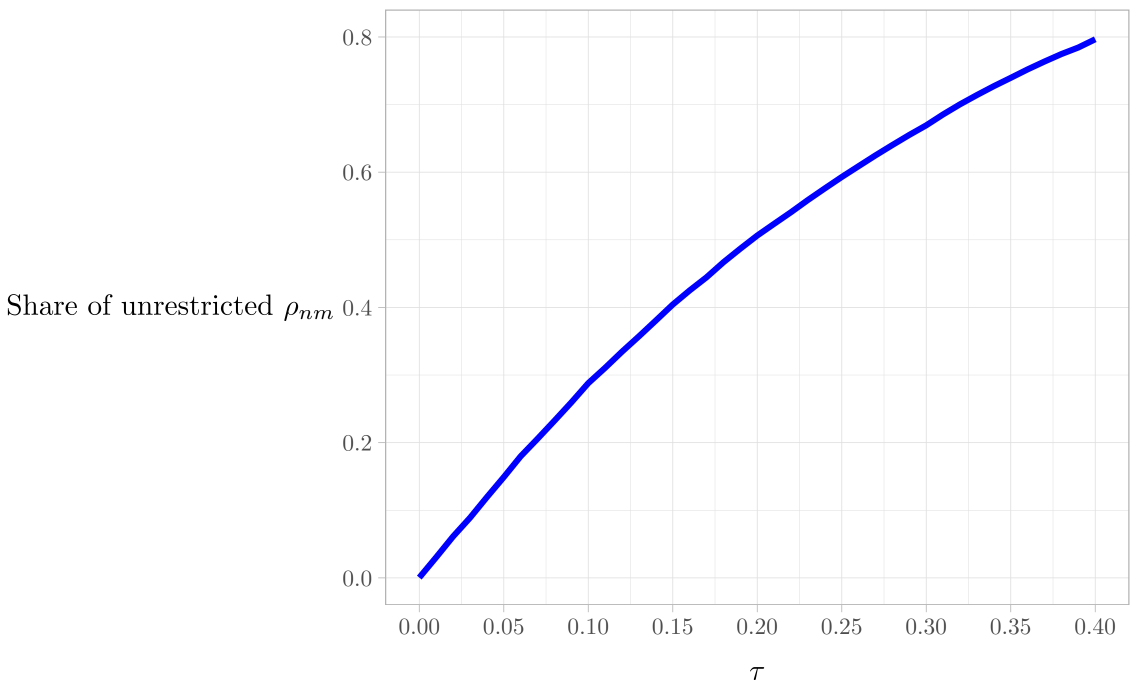
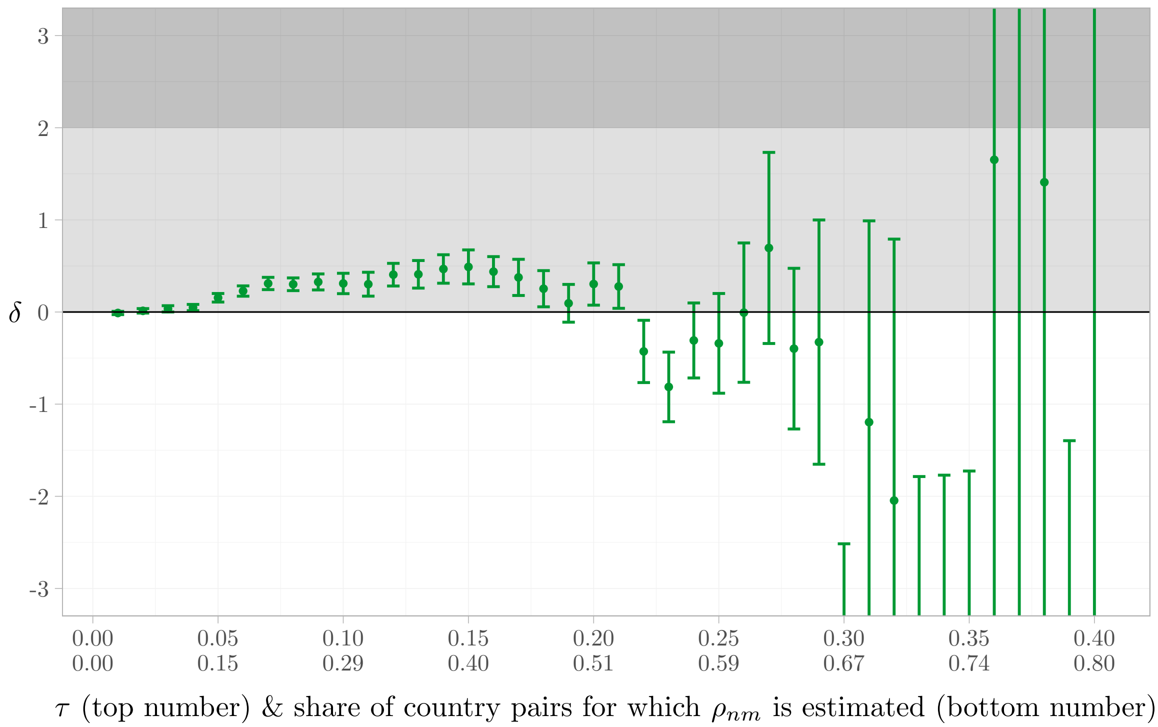


Figure 2 plots the threshold distance τ on the horizontal axis and the number of unrestricted correlation coefficients ρ_{nm} relative to the total number of country pairs on the vertical axis (there are 10,585 country pairs as Nunn has the necessary data for 146 countries). When τ is very small, the condition $\rho_{nm} = 0$ is assumed for all country pairs $n \neq m$ (all countries have different institutional quality in the Nunn data). Hence, the number of unrestricted ρ_{nm} relative to the total number of country pairs is 0. This corresponds to the implicit assumption in the cross-industry cross-country literature. For $\tau = 0.2$, around half the ρ_{nm} with $n \neq m$ are unrestricted and must therefore be estimated. As τ goes to 0.4, 80% of the ρ_{nm} are unrestricted and must be estimated. For values of τ strictly larger than 0.4, $\mathbf{G}^{\tau}$ no longer has full rank and $\boldsymbol{\mu}^{\tau}$ and $\boldsymbol{\lambda}^{\tau}$ cannot be determined. Hence, $\boldsymbol{\Omega}^{\tau}$ cannot be identified.

Figure 3 summarizes our results for δ. Estimates are shown as green dots and 95% confidence intervals are marked as green lines. The area shaded in light grey marks values of δ that according to Proposition 1 result in an attenuation bias of the standard benchmarking estimator. The area shaded in darker grey marks values of δ that according to Proposition 1 result in an amplification bias of the standard benchmarking estimator. For very small values of τ, we obtain $\delta = 0$. This is unsurprising as the condition $\rho_{nm} = 0$ for $n \neq m$ is assumed for all country pairs in this case (as implicitly assumed in the cross-industry cross-country

Figure 3: Estimates of δ for Nunn's baseline specification.

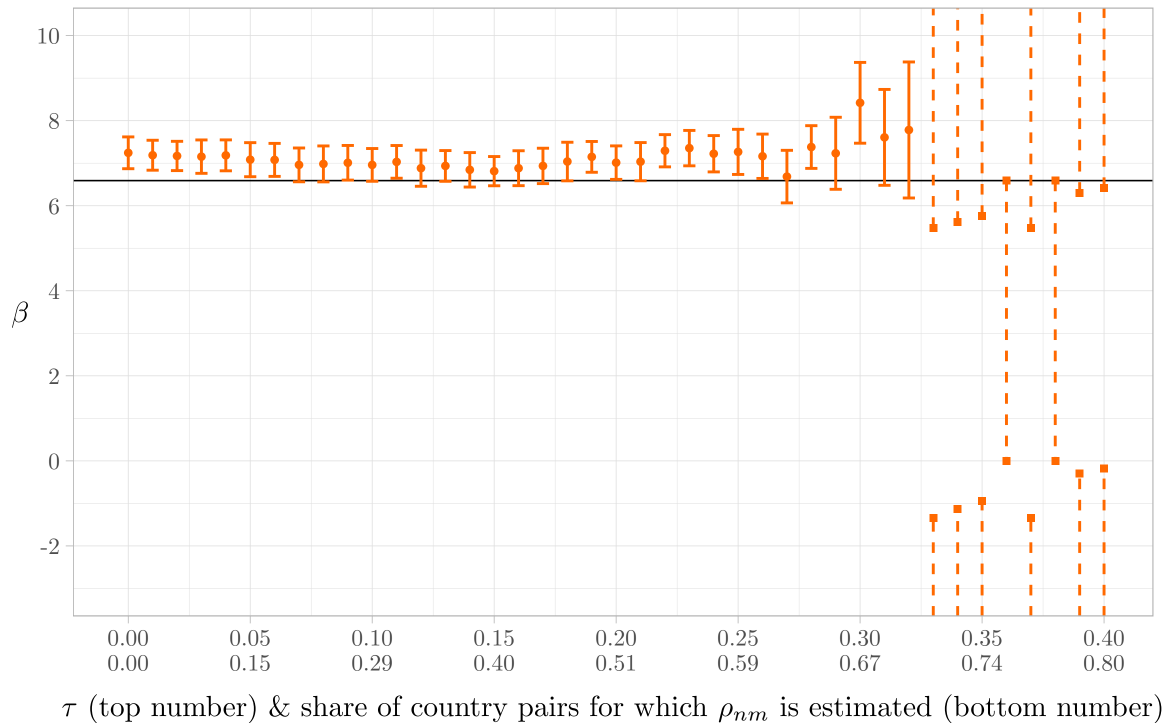


literature). Hence, $A = B = 0$ in (9)–(10) and $\delta = 0$ in (18). Estimates of δ remain very small for values of τ smaller than 0.02. Point estimates are between -0.01 and $+0.01$ and 95% confidence intervals include 0. Hence, we cannot reject $\delta = 0$. According to Proposition 1, $\delta = 0$ implies that the standard benchmarking estimator is subject to an attenuation bias. According to Proposition 2, $\delta = 0$ implies that we can estimate β as $\beta = \kappa b$ with κ given in (27). Figure 4 shows our point estimates for β as orange dots and 95% confidence intervals as orange lines. Point estimates are around 7.2, around 10% larger than Nunn's estimate of 6.6 obtained with the standard benchmarking estimator (marked by the horizontal black line).[22] The 95% confidence intervals of our estimates are between 6.7 and 7.7.

For values of τ between 0.03 and 0.21, point estimates of δ in Figure 3 are between 0.04 and 0.49. The 95% confidence bands are strictly between 0 and 1, except for $\tau = 0.19$. The data therefore support values of δ greater than 0 but below 1. Proposition 1 implies that for $0 \leq \delta \leq 1$, the standard benchmarking estimator is subject to an attenuation bias. Proposition 2 implies that for $0 \leq \delta \leq 1$, we can estimate β as $\beta = \kappa b$. Figure 4 shows our estimates for β. Point estimates are between 7.2 and 6.8. Hence, the difference with Nunn's benchmarking estimate of 6.6 is smaller than what we obtained for very small τ. This is because very small values for τ imply $\delta = 0$, and the bias of the standard benchmarking

[22] The estimate reported by Nunn differs because it is standardized. We report non-standardized estimates throughout.

Figure 4: Estimates of β for Nunn's baseline specification.



estimator is solely shaped by a force generating attenuation in this case. When $\delta > 0$, the bias of the standard benchmarking estimator is also shaped by a force countervailing the attenuation bias. The 95% confidence intervals of our estimates lie between 6.5 and 7.6.

For values of τ between 0.22 and 0.32, estimates of δ in Figure 3 are negative, except for $\tau = 0.27$. As a result, we cannot use Proposition 2 to estimate β. However, we can still estimate β as $\beta = \kappa b$ as our point estimates of κ satisfy the condition for exact identification in Proposition 3.[23] Figure 4 shows our estimates for β. Point estimates are between 6.6 and 8.5 and therefore up to 30% larger than Nunn's standard benchmarking estimate of 6.6 (the horizontal black line). The 95% confidence intervals of our estimates lie between 6 and 9.5.

For values of τ between 0.33 and 0.4, our estimates of δ in Figure 3 become very noisy. The range of 95% confidence intervals varies between 6 and 82 (we do not show the full intervals as this would make the figure unreadable). This likely reflects that for $\tau \geq 0.33$, the correlation coefficients ρ_{nm} with $n \neq m$ of at least 70% of the 10,585 country pairs are being estimated. Point estimates of δ for τ between 0.33 and 0.4 are mostly negative. As the necessary and sufficient condition for exact identification of β in Proposition 3 is not satisfied, we can only establish the bounds in Proposition 5. Figure 4 illustrates the values

[23]For simplicity, we are evaluating the condition in (28) based on the point estimates of δ and κ. A more complete approach would be to test the condition. To do so, note that the two cases in (28) can be combined in a single condition $\delta(\kappa - 1) + 1 \geq 0$. Bootstrapping the 95% confidence levels of the left-hand side of the inequality would allow testing this condition.

of β consistent with these bounds as dashed lines delimited by squares.

Overall, our estimation approach applied to Nunn's baseline specification yields estimates that are close to Nunn's even when we impose little structure on the cross-country heterogeneity in technological industry characteristics, i.e. as many as 70% of the ρ_{nm} with $n \neq m$ are unrestricted and hence estimated. Sometimes this is because the countervailing forces generating an attenuation and amplification bias of the standard benchmarkiong estimator partly offset each other. Our estimates become very noisy and/or exact identification becomes impossible when more than 70% of the ρ_{nm} are unrestricted (the theoretical limit to identification is when 80% of the ρ_{nm} are unrestricted).

5.2.2 Results with Controls for Human and Physical Capital

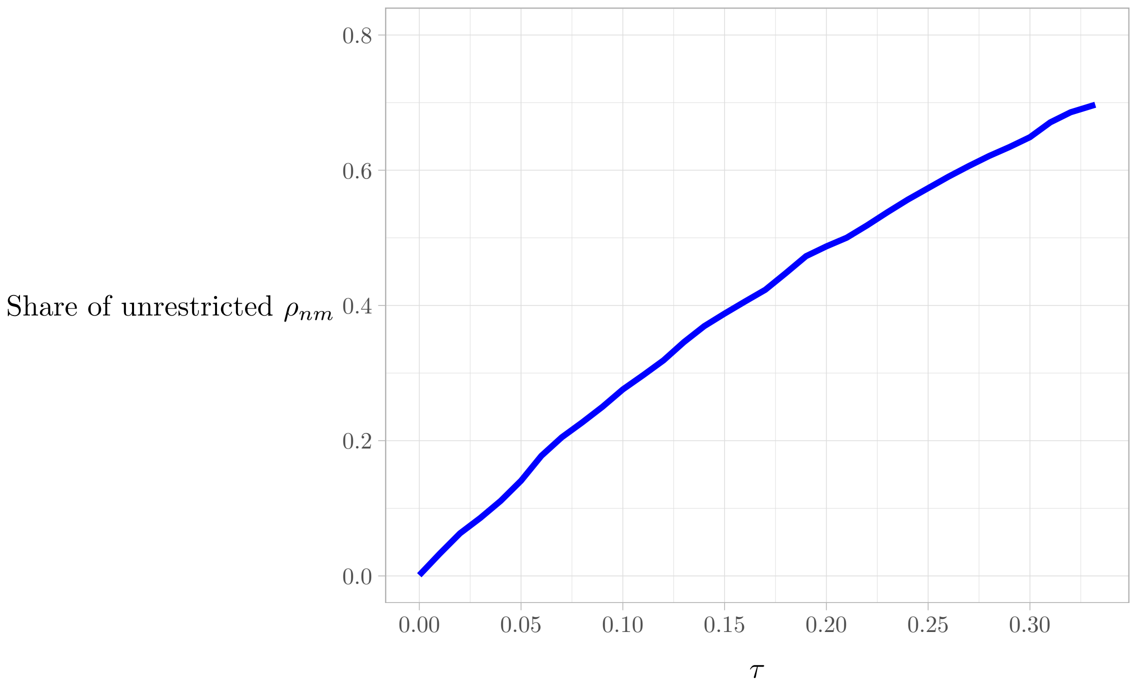
Building on Romalis (2004) and other studies in international trade, Nunn also presents results controlling for the effect of human and physical capital on comparative advantage. He does so by augmenting his baseline specification with an interaction between country-level human capital and the human-capital-intensity of industries as well as an interaction between country-level physical capital and the physical-capital-intensity of industries.

We reestimate Nunn's specification controlling for the effect of human and physical capital on comparative advantage using the new benchmarking estimator. The implementation follows the same steps as for Nunn's baseline specification, except that least-squares estimation of (20) accounts for the effect of human and physical capital following Nunn.

Figure 5 plots the threshold distance τ on the horizontal axis and the number of unrestricted correlation coefficients ρ_{nm} with $n \neq m$ relative to the total number of country pairs on the vertical axis (there are only 2,415 country pairs in this specification, as Nunn has the necessary data for fewer countries). The ratio starts at 0 when τ is very small. This corresponds to the implicit assumption in the cross-industry cross-country literature. For $\tau = 0.2$, around half the ρ_{nm} are unrestricted and must therefore be estimated. As τ goes to 0.31, 70% the ρ_{nm} are unrestricted and must therefore be estimated. For larger values of τ, $\boldsymbol{G}^{\tau}$ no longer has full rank. Hence, $\boldsymbol{\Omega}^{\tau}$ can no longer be identified.

Figure 6 summarizes our results for δ. The area shaded in light grey continues to mark values of δ that according to Proposition 1 result in an attenuation bias of the standard benchmarking estimator. The area shaded in darker grey marks values of δ that according to Proposition 1 result in an amplification bias of the standard benchmarking estimator. For values of τ smaller than 0.02, point estimates of δ are between -0.01 and $+0.01$ and 95% confidence intervals include 0. Hence, we cannot reject $\delta = 0$. According to Proposition 1, $\delta = 0$ implies that the standard benchmarking estimator is attenuated. According to Proposition 2, $\delta = 0$ implies that we can estimate β as $\beta = \kappa b$. This yields estimates of β around 7, see Figure 7. These estimates are about 10% larger than Nunn's point estimate

Figure 5: Share of unrestricted ρ_{nm} with $n \neq m$ as a function of τ for Nunn's specification with controls for human and physical capital.



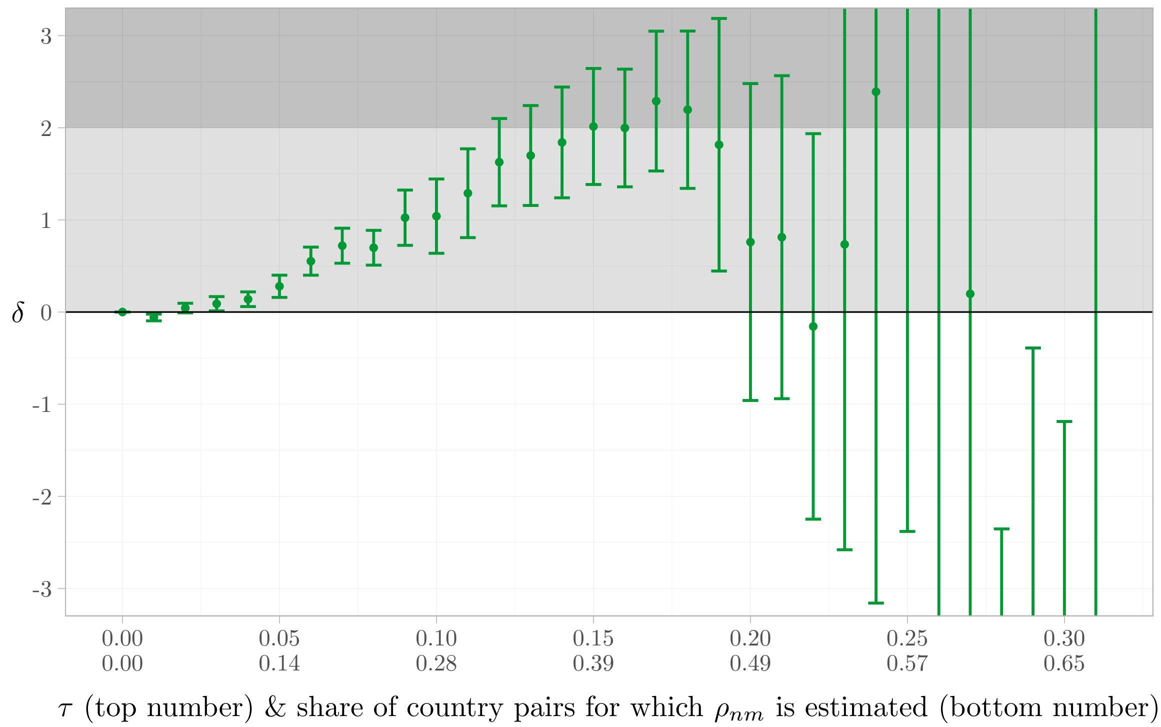
of 6.4 obtained with the standard benchmarking estimator (marked by the horizontal black line). The 95% confidence intervals of our estimates for β lie between 6.4 and 7.5.

For values of the threshold distance τ between 0.03 and 0.11, point estimates of δ in Figure 6 are between 0.09 and 1.3. The 95% confidence intervals are between 0 and 2. Hence, the data support values of δ between 0 and 2. According to Proposition 2, $0 \leq \delta \leq 2$ implies that we can estimate β as $\beta = \kappa b$. This yields the estimates of β in Figure 7. These are sometimes above Nunn's estimate of 6.4 and sometimes below. This makes sense as according to Proposition 1, the standard benchmarking estimator is subject to an attenuation bias when δ is between 0 and 1 and subject to an amplification bias when δ is greater than 1. The 95% confidence intervals of our estimates of β are between 5.7 and 7.5.

When τ is between 0.12 and 0.19, estimates of δ in Figure 6 are between 1.6 and 2.3. Hence, the standard benchmarking estimator is subject to an amplification bias according to Proposition 1. The condition for exact identification in Proposition 3 is always satisfied and we can therefore estimate β as $\beta = \kappa b$. Estimates of β in Figure 7 are between 5.9 and 4.9, up to 25% smaller than Nunn's estimate of 6.4. The 95% confidence intervals of our estimates lie between 6.5 and 4.2.

For values of τ between 0.2 and 0.23, estimates of δ in Figure 6 are generally between 0 and 1. According to Proposition 2, $0 \leq \delta \leq 1$ implies that we can estimate β as $\beta = \kappa b$. Our point estimates of β in Figure 7 are between 6.5 and 6.7, only slightly larger than Nunn's

Figure 6: Estimates of δ for Nunn's specification with controls for human and physical capital.

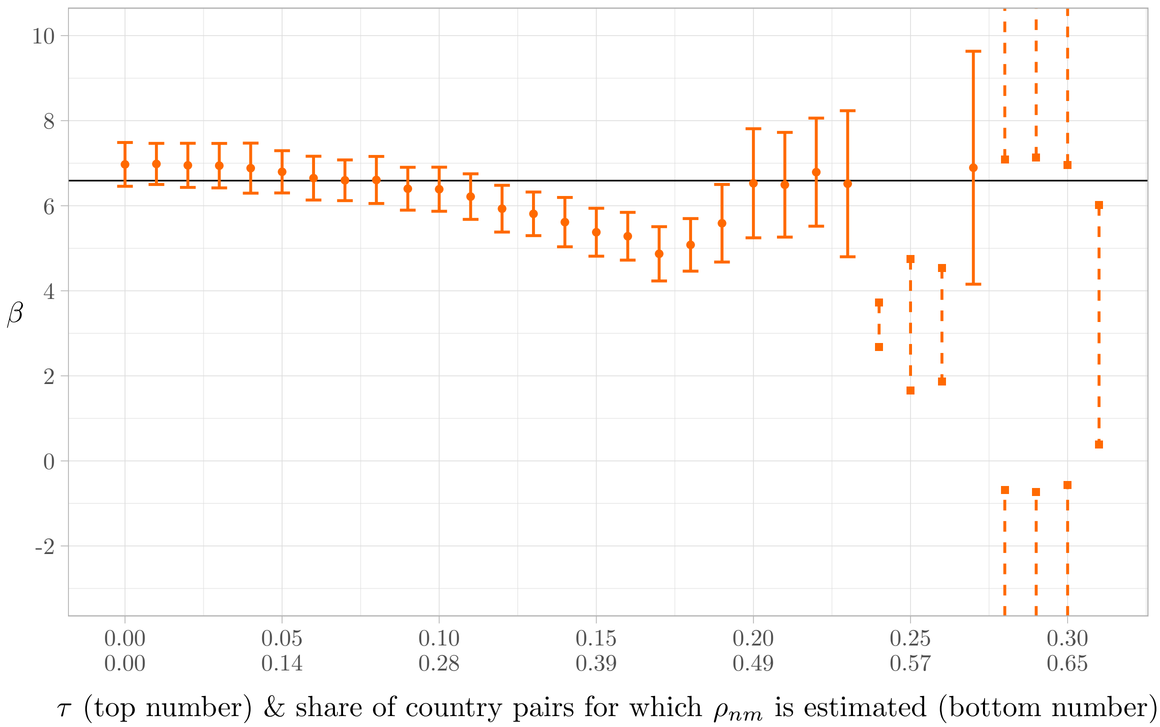


estimate of 6.4. The 95% confidence intervals of our estimates lie between 4.8 and 7.8.

For values of τ between 0.24 and 0.31, our estimates of δ are very noisy. The range of 95% confidence intervals varies between 11 and 42 (we do not show the full intervals as the figure would become unreadable). This likely reflects that we are approaching the limits of identification as for $\tau \geq 0.24$, the correlation coefficients ρ_{nm} of at least 55% of the 2,415 country pairs are being estimated.

Overall, our estimation approach applied to Nunn's specification with controls for human and physical capital yields results that are similar to Nunn's even when as many as 55% of the ρ_{nm} with $n \neq m$ are unrestricted and hence estimated. When there are larger discrepancies between our estimates and those of Nunn, the forces generating an amplification bias of the standard benchmarking estimator dominate those generating an attenuation bias. As a result, our estimates tend to indicate smaller effects than Nunn's. Our estimates become noisy and/or exact identification becomes impossible when more than 55% of the ρ_{nm} are unrestricted (the theoretical limit to identification is when 70% of the ρ_{nm} are left unrestricted).

Figure 7: Estimates of β for Nunn's specification with controls for human and physical capital.



6 Conclusion

Using industry data to examine the economic effects of cross-country differences in financial development, institutional quality, human capital, and other potential determinants of aggregate economic activity is attractive. It permits testing whether the impact of, say, financial underdevelopment or malfunctioning institutions is strongest in the industries where it should be theoretically. This helps bringing empirical work closer to the mechanisms emphasized by economic theory. But the cross-industry cross-country approach is not without pitfalls. As the theoretically relevant technological industry characteristics are unobservable in most countries, they must be proxied by industry characteristics in a benchmark country. That this can lead to an attenuation bias is unsurprising and acknowledged in the literature.

What appears not understood is that the estimation approach used in the cross-industry cross-country literature can also lead to an amplification bias or entirely spurious effects. Amplified or spurious estimates can arise when technologically more similar countries are also more similar in other characteristics. The size of the amplification bias and of spurious effects depends on how much more similar countries are technologically to the benchmark country as they become more similar in other characteristics. As a result, these biases can be sizable if there is a drop-off in technological similarity with the benchmark country as countries become less similar in, for example, their levels of financial development or institutional quality, even if countries are on average quite similar to the benchmark country.

As the estimation approach in the cross-industry cross-country literature does not identify the effect of interest when there is cross-country heterogeneity in technological industry characteristics, it is important to develop alternatives. To do so, we first provided an analysis of identification. We then showed how the new approach to identification can be implemented by reestimating the effect of institutional quality on comparative advantage in industries that rely on relationship-specific inputs in Nunn (2007). Our estimates tend to be similar to Nunn's even when we impose little structure on cross-country heterogeneity in technological industry characteristics.

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

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, Emmanuel Farhi, and Enisse Kharroubi (2015). Liquidity and Growth: The Role of Counter-cyclical Interest Rates. BIS Working Papers No. 489.

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 (2015). Patent Rights, Product Market Reforms, and Innovation. *Journal of Economic Growth*, *20*(3), pp. 223-262.

Aghion, Philippe, Antonin Bergeaud, Gilbert Cette, Remy Lecat, and Helene Maghim (2018). The Counteracting Effects of Credit Constraints on Productivity: Theory and Evidence. Working Paper.

Aizenman, Joshua and Vladyslav Sushko (2011). Capital Flow Types, External Financing Needs, and Industrial Growth: 99 Countries, 1991-2007. NBER Working Paper No. 17228.

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

Alfaro Laura, Paola Conconi, Harald Fadinger, and Andrew F. Newman (2016). Do Prices Determine Vertical Integration? *Review of Economic Studies*, *83*(3), pp. 855–888.

Avdjiev, Stefan, Valentina Bruno, Catherine Koch, and Hyun Song Shin (2018). The Dollar Exchange Rate as a Global Risk Factor: Evidence from Investment. BIS Working Papers No. 695.

Avharya, Viral and Zhaoxia Xu (2017). Financial Dependence and Innovation: The

Case of Public Versus Private Firms. *Journal of Financial Economics*, *124*(2), pp. 223-243.

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

Bassanini, Andrea and Federico Cingano (2019). Before It Gets Better: The Short-Term Employment Costs of Regulatory Reforms. *Industrial Labor Relations Review*, *72*(1), pp. 127-157.

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

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.

Bena, Jan and Ting Xu (2017). Competition and Ownership Structure of Closely Held Firms. *Review of Financial Studies*, *30*(5), pp. 1583-1626.

Bernard, Andrew and Charles I. Jones (1996). Comparing Apples to Oranges: 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: Evidence from the 1985 French Banking Act. *Journal of Finance*, *62*(2), pp. 597-628.

Bernstein, Shai, Josh Lerner, and Filippo Mezzanotti (2017). Private Equity and Financial Fragility during the Crisis. NBER Working Paper No. 23626.

Beverelli, Cosimo, Matteo Fiorini, and Bernard Hoekman (2017). Services Trade Policy and Manufacturing Productivity: The Role of Institutions. *Journal of International Economics*, *104*(1), pp. 166-182.

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 (2016). Sources of Comparative Advantage in Polluting Industries. Working Paper.

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 (2017). How Collateral Laws Shape Lending and Sectoral Activity. *Journal of Financial Economics*, *123*(1), pp. 163-188.

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. Amsterdam: North Holland.

Cecchetti, Steven G. and Enisse Kharroubi (2018). Why Does Credit Growth Crowd Out Real Economic Growth? NBER Working Paper No. 25079.

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.

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.

Chauvet, Lisa and Helene Ehrhart (2018). Aid and Growth: Evidence from Firm-Level Data. *Journal of Development Economics*, *135*, pp. 461-477.

Choi, Sangyup, Davide Furceri, Yi Huang, and Prakash Loungani (2018). Aggregate Uncertainty and Sectoral Productivity Growth: The Role of Credit Constraints. *Journal of International Money and Finance*, *88*, pp. 314-330.

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

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 and Paolo Pinotti (2016). Trust, Firm Organization, and the Pattern of Comparative Advantage. *Journal of International Economics*, *100*(3), pp. 1-13.

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.

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

Claessens, Stijn and Neeltje van Horen (2017). The Role of Foreign Banks in Trade. Bank of England Staff Working Paper No. 656.

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.

Conley, Timothy (2010). Spatial Econometrics. In *Microeconometrics (The New Palgrave Economics Collection)* edited by Lawrence E. Blume and Stephen N. Durlauf. London: Palgrave Macmillan.

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

Crin, Rosario and Laura Oglirari (2017). Financial Imperfections, Product Quality, and International Trade. *Journal of International Economics*, *104*(1), pp. 63-84.

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.

Demirguk-Kunt, Aslia, Balint L. Horvath, and Harry Huizinga (2017). How Does Long-Term Finance Affect Economic Volatility? *Journal of Financial Stability*, *33*, pp. 41-59.

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: Evidence from the Great Recession. *Journal of Monetary Economics*. *75*(1), pp. 89-105.

Edmans, Alex, Sudarshan Jayaraman, and Jan Schneemeier (2017). The Source of Information in Prices and Investment-Price Sensitivity. *Journal of Financial Economics*, *126*(1), pp. 74-96.

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

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

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.

Fort, Teresa C. (2017). Technology and Production Fragmentation: Domestic Versus Foreign Sourcing. *Review of Economic Studies*, *84*(2), 650-687.

Gorton, Gary, Toomas Laarits and Tyler Muir (2019). 1930: First Modern Crisis. NBER Working Paper No. 25452.

Gupta, Nandini and Kathy Yuan (2009). On the Growth Effect of Stock Market Liberalizations. *Review of Financial Studies*, *22*(11), pp. 4715-4752.

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 the Chilean 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 No. 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.

Larrain, Mauricio and Sebastian Stumpner (2017). Capital Account Liberalization and Aggregate Productivity: The Role of Firm Capital Allocation. *Journal of Finance*, *72*(4), pp. 1825-1858.

Lei, Jin, Jiaping Qiu, and Chi Wan (2018). Asset Tangibility, Cash Holdings, and Financial Development. *Journal of Corporate Finance*, *50*(3), pp. 223-242.

Levine, Ross, Chen Lin, and Wensi Xie (2016). Spare Tire? Stock Markets, Banking Crises, and Economic Recoveries. *Journal of Financial Economics*, *120*(1), pp. 81-101.

Levine, Ross, Chen Lin, and Wensi Xie (2018). The Origins of Financial Development: The African Slave Trade and Modern Finance. NBER Working Paper No. 23800.

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.

Mueller, Holger M. and Thomas Philippon (2011). Family Firms and Labor Relations. *American Economic Journal: Macroeconomics*, *3*(2):, pp. 218-245.

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. Amsterdam: North Holland.

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.

Pierce, Lamar and Jason A. Snyder (2018). The Historical Slave Trade and Firm Access to Finance in Africa. *Review of Financial Studies*, *31*(1), pp. 142-174.

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), pp. 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*. Cambridge: MIT Press.

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
Finance and Industry Growth					
1	Finance and growth	Rajan and Zingales (1998)	Industry dependence on external finance [ratio of capital expenditures minus cash flow over capital expenditures]	Country financial development [market capitalization, private credit, measure of accounting standards]	Sectors that depend for inherent technological reasons more on external sources of finance (debt and equity), as compared to internal sources (retained earnings), grow faster in financially developed countries
2	Finance and growth	Claessens and Laeven (2003)	Industry intangible intensity [ratio of intangible assets to net fixed assets]	Country-level property rights protection [index of intellectual property rights, patent rights, risk of expropriation]	Sectors with an asset mix tilted towards intangibles grow faster in countries with better property rights
3	Finance and growth	Fisman and Love (2003)	Industry dependence on trade credit [accounts payable to total assets]	Country financial development [market capitalization, private credit, measure of accounting standards]	Industries with higher reliance on trade credit grow faster in countries with weaker financial institutions
4	Finance and growth	Fisman and Love (2007)	Industry growth opportunities [sales growth]	Country financial development [sum of domestic credit to private sector and market capitalization as a share of GDP]	Industries with better growth opportunities grow faster in more financially developed countries
5	Finance and growth	Beck, Demirguc-Kunt, Laeven, and Levine (2008)	Industry share of small firms [percentage of firms in each sector with less than 5, 10, 20, and 100 employees]	Country financial development [private credit to GDP]	Industries with a larger share of small firms grow faster in more financially developed countries
6	Firm size and growth	Pagano and Schivardi (2003)	Sector R&D intensity [share of R&D personnel in total employment, ratio of R&D to total investment and value added]	Average firm size of firm in sector in country [measured by employment]	Sectors with larger average firm size grow faster; particularly in R&D intense sectors
7	Financial dependence and business cycles	Braun and Larrain (2005)	Industry dependence on external finance	Recession in country c at time t	Industries that are more dependent on external finance are hit harder during recessions
8	Credit constraints, entry	Aghion, Fally and Scarpetta (2007)	Industry dependence on external finance	Country financial development [sum of private credit and stock market capitalization as a share of GDP, state ownership of banks]	More small firms enter in more externally dependent sectors in more financially developed countries
9	Finance and R&D investment	Brown, Martinsson, and Petersen (2013)	Industry dependence on external finance	Country financial development [value of IPOs as a share of GDP, accounting standards, anti-self-dealing index of shareholder protection]	Firms in more externally financially dependent industries invest more in R&D in more financially developed countries and in countries with stronger shareholder protection

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
10	Finance and innovation	Hsu, Tian, and Xu (2014)	Industry dependence on external finance and industry high-tech intensity	Country financial development [stock market capitalization, bank credit]	High-tech sectors that depend more on external sources of finance innovate more in financially developed countries
11	Finance and innovation	Acharya and Zu (2017)	Industry dependence on external finance	Public/Private Firm Indicator in the United States	Listed firms spend more on R&D in external-finance-dependent sectors
12	Firms' cash holdings, financial development, and firm growth	Lei, Qiu, and Wan (2018)	Industry asset tangibility	Private credit to GDP, contract enforcement, accounting standards, and log GDP p.c.	Sectors with a smaller proportion of tangible assets grow faster in countries with more developed financial markets
13	Access to long-term finance and volatility	Demirguk-Kunt, Horvath, and Huizinga (2017)	Sectoral measure of loan maturity	Various proxies of financial development and institutional quality	Financial development reduces firm growth volatility especially in external-finance-dependent sectors
14	Role of insider trading enforcement legislation on investment	Edmans, Jayaraman, and Schneemeir (2017)	Industry dependence on external finance	Insider trading enforcement legislation	The investment-Tobin's Q sensitivity increases after the enforcement of insider trading legislation in finance-dependent sectors and especially in emerging markets
15	Collateral laws and lending (loan-to-value)	Calomiris, Larrain, Liberti, and Sturgess (2017)	Sectoral index of real estate intensity	Laws shaping collateral and contract enforcement	Weak movable collateral laws create distortions in the allocation of resources that favor immovable-based production and investment
16	Real effects of banking crises	Dell'Ariccia, Detragiache, and Rajan (2008)	Industry dependence on external finance	Banking crisis in country c at time t	Sectors relatively more dependent on external finance perform worse during banking crises
17	Banking crises and exports	Iacovone and Zavacka (2009)	Industry dependence on external finance	Banking crisis in country c at time t	During a crisis, exports of sectors more dependent on external finance grow relatively less than those of other sectors
18	Investment effect of the subprime mortgage crisis	Duchin, Ozbas, and Sensoy (2010)	Industry dependence on external finance	Before/after sub-prime crisis	Decline in corporate investment is sharpest in industries with high external financial dependence
19	Transmission of financial crises	Claessens, Tong, and Wei (2012)	Industry dependence on external finance and trade sensitivity [global GDP elasticity of global exports at 3-digit sector level]	Country trade openness and fiscal and monetary policy	Crisis hit firms more sensitive to trade and business cycles hardest, especially in countries more open to trade

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
20	Firm growth and bank recapitalization	Laeven and Valencia (2013)	Industry dependence on external finance	Country bank recapitalization policies [committed amounts of public recapitalization funds]	Growth of finance dependent firms is disproportionately positively affected by bank recapitalization
21	Capital account liberalization, capital allocation, and productivity	Larrain and Stumpner (2018)	Industry dependence on external finance	Financial (capital account) liberalization	Within-sector misallocation (dispersion in marginal product of capital) falls when countries open their capital markets, especially in external finance dependent sectors
22	Monetary policy and growth	Aghion, Farhi, and Kharroubi (2015)	Industry credit or liquidity constraints [asset tangibility measured by value of net property, plant and equipment to total assets for credit constraints; labor-cost to sales for liquidity constraints]	Degree of counter-cyclicality of short-term interest rates [coefficient on output gap in regression with ST-rates on LHS]	Credit or liquidity constrained industries grow more quickly in countries with more counter-cyclical short-term interest rates
23	Fiscal policy and industry growth	Aghion, Hemous, and Kharroubi (2014)	Industry dependence on external finance	Countercyclicality of country fiscal policies [coefficient on output gap in regression with fiscal balance to GDP on LHS]	More externally dependent industries grow faster in countries that implement more countercyclical fiscal policies
24	Financial expansion (credit growth) and crowding out of output growth	Cecchetti and Kharoubi (2018)	Industry asset tangibility and industry R&D intensity	Credit growth	Credit growth disproportionately harms output per worker growth in industries that have either less tangible assets or are more R&D intensive
25	Dollar exchange rate and investment in emerging markets	Addjiev, Bruno, Koch and Shin (2018)	Industry dependence on external finance	Nominal and real exchange rates in emerging markets	A US dollar appreciation reduces investment in external finance dependent sectors in emerging markets implying a global dollar supply effect
26	Determinants of vertical integration	Alfaro, Conconi, Fadinger, and Newman (2016)	Industry external finance dependence	Financial development (and legal quality)	Financial development is associated with a higher level of vertical integration in external finance dependent sectors

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
International Trade and Industrial Specialization					
27	Factor proportions and trade	Romalis (2004)	Industry factor intensities in skilled labour, unskilled labour, and physical capital	Country factor endowments [human capital, physical capital, labour]	Countries specialize in industries that intensively use factors that (a) they are already abundant in; (b) they are accumulating rapidly
28	Human capital and growth	Ciccone and Papaioannou (2009)	Industry skill intensity [average years of employee schooling, share of high-school and college graduates]	Country initial human capital [average years of schooling]	Countries with higher initial education levels grew faster in schooling-intensive industries
29	Institutions and trade	Levchenko (2007)	Industry institutional dependence [concentration-Herfindahl index of intermediate input use]	Country institutional quality [rule of law]	Countries with better institutions have a greater share of US imports in more institutionally dependent sectors
30	Institutions and trade	Nunn (2007)	Industry contract intensity-complexity [reflecting relationship-specific investments]	Quality of contract enforcement and the judiciary [perception based rule of law index]	Countries with good contract enforcement specialize in goods for which relationship-specific investments are most important
31	Institutions, trade and organizational choice	Ferguson and Formai (2013)	Industry vertical integration-propensity and industry contract intensity	Country judicial quality [rule of law]	Benefits of judicial quality [high quality contractual institutions] for exports of contract-intensive goods are smaller in industries where firms are more likely to be integrated with their input suppliers
32	Institutions and comparative advantage	Nunn and Trefler (2014)	Industry cost sensitivity to quality of contracting institutions	Country quality of contracting institutions	Institutional sources of comparative advantage [as reflected by the interaction of country-level rule of law with industry-level contract intensity] are quantitatively as important as the impact of human capital and physical capital
33	Trade policy in services and productivity of downstream manufacturing	Beverelli, Fiorini, and Hoekman (2017)	Industry reliance on services as intermediate inputs	Index reflecting restrictiveness on trade in services; control of corruption	lower services trade restrictiveness is associated with higher downstream manufacturing labor and total-factor productivity, with the estimated effect increasing with country-level institutional capacity
34	Financial liberalization and trade	Manova (2008)	Industry dependence on external finance and industry asset tangibility [share of net property, plant and equipment in total book-value assets]	Time-varying country equity-market openess and liberalization	Liberalization increases exports disproportionately in sectors more dependent on outside finance or using fewer collateralized assets
35	Credit constraints and trade	Manova (2013)	Industry dependence on external finance and industry asset tangibility	Country financial development [private credit to GDP]	More financially developed countries export more in sectors more dependent on outside finance or using fewer collateralized assets

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
36	Finance and choice of export destinations	Chan and Manova (2015)	Industry dependence on external finance and industry asset tangibility	Country financial development [private credit to GDP]	More financially developed countries have more trading partners and particularly so in financially dependent sectors
37	Credit constraints and trade	Manova, Wei, and Zhang (2015)	Sector financial vulnerability [external financial dependence, asset tangilibity, inventory/sales ratio, reliance on trade credit]	Firm indicators for JV, MNC affiliates, firms with foreign ownership	Foreign affiliates and JVs in China have better export performance than private domestic firms in financially more vulnerable sectors
38	Financial frictions and product quality in international trade	Crino and Oglirari (2017)	Industry measures of financial vulnerability (asset tangibility, external-finance-dependence, capital intensity)	Financial development (private credit)	Financial development shapes comparative advantage in quality goods. The positive effect of financial development on the quality of exports is especially strong in finance-dependent sectors, in sectors with intangible assets, and capital intensive sectors
39	Role of foreign banks on trade	Claessens, Hassib, and van Horen (2017)	Industry dependence on external finance	Foreign banks from importing countries	For emerging markets, greater local foreign bank presence, especially from the importing country, is associated with higher exports in sectors more dependent on external finance
40	Employment protection and investment	Cingano, Leonardi, Messina and Pica (2010)	Sector worker reallocation intensity [average of normalized firm changes in employment in a country-industry cell]	Country employment protection legislation [OECD produced weighted average of 18 basic items]	EPL reduces investment in high reallocation- relative to low reallocation-sectors
41	Volatility, labour market flexibility and specialization	Cunat and Melitz (2012)	Volatility of firm output growth [standard deviation of annual growth rate of firm sales]	Country labour market flexibility [hiring-costs, firing costs, and restrictions on changing working hours as captured by World Bank index]	Exports of countries with more flexible labor markets are biased towards high-volatility sectors
42	Labor relations and family firms	Mueller and Philippon (2011)	Industry labor intensity	Labor market regulation (cooperative labor relations)	Sclerotic labor market regulation and institutions increase the share of family firms in labor intensive sectors
43	Labour markets, education and trade	Tang (2012)	Industry firm-specific skill intensity [estimated from Mincer wage regression with interaction of worker job tenure with industry dummy]	Country labour market protection	Countries with more protective labour laws export more in firm-specific skill intensive sectors at both intensive and extensive margins
44	Labour market institutions and innovation	Griffiwth and Macartney (2014)	Industry propensity to adjust to external labour market [layoff rate for 3-digit industry above or below the median layoff rate]	Country employment protection legislation [weighted sum of sub-indicators for regular and temporary contracts and collective dismissals]	Fewer radical innovations are done by high-layoff industries in countries with high EPL

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
45	Pollution and comparative advantage	Broner, Bustos, and Carvalho (2016)	Industry pollution intensity [EPA-computed total air pollution per unit of output]	Country laxity of air pollution regulation [proxied by outcome measure: grams of lead content per liter of gasoline]	Countries with laxer environmental regulation have a comparative advantage in polluting industries
46	Natural resources and comparative advantage	Debaere (2015)	Sector water intensity [sector water withdrawals both direct and indirect (inputs) from US Geological Survey]	Country water resources [volume of renewable fresh water per capita]	Relatively water abundant countries export more water-intensive products
	Other Applications				
47	Vertical vs horizontal, intra vs inter industry FDI	Alfaro and Charlton (2009)	Industry skill intensity [ratio of non-production to total workers]	Country skill abundance [average years of schooling]	Vertical FDI appears driven by comparative advantage at 2-digit level but not at 4-digit level
48	Boundaries of the firm	Costinot, Oldenski and Rauch (2011)	Sector task-routineness [importance of "making decisions and solving problems" for occupations within sectors]		Less-routine sectors have a higher share of intra-firm trade
49	Sourcing of goods of different complexity	Carluccio and Fally (2012)	Product complexity [measured with different indicators of R& D expenditures]	Country financial development [private credit to GDP]	Complex goods are more likely sourced from more financially developed countries
50	Offshoring	Basco (2014)	Industry R&D intensity [average industry R&D expenditure]	Country financial development [share of domestic credit to private sector over GDP]	More R&D intense industries use more intermediate inputs (offshore more) in more financially developed countries
51	Infrastructure and FDI	Blyde and Molina (2015)	Industry dependence logistic services [firm-in-industry willingness to pay for air shipping to avoid an additional day of ocean transport]	Country logistic infrastructure [number of ports and airports above a certain size normalized by country population]	Countries with better logistic infrastructure attract more vertical FDI in more time-sensitive industries
52	Corruption and innovation	Paunov (2016)	Industry usage intensity of quality certificates and patents [share of firms holding quality certificates; fractional patent count to value added]	Country corruption [share of firms reporting gift required to obtain operating license]	Firms in industries with greater reliance on quality certificates own less such certificates in more corrupt countries

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
53	Technology on outsourcing and production fragmentation	Fort (2017)	Industry use of advanced design and manufacturing software	Electronic networks at the firm level	firm's adoption of communication technology is associated with an increase in its probability of fragmentation. The effect of firm technology is higher, relative to the mean, in industries with production specifications that are easier to codify in an electronic format
54	Regulation and entry	Klapper, Laeven, and Rajan (2006)	Industry natural propensity to high entry [fraction of firms in industry that is one or two years old]	Country entry regulation [cost of business registration; in per capita GNP, time, or procedures]	Costly regulations reduce firm creation, especially in industries with naturally high entry
55	Determinants of vertical integration	Acemoglu, Johnson, and Mitton (2009)	Industry capital intensity as a proxy for vulnerability to holdup problems [fixed assets to sales]	Country-level contracting costs [procedural complexity, contract enforcement procedures, legal formalism]	Firms in more capital-intense industries are more vertically integrated in countries with higher contracting costs
56	Competition and Ownership Structure	Bena and Xu (2017)	Industry external finance sensitivity	Change in import penetration at the country-industry level	The effect of competition on ownership dispersion is higher is larger in sensitive to external finance sectors
57	Regulatory reforms and short-term employment costs	Bassanini and Cingano (2019)	Industry worker dismissal rate (in the US)	Employment protection legislation and product market regulation and business cycle conditions	Employment in dismissal-intensive sectors falls considerably more in years of labor and product market reform
58	Uncertainty and Total Factor Productivity	Choi, Fuceri, Huang, and Loungani (2018)	Sectoral dependence of external finance and industry asset tangibility	Uncertainty (based on stock market volatility)	Uncertainty reduces productivity in external-finance-dependence sectors and sectors with intangible assets
59	Aid and manufacturing growth	Rajan and Subramanian (2011)	Industry sensitivity to exchange rate appreciation [industry ratio of exports to value above or below the median]	Country receipts of foreign aid	Industries more sensitive to exchange rate appreciations grew relatively more slowly in countries receiving larger aid inflows
60	Aid and firm growth	Chauvet and Ehrhart (2018)	Industry reliance on exports, contract intensity, extremal-finance-dependence, transport-intensity, and reliance on electricity	Foreign aid	Aid spur firm growth in external finance dependent sectors and industries that use intensively electricity and rely on transportation infrastructure
61	The legacy of Africa's slave trades on finance	Pierce and Snyder (2017)	Industry dependence on sales credit	Slave trades as a share of country land area	Lower firm credit in sectors that depend on intensively on sales credit

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
62	The legacy of Africa's slave trades on firm's financial constraints and investment	Levine, Lin, and Xie (2018)	Industry dependence on external finance and sectoral capital intensity	Slave trades as a share of countries' land area and population	Firms in countries affected the most from African slave trades get lower levels of bank credit (for investment and working capital); this effect is especially strong for firms in capital intensive and external finance dependent sectors
63	International financial flows and growth	Aizenman and Sushko (2011)	Industry dependence on external finance	Portfolio equity, debt, and FDI inflows in country c at time t	Equity inflows have negative aggregate growth impact but positive impact in more financially constrained industries; FDI inflows have positive impact, both at the aggregate level and more external finance dependent industries
64	Human capital and trade	Bombardini, Gallipoli, and Putato (2012)	Industry skill substitutability [residual wage dispersion; rankings on teamwork, impact on co-woker output and communication / contact]	Country skill dispersion [within-country standard deviation of log scores on standardised tests]	Countries with more dispersed skill distributions export more in sectors with high substitutability of workers' skills
65	Business risk and growth	Michelacci and Schivardi (2013)	Sector idiosyncratic risk [sectoral component of volatility of firm stock returns]	Country lack of diversification opportunities [importance of family firms in the economy; share of widely held firms in the economy]	OECD countries with low levels of risk diversification opportunities perform relatively worse in sectors with high idiosyncratic risk
66	Capital account opening and inequality	Larrain (2014)	Industry dependence on external finance and capital-skill complementarity [external financial dependence as Rajan and Zingales (1998); capital intensity elasticity of skilled wage share]	Timing of country capital account opening	Capital account opening increases sectoral wage inequality, particularly in industries with both high external finance dependence and strong capital-skill complementarity
67	Intellectual property rights and innovation	Aghion, Howitt, and Prantl (2015)	Industry reliance on patents [R&D expenditure to nominal value added; patent count]	EU wide product market reform interacting with country-level strength of patent rights [data on patent law reforms]	1992 EU product market reform led to more innovation in countries with stronger patent protection and in particular in industries relying more on patents

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
68	Entry and access to finance	Cetorelli and Strahan (2006)	Industry external financial dependence	Degree of concentration in local banking markets [two policy variables on within-state branching and inter-state-banking restrictions; deposit Herfindahl concentration index]	Sectors with greater external financial dependence have larger and fewer firms in more concentrated local banking markets
69	Real effects of banking deregulation	Bertrand, Schoar, and Thesmar (2007)	Industry reliance on bank financing [all debt excluding trade credit and bonds over total outside financing (debt and book value of equity)]	Before/after 1985 French bank reform	Industries more reliant on bank financing before 1985 deconcentrated and experienced faster employment growth post bank-reform
70	Corporate tax reform and growth	Hsieh and Parker (2007)	Industry dependence on external finance	Before / after 1984 Chilean corporate tax reform	Post-reform investment boom occurred primarily in industries more dependent on external finance
71	Credit constraints and cylicality of R&D investment	Aghion, Askenazy, Berman, Cette, and Eymard (2012)	Industry dependence on external finance or asset tangility	Business cycle in France	For industries more reliant on external finance or with low asset tangibility, R&D investment is countercyclical without credit constraitns, and becomes pro-cyclical with tighter credit constraints
72	Institutions and trade in China	Feenstra, Hong, Ma, and Spencer (2013)	Industry reliance on contracts [from Nunn (2007), differentiation of intermediate inputs]	Cross-provincial variation in institutional quality in China [court efficiency as measured by overall quality, delays of verdicts and court costs]	Institutions matter more for processing trade and foreign firms, both of which rely more on contracts
73	Firm growth and access to finance in Morocco	Fafchamps and Schündeln (2013)	Sectoral growth opportunities [value added growth 1998-2003]	Local bank availability [dummy = 1 if local commune has a bank]	Firms in sectors with better growth opportunities grow faster in localities with bank availability
74	Unemployment, recessions and financing constraints	Duygan-Bump, Levkov, and Montriol-Garriga (2015)	Industry dependence on external finance	US recessions 90-91, 2001, 2007-2009	Workers in small firms are more likely to become unemployed if they work for firms in industries with high dependence on external finance during recessions in which loan supply contracts
75	Trade credit chains and corporate failure	Jacobson and von Schedvin (2015)	Industry dependence on external finance and liquidity [latter measured by inventory/ sales ratio]	Failure of trade credit debtors in Sweden	Propagation of corporate failure from trade-debtor to creditor is particularly severe in finanically constrained industries

#	Topic	Paper	Industry Characteristic	Country Characteristic	Main Finding
76	Trust, firm organization, and comparative advantage	Cingano and Pinotti (2016)	Industry need on delegation in the production process	Trust	European countries with higher mean levels of trust export more and specialize more in delegation-intensive sectors. Also Italian regions with high levels of trust specialize in delegation-requiring sectors

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

Appendix

Antonio Ciccone*
University of Mannheim, Barcelona GSE, and CEPR

Elias Papaioannou*
London Business School and CEPR

May 2019

*Emails: antonio.ciccone@uni-mannheim.de and eliaspapaioannou@london.edu.

1 Detailed Derivation of Equation (8) in the Main Text

Using (2) in (1) in the main text yields that the demeaned outcome in the numerator of (7) can be written as $y_{in} - \overline{y}_i - \overline{y}_n + \overline{y} = \beta(z_i - \overline{z})(x_n - \overline{x}) + \upsilon_{in}$. Here z_i is the global technological industry characteristic of industry i, $\overline{z}$ is the average technological industry characteristic across all industries, and

$$\upsilon_{in} = u_{in} - \overline{u}_n - \overline{u}_i + \overline{u} \tag{A1}$$

with

$$u_{in} = (\alpha + \beta x_n)\varepsilon_{in}, \tag{A2}$$

where $\overline{u}_n$ is the average of u_{in} across industries i for country n, $\overline{u}_i$ is the average of u_{in} across countries n for industry i, and $\overline{u}$ is the average of u_{in} both across countries and across industries. Substituting $y_{in} - \overline{y}_i - \overline{y}_n + \overline{y} = \beta(z_i - \overline{z})(x_n - \overline{x}) + \upsilon_{in}$ in (7) yields

$$\widehat{b} = \beta \frac{\frac{1}{I}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})(z_i - \overline{z})}{\frac{1}{I}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})^2} + \frac{\frac{1}{N}\frac{1}{I}\sum_{n=1}^{N}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})(x_n - \overline{x})\upsilon_{in}}{\frac{1}{N}\frac{1}{I}\sum_{n=1}^{N}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})^2(x_n - \overline{x})^2}. \tag{A3}$$

Note that the first ratio on the right-hand side of (A3) does not involve $\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2$ as this term cancels out.

Using (2), we can write demeaned US industry characteristics in terms of global and US-specific industry characteristics: $z_{iUS} - \overline{z}_{US} = (z_i - \overline{z}) + (\varepsilon_{iUS} - \overline{\varepsilon}_{US})$. Substituting in (A3) yields

$$\widehat{b} = \beta \frac{\frac{1}{I}\sum_{i=1}^{I}(z_i - \overline{z})^2 + \frac{1}{I}\sum_{i=1}^{I}(z_i - \overline{z})(\varepsilon_{iUS} - \overline{\varepsilon}_{US})}{\frac{1}{I}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})^2} + \frac{\frac{1}{N}\frac{1}{I}\sum_{n=1}^{N}\sum_{i=1}^{I}(z_i - \overline{z})(x_n - \overline{x})\upsilon_{in}}{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2\frac{1}{I}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})^2} \tag{A4}$$

$$+ \frac{\frac{1}{N}\frac{1}{I}\sum_{n=1}^{N}\sum_{i=1}^{I}(\varepsilon_{iUS} - \overline{\varepsilon}_{US})(x_n - \overline{x})\upsilon_{in}}{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2\frac{1}{I}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})^2}.$$

We will now discuss the probability limit as I goes to infinity of each of the three ratios on the right-hand side of (A4).

To begin with, we show that the probability limit of the first ratio on the right-hand side of (A4) is $\beta(1 - \phi)$. To see this note that the second term in the numerator can be written as $\frac{1}{I}\sum_{i=1}^{I}(z_i - \overline{z})(\varepsilon_{iUS} - \overline{\varepsilon}_{US}) = \frac{1}{I}\sum_{i=1}^{I} z_i\varepsilon_{iUS} - \overline{z}\frac{1}{I}\sum_{i=1}^{I}\varepsilon_{iUS}$. As z_i is i.i.d., the standard version of the law of large numbers yields that the probability limit of $\overline{z}$ is $E(z_i)$. Using the law of large numbers for independent random variables with the same expectation and bounded variances we obtain probability limits for the two averages across industries, $\frac{1}{I}\sum_{i=1}^{I} z_i\varepsilon_{iUS}$ and $\frac{1}{I}\sum_{i=1}^{I}\varepsilon_{iUS}$. The probability limit of the first average is equal to $Ez_i\varepsilon_{iUS} = Ez_iE\varepsilon_{iUS} = 0$, as z_i is independent of all other model elements and $E\varepsilon_{iUS} = 0$. The probability limit of the second average is $E\varepsilon_{iUS} = 0$. Thus, $\frac{1}{I}\sum(z_i - \overline{z})(\varepsilon_{iUS} - \bar{\varepsilon}_{US})$ vanishes in the limit as I goes to infinity. Moreover, the probability limitsof $\frac{1}{I}\sum_{i=1}^{I}(z_{iUS} - \overline{z}_{US})^2$ and $\frac{1}{I}\sum_{i=1}^{I}(z_i - \overline{z})^2$ are $Var(z_{US})$ and $Var(z_i)$ respectively. Eventually, (5) in the main text implies $1 - \phi = Var(z_i)/Var(z_{US})$.

Next, we show that the probability limit, as I goes to infinity, of the second ratio on the right-hand side of (A4) is zero. Using (A1), the numerator of this ratio can be written as

$$\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})\left[\frac{1}{I}\sum_{i=1}^{I}(z_i - \overline{z})(u_{in} - \overline{u}_n - \overline{u}_i + \overline{u})\right] \tag{A5}$$

and the square bracket can be written as

$$\frac{1}{I}\sum_{i=1}^{I} z_i(u_{in} - \overline{u}_i) - \overline{z}\frac{1}{I}\sum_{i=1}^{I}(u_{in} - \overline{u}_i) - (\overline{u}_n - \overline{u})\frac{1}{I}\sum_{i=1}^{I} z_i + \overline{z}(\overline{u}_n - \overline{u}). \tag{A6}$$

All weighted sums across industries in (A6) are sums of independent random variables with equal expectation and bounded variances. Hence, the law of large numbers implies that the probability limit of the first weighted sum is $Ez_i(u_{in} - \overline{u}_i) = Ez_iE(u_{in} - \overline{u}_i) = 0$, where we use that global industry characteristics z_i are independent of all other model elements and that $E(u_{in} - \overline{u}_i) = Eu_{in} - E\overline{u}_i = 0$. The probability limits of the second and third weighted sum are $E(u_{in} - \overline{u}_i) = Eu_{in} - E\overline{u}_i = 0$ and Ez_i respectively. Again, as z_i is i.i.d., the probability limit of $\overline{z}$ is $E(z_i)$. Moreover, the terms $\overline{u}_n$ and $\overline{u}$ in (A6) go to zero in probability, as $E\overline{u}_n = E\overline{u} = 0$ and the variances $Var(\overline{u}_n) = \frac{1}{I}(\alpha + \beta x_n)^2\sigma^2$ and $Var(\overline{u}) = Var(\frac{1}{I}\sum_{i=1}^{I}\overline{u}_i) = \frac{1}{I}Var(\overline{u}_i) = \frac{1}{I}\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}(\alpha + \beta x_n)(\alpha + \beta x_m)\rho_{nm}\sigma^2$ go to zero as I goes to infinity. Hence, all terms in (A6) vanish in the limit. At the same time, the denominator of the second ratio in (A4) goes to some strictly positive number as I goes to infinity. On the other hand, the denominator of this ratio goes to some strictly positive number as I goes to infinity. Hence, the second ratio in (A4) vanishes in the probability limit.

Collecting the results we have so far, as I goes to infinity, the probability limit of (A4) is

$$b = (1-\phi)\beta + \frac{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x}) \operatorname*{plim}_{I\to\infty} \frac{1}{I}\sum_{i=1}^{I}(\varepsilon_{iUS} - \overline{\varepsilon}_{US})\upsilon_{in}}{Var(z_{US})\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2} \tag{A7}$$

where we rewrote the numerator of the last term in (A4) in terms of an outer sum across countries and an inner sum across industries. The key term in (A7) is the term behind the probability limit (plim). Using (A1), this term can be written as

$$\frac{1}{I}\sum_{i=1}^{I}(\varepsilon_{iUS} - \overline{\varepsilon}_{US})(u_{in} - \overline{u}_i) - (\overline{u}_n - \overline{u})\frac{1}{I}\sum_{i=1}^{I}(\varepsilon_{iUS} - \overline{\varepsilon}_{US}). \tag{A8}$$

The second term in (A8) is equal to zero, as $\overline{\varepsilon}_{US} = \frac{1}{I}\sum_{i=1}^{I}\varepsilon_{iUS}$. The first term can be written as

$$\frac{1}{I}\sum_{i=1}^{I}(\varepsilon_{iUS} - \overline{\varepsilon}_{US})(u_{in} - \overline{u}_i) = \frac{1}{I}\sum_{i=1}^{I}\varepsilon_{iUS}(u_{in} - \overline{u}_i) - \overline{\varepsilon}_{US}(\overline{u}_n - \overline{u}). \tag{A9}$$

As $E\bar{\varepsilon}_{US} = E\bar{u}_n, = E\bar{u} = 0$ and the variances $Var(\overline{\varepsilon}_{US}) = \frac{1}{I}\sigma^2$, $Var(\overline{u}_n) = \frac{1}{I}(\alpha + \beta x_n)^2\sigma^2$, and $Var(\overline{u}) = Var(\frac{1}{I}\sum_{i=1}^{I}\overline{u}_i) = \frac{1}{I}Var(\overline{u}_i) = \frac{1}{I}\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}(\alpha + \beta x_n)(\alpha + \beta x_m)\rho_{nm}\sigma^2$ vanish as I tends towards infinity, the second term on the right-hand side of (A9) goes to zero in probability. Making use of the law of large numbers for independent random variables of equal expectation and bounded variance the probability limit of the first term on the right-hand side of (A9) is

$$E\varepsilon_{iUS}(u_{in} - \overline{u}_i) = (\alpha + \beta x_n)E\varepsilon_{iUS}\varepsilon_{in} - \frac{1}{N}\sum_{n=1}^{N}(\alpha + \beta x_n)E\varepsilon_{iUS}\varepsilon_{in}. \tag{A10}$$

Noting that $\sigma^2\rho_{nUS} = E\varepsilon_{iUS}\varepsilon_{in}$, we have

$$E\varepsilon_{iUS}(u_{in} - \overline{u}_i) = (\alpha + \beta x_n)\sigma^2\rho_{nUS} - \frac{1}{N}\sum_{n=1}^{N}(\alpha + \beta x_n)\sigma^2\rho_{nUS}. \tag{A11}$$

Using this, the numerator of the second term on the right-hand side of (A7) is

$$\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})(\alpha + \beta x_n)\sigma^2\rho_{nUS} - \left(\frac{1}{N}\sum_{n=1}^{N}(\alpha + \beta x_n)\sigma^2\rho_{nUS}\right)\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x}). \tag{A12}$$

As $\frac{1}{N}\sum_{n=1}^{N} x_n = \overline{x}$, the second term in (A12) is zero. Substituting the first term in (A12) for the numerator in (A7) yields

$$b = (1-\phi)\beta + \left(\frac{\sigma^2}{Var(z_{US})}\right)\frac{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})(\alpha + \beta x_n)\rho_{nUS}.}{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x}^2)}. \tag{A13}$$

Using the definitions for A from (9) and for B from (10) in the main text as well as the fact that $\phi = \sigma^2/(\sigma^2 + Var(z_i)) = \sigma^2/Var(z_{US})$, rewriting (A13) yields (8).

2 Detailed Derivation of Equation (36) in the Main Text

We are interested in the probability limit of $\frac{1}{I}\sum_{i=1}^{I}\widehat{u}_{in}\widehat{u}_{im}$ as the number of industries I goes to infinity, where

$$\widehat{u}_{in} = \upsilon_{in} - (x_n - \overline{x})\sum_{k=1}^{N}\psi_k\upsilon_{ik}, \tag{A14}$$

ψ_k is the least-squares regression weight defined in (35) in the main text, and

$$\upsilon_{in} = u_{in} - \overline{u}_n - \overline{u}_i + \overline{u}. \tag{A15}$$

In (A15), $\overline{u}_n$ is the average of u_{in} across industries i for country n, $\overline{u}_i$ is the average of u_{in} across countries n for industry i, and $\overline{u}$ is the average of u_{in} both across countries and across industries. Making use of (A14),

$$\frac{1}{I}\sum_{i=1}^{I}\widehat{u}_{in}\widehat{u}_{im} = \frac{1}{I}\sum_{i=1}^{I}\upsilon_{in}\upsilon_{im} - (x_n - \overline{x})\sum_{k=1}^{N}\psi_k\left(\frac{1}{I}\sum_{i=1}^{I}\upsilon_{ik}\upsilon_{im}\right) \\ - (x_m - \overline{x})\sum_{k=1}^{N}\psi_k\left(\frac{1}{I}\sum_{i=1}^{I}\upsilon_{ik}\upsilon_{in}\right) \\ + (x_n - \overline{x})(x_m - \overline{x})\sum_{k=1}^{N}\sum_{g=1}^{N}\psi_g\psi_k\left(\frac{1}{I}\sum_{i=1}^{I}\upsilon_{in}\upsilon_{im}\right). \tag{A16}$$

As (A16) reveals, a key term to determine the probability limit of $\frac{1}{I}\sum_{i=1}^{I}\widehat{u}_{in}\widehat{u}_{im}$ is the probability limit of

$$\frac{1}{I}\sum_{i=1}^{I}\upsilon_{in}\upsilon_{im}. \tag{A17}$$

We show that this probability limit is $\omega_{nm} - \overline{\omega}_n - \overline{\omega}_m + \overline{\omega}$, where ω_{nm} is the covariance $Eu_{in}u_{im}$ defined in (23) in the main text, $\overline{\omega}_p$ denotes the average of ω_{pq} across q, i.e. $\overline{\omega}_p = \frac{1}{N}\sum_{q=1}^{N}\omega_{pq}$, and $\overline{\omega}$ is the average of ω_{pq} across q and p, i.e. $\overline{\omega} = \frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N}\omega_{pq}$. To this end, it is useful to use (A1) to rewrite (A17) as the weighted sum of four terms:

$$\frac{1}{I}\sum_{i=1}^{I}\upsilon_{in}\upsilon_{im} = \frac{1}{I}\sum_{i=1}^{I}(u_{in} - \overline{u}_i)(u_{im} - \overline{u}_i) + (\overline{u}_n - \overline{u})(\overline{u}_m - \overline{u}) \\ - (\overline{u}_m - \overline{u})\frac{1}{I}\sum_{i=1}^{I}(u_{in} - \overline{u}_i) - (\overline{u}_n - \overline{u})\frac{1}{I}\sum_{i=1}^{I}(u_{im} - \overline{u}_i). \tag{A18}$$

All $(\overline{u}_n - \overline{u})$-terms on the right-hand side of (A18) go to zero in probability as the number of industries I goes to infinity. To see this, note that $E(\overline{u}_n - \overline{u}) = 0$ and that the variance $Var(\overline{u}_n - \overline{u})$ goes to zero as the number of industries I goes to infinity. This can be verified by writing the variance as

$$E(\overline{u}_n - \overline{u})^2 = E\overline{u}_n^2 - 2E\overline{u}_n\overline{u} + E\overline{u}^2. \tag{A19}$$

Now, the three terms on the right-hand side of (A19) can be respectively written as

$$E\overline{u}^2 = E\left(\frac{1}{I}\sum_{i}^{I}\overline{u}_i\right)^2 = \frac{1}{I}E\overline{u}_i^2 = \frac{1}{I}\frac{1}{N^2}\sum_{g=1}^{N}\sum_{k=1}^{N}\omega_{gk}, \tag{A20}$$

$$E\overline{u}_n^2 = E\left(\frac{1}{I}\sum_{j=1}^{I}u_{jn}\right)^2 = \frac{1}{I}\omega_{nn}, \tag{A21}$$

and

$$2E\overline{u}_n\overline{u} = 2\frac{1}{N}\sum_{k=1}^{N} E\overline{u}_n\overline{u}_k = 2\frac{1}{N}\frac{1}{I}\sum_{k=1}^{N}\omega_{nk}. \tag{A22}$$

Therefore, all three terms vanish in the limit as the number of industries I goes to infinity.

The terms on the right-hand side of (A18) that involve weighted sums across industries can be analyzed using the law of large numbers for independent random variables with the same expectation and bounded variances. Thus, the probability limit of

$$\frac{1}{I}\sum_{i=1}^{I}(u_{im} - \overline{u}_i) \tag{A23}$$

is $E(u_{im} - \overline{u}_i) = Eu_{im} - E\overline{u}_i = 0$. Combined with the properties of the term $\overline{u}_n - \overline{u}$ discussed in (A19)–(A22), this implies that the probability limit of all terms on the right-hand side of (A18) except the first one is zero. By another application of the law of large numbers for independent random variables of equal expectation and bounded variance implies, the probability limitof

$$\frac{1}{I}\sum_{i=1}^{I}(u_{in} - \overline{u}_i)(u_{im} - \overline{u}_i) \tag{A24}$$

is $E(u_{in} - \overline{u}_i)(u_{im} - \overline{u}_i)$, which can be further calculated to be

$$E(u_{in} - \overline{u}_i)(u_{im} - \overline{u}_i) = \omega_{nm} - \overline{\omega}_n - \overline{\omega}_m + \overline{\omega}. \tag{A25}$$

Hence, it follows that, as the number of industries I goes to infinity, the probability limit of $\frac{1}{I}\sum_{i=1}^{I} \upsilon_{in}\upsilon_{im}$ is $\omega_{nm} - \overline{\omega}_n - \overline{\omega}_m + \overline{\omega}$.

Returning to the analysis of (A16), we have just shown the probability limit of the first term to be equal to

$$\omega_{km} - \overline{\omega}_k - \overline{\omega}_m + \overline{\omega} \tag{A26}$$

as the number of industries I tends to infinity. The probability limit of the second term in (A16) is

$$(x_n - \overline{x})\sum_{k=1}^{N}\psi_k(\omega_{km} - \overline{\omega}_k - \overline{\omega}_m + \overline{\omega}) = (x_n - \overline{x})\sum_{k=1}^{N}\psi_k(\omega_{km} - \overline{\omega}_k) \tag{A27}$$

where we have once again substituted $\omega_{nm} - \overline{\omega}_n - \overline{\omega}_m + \overline{\omega}$ for the probability limit of $\frac{1}{I}\sum_{i=1}^{I} \upsilon_{in}\upsilon_{im}$ and made use of $\sum_{k=1}^{N}\psi_k = 0$. The probability limit of the third term in (A16) is equal to (A27) with n and m switched. Finally, the probability limit of the last term in (A16) is

$$(x_m - \overline{x})(x_n - \overline{x}) \sum_{k=1}^{N} \sum_{g=1}^{N} \psi_g \psi_k (\omega_{kg} - \overline{\omega}_k - \overline{\omega}_g + \overline{\omega}) \tag{A28}$$
$$= (x_m - \overline{x})(x_n - \overline{x}) \sum_{k=1}^{N} \sum_{g=1}^{N} \psi_g \psi_k \omega_{kg},$$

where we made use of $\sum_{k=1}^{N} \psi_k = 0$ again. Collecting the results in (A26)-(A28) yields that, as the number of industries I goes to infinity, the probability limit of $\frac{1}{I} \sum_{i=1}^{I} \widehat{u}_{in} \widehat{u}_{im}$ is

$$\omega_{nm} - \overline{\omega}_n - \overline{\omega}_m + \overline{\omega} - (x_m - \overline{x}) \sum_{k=1}^{N} \psi_k (\omega_{kn} - \overline{\omega}_k) \tag{A29}$$
$$- (x_n - \overline{x}) \sum_{k=1}^{N} \psi_k (\omega_{km} - \overline{\omega}_k)$$
$$+ (x_m - \overline{x})(x_n - \overline{x}) \sum_{k=1}^{N} \sum_{g=1}^{N} \psi_g \psi_k \omega_{kg}.$$

Defining

$$\mu_n = \overline{\omega}_n - \frac{1}{2} \overline{\omega} \tag{A30}$$

$$\lambda_n = \sum_{k=1}^{N} \psi_k (\omega_{kn} - \overline{\omega}_k) - \frac{1}{2} (x_n - \overline{x}) \sum_{k=1}^{N} \sum_{g=1}^{N} \psi_g \psi_k \omega_{kg} \tag{A31}$$

(A29) can be rewritten as

$$\omega_{nm} - \mu_n - \mu_m - (x_m - \overline{x}) \lambda_n - (x_n - \overline{x}) \lambda_m \tag{A32}$$

which is the right-hand side of (36) in the main text.

It remains to be shown that, as claimed in the main text, $\sum_{n=1}^{N} \lambda_n = 0$. However, this follows immediately from the fact that $\frac{1}{N} \sum_{n=1}^{N} x_n = \overline{x}$ and $\frac{1}{N} \sum_{n=1}^{N} \omega_{kn} = \overline{\omega}_k$.

3 Show that Equation (36) in the Main Text Does Not Determine ω_{nm} for Arbitrary $\boldsymbol{\Omega}$

Using standard results in econometrics it can be shown that it is impossible to identify the elements ω_{nm} from the π_{nm} in (36) in the main text for an arbitrary variance-covariance matrix $\boldsymbol{\Omega}$. To do so, we collect the π_{nm} in a $N \times N$ matrix $\boldsymbol{\Pi}$ and note that the equation system in (36) can be rewritten in matrix form as

$$\mathbf{\Pi} = \mathbf{M\Omega M} \tag{A33}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{P}$, $\mathbf{I}$ is a square identity matrix of size N, $\mathbf{P}$ is the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{X} = (1, \mathbf{x})$ with 1 being a column vector of length N and $\mathbf{x}' = (x_1, \ldots, x_N)$. The key issue then becomes whether the equation system in (A33) determines the symmetric variance-covariance matrix $\mathbf{\Omega}$ for given $\mathbf{\Pi}$ and $\mathbf{M}$. Using the fact that $\mathbf{P}$ is a projection matrix, i.e. $\mathbf{PX} = \mathbf{X}$ and thus $\mathbf{MX} = 0$, it is easy to show that $\mathbf{\Omega}$ cannot be determined. Indeed, if $\mathbf{\Omega}$ solves (A33) then so does any $\widetilde{\mathbf{\Omega}} = \mathbf{\Omega} + \mathbf{XD} + \mathbf{D}'\mathbf{X}' + \mathbf{XEE}'\mathbf{X}'$, where $\mathbf{D}$ and $\mathbf{E}$ are arbitrary $2 \times N$ matrices. Hence, (A33) does not identify $\mathbf{\Omega}$.

Next, we verify that equation (36) can indeed be rewritten as $\mathbf{\Pi} = \mathbf{M\Omega M}$. Using the definitions introduced above, we can rewrite $\mathbf{\Pi} = \mathbf{M\Omega M}$ as

$$\begin{aligned} \mathbf{\Pi} = \mathbf{\Omega} &- \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega} - \mathbf{\Omega X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &+ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \end{aligned} \tag{A34}$$

The first step to show that this corresponds to (36) in the main text is to write $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ as

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \left(\sum_{k=1}^{N}(x_k - \overline{x})^2\right)^{-1} \begin{pmatrix} \frac{1}{N}\sum\limits_{k=1}^{N} x_k^2 - x_1\overline{x} & x_1 - \overline{x} \\ \vdots & \vdots \\ \frac{1}{N}\sum\limits_{k=1}^{N} x_k^2 - x_N\overline{x} & x_N - \overline{x} \end{pmatrix} \tag{A35}$$

and $\mathbf{X}'\mathbf{\Omega}$ as

$$\mathbf{X}'\mathbf{\Omega} = \begin{pmatrix} N\overline{\omega}_1 & \ldots & N\overline{\omega}_N \\ \sum\limits_{k=1}^{N} x_k\omega_{k1} & \ldots & \sum\limits_{k=1}^{N} x_k\omega_{kN} \end{pmatrix}, \tag{A36}$$

where ω_{nm} is the typical element of $\mathbf{\Omega}$ and $\overline{\omega}_p$ denotes the average of ω_{pq} across q. Hence the typical element of the matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}$ in (A34) is

$$\left(\sum_{k=1}^{N}(x_k - \overline{x})^2\right)^{-1} \left[\left(\sum_{k=1}^{N} x_k^2 - N\overline{x}^2\right)\overline{\omega}_m - (x_n - \overline{x})\overline{x}N\overline{\omega}_m + (x_n - \overline{x})\sum_{k=1}^{N} x_k\omega_{km}\right] \tag{A37}$$

or, collecting terms,

$$\overline{\omega}_m + (x_n - \overline{x})\sum_{k=1}^{N} \psi_k\omega_{km} \tag{A38}$$

where ψ_k is the least-squares regression weight:

$$\psi_k = \frac{x_k - \overline{x}}{\sum_{m=1}^{N}(x_m - \overline{x})^2}. \tag{A39}$$

As $\mathbf{\Omega X}(\mathbf{X'X})^{-1}\mathbf{X'}$ in (A34) is the transpose of $\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'\Omega}$, the typical element of $\mathbf{\Omega X}(\mathbf{X'X})^{-1}\mathbf{X'}$ is

$$\overline{\omega}_n + (x_m - \overline{x})\sum_{k=1}^{N}\psi_k\omega_{kn}. \tag{A40}$$

What is left is to determine the typical element of $\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'\Omega X}(\mathbf{X'X})^{-1}\mathbf{X'}$ in (A34). The typical element of $\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$ is

$$\left(\sum_{k=1}^{N}(x_k - \overline{x})^2\right)^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}x_k^2 - x_n\overline{x} + (x_n - \overline{x})x_m\right) \tag{A41}$$

or

$$\left(\sum_{k=1}^{N}(x_k - \overline{x})^2\right)^{-1}\left(\frac{1}{N}\sum_{k=1}^{N}(x_k - \overline{x})^2 + (x_n - \overline{x})(x_m - \overline{x})\right). \tag{A42}$$

Multiplying $\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'\Omega}$, the typical element of which is given by (A38), with $\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$, the typical element of which is given by (A42), yields

$$\left(\sum_{p=1}^{N}(x_p - \overline{x})^2\right)^{-1}\left[\sum_{g=1}^{N}\left(\overline{\omega}_g + (x_n - \overline{x})\sum_{k=1}^{N}\psi_k\omega_{kg}\right)\left(\frac{1}{N}\sum_{k=1}^{N}(x_k - \overline{x})^2 + (x_g - \overline{x})(x_m - \overline{x})\right)\right] \tag{A43}$$

as typical element of $\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'\Omega X}(\mathbf{X'X})^{-1}\mathbf{X'}$. This can be further rewritten as

$$\sum_{g=1}^{N}\left(\overline{\omega}_g + (x_n - \overline{x})\sum_{k=1}^{N}\psi_k\omega_{kg}\right)\left(\frac{1}{N} + \psi_g(x_m - \overline{x})\right) \tag{A44}$$

or as

$$\overline{\omega} + (x_n - \overline{x})\sum_{k=1}^{N}\psi_g\overline{\omega}_g + (x_m - \overline{x})\sum_{k=1}^{N}\psi_g\overline{\omega}_g + (x_n - \overline{x})(x_m - \overline{x})\sum_{k=1}^{N}\sum_{k=1}^{N}\psi_k\psi_g\omega_{kg}. \tag{A45}$$

Collecting terms in (A38), (A40), and (A45) nd using the fact that the typical element of $\mathbf{\Omega}$ in (A34) is ω_{nm} yields that the typical element of the right-hand side of (A34) is

$$\omega_{nm} - \overline{\omega}_n - \overline{\omega}_m + \overline{\omega} - (x_m - \overline{x}) \sum_{k=1}^{N} \psi_k(\omega_{kn} - \overline{\omega}_k) \tag{A46}$$

$$-(x_n - \overline{x}) \sum_{k=1}^{N} \psi_k(\omega_{km} - \overline{\omega}_k) + (x_m - \overline{x})(x_n - \overline{x}) \sum_{k=1}^{N} \sum_{g=1}^{N} \psi_g \psi_k \omega_{kg}.$$

This is identical to (A29). As shown above, rewriting (A29) as (A30) yields the right-hand side of equation (36). Hence, (36) in the main text can be written as $\mathbf{\Pi} = \mathbf{M\Omega M}$.

4 Proof of Proposition 2

To prove the proposition it is useful to define $\phi = \sigma^2/\sigma_{US}^2$. As $0 \leq \sigma^2 < \sigma_{US}^2$, it follows that $\phi \in [0, 1)$. Recall that the two solutions for q in (26) in the main text are β and $\phi(\delta - 1)\beta$, implying $q_1 + q_2 = [1 + \phi(\delta - 1)]\beta$. Hence, the two solutions for q divided by $q_1 + q_2$ are $1/[1 + \phi(\delta - 1)]$ and $\phi(\delta - 1)/[1 + \phi(\delta - 1)]$. This implies that if $\delta \in [0, 2]$, then $\kappa = 1/[1 + \phi(\delta - 1)]$. Hence, using (17) in the main text, $\kappa b = b/[1 + \phi(\delta - 1)] = \beta$.

5 Proof of Proposition 3

For $\delta \in [0, 2]$, see the proof of Proposition 2. To prove it for other values of δ, it is useful to distinguish the cases $\delta > 2$ and $\delta < 0$. We continue to use the definition $\phi = \sigma^2/\sigma_{US}^2$ with $\phi \in [0, 1)$ as $Var(z_i) > 0$ implies that $0 \leq \sigma^2 < \sigma_{US}^2$.

Recall that the two solutions for q in (26) in the main text are β and $\phi(\delta - 1)\beta$, implying $q_1 + q_2 = [1 + \phi(\delta - 1)]\beta$. Hence, the two solutions for q divided by $q_1 + q_2$ are $1/[1 + \phi(\delta - 1)]$ and $\phi(\delta - 1)/[1 + \phi(\delta - 1)]$. Clearly, $1 + \phi(\delta - 1) \geq 0$ for $\delta > 2$. Therefore, the definition of κ in (27) implies

$$\begin{aligned} \kappa &= \tfrac{1}{1+\phi(\delta-1)} \quad \text{if } \phi(\delta - 1) \leq 1 \\ \kappa &= \tfrac{\phi(\delta-1)}{1+\phi(\delta-1)} \quad \text{if } \phi(\delta - 1) > 1. \end{aligned} \tag{A47}$$

Using the notation $\kappa(\phi)$ to capture that κ is a function of ϕ, this can be written as

$$\kappa(\phi) = \begin{cases} \frac{1}{1+\phi(\delta-1)} & \text{if } \phi \in \left[0, \frac{1}{\delta-1}\right] \\ \frac{\phi(\delta-1)}{1+\phi(\delta-1)} & \text{if } \phi \in \left[\frac{1}{\delta-1}, 1\right) \end{cases} \tag{A48}$$

where $0 < 1/(\delta - 1) < 1$. The function $\kappa(\phi)$ is illustrated in Figure A1. $\kappa(\phi)$ is strictly decreasing in ϕ up to the point where $\phi = 1/(\delta - 1) < 1$, and is strictly increasing in ϕ from

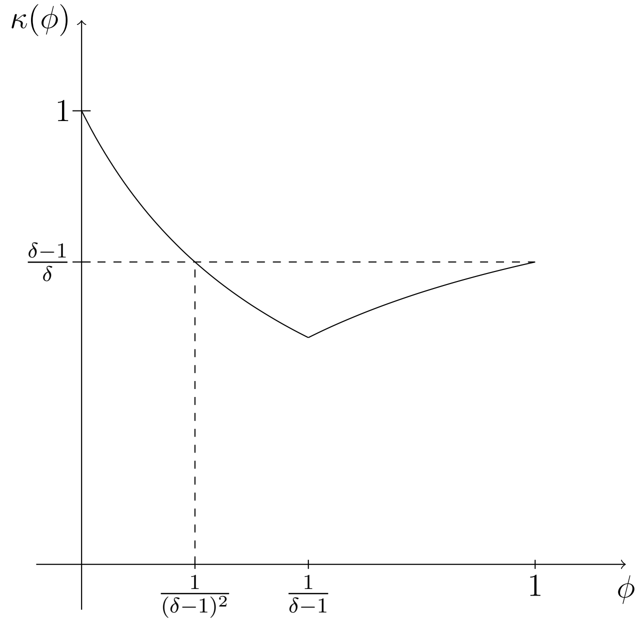


Figure A1: The shape of $\kappa(\phi)$ for $\delta > 2$.

that point on. Moreover, $\kappa(1) = (\delta - 1)/\delta$. As $\kappa(\phi)$ is strictly increasing for $\phi > 1/(\delta - 1)$, we get that $\kappa(\phi) < (\delta - 1)/\delta$ for all $\phi \in [1/(\delta - 1), 1)$.

For $\delta > 2$, the relevant version of condition (28) in Proposition 3 is

$$\kappa \geq \frac{\delta - 1}{\delta}. \tag{A49}$$

It can therefore never be satisfied for $\phi \in (1/(\delta-1), 1)$. Put differently, the relevant condition in the proposition can be satisfied only if $\phi \in [0, 1/(\delta-1)]$. For ϕ in this range, (A48) implies $\kappa(\phi) = 1/[1+\phi(\delta-1)]$ and the condition in (A49) is satisfied if $\phi \leq 1/(\delta-1)^2$. Summarizing, when $\delta > 2$, the relevant condition in Proposition 3 is satisfied if and only if ϕ satisfies

$$\phi(\delta - 1)^2 \leq 1. \tag{A50}$$

As $\kappa = 1/[1 + \phi(\delta - 1)]$ for ϕ in this range, the claim $\beta = \kappa b$ in Proposition 3 follows from rewriting (17) in the main text as $b = [1 + \phi(\delta - 1)]\beta$.

When $\delta < 0$, the two solutions for q divided by $q_1 + q_2$, $1/[1+\phi(\delta-1)]$ and $\phi(\delta-1)/[1+\phi(\delta - 1)]$, imply that κ in Proposition 3 is

$$\begin{array}{ll} \kappa = \frac{1}{1+\phi(\delta-1)} & \text{if } \phi(\delta-1) \geq -1 \\ \kappa = \frac{\phi(\delta-1)}{1+\phi(\delta-1)} & \text{if } \phi(\delta-1) < -1 \end{array}. \tag{A51}$$

Or, using the notation $\kappa(\phi)$ to capture that κ is a function of ϕ:

$$\kappa(\phi) = \begin{cases} \frac{1}{1+\phi(\delta-1)} & \text{if } \phi \in \left[0, -\frac{1}{\delta-1}\right] \\ \frac{\phi(\delta-1)}{1+\phi(\delta-1)} & \text{if } \phi \in \left[-\frac{1}{\delta-1}, 1\right) \end{cases} \tag{A52}$$

where $0 < -1/(\delta - 1) < 1$. The function $\kappa(\phi)$ is illustrated in figure A2. For $\phi < -\frac{1}{\delta-1}$,

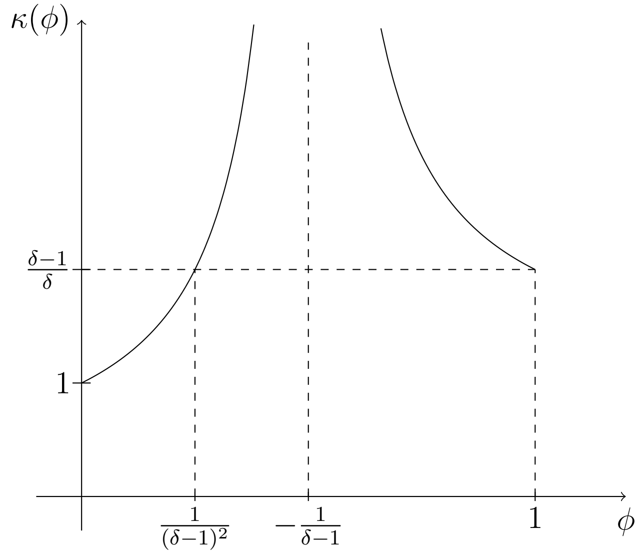


Figure A2: The shape of $\kappa(\phi)$ for $\delta < 0$.

κ is strictly increasing in ϕ. For values of ϕ larger than $\phi = -1/(\delta - 1)$, $\kappa(\phi)$ is strictly decreasing. Furthermore, $\kappa(1) = (\delta - 1)/\delta$. As a result, we get that $\kappa(\phi) > (\delta - 1)/\delta$ for $\phi \in (-1/(\delta - 1), 1)$. For $\delta < 0$, the relevant version of condition (28) is

$$\kappa \leq \frac{\delta - 1}{\delta}. \tag{A53}$$

For $\phi \in (-1/(\delta - 1), 1)$, it can never be satisfied. Put differently, the condition in (A52) can be satisfied only if $\phi \in [0, -1/(\delta - 1)]$. For ϕ in this range, (A51) implies $\kappa = 1/[1 + \phi(\delta - 1)]$ and hence that (A52) is satisfied if $\phi(\delta - 1)^2 \leq 1$. Summarizing, when $\delta < 0$, the condition in Proposition 3 is satisfied if and only ϕ satisfies

$$\phi(\delta - 1)^2 \leq 1. \tag{A54}$$

As we have $\kappa = 1/[1 + \phi(\delta - 1)]$ for ϕ in this range, the claim $\beta = \kappa b$ in Proposition 3 follows from rewriting (17) in the main text as $b = [1 + \phi(\delta - 1)]\beta$ once again.

It remains to be shown that if the condition in Proposition 3 is not satisfied, then the parameters b, η, and δ do not allow us to determine which of the two solutions for q in (26) in the main text identifies β. Consider first the case $\delta > 2$. In this case, κ as defined in (27) is given by (A48). To capture that κ in (A48) is a function of ϕ, we use the notation $\kappa(\phi)$. It is straightforward to establish that if $\delta > 2$ and the value of κ' implied by the two solutions for q in (26) satisfies $\kappa' < (\delta - 1)/\delta$ – that is, the condition relevant for the case $\delta > 2$ in Proposition 3 (also stated in (A49)) is not satisfied – then the equation $\kappa(\phi) = \kappa'$ has two solutions for ϕ that satisfy $\phi \in [0, 1)$. Moreover, one of the two solutions for ϕ is smaller than $1/(\delta - 1)$ and the other solution for ϕ is larger than $1/(\delta - 1)$. As a result, $\beta = \kappa b$ for one of the solutions (the solution for ϕ smaller $1/(\delta - 1)$) and $\beta = (1 - \kappa)b$ for the other solution. As both solutions for q in (26) are consistent with the parameters b, η, and δ, and both solutions yield that the implied ϕ satisfies $\phi \in [0, 1)$, it is impossible to know which of the two solutions for q in (26) identifies β. The proof for the case $\delta < 0$ is analogous.

6 Proof of Proposition 4

In proving Proposition 3 we have shown that the condition in (28) holds if and only if $(\delta - 1)^2\sigma^2/\sigma^2_{US} \leq 1$.

7 Proof of Proposition 5

From Proposition 4, we know that the condition in (28) is not satisfied if and only if $\phi(\delta - 1)^2 > 1$. In these circumstances we only know that β is one of the two solutions for q in (26), that is $\beta \in \{q_1, q_2\}$. As $q_1 + q_2 = b$, this implies that $\beta/b \in \{q_1/(q_1 + q_2), q_2/(q_1 + q_2)\}$. Or, making use of the definition for κ in (27) in the main text, $\beta/b \in \{\kappa, 1 - \kappa\}$.

When $\delta > 2$, it follows from (A47) that for $\phi(\delta - 1)^2 > 1$ or, equivalently, for $\phi \in (1/(\delta - 1)^2, 1)$: $\kappa < (\delta - 1)/\delta$. This in turn implies that $1 - \kappa > 1/\delta$. As $(\delta - 1)/\delta > 1/\delta$ when $\delta > 2$, it follows that $\beta/b \in \{\kappa, 1 - \kappa\}$ implies $\beta/b \in (1/\delta, (\delta - 1)/\delta)$. This establishes the part of the proposition that applies to $\delta > 2$.

When $\delta < 0$, it follows from (A51) that for $\phi(\delta - 1)^2 > 1$ or, equivalently, for $\phi \in (1/(\delta - 1)^2, 1)$: $\kappa > (\delta - 1)/\delta$. This in turn implies that $1 - \kappa < 1/\delta$. As $(\delta - 1)/\delta > 1/\delta$ when $\delta < 0$, it follows that $\beta/b \in \{\kappa, 1 - \kappa\}$ implies $\beta/b \notin [1/\delta, (\delta - 1)/\delta]$. This establishes the part of the proposition that applies to $\delta < 0$.