Essays on Firm Performance, Innovation, and Cross-Border Economic Activity

A dissertation submitted to the H. John Heinz III College in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Public Policy and Management

by

Matej Drev

Dissertation Committee:

Ashish Arora

Lee G. Branstetter (Chair)

Brian Kovak

Carnegie Mellon University

Pittsburgh, Pennsylvania

May 2013

Table of Contents

Introduction

This thesis is composed of three essays that explore different facets of firm performance, innovation, and cross-border economic activity.

The first essay documents a systematic shift in the nature of innovation in information technology (IT) towards increasing dependence on software. Using a broad panel of US and Japanese publicly listed IT firms in the period 1983-2004, it shows this change in the nature of IT innovation had differential effects on the performance of the IT industries in the United States and Japan, resulting in US firms increasingly outperforming their Japanese counterparts, particularly in more software-intensive sectors. It also provides suggestive evidence that human resource constraints played a role in preventing Japanese firms from adapting to the documented shift in IT innovation.

The second essay asks whether the United States have a comparative advantage in applications-related software research. It classifies software patents into downstream and upstream software inventions based on a unique classification algorithm, then offers empirical evidence that downstream software research is disproportionally concentrated in the United States, and that U.S. firms are significantly less likely to locate downstream software research projects offshore than upstream research projects. It also explores self-citation and co-invention patterns of software patents and provides suggestive evidence that U.S. firms may use intra-firm knowledge flows to mitigate challenges of conducting downstream software research remotely. Finally, it explores the sources for the observed U.S. advantage in downstream software research and provides initial empirical evidence supporting the hypothesis that it is at least partially due to the relative abundance of lead users of software within the United States.

The final essay uses a rich panel dataset of Slovenian firms in the period 1994-2010 to examine how receiving foreign investment impacts the subsequent performance and behavior of local firms. Using a variety of propensity score based estimation techniques, it shows that foreign investment leads recipient firms to subsequently significantly expand the scale and scope of their activities. In addition, the essay explores how heterogeneity in investor origin modulates the effects of foreign investment, and it shows that investor origin heterogeneity is indeed important for understanding local firms' ex post performance, the scale of their operations, the scope of their product mix and their geographical presence in export markets. It finds, for instance, that firms receiving investment from advanced country investors subsequently broaden the scope of their product mix and the number of export destinations they serve, while those receiving investment from developing country investors decrease their scope in terms of product space and geographical coverage. The empirical analysis is motivated with a theoretical model in which local firms endogenously chose their product mix and export destinations. The model details how receiving foreign investment affects the way firms alter their ex-post behavior, and then shows that predictions of the model align closely with the empirical results. The findings in this essay suggest that incorporating investor heterogeneity and the multi-product and multi-destination nature of firms yields important insights for furthering our understanding of how foreign investment impacts recipient firms.

Acknowledgements

The research in the first essay was supported by the Software Industry Center at Carnegie Mellon University and benefitted from the research assistance of Ms. Kanako Hotta of UCSD. I acknowledge with gratitude useful comments from Hiroyuki Chuma, Anthony D'Costa, Kyoji Fukao, Shane Greenstein, Susumu Hayashi, Takao Kato, Toshiaki Kurokawa, Mark Kryder, Koji Nomura, Robert Cole, Jeffrey Smith, David Weinstein and participants in the 2009 Spring Meeting of the NBER Productivity Program, the 2009 NBER Japan Project Conference, and the September 2009 Meeting of the Japan Economic Seminar at Columbia University.

The research in the third essay was made possible by data made available for this research project by the Statistical Office of the Republic of Slovenia and the Bank of Slovenia. I would like to thank Natasa Zidar of the Statistical Office of the Republic of Slovenia and Eva Sencar-Svetina of the Bank of Slovenia for their help throughout the data acquisition process.

Essay 1: "Going Soft: How the Rise of Software Based Innovation led to the Decline of Japan's IT Industry and the Resurgence of Silicon Valley"

(with Ashish Arora and Lee G. Branstetter)

Introduction

The surge of innovation in Information Technology (IT) is one of the great economic developments of the last two decades. This period also coincides with the unexpected resurgence of the United States IT sector, belying the gloomy predictions about the US IT industry popular in the late 1980s and early 1990s (e.g. Cantwell, 1992; Arrison and Harris, 1992). In this essay, we argue that these two developments are closely related.

We present evidence that the IT innovation process is increasingly software intensive: non-software IT patents are significantly more likely to cite software patents, even after controlling for the increase in the pool of citable software patents. We also see substantial differences across IT sub-sectors in the degree to which innovation is software intensive. We exploit these differences to sharpen our empirical analysis.

If the innovation process in IT has indeed become more dependent on software competencies and skills, then firms better able to use software advances in their innovation process will benefit more than others. Indeed, we argue that the shift in software intensity of IT innovation has differentially benefited American firms over their Japanese counterparts. Our results from a sizable unbalanced panel of the largest publicly traded IT firms in US and Japan for the period 1983-2004 show that US IT firms have started to outperform their Japanese counterparts, both as measured by productivity of their innovative activities, and as measured by the stock market valuation of their R&D.[1]

The timing and the concentration of this improvement in relative performance appears to be systematically related to the software intensity of IT innovation. We show that the relative strength of American firms tends to grow in the years after the rise in software intensity had become well established. Furthermore, the relative improvement of the U.S. firms is greatest in the IT sub-sectors in which the software intensity of innovation is the highest. Finally, much of the measured difference in financial performance disappears when we separately control for the software intensity of IT innovation at the firm level.

Why were U.S. firms better able to take advantage of the rising software intensity of IT innovation? Bloom et al. (forthcoming) argue that superior American management allows U.S. multinationals to derive a greater productivity boost out of a given level of IT investment than their European rivals. In the context of our study, we find evidence that the openness of America's labor market to foreign software engineers may have played a key role in alleviating for American firms what was likely to have been a global shortage of skilled software engineers during the 1990s. When Japanese firms undertake R&D and product development in the U.S., it appears to be much more software intensive than similar activity undertaken in Japan. These results highlight the importance of local factor market conditions in shaping the geography of innovation.

This essay is structured as follows. Section II documents the existence of a shift in the technological trajectory of IT, Section III empirically explores its implications for innovation performance of US and Japanese IT firms, and Section IV discusses the possible explanations for the trends we observe in our data. We conclude in Section V with a summary of the key results and suggestions for future work.

The Changing Technology of Technological Change in IT

A survey of the computer and software engineering literature points to an evident increase in the role of software for successful innovation and product development in the IT industry. The share of software costs in product design has increased steadily over time (Allan et al, 2002) and software engineers have become more important as high-level decision-makers at the system design level in telecommunications, semiconductors, hardware, and specialized industrial machinery (Graff, Lormans, and Toetenel, 2003). Graff, Lormans, and Toetenel (2003) further argue that software will increase in importance in a wide range of products, such as mobile telephones, DVD players, cars, airplanes, and medical systems. Industry observers claim that software development and integration of software applications has become a key differentiating factor in the mobile phone and PDA industry (Express Computer, 2002). A venture capital report by Burnham (2007) forcefully argues that that the central value proposition in the computer business has shifted from hardware to systems and application software.

Similarly, De Micheli and Gupta (1997) assert that hardware design is increasingly similar to software design, so that the design of hardware products requires extensive software expertise. Gore (1998) argues that peripherals are marked by the increasing emphasis on the software component of the solution, bringing together hardware and software into an integrated environment.[2] Kojima and Kojima (2007) suggest that Japanese hardware manufacturers will face increasing challenges due to the rising importance of embedded software in IT hardware products. In sum, there is broad agreement among engineering practitioners and technologists that software has become more important in IT. In the next section, we validate this assertion formally, using data on citation patterns of IT patents.

Measuring the Shift in the Technology of Technological Change in IT

Approach

If innovation in IT truly has come to rely more heavily on software, then we should observe that more recent cohorts of IT patents cite software technologies with increasing intensity, and this should be the case even when we control for the changes over time in the volume of IT and software patenting. We therefore use citations by *non-software* IT patents to software patents as a measure of the software intensity of IT innovation.

Patents have been used as a measure of innovation in mainstream economic research at least since the early 1960s. Though subject to a variety of limitations, patent citations are frequently used to measure knowledge flows (Jaffe and Trajtenberg, 2002). Following Caballero and Jaffe (1993) and Jaffe and Trajtenberg (1996, 2002), we use a citation function model in which we model the probability that a particular patent, *p*, applied for in year *t*, will cite a particular patent, *P*, granted in year *T*. This probability is determined by the combination of an exponential process by which knowledge diffuses and a second exponential process by which knowledge becomes superseded by subsequent research (Jaffe and Trajtenberg, 2002). The probability, Pr(*p,P*), is a function of the attributes of the citing patent p and the the cited patent P, $\alpha(p, P)$, and the time lag between them (t-T), as depicted below:

$$\Pr(p,P) = \alpha(p,P)\cdot \exp(-\beta_1(t-T)\cdot(1-\exp(-\beta_2(t-T)) \qquad (1)$$

We sort all potentially citing patents and all potentially cited patents into cells corresponding to the attributes of patents. The attributes of the citing patents comprise the citing patent's grant year, its geographic location, and its technological field (IT, software). The attributes of the cited patents are the cited patent's grant year, its geographic location, and its technological field. Thus, the expected number of citations from a particular group of citing patents to a particular group of cited patents can be expressed as the following:

$$E(c_{abcdef}) = n_{abc} \cdot n_{def} \cdot \alpha_{abcdef} \cdot \exp(-\beta_1(t-T) \cdot (1-\exp(-\beta_2(t-T)) \qquad (2)$$

where the dependent variable measures the number of citations made by patents with grant year (a), geographic location (b), and technological field (c) to patents with grant year (d), geographic location (e), and technological field (f). The alpha terms are multiplicative effects estimated relative to a benchmark or "base" group of citing and cited patents, and n_{abc} and n_{def}. is the number of patents in the respective categories. Rewriting equation (2) gives us the Jaffe – Trajtenberg (2002) version of the citation function, expressing the average number of citations from one category patent to another:

$$p(c_{abcdef}) = \frac{E(c_{abcdef})}{n_{abc} \cdot n_{def}} = \alpha_{abcdef} \cdot \exp(-\beta_1(t-T) \cdot (1-\exp(-\beta_2(t-T)) \qquad (3)$$

Adding an error term, we can estimate this equation using the nonlinear least squares estimator. The estimated equation thus becomes the following:

$$p(c_{abcdef}) = \alpha_a \cdot \alpha_b \cdot \alpha_c \cdot \alpha_d \cdot \alpha_e \cdot \alpha_f \cdot \exp(-\beta_1(t-T) \cdot (1-\exp(-\beta_2(t-T)) + \varepsilon_{abcdef} \qquad (4)$$

In estimating equation (4) we adjust for heteroskedasticity by weighting the observations by the square root of the product of potentially cited patents and potentially citing patents corresponding to the cell, that is

$$w = \sqrt{(n_{abc}) \cdot (n_{def})} \qquad (5)$$

Data

We use patents granted by the United States Patent and Trademark Office (USPTO) between 1983 and 2004. We use the geographic location of the first inventor to determine the "nationality" of the patent. We identify IT patents, broadly defined, using a classification system based on USPTO classes, developed by Hall, Jaffe, and Trajtenberg (2001). They classified each patent into 36 technological subcategories. We applied their system and identified IT patents as those belonging to any of the following categories: computers & communications, electrical devices, or semiconductor devices. We obtained these data from the most recent version of the NBER patent dataset, which covers patents granted through the end of 2006.

Next, we identified software related patents, which is a challenge in itself. There have been three significant efforts to define software patents. Graham and Mowery (2003) defined software patents as an intersection of those falling within a narrow range of International Patent Classification (IPC) classes and those belonging to packaged software firms. This created a sample that omitted large numbers of software patents, according to Allison et al, (2006).

The second effort was that of Bessen and Hunt (2007), who defined a software invention as one in which the data processing algorithms are carried out by code either stored on a magnetic storage medium or embedded in chips. They rejected the use of official patent classification systems, and used a keyword search method instead. They identified a small set of patents that adhered to their definition, and then used a machine learning algorithm to identify similar patents in the patent population, using a series of keywords in the patent title and abstract. Recently, Arora et al. (2007) used a similar approach that connects the Graham-Mowery and Bessen-Hunt definitions.[3]

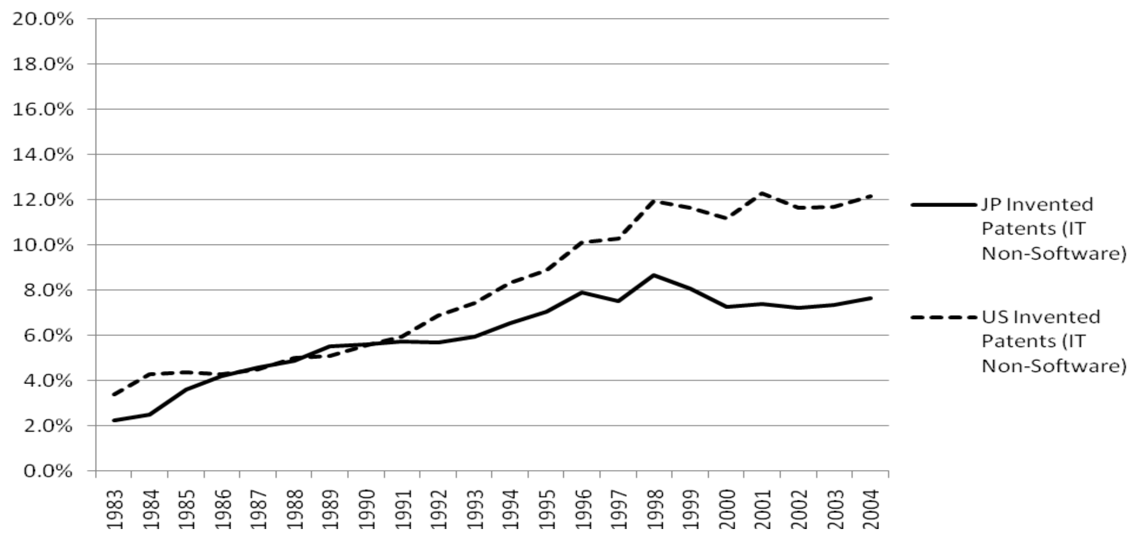
We used a combination of broad keyword-based and patent class strategies to identify software patents. First, we generated a set of patents, granted after January 1st 1983 and before December 31st 2004 that used the words "software" or "computer program" in the patent document. Then, we defined the population of software patents as the intersection of the set of patents the query returned and IT patents broadly defined as described above, granted in the period 1980-2006. This produced a dataset consisting of 106,379 patents.

These data are potentially affected by a number of biases. Not all inventions are patented, and special issues are raised by changes in the patentability of software over the course of our sample period – making it all the more important to control for the expansion in the pool of software patents over time, as we do. We also rely on patents generated by a single authority – the USPTO – to measure invention for both U.S. and Japanese firms. However, Japanese firms have historically been among the most enthusiastic foreign users of the U.S. patent system. Evidence suggests that the U.S. patents of Japanese firms are a reasonably accurate proxy of their inventive activity (Branstetter, 2001; Nagaoka, 2007). This is particularly true in IT, given the importance of the U.S. market in the various components of the global IT industry.

Results

Figure 1 shows trends over time in the fraction of total (non-software) IT patents' citations going to software patents. While the trends for both Japanese and U.S. firms rise significantly over the 1990s, then level off a bit in the 2000s, the measured gap between Japanese and U.S. firms rises substantially over the period. A one-tailed t-test reveals that these differences are statistically significant at conventional levels for every year of interest. However, this analysis does not take into account a variety of other factors, thus we turn next to parametric analysis.

Figure 1: Software Intensity of Non-Software IT Patents, Fraction of IT Patent Citations Made to Software Patents



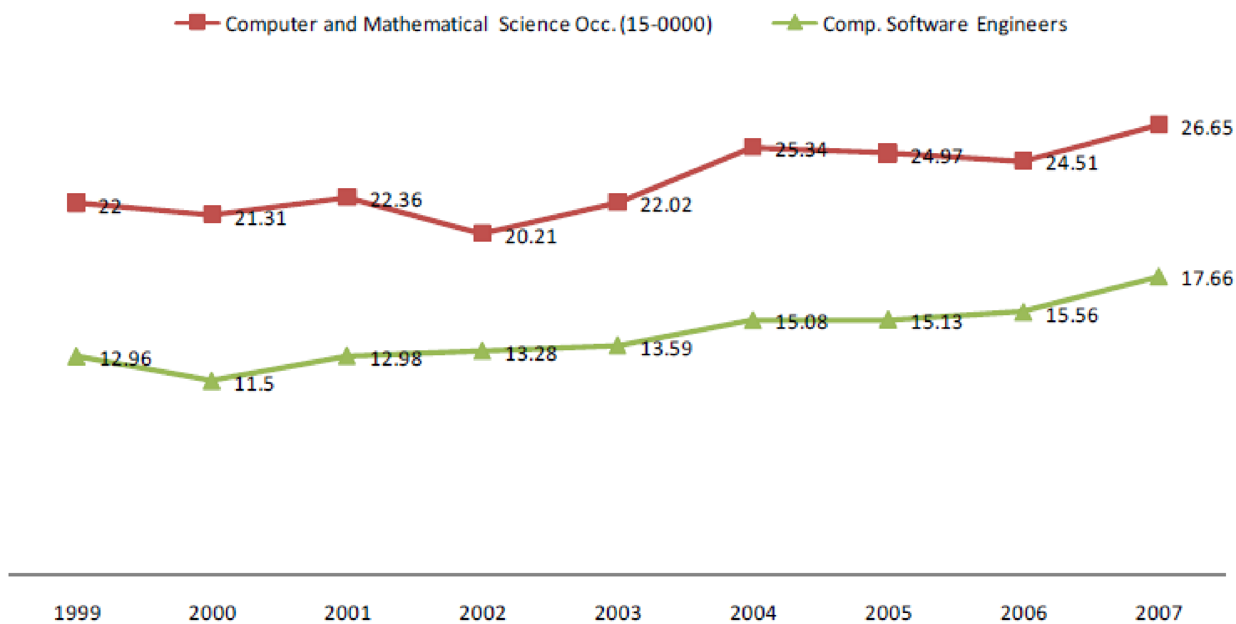
The unit of analysis in Table I is an ordered pair of citing and cited patent classes. Our regression model is multiplicative, so a coefficient of 1 indicates no change relative to the base category. Our coefficients are reported as deviations from 1. The software patent dummy is large, positive, statistically significant, and indicates that IT patents in the 1990s are 9.42 times more likely to cite software patents than prior IT patents, *controlling for the sizes of available IT and software patent pools*. The second specification in Table I includes only software patents in the population of possibly cited patents. The coefficients on the citing grant years show a sharp increase in citation probabilities from 1991 to 2003. An IT patent granted in 1996 is 1.85 times more likely to cite a software patent than an IT patent granted in 1990. Furthermore, an IT patent granted in 2003 is almost 3.2 times more likely to cite a software patent than that granted in 1990. Comparing this trend to that of the specification in the left-hand column of Table I, we see that this trend is much more pronounced, suggesting that software patents are becoming increasingly important for IT innovation. In Table I, we also explore citation differences between

Japanese and non-Japanese invented IT inventions. The specification in the left-hand column indicates that Japanese invented IT patents are 31 percent less likely to cite other IT patents than non-Japanese IT patents. However, they are also much less likely to cite software patents than non-Japanese IT patents. This result is corroborated by the regression in the right-hand column, where the coefficient on the Japanese dummy again shows that Japanese invented IT patents are significantly less likely to cite software patents than non-Japanese patents.

The citation function results were subjected to a number of robustness checks. Concerned that our results might be driven by large numbers of U.S.-invented software patents appearing in the more recent years of our sample, we estimated the propensity of U.S. IT patents to cite software patents generated outside the U.S. and found a rise in this propensity qualitatively similar to that depicted in Table 1. We also directly controlled for the disproportionately high likelihood that patents cite patents from the same country, but our result that Japanese IT hardware patents are systematically less likely to cite software over time was robust to this. Finally, concerned that this result might be observed at least partially due to traditionally stronger university-industry ties in the United States[4], we also estimated a version of the citations function in which we excluded all university-assigned patents and those citing them, and found our results to be robust to this as well.

The U.S. Bureau of Labor Statistics data on U.S. employment by occupation and industry from 1999-2007[5] reveal trends consistent with a rising importance of software in IT innovation. For instance, Figure 2 illustrates how two measures of the share of software engineers in total employment in the computer and peripheral equipment manufacturing industry have trended upward over time. We see similar trends in other IT subsectors as well. The share is highest in computers and peripherals, lowest in audio and visual equipment manufacturing, and at intermediate levels in semiconductors. Interestingly, the relative share of software engineers in total employment across subsectors appears to accord with patent citation-based measures of software intensity.

Figure 2: Trends in Software Engineering Employment



Source: Bureau of Labor Statistics, Occupational Employment Survey, 1999-2007

Note: Data include domestically employed H1-B Visa holders

Comparing US and Japanese Firm-Level Innovation Performance in IT

Our citation function results suggest that there has been a shift in the nature of technical change within IT – invention has become much more software intensive. Our results also suggest that U.S. firms have more actively incorporated software into their inventive activity than have Japanese firms. If this is true, then it is reasonable to expect that changes in the relative performance of Japanese and American firms may be related to the software intensity of the industry segments in which they operate. In segments of IT where innovation has become most reliant on software, we should expect to see American firms improve their relative innovation performance relative to Japanese firms. In segments of IT where innovation does not draw heavily on software, we would expect less of an American resurgence. As we shall see, two very different measures of relative performance show exactly this pattern.

We use two of the most commonly employed empirical approaches to compare firm-level innovation performance of US and Japanese IT firms: the innovation (patent) production function and the market valuation of R&D. While the former approach relates R&D investments to patent counts and allows us to study the patent productivity of R&D, the second approach relates R&D investment to the market value of the firm and explores the impact of R&D on the value of the firm (Tobin's Q).

Patent Production Function

This approach builds on Pakes and Griliches (1984) and Hausman, Hall, and Griliches (1984). We use a log-log form of the patent production function.

$$P_{it} = r_{it}^{\beta} \phi_{it} e^{\varphi JP_i} \tag{6}$$

where

$$\phi_{it} = e^{\sum_c \delta_c D_c} \tag{7}$$

In equation (6), P_{it} are patents taken out by firm i in period t, r_{it} are research and development expenditures, JP_i indicates if the firm is Japanese, and Φ's represent innovation-sector-specific technological opportunity and patenting propensity differences D across c different innovation sectors as specified in (7). Substituting (7) into (6), taking logs of both sides, and expressing the sample analog we obtain the following:

$$p_{it} = \beta r_{it} + \sum_c \delta_c D_c + \varphi JP_i + \mu_{it} \tag{8}$$

where p_{it} is the natural log of new patents (flow) and the error term which is defined below.

$$\mu_{it} = \xi_i + u_{it} \tag{9}$$

We allow the error term in (9) to contain a firm-specific component, ξ_i, which accounts for the intra-industry firm-specific unobserved heterogeneity, and an *iid* random disturbance, u_{it}. The presence of the firm-specific error component suggests using random or fixed effect estimators. Since the fixed effects estimator precludes time-invariant regressors, including the firm origin indicator, we feature the pooled OLS and random effects estimators, and use the fixed effects estimator as a robustness check.

Private Returns to R&D and Tobin's Q

Griliches (1981) pioneered the use of Tobin q regressions to measure the impact of R&D on a firm's economic performance (see Hall (2000) for a detailed review). We can represent the market value V of firm i at time t as a function of its assets:

$$V_{it} = f(A_{it}, K_{it}) \tag{10}$$

where A_{it} is the replacement cost of the firm's tangible assets, typically measured by their book value, and K_{it} is the replacement value of the firm's technological knowledge, typically measured by stocks of R&D expenditures[6]. We follow the literature, which assumes that the different assets enter into the equation additively:

$$V_{it} = q_t (A_{it} + \beta * K_{it})^{\sigma} \tag{11}$$

where q_t is the average market valuation coefficient of the firm's total assets, β is the shadow value of the firm's technological knowledge measuring the firm's private returns to R&D, and σ is a factor measuring returns to scale. Again, following standard practice in the literature (e.g. Hall and Oriani, 2006), we assume constant returns to scale (σ = 1). Then, by taking natural logs on both sides of (11) and subtracting ln A_{it}, we obtain the following expression that relates a firm's technological knowledge to its value above and beyond the replacement cost of its assets, Tobin's Q:

$$\ln Q_{it} = \ln(\frac{V_{it}}{A_{it}}) = \ln q_t + \ln\left[1 + \beta_t * \left(\frac{K_{it}}{A_{it}}\right)\right] \tag{12}$$

Following Hall and Kim (2000) and others, we estimate a version of (12) using the nonlinear least squares estimator, with time dummies and a firm origin indicator. We were unable to estimate a specification with firm-fixed effects because the NLS algorithms did not converge. As a robustness check, we estimated a linearized version of (12) with fixed effects.

Data and Variables

Sample

Our sample consists of large publicly traded IT companies in the United States and Japan, observed from 1983 to 2004.[7] We obtained the sample of US firms from historical lists of constituents of Standard & Poor's (S&P) US 500 and S&P 400 indices. The resulting set of firms was refined using Standard & Poor's Global Industry Classification Standard (GICS) classification[8] so that only firms appearing in "electronics", "semiconductors", "IT hardware" and "IT software and services" categories remained in the sample. This initial set of approximately 290 firms was narrowed further as follows: (a) only firms with least 10 patents in between 1983-2004 were retained, (b) US firms in "IT software and services" were removed to achieve compatibility,[9] and (c) only firms for which at least 3 consecutive years of R&D investment and sales data were available were kept in the sample. This yielded an unbalanced panel of 133 US IT firms.

The initial sample of 154 large publicly traded Japanese IT firms derived from the Development Bank of Japan (DBJ) database[10] was supplemented by an additional 34 firms included in Standard & Poor's Japan 500 index as of January 1st 2003[11] that belong to either "electronics", "semiconductors", "IT hardware", or "IT software and services".

We winnowed the sample by (a) dropping all firms without at least 10 patents in the observed period, (b) dropping Nippon Telephone and Telegraph, and most significantly, (c) all firms for which at least three consecutive years of R&D investment and positive output data were not available. This produced a final sample of 77 Japanese IT firms.

Collectively, the Japanese and U.S. firms in our sample accounted for over 70% of total U.S. IT patenting by Japanese and U.S. firms, respectively, in the late 1990s and early 2000s, confirming that we are capturing a large majority of private sector innovative activity in this domain.[12]

Locating Firms in Software Intensity Space

To explore how innovation performance differentials between US and Japanese firms vary with software intensity, we classify firms into industry segments. GICS provided us with a classification of US firms in our sample into four sectors – "electronics", "semiconductors", "IT hardware", and "IT software and services". Japanese firms were classified manually using the two-digit GSIC classification data from the S&P Japan 500 along with data from Japan's Standard Industrial Classification (JSIC), supplemented by data from Google Finance, Yahoo! Finance and corporate websites.

We construct two separate measures of software intensity, both of which suggest a similar ranking of IT subsectors. First, we use the shares of software patents in total patents taken out by the firms, averaged across firms in an industry category. Second, we calculate the fraction of citations to software patents by non-software IT patents, averaged across firms in a sample category. Table II presents summary statistics for both these measures of software intensity. As expected, electronics is the least software intensive, followed by semiconductors and IT hardware. A two-sided test for the equality of means rejects that the intensities are the same in any pair of sectors when we use the share of software patents as our measure. The second measure, citations to software patents, yields similar results, albeit at lower levels of significance in some cases. Tables III and III-2 calculate the industry averages of our measures of software intensity separately for U.S. and Japanese firms. In general, the ranking of industries in terms of software intensity suggested by the overall sample apply to the country-specific subsamples as well.[13] Japanese firms are disproportionately located in less software intensive sectors, and within those sectors, are less software intensive than their US counterparts.

Taking the assignment of firms to the different IT industries as given[14], we test whether US firms outperform Japanese firms, and whether this performance gap is more marked in IT industries that are more software intensive.

Construction of Variables

Patent Counts: Patent data for our sample of firms were collected from the updated NBER patent dataset containing patents granted by the end of 2006. Compustat firm identifiers were matched with assignee codes based on the matching as constructed and available on the NBER's Patent Data Project website.[15] The matching algorithm for Japanese firms was based on a Tokyo Stock Exchange (TSE) code - assignee code concordance previously used in Branstetter (2001), but was manually updated by matching strings of firm names and strings of assignee names as reported by the USPTO.

R&D Investment: Annual R&D expenditure data for US firms were collected from Compustat, and a set of self-reported R&D expenditure data for Japanese firms were collected from annual volumes of the Kaisha Shiki Ho survey.[16] We deflated R&D expenditures following

Griliches (1984), and constructed a separate R&D deflator for US and Japanese firms that weigh the output price deflator for nonfinancial corporations at 0.51 and the unit compensation index for the same sector at 0.49. Using data on wage price indexes for service-providing and goods-producing employees,[17] we constructed a single unit compensation index for each country, and then applied the proposed weights and appropriate producer price indexes to compute the R&D deflators and deflate the R&D expenditure flows.

R&D stocks: We calculated R&D capital stocks from R&D expenditure flows using the perpetual inventory method, with a 15% depreciation rate.[18] We used 5 pre-sample years of R&D expenditures to calculate the initial stocks.[19]

Market Value of the Firm: Market value of a firm equals the sum of market value of its equity and market value of its debt (Perfect and Wiles, 1994). Market value of equity equals the sum of the value of outstanding common stock and the value of outstanding preferred stock. The value of outstanding common (preferred) stock equals the number of outstanding common (preferred) shares multiplied by their price. For US firms, we used year-close prices, year-close outstanding share numbers, and year-close liquidating values of preferred capital. For Japanese firms, the only available share price data were year-low and year-high prices, and we used the arithmetic mean of the two to obtain share price for each firm-year combination. In addition, preferred capital data was not available for Japanese firms, which should not create problems as long as preferred capital does not systematically vary with time and across technology sectors. For market value of debt we used total long-term debt and debt in current liabilities. For Japanese firms, we used fixed liabilities as a proxy for the value of long-term debt and short-term borrowings as a proxy for the value of short-term debt.[20]

Replacement Cost of Assets: The replacement cost of the firm's assets is the deflated year-end book values of total assets[21] where the deflator is a country-specific capital goods deflator obtained from the Bureau of Labor Statistics and the Statistics Bureau of Japan, respectively.

Patent Production Function Results

Figure 3 compares the number of patents per firm for the US and Japanese firms in our sample. We observe that Japanese firms obtain more non-software IT patents than their US counterparts. Between 1983 and 1988, the average number of non-software IT patent applications were almost identical for Japanese and US firms. Between 1988 and 1993, patent applications by Japanese firms outpaced those of US firms, after which both grew at a similar pace. By contrast, Japanese firms file fewer software patents than their US counterparts, and the difference has grown steadily since the late 1980s, and especially after the mid 1990s.

Figure 3: Average Number of non-software IT and Software Patents per Firm

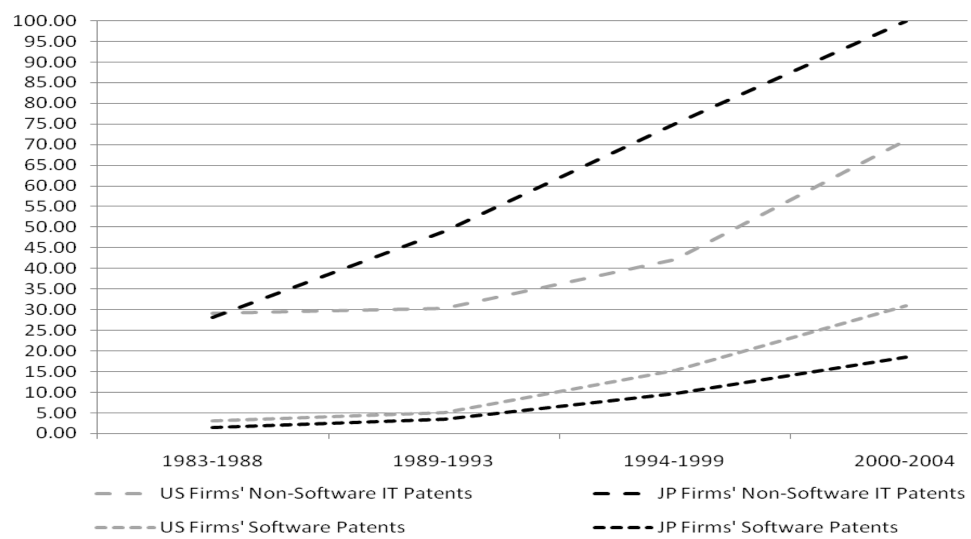
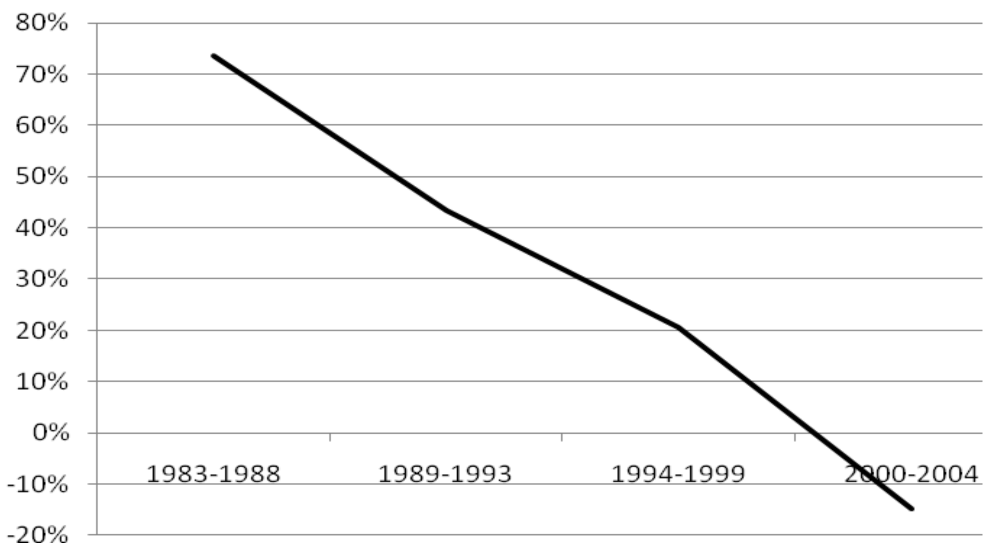


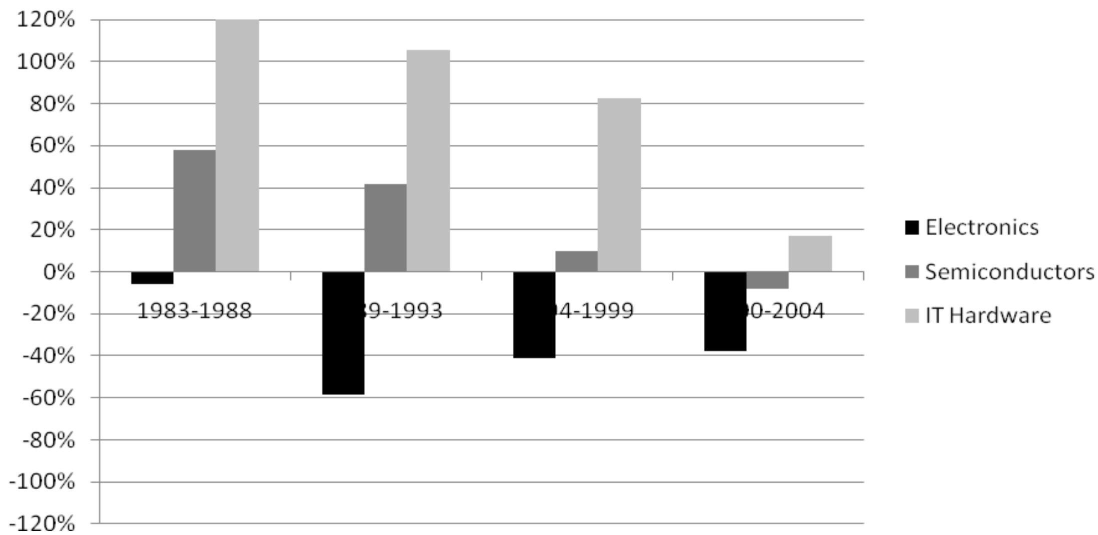
Table V reports the estimates of the patent production functions of U.S. and Japanese IT firms. Our first key result is presented in Figure 4 below, which plots the pooled OLS average difference in log patent production per dollar of R&D, between Japanese and US firms in our sample through time, controlling for time and sector dummies. We see that R&D spending by Japanese firms was 70% more productive than that of their US counterparts during 1983-1988, but became less and less productive from 1989-1993 onwards. This trend accelerated in the 1990s and early 2000s, with Japanese IT firms producing 20% fewer patents, controlling for the level of R&D spending, than their US counterparts in the period 2000-2004.

Figure 4: Average Japan-US Productivity Differences, Entire Sample



Based on results from Table V. Reported are pooled OLS estimation coefficients.

Figure 5: Average Japan-US Productivity Differences, By Software Intensity Sector



Based on results from Table V. Reported are selected pooled OLS estimation coefficients.

Figure 5 reports Japan-U.S. differences in patent output controlling for R&D input by IT sector. In electronics, previously shown to be the least software intensive, and where average software intensity is similar between US and Japanese firms, Japanese firms have been less productive in patent production in the 1980s and early 1990s, but have been catching up to their US counterparts in the mid-to-late 1990s and early 2000s.[22] On the other hand, in semiconductors and IT hardware, which have significantly higher software intensity than electronics, and where average software intensity of US firms is greater than of Japanese firms, Japanese firms exhibited higher productivity in the mid 1980s, started losing their advantage by the turn of the 1990s, and started to lag behind their US counterparts in the mid to end 1990s and early 2000s.[23]

Most of the results in Table V are statistically significant at the 5% level and become more statistically significant in more recent time periods. In addition, the results are robust to changes estimation techniques and measures. Random effects and fixed effects estimates are similar, suggesting that our results are not driven by unobserved firm-specific research productivity or patent propensity differences.. The dependent variable in these estimations is the log of total patents applied for by firm *i* in year *t*. Unreported estimations show that the results are very similar if we use instead the log of IT patents, or the log of IT patents excluding software patents, or if we weight patents by subsequent citations or by the number of claims.

Accounting for Alternative Hypotheses

The collapse of the Japanese bubble economy at the end of the 1980s. The shift in relative performance parallels the slowdown in the Japanese domestic economy at the end of the 1980s. This domestic slowdown could have led to lower levels of R&D expenditure by Japanese firms. However, a simple recession induced decline in R&D investment cannot explain our results. We are estimating the *productivity* of R&D in producing patents, rather than the number of patents produced. If Japanese firms sought cost savings by eliminating marginal R&D projects, measured productivity should be higher, not lower. Budget pressures could have also led Japanese firms to change their patent propensity, filing fewer but higher quality patents outside Japan. However, estimates using citation weighted patents yield results similar to those reported above. More fundamentally, no simple story about a post-bubble slowdown in the domestic economy can explain the observed pattern, wherein the relative decline in productivity is greater in more software intensive segments.

The appreciation of the yen after 1985. The yen appreciated sharply in the mid-1980s and remained much stronger through the mid-to-late 1990s.[24] These exchange rate shifts lowered the international competitiveness of Japan-based manufacturing. However, we do not think that exchange rate shifts are driving our results. All the segments of the Japanese IT industry confronted the same yen-dollar exchange rate, yet the relative innovative performance of the different segments varied in ways that are difficult to explain based on exchange rate considerations alone. For example, the Japanese electronics sector is arguably the one most likely to be affected by an appreciating currency; electronics had a much larger "commodity" share in total output, as compared to semiconductors and hardware. However, it is electronics in which Japan's relative performance strengthened the most.

Strong venture capital in America, weak venture capital in Japan. Kortum and Lerner (2001) provide evidence of the strong role played by venture capital backed firms in the acceleration of innovation in the United States in the 1990s. Recent Japanese scholarship (Hamada, 1996, Goto, 2000, Goto and Odagiri, 2003) stresses the relative weakness of venture capital in Japan as an impediment to the growth of science-based industries. While it is certainly true that new firms adept at software-based innovation entered the market in the mid-to-late 1990s, often with backing from venture capitalists, our results do not depend on their inclusion in the sample. For instance, we get similar results if we remove all U.S. firms that went public after the Netscape IPO, widely regarded as the start of the VC fuelled boom in the U.S.

Strong university-industry linkages in the U.S., weak linkages in Japan. Goto (2000), Nagaoka (2007), and many others have suggested that weaker Japanese universities and weaker mechanisms for university-industry technology transfer impede growth in Japan's science-based industries. We acknowledge the importance of these linkages. However, if university-generated inventions were an important element in the transformation of the U.S. IT sector, then corporate patents citing these university-generated inventions should be especially important in generating our empirical results. We delete all university-owned inventions and all corporate patents citing university-owned inventions from our data; the results do not change.

Technology standards and market dominance. Japanese scholars, such as Tanaka (2003), have suggested that the increasing dominance of U.S. IT firms since the 1990s is driven largely by U.S. ownership of key technology standards in the industry. Though owning a major technology standard may be beneficial, we can delete from our sample all U.S firms that could plausibly be described as owners of a major IT technology standard without altering our results. The most (in)famous standard owner, Microsoft, is never included in the sample: We do not include firms from the packaged software industry, because there are very few publicly traded Japanese firms in that segment. [25] If we were to include the packaged software firms such as Oracle and Google, the productivity differences would be even more favorable to the US.

The same arguments may apply to the decline of one of Japan's important technology standards. Throughout the 1980s, the Japanese firm NEC dominated the sales of personal computers in Japan. NEC pioneered the development of a PC capable of handling Japan's complex written language. The popularity of the NEC standard created a virtuous cycle in which Japanese software firms and game developers focused their efforts on NEC-compatible products, reinforcing NEC's market dominance. In 1991, a consortium led by IBM Japan introduced DOS/V, an operating system that allowed IBM-compatible PCs to handle the Japanese language without any additional IT hardware.[26]

The introduction of this software ended NEC's market dominance, and allowed a new group of firms to gain market share. The firm most obviously affected by DOS/V is NEC, and our results are robust to the exclusion of NEC. Insofar as the introduction of DOS/V reduced R&D by other Japanese IT firms by shrinking their markets, this may be reflected in our Tobin's q results. However, to the extent that this market compression induced firms to reduce R&D

spending, they should have cut the marginal projects first, suggesting, if anything, and increase in R&D productivity rather than the decrease that we see in the data.

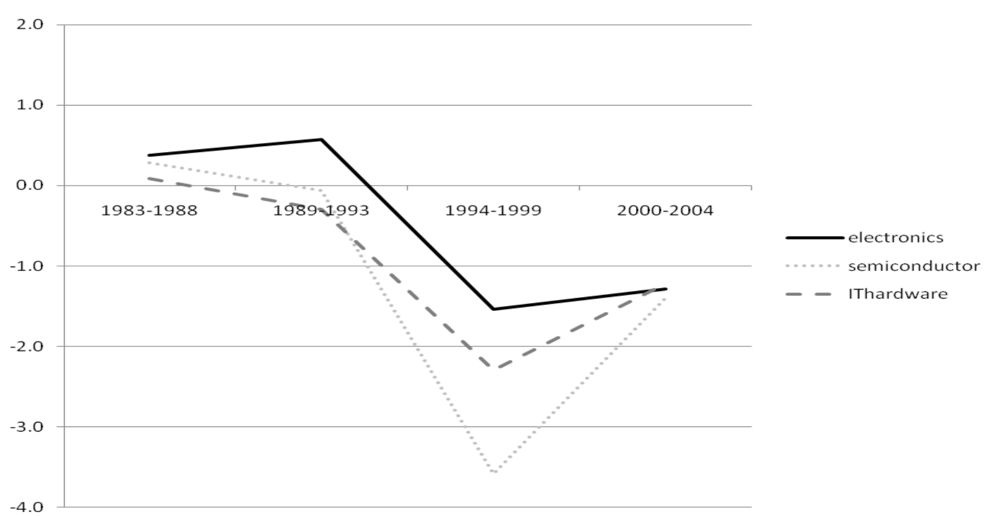
Results Based on Private Returns to R&D

We begin by plotting the average difference in Tobin's Q between our sample of US and Japanese firms through time, shown in Figure 6 below. We observe that Japanese firms, on average, have had higher Q values than US firms in the mid 1980s and early 1990s. These differences diminished with the bursting of the Japanese economic bubble at the dawn of the 1990s, and Japanese Q values have lagged throughout the 1990s, especially in semiconductors, and to a lesser extent, also in IT hardware, before recovering somewhat in the early 2000s with the bursting of the U.S. stock market bubble. Thus trends in average Tobin's Q values generally parallel those in patent production.

Moving beyond the descriptive analysis, we regress Tobin's Q on the ratio of R&D stocks by total assets to estimate private returns to R&D (shadow value of R&D). Table IV reports estimates of equation (12) by period using nonlinear least squares. It shows that the shadow price of R&D/Assets for US firms was close to zero and not statistically significant in most periods, but rose to positive and statistically significant levels by the mid-to-late 1990s. On the other hand, the coefficient on R&D/Assets for Japanese firms has not followed this trend. It has hovered just above zero in the 1980s but dropped significantly by the mid 1990s and early 2000s. In these periods it was much lower than that of US firms, with the difference statistically significant at the 5% level. This is consistent with what we observed when plotting the values of Tobin's Q through time, except that we do not observe much of a positive pullback for Japanese firms in the early and mid 2000s.

Interestingly, this "reversal of fortune" for the market valuation of U.S. firm R&D appears to be sensitive to the inclusion of a direct measure of software intensity. Table IV-2 reports the results of a regression in which we add a variable representing firm-level software intensity, and also interact it with R&D/Assets. This additional regressor significantly alters our results. The R&D/Assets coefficient for U.S. firms is lower than before, while the differences between US and Japanese firms disappear and, in some periods, reverse with the inclusion of an indicator of firm-level software intensity. These results support the view that the relative increase in U.S. performance is related to software intensity.

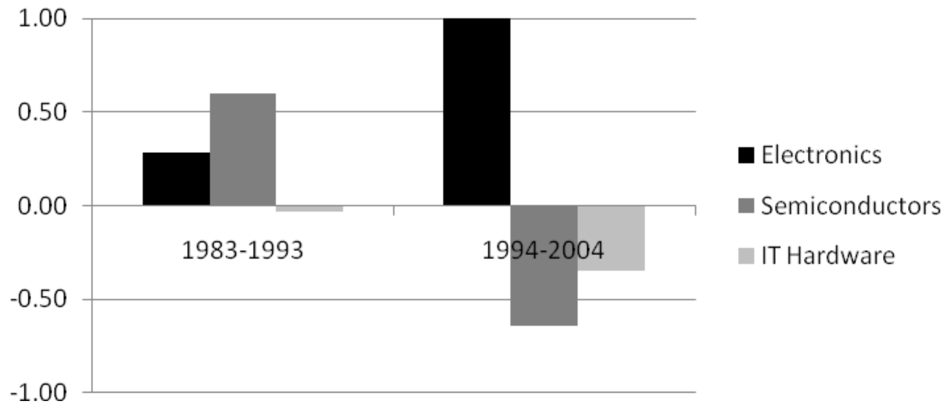
Figure 6: Average Difference in a Raw Measure of Tobin's Q, By Sector



Tobin's Q as calculated in the database, averaged across sector. Calculated as US average subtracted from JP average.

Figure 7 compares private returns to R&D for Japanese and US firms by IT sector. As with patent productivity, we find that results differ by sector. In electronics, the least software intensive sector, the Japanese firms started off with a small advantage in the 1980s, before increasing it substantially by the mid 1990s. The reverse is true in IT hardware, the most software-intensive sector. We report detailed regression results in Tables VII-VIII .[27]

Figure 7: Average Difference in Private Returns to R&D, By Sector



Shadow values of R&D as estimated by OLS/FE in Table VII. Calculated as US average subtracted from JP average.

We conducted several robustness checks. We first estimated versions of (12) using NLS and FE estimators, where we directly estimated time trends for private returns to R&D separately for US and Japanese firms. Table VI shows that the direction of the trends remains unperturbed. Private returns to R&D for Japanese firms linger, as before, around 0, and show a slight negative trend over time, while private returns to R&D for US firms show a marked and statistically significant positive trend. In Tables VII-VIII, we report both estimates of the linear approximation using firm fixed effects and estimates obtained using nonlinear least squares. Again, we observe that the signs of the coefficients remain qualitatively unchanged.

As in the previous section, we consider our results alongside alternative explanations. We estimated versions of (12) by excluding VC-backed entrants from our sample, and found little qualitative change in our results. Similarly, we re-estimated our regressions by excluding firms who owned major technological standards during the sample period (as well as to the exclusion of NTT), and again found little change in our results.

In order to directly test the robustness of our results to changes in industry group assignment of firms, we estimated a linearized version of the regression where we assigned firms in our sample into groups of the same sizes as those suggested by the industry classification, but based on both firm-level shares of software patents and firm-level shares of citations directed towards software patents. We found our results to be qualitatively robust to this exercise that allowed us to estimate the regressions without imposing possibly restrictive assumptions about firm industry assignments. Finally, we estimated a version where we split US and Japanese firms into quartiles according to the firm-level share of software patents in total patents. We observe that US firms' private returns to R&D increase with software intensity, while they fall in the case of Japanese firms. Interestingly, we also observe that US firm's private returns to R&D increase with the software intensity of the sector when they are also in the top quartile of software intensity. The same is true for Japanese firms. Conversely, private returns to R&D decrease with the software intensity of the sector for firms located in the bottom quartile of software intensity.

This essay is focused on innovation in the IT sector and the market returns to IT innovation in that sector, rather than IT production. However, our findings are consistent with reported industry-level productivity trends. Specifically, Jorgenson and Nomura's (2007: p 26, fig 9) show that in both computers and electronic components, an initially more productive Japanese industry is sharply overtaken by its U.S. counterpart in TFP over the course of the 1990s.[28]

Discussion

This essay documents three facts. First, IT innovation has become more software intensive. Second, Japanese firms rely less on software knowledge in IT hardware invention than their US counterparts (and produce significantly fewer software inventions). Third, the innovation performance of Japanese IT firms is increasingly lagging behind, particularly in software intensive sectors. Together, they point to a link between the changing technology of technical change in IT and an inability of Japanese firms to respond adequately to the shift..[29]

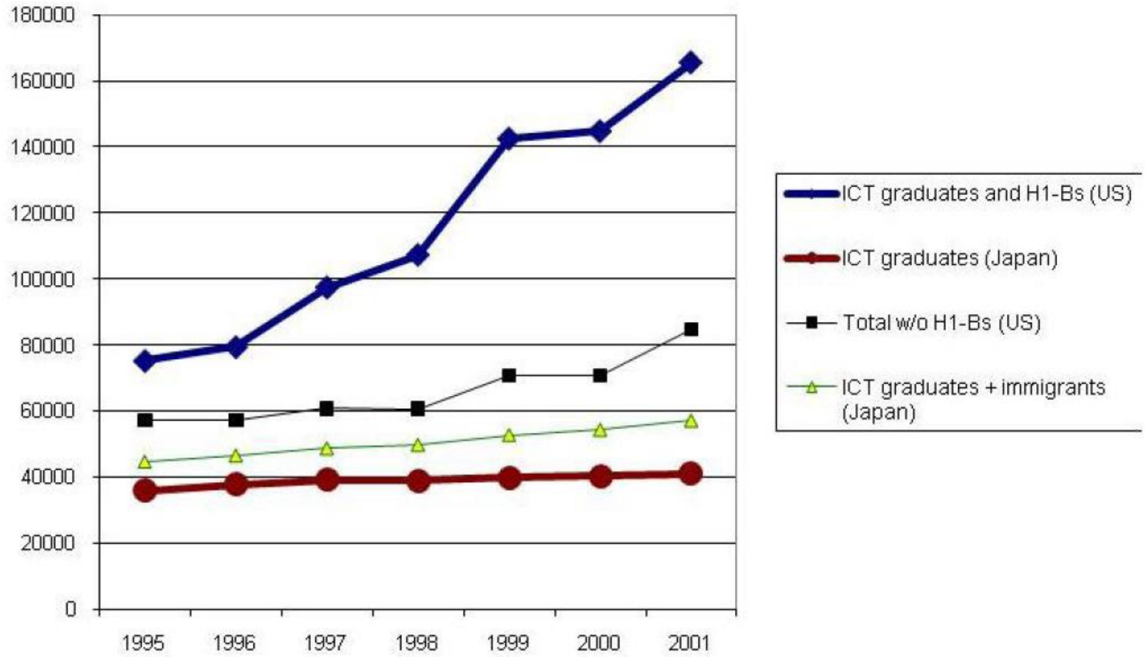
What prevented Japanese firms from using software advances as effectively as U.S. firms? There are at least two explanations. The first is a resource constraint argument: U.S.-based firms have access to a much larger pool of software engineers than do their Japanese counterparts. Japanese firms have not yet been able to overcome their national labor resource constraints by offshoring their software-intensive R&D. The second explanation is one rooted in the failure of Japanese managers to understand and adequately respond to the changing nature of technological change in IT.

Many studies have pointed out the persistent shortages of software engineers in Japan, dating back to the 1970s and 1980s.[30] This longstanding weakness did not prevent Japanese firms from acquiring a strong market position in IT in the 1980s, but it may have become more important as IT hardware product development became steadily more software-intensive.[31] The level of local human capital might not be a constraint if knowledge flowed freely across countries. However, tapping into foreign knowledge pools can be difficult (Jaffe, Trajtenberg, and Henderson 1993), especially for Japanese firms.[32] Belderbos (2001), Odagiri and Yasuda (1997), and Belderbost, Fukao, and Kwon (2006) document the relatively limited extent of Japanese R&D activity outside Japan during the period under consideration. Japan's relatively restrictive immigration laws and its long history as an ethnically homogenous society mitigate against large-scale importation of skilled labor.[33]

The available data make it difficult to precisely quantify the differences in software human resources between the U.S. and Japan, but the gap between the two is clearly large. Figure 8 presents data from several sources comparing the flows of new (potential) domestic IT workers during the crucial years from the mid-1990s through the early 2000s.[34] .Due to differences in reporting conventions, we aggregate over IT software and hardware related disciplines to produce a count of total IT bachelors, masters, and Ph.D. level graduates for both countries. We use data reported by Lowell (2000) and Kirkegaard (2005) to estimate the number of temporary workers joining the U.S. labor force in "computer-related fields" under the auspices of an H-1B visa. In Figure 8, we assume that half of all foreign workers newly admitted to Japan as "researchers," "engineers," or "intracompany transferees" are employed as IT workers in Japan – a far larger fraction than plausibly holds true in reality.[35]

Figure 8: ICT Human Resources, U.S. vs. Japan

(ICT graduates and H1-B immigrants into computer-related professions, 1995-2001)



Arora, Branstetter, and Drev (2010) describe these data (and their shortcomings) in greater detail.[36] Despite these caveats, the picture painted by Figure 8 is quite striking: the flow into the domestic IT labor pool grew much faster in the U.S. compared to Japan. In 1995, the inflows into the domestic IT labor pool in the U.S. were about 68% greater than those in Japan. By 2001, the inflows in the U.S. were nearly three times bigger than those in Japan, with the difference being driven largely by H-1Bs. In some of the latter years of the sample period, the U.S. was importing more IT specialists per year than it was graduating from all IT-related bachelors, masters, and doctoral programs combined. Of course, firms are not confined to their domestic labor pool. Accounting for the level of software offshoring in the U.S. and Japan is even harder, but the available data suggest that consideration of software offshoring would *significantly increase* the resource gap implied by Figure 8 (Arora, Branstetter, and Drev, 2010).

In other words, imports of workers and software offshoring may have been a critical source of advantage for U.S. based firms. Relatively few of these imported experts may have been software architects of the highest order, capable of undertaking transformative innovation. However, creating, testing, and implementing software for IT innovation required both fundamental innovators and programmers undertaking more routine and standardized kinds of software engineering. America's ability to tap into an increasingly abundant (and increasingly foreign) supply of the latter may have raised the productivity of the former and enabled American firms to outpace their rivals. Arora, Branstetter, and Drev (2010) present a simple model in which a more abundant supply of software engineers capable of routine coding and testing raises the productivity of highly skilled software innovators, and show how it could imply results for the relative research productivity of Japanese and U.S. IT firms that are similar to those documented in this essay.

An alternative hypothesis posits that Japan's relative decline in innovative productivity was driven by the failure of Japanese IT managers to appreciate and respond to the rising importance of software in IT product development. A stream of the recent management literature has focused on how managerial mindsets, formed through years of experience, affect the (in)ability of firms to make strategic shifts when firm environments change (Bettis and Hitt, 1995). In the economics literature, Nick Bloom, John Van Reenen, and their co-authors have shown that persistent performance differences across firms based in different countries could be driven by differences in management practices (e.g., Bloom, Sadun, and Van Reenen, forthcoming; Bloom and Van Reenen, 2010; Bloom and Van Reenen, 2007). The papers also show that multinationals tend to bring their management practices, both good and bad, with them when they set up subsidiaries abroad.

Distinguishing Between Possible Hypotheses

These two possible explanations yield different predictions regarding what types of innovative activities Japanese firms should undertake in Japan and abroad. If they are constrained by their software human resources at home, then Japanese firms will have the incentive to tap into foreign knowledge and expertise by setting up software intensive R&D facilities abroad. On the other hand, if differences in relative performance are because Japanese managers downplay or ignore the importance of software, then the research output of Japanese overseas subsidiaries ought also to be less software intensive than their American counterparts.

Because Japanese and U.S. firms conduct IT R&D (and generate patents associated with that activity) at home and in the other country, we can submit these two hypotheses to a test. What we observe is consistent with the resource constraint hypothesis. The share of software patents in total patents invented in Japan by Japanese parent firms in our sample is 6%, as reported in Figure 9-1. However, the share of software patents in total patents invented in the US by Japanese firms is significantly higher – 24%. This surpasses even the share of software patents in total patents invented in the US by US-based IT firms, which is approximately 17%. This suggests Japanese firms are disproportionally likely to engage in software innovation abroad. In addition, as shown in Figure 9-2, patents invented in the U.S. by the subsidiaries of Japanese firms are far more likely to cite software innovation than those invented in Japan -- and they are even more likely to cite software than the comparable patents of U.S.-based firms. As reported in Figures 9-3 and 9-4, these patterns hold when we focus on individual sectors – electronics, semiconductors, IT hardware - but are strongest in IT hardware. It is almost as if Japanese firms are trying to work around the constraints in their home market by choosing a very software-intensive style of innovation in the U.S., where the resources exist to support it.

Bloom et al. (forthcoming) present a compelling case that superior American firm management practices may be important in explaining why American firms *deploy* IT more effectively than their foreign rivals. In this essay, we find evidence that human resource constraints may be important in explaining the success of American firms in *creating* new IT products. In general, the role of international differences in access to human resources and the *interaction* of these differences with local management practices would appear to be an interesting and fruitful area for further research.

Conclusions, Implications and Next Steps

In this essay, we document the existence of a software-biased shift in the innovation process in information technology. Although widely acknowledged in the computer and software engineering literature, this shift has received very little prior attention from economists or management scholars.[37] We provide evidence on the economic importance of this shift by studying how it affected the innovation performance of IT firms in the United States and Japan. We show that this shift has resulted in a deterioration of the relative innovation performance of Japanese firms, and we find that this effect is more pronounced in software intensive sectors. This pattern of relative deterioration and its concentration in software-intensive sectors is robust to controls for the different levels of development of venture capital and formal mechanisms for university-industry technology transfer in the two countries and to controls for disproportionately American ownership of key technology standards. Our findings thus provide a largely new explanation for the precipitous global decline of one of Japan's once leading industrial sectors – another development that has received relatively little attention from mainstream economists.

Finally, we provide evidence that suggests that a constrained supply of software knowledge and skills in Japan might explain the relatively weaker innovation performance of

Japanese IT firms in the 1990s. These findings are particularly interesting in light of a growing literature that explores linkages between factor endowments, technological change, and industry performance (e.g. Acemoglu, 2002; Dudley and Moenius, 2007), and may provide a useful complement to the growing literature that links the superior performance of American firms in some contexts to superior management practices (Bloom and Van Reenen, 2010).

References

Acemoglu, Daron, “Technical Change, Inequality, and the Labor Market,” Journal of Economic Literature 40:1 (2002), 7-72.

Allan, Alan et al., “2001 Roadmap for Semiconductors,” Computer 35:1 (2001), 42-53.

Anchordoguy, Marie, “Japan’s Software Industry: A Failure of Institutions,” Research Policy 29:3 (2000), 391-408.

Arora, Ashish, Lee. G. Branstetter, and Matej Drev, "Going Soft: How the Rise of Software Based Innovation Led to the Decline of Japan's IT Industry and the Resurgence of Silicon Valley," NBER Working Paper #16156 (2010).

Arora, Ashish, and Alfonso Gambardella, “The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labor,” Research Policy 23:5 (1994), 523-532.

Arora, Ashish, Matej Drev, Chris Forman, and Mustafa D. Alpman, “A Note on the Classification of Software Patents,” Unpublished Research Note (2007).

Arrison, Thomas S. et al (Eds.), Japan’s Growing Technological Capability – Implications for the U.S. Economy (Washington, D.C.: National Academy Press, 1992).

Arrison, Thomas S., and Martha Caldwell Harris, “Japan’s Growing Technological Capability and Implications for the U.S. Economy – An Overview,” in Arrison, Thomas S. et al (Eds.), Japan’s Growing Technological Capability – Implications for the U.S. Economy (Washington, D.C.: National Academy Press, 1992).

Arthur, Brian W., “Competing Technologies, Increasing Returns, and Lock-in By Historical Events,” The Economic Journal, 99:384 (1989), 116-131.

Athreye, Suma S., “The Indian Software Industry,” in Arora, Ashish, and Alfonso Gambardella (Eds.), From Underdogs to Tigers: The Rise and Growth of the Software Industry in Brazil, China, India, Ireland, and Israel (Oxford, U.K.: Oxford University Press, 2005).

Belderbos, Rene, “Overseas Innovation by Japanese Firms: An Analysis of Patent and Subsidiary Data,” Research Policy, 30:2 (2001), 313-332.

Bessen, James, and Robert M. Hunt, “An Empirical Look at Software Patents,” Journal of Economics & Management Strategy, 6:1 (2007), 157-189.

Bettis, Richard A., and Michael A. Hitt, “The New Competitive Landscape,” Strategic Management Journal, 16:S1 (1993), 7-19.

Bloom, Nick, and John Van Reenen, "Measuring and Explaining Management Practices Across Firms and Countries," Quarterly Journal of Economics, 122:4 (2007), 1351-1408.

Bloom, Nick, and John Van Reenen, "Why Do Management Practices Differ Across Firms and Countries?" Journal of Economic Perspectives, 24:1 (2010), 203-224.

Bloom, Nick, Raffaella Sadun, and John Van Reenen, “Americans Do I.T. Better: US Multinationals and the Productivity Miracle,” American Economic Review, 102:1 (2012), 167-201.

Brander, James A., and Barbara J. Spencer, 1983. “International R&D Rivalry and Industrial Strategy,” Review of Economic Studies, 50:4 (1983), 707-722.

Branstetter, Lee G., “Are Knowledge Spillovers International or Intranational in Scope? Microeconomic Evidence from the U.S. and Japan,” Journal of International Economics, 53:1 (2001), 53-79.

Branstetter, Lee G., and Nakamura, Yoshiaki, “Is Japan’s Innovation Capacity in Decline?” in Kashyap, Anil, Magnus Blomstrom, Jennifer Corbett, and Fumio Hayashi, (Eds.), Structural Impediments to Growth in Japan (Chicago: University of Chicago Press and NBER, 2003).

Branstetter, Lee G., “Is Foreign Direct Investment a Channel of Knowledge Spillovers? Evidence from Japan’s FDI in the United States,” Journal of International Economics, 68:2 (2006), 53-79.

Burnham, Brad, “What’s Next?” Union Square Ventures [computer file] http://www.unionsquareventures.com/2007/01/whats_next.html (accessed April 2008).

Caballero, Ricardo J., and Adam B. Jaffe, “How High are the Giants’ Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth,” NBER Macroeconomics Annual, 8 (1993), 15-74.

Cantwell, John, “Japan’s Industrial Competitiveness and the Technological Capability of Leading Japanese Firms,” in Arrison, Thoms S. et al (Eds.), Japan’s Growing Technological Capability – Implications for the U.S. Economy (Washington, D.C.: National Academy Press, 1992).

Chuma, Hiroyuki, and Hashimoto, Norikazu, “Moore’s Law, Increasing Complexities, and the Limits of Organization: Modern Implications of the Japanese DRAM Era,” in Itami, Hiroyuki et al (Eds.), Dynamics of Knowledge, Corporate Systems, and Innovation (New York: Springer, 2007).

Cockburn, Iain M., and Megan J. MacGarvie, “Patents, Thickets, and the Financing of Early-Stage Firms: Evidence from the Software Industry,” Journal of Economics & Management Strategy, 18:3 (2009), 729-773.

Cohen, Wesley M., and Daniel A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation", Administrative Science Quarterly, 35:1 (1990), 128-152.

Cole, Robert E., 2006. "Software's Hidden Challenges," in Whittaker, D. Hugh, and Robert E. Cole (Eds.), Recovering from Success: Innovation and Technology Management in Japan (London: Oxford University Press, 2006).

Cole, Robert E., and Fushimi, Shinya, "The Japanese Enterprise Software Industry," in Miyoshi, Hiroaki, and Yoshifumi Nakata (Ed.), Have Japanese Firms Changed? (London: Palgrave MacMillan, 2011).

Cusumano, Michael A., "Japan's Software Factories: A Challenge to U.S. Management," (New York: Oxford University Press, 1991).

Cusumano, Michael A., "The Puzzle of Japanese Software," Communications of the ACM, 48:7 (2005), 25-27.

De Micheli, Giovanni, and Rajesh K. Gupta, "Hardware/Software Co-Design," Proceedings of the IEEE, 85:3 (1997), 349-365.

Dudley, Leonard, and Moenius, Johannes, "The Great Realignment: How Factor-Biased Innovation Reshaped Comparative Advantage in the U.S. and Japan, 1970-1992," Japan and the World Economy, 19:1 (2007), 112-132.

Express Computer, "Pricing and Apps Will Drive PDA Growth," December 23rd 2002.

Finan, William F., and Williams, Carl, "Implications of Japan's 'Soft Crisis': Forcing New Directions for Japanese Electronics Companies," in Arrison, Thomas L. et al (Eds.), Japan's Growing Technological Capability – Implications for the U.S. Economy (Washington, D.C.: National Academy Press, 1992).

Fransman, Martin, "Japan's Computer and Communications Industry: The Evolution of Industrial Giants and Global Competitiveness," (Oxford, UK: Oxford University Press, 1995).

Gore, Tony, "OMI – Developments in Processor Cores and Peripherals," in Roger, Jean-Yves (Ed.), Technologies for the Information Society: Developments and Opportunities (IOS Press, 1998).

Goto, Akira, "Japan's National Innovation System: Current Status and Problems," Oxford Review of Economic Policy, 16:2 (2000), 103-113.

Goto, Akira, and Odagiri, Hiroyuki, (Eds.), "Science-Based Industries," (Tokyo: NTT Publishing, 2003).

Graff, Bas, Lormans, Marco, and Toetenel, Hans, "Embedded Software Engineering: The State of the Practice," IEEE Software, 20:6 (2003), 61-69.

Graham, Stuart J.H., and Mowery, David D., "Intellectual Property Protection in the U.S. Software Industry," in Cohen, Wesley M. et al (Eds.), Patents in the Knowledge-Based Economy (Washington, D.C.: The National Academies Press, 2003).

Griliches, Zvi, "Market Value, R&D, and Patents," Economic Letters, 7:2 (1981), 183-187.

Griliches, Zvi, and Jacques Mairesse, "Productivity and R&D at the Firm Level," in Griliches, Zvi (Ed.), R&D, Patents, and Productivity (Chicago: University of Chicago Press and NBER, 1984).

Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey Part I," NBER Working Paper #3301 (1990).

Grossman, Gene M., and Elhanan Helpman, "Innovation and Growth in the Global Economy," (Cambridge, MA: MIT Press, 1991).

Hall, Bronwyn H., “Innovation and Market Value,” in Barrell. Ray, Geoff Mason, and Mary O’Mahoney (Eds.), Productivity, Innovation, and Economic Performance (Cambridge: Cambridge University Press, 2000).

Hall, Bronwyn H., and Daehwan Kim, “Valuing Intangible Assets: The Stock Market Value of R&D Revised,” University of California at Berkeley Working Paper (2000).

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, “The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools,” NBER Working Paper #8498 (2001).

Hall, Bronwyn H., and Megan J. MacGarvie, “The Private Value of Software Patents,” Research Policy, 39:7 (2010), 994-1009.

Hamada, Yasuyuki, “Venture Capital in Japan: Strategic Investment for the Future” (Tokyo: Nikkei Shimbun Press, 1996).

Hamada, Koichi, and Yasushi Okada, "Monetary and International Factors Behind Japan's Lost Decade," Journal of the Japanese and International Economies, Vol. 23:2 (2009), 200-219.

Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches, “Econometric Models for Count Data with an Application to the Patents – R&D Relationship,” NBER Technical Working Paper #17 (1984).

Hoshi, Takeo, Anil Kashyap, and David Scharfstein, “Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups,” Quarterly Journal of Economics, 106:1 (1991), 33-60.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle, “How Much Does Immigration Boost Innovation,” NBER Working Paper #14312 (2008).

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” Quarterly Journal of Economics, 108:3 (1993), 577-598.

Jaffe, Adam B., and Manuel Trajtenberg, “Flows of Knowledge from Universities and Federal Labs: Modeling the Flow of Patent Citations over Time and across Institutional and Geographic Boundaries,” NBER Working Paper #5712 (1996).

Jaffe, Adam B., and Manuel Trajtenberg, “Patents, Citations, and Innovations: A Window on the Knowledge Economy” (Boston: MIT Press, 2002).

Japanese Ministry of Justice, “Annual Report on Statistics of Legal Migrants”, Various Issues.

Jorgenson, Dale W., and Koji Nomura, "The Industry Origins of Japanese Economic Growth," Journal of the Japanese and International Economies, 19:4 (2005), 482-542.

Jorgenson, Dale W., and Koji Nomura, “The Industry Origins of the U.S.-Japan Productivity Gap,” Economic Systems Research, 19:3 (2007), 315-342.

Kerr, William R., and William F. Lincoln, “The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention,” HBS Working Paper (2008).

Kirkegaard, Jacob F., “Outsourcing and Skill Imports: Foreign High-Skilled Workers on H-1B and L-1 Visas in the United States,” Institute for International Economics Working Paper #05-15 (2005).

Kojima, Sakura, and Makoto Kojima, “Making IT Offshoring Work for the Japanese Industries,” in Meyer, Bertrand, and Mathai Joseph (Eds.), SEAFOOD (Software Engineering for Offshore and Outsourced Development) Conference Proceedings, LNCS 4716 (2007), 67-82.

Kortum, Samuel, and Josh Lerner, “Assessing the Contribution of Venture Capital to Innovation,” RAND Journal of Economics, 31:4 (2001), 674-692.

Krugman, Paul, “Import Protection as Export Promotion: International Competition in the Presence of Oligopoly and Economies of Scale,” in Kierzkouski, Henryk (Ed.), Monopolistic Competition and International Trade (Oxford: Oxford University Press, 1984).

Kurokawa, Toshiaki, and Susumu Hayashi, “Japan’s Critical Issues on IT Human Resource,” Science & Technology Trends Quarterly Review, 30 (2008) 23-40.

Levinthal, Daniel A. and James G. March, “The Myopia of Learning,” Strategic Management Journal, 14:S2 (1993), 95-112.

Lowell, Lindsay B., “H-1B Temporary Workers: Estimating the Population,” The Center for Comparative Immigration Studies Working Paper #12 (2000).

Mansfield, Edwin, “The Speed and Cost of Industrial Innovation in Japan and the United States: External vs. Internal Technology,” Management Science, 34:10 (1988), 1157-1168.

Ministry of Internal Affairs and Communications (MIC), Government of Japan, “Research Report on High-Level ICT Human Resource Development,” MIC Website (accessed 2008).

Nagaoka, Sadao, “Assessing the R&D Management of a Firm in Terms of Speed and Science Linkage: Evidence from the U.S. Patents,” Journal of Economics, Management, and Strategy, 16:1 (2007), 129-156.

Odagiri, Hiroyuki, and Hideto Yasuda, “Overseas R&D Activities of Japanese Firms,” in Goto, Akira, and Hiroyuki Odagiri (Eds.), Innovation in Japan (Oxford: Clarendon Press, 1997).

Organization for Economic Cooperation and Development, “OECD Information Technology Outlook: ICTs and the Information Economy,” (Paris, 2002).

Pakes, Ariel, and Zvi Griliches, “Patent and R&D at the Firm Level: A First Look,” NBER Working Paper #561 (1984).

Perfect, Steven B., and Kenneth W. Wiles, "Alternative Constructions of Tobin's Q: An Empirical Comparison," Journal of Empirical Finance, 1:3-4 (1994), 313-341.

Tanaka, Tatsuo, "The Software Industry," in Goto, Akira, and Hiroyuki Odagiri (Eds.), Science-Based Industries (Tokyo: NTT Publishing, 2003)

.

Footnotes

[1]These results parallel the findings of Jorgenson and Nomura (2007), who demonstrate that Japanese TFP rose rapidly for decades, converging to U.S. levels, but then began diverging from it around 1995. Their industry level analysis suggests that a change in the relative performance of the IT-producing industries (which we study in this paper) and the IT-using industries were particularly important in driving the shift from convergence to divergence. Jorgenson and Nomura do not attempt to explain the mechanisms behind divergence in productivity.

[2]Personal discussions with Mark Kryder, former CTO of Seagate, confirmed that software has become an increasingly important driver of product functionality and product differentiation in the hard disk drive industry.

[3]Allison et al. (2006) rejected the use of both the standard classification system and keyword searches, resorting to the identification of software patents by reading through them manually. Although potentially more accurate, this method is inherently subjective and not scalable.

[4]See Goto (2000) and Nagaoka (2007) for a more detailed discussion.

[5]Methodological changes in the survey make it difficult to track occupational employment in the U.S. IT industry in a consistent way over time, particularly in comparing the periods before and after 1999.

[6]The construction of variables is explained in greater detail in subsequent sections.

[7]We use the NBER Patent Database, which currently incorporates all patents granted through 2006. Since our empirical specifications use patents dated by the date of application, and since can patents take more than two years to work their way through the USPTO evaluation process, we are currently unable to extend our data past 2004.

[8]GICS, the Global Industry Classification System, is constructed and managed by Moody's in collaboration with Compustat.

[9]NTT is the only Japanese firms in "IT services and software" in our sample.

[10]We thank the Columbia Business School Center on the Japanese Economy and Business for these data.

[11]January 1st, 2003 was the date of creation of this index.

[12]Figuring out what fraction of total IT production is accounted for by our firms is harder, because of the far-reaching globalization of IT production by the late 1990s. According to the OECD, in 1999, the top 10 IT U.S. firms in our sample had global revenues greater than the entire amount of IT production in the U.S. in that year. The picture is similar for our Japanese firms, who have also taken increasing advantage of opportunities to offshore production.

[13]Depending on the measure, tests of equality are not always statistically significant when we disaggregate it by country of origin. When Japanese software intensity is measured by citations to software in non-software patents, electronics is (insignificantly) more software intensive than semiconductors.

[14]Our main results are robust to using firm-level software intensity assignments instead of industry classifications.

[15]Downloaded from the following link: https://sites.google.com/site/patentdataproject/ (5/15/2011)

[16]*Kaisha Shiki Ho* (Japan Company Handbooks) is an annual survey of Japanese firms, published by the Japanese equivalent of Dow Jones & Company, *Toyo Keizai* Inc. We thank Ms. Kanako Hotta for assistance in obtaining these data from the collections at the School of International Relations and Pacific Studies of the University of California at San Diego.

[17]We obtained these data from the Bureau of Labor Statistics and Statistics Bureau of Japan, respectively.

[18]See Griliches and Mairesse (1984) and Hall (1990) for a detailed description and discussion of this methodology. We used several depreciation rates between 10% and 30%, with little change in the results.

[19]When the expenditure data was not available, we used first 5 years of available R&D expenditure data, "backcast them" using linear extrapolation, and calculated the initial R&D capital stock based on the projected R&D expenditures.

[20]Perfect and Wiles (1994) suggests that the measurement error in using book value of debt is modest.

[21]Perfect and Wiles (1994) note that different calculation methodologies do result in different absolute replacement cost values, but do not seem to bias coefficients on R&D capital.

[22]In the mid-2000s, Japanese electronics firms received a boost from the rapidly growing sale of so-called digital appliances, such as DVD recorders, digital cameras, and LCD televisions. Industry observers, such as Ikeda (2003), warned of imminent commoditization of these new products – a prediction that has been born out in the latter years of the decade.

[23]An earlier version of the paper used data that ended in the late 1990s, raising the possibility that our results were driven by the late 1990s IT bubble. Extension of our data into the mid-2000s shows that this is not the case. We thank an anonymous referee for pushing us to extend these data.

[24]See Jorgenson and Nomura (2005) and Hamada and Okada (2009) for a discussion of the impact of exchange rate movements on Japanese industry and the overall economy.

[25]Towards the end of the 1990s, a small number of publicly listed firms, such as Softbank, that we could classify as software firms appeared on the Tokyo Stock Exchange. Motohashi (2009) uses a different data set to explore productivity trends in the Japanese software industry, but does not attempt an international comparison.

[26]We thank an anonymous referee for stressing the importance of this event. Jorgenson and Nomura (2005) discuss this event and show that the pace of IT price declines in Japan accelerates after the introduction of DOS/V.

[27]In unreported estimates, we obtain similar results if we divide our sample into the following periods, 83-88, 89-93, 94-99, and 2000-2004.

[28]Interestingly, Jorgenson and Nomura find quite different trends in the communications equipment industry. The firms in our sample include many major Japanese manufacturers of communications equipment, but as one of many lines of business. Given our data, we cannot separately analyze the communications equipment business units of IT firms.

[29]As we were writing this paper, we became aware of the work of Cole (2006) and Cole and Fushimi (2011), who use narrative history and interviews with practitioners to suggest that the changing fortunes of the U.S. and Japanese IT industries are linked to the superior ability of American firms to exploit software advances in their new product development. Our quantitative analysis is broadly consistent with their interview-based description.

[30]Finan and Williams (1992) and Cusumano (1991, 2005) discuss the scarcity of software engineers, as do Fransman (1995), the Japanese Ministry of Internal Affairs and Communications (2005), and Kurokawa and Hayashi (2008).

[31]Some Japanese firms, most notably in videogames, have maintained a strong international market positions in software-intensive segments of IT. However, videogames sales are driven by

artistic factors as well as purely technological ones, and Japanese developers have a rich local cultural tradition of *manga* (a Japanese art form akin to comic books in the West) and *anime* (animated films) to draw upon.

[32]Branstetter (2006) finds a positive but limited impact of U.S. R&D centers on the research productivity of Japanese firms' home R&D operations. Anchordoguy (2000) argues that tapping into foreign pools of software knowledge was especially difficult for Japanese firms, given language barriers and differences in labor market practices.

[33]Kojima and Kojima (2007) examine the available data on Japanese offshoring of software development to other countries. While the data are highly problematic, they suggest a very low level of offshoring relative to the U.S. – something as low as 5-10% of the U.S. level – even by the mid-2000s.

[34]U.S. data are from the NSF's SESTAT survey (http://www.nsf.gov/statistics/recentgrads/) and the annual Survey of Earned Doctorates http://www.nsf.gov/statistics/doctorates/. Data for Japan is taken from the Japanese Ministry of Education, Sports, and Welfare's Basic School Survey. We thank Professor Kyoji Fukao of Hitotsubashi University and Professor Takao Kato of Colgate University and Professor Anthony D'Costa of Copenhagen Business School for helping us identify and obtain the Japanese data sources used in this paper.

[35]Japanese statistics track newly registered foreign workers across a number of broad categories including "researchers," "engineers," and "intracompany transferees." These data are reported annually in the *Shutsu Nyukoku Kanri Toukei Nenpo* (Annual Report of Statistics on Legal Migrants), published by the Japanese Ministry of Justice.

[36]Only a fraction of IT graduates will enter employment in IT industries in the countries in which they study, and only a fraction of those who obtain employment in the IT industry will be

engaged in research. Likewise, our estimates of H-1B temporary workers include individuals employed in IT companies as well as individuals working for banks and insurance companies, and only a fraction of the H-1Bs employed in IT companies are involved in research. These data track (potential) new entrants to the IT workforce, not the total stocks of workers available for employment in the sector.

[37]The growing literature on software patents has examined the impact of software patentability on R&D and the impact of software patents on venture firm financing, but it has not yet addressed the impact of software technology on innovation elsewhere in IT. See Bessen and Hunt (2007), Hall and MacGarvie (2010), and Cockburn and MacGarvie (2009).

Tables and Figures

Table I: Citation Function Results

	Full Sample		Citations to Software Patents Only	
	Coefficient	Std. Error	Coefficient	Std. Error
Citing Grant Year				
1991	0.4549 **	0.1760	0.5013 ***	0.1662
1992	0.6572 ***	0.1783	0.7418 ***	0.1716
1993	0.7317 ***	0.1683	0.8482 ***	0.1645
1994	1.0131 ***	0.1750	1.2010 ***	0.1752
1995	1.2123 ***	0.1717	1.4509 ***	0.1742
1996	1.5258 ***	0.1722	1.8499 ***	0.1779
1997	1.5966 ***	0.1548	1.9673 ***	0.1619
1998	1.7073 ***	0.1378	2.1389 ***	0.1462
1999	1.6623 ***	0.1156	2.1203 ***	0.1239
2000	1.5740 ***	0.0960	2.0478 ***	0.1039
2001	2.1979 ***	0.0966	2.8943 ***	0.1072
2002	2.3529 ***	0.0915	3.1451 ***	0.1029
2003	2.3546 ***	.	3.1691 ***	.
Cited Grant Year				
1990	-0.0958 ***	0.0197	-0.1078 ***	0.0174
1991	-0.3330 ***	0.0191	-0.3621 ***	0.0165
…	…	…	…	…
2001	-0.8881 ***	0.0157	-0.9138 ***	0.0112
2002	-0.9167 ***	0.0191	-0.9367 ***	0.0137
Citing Patent Type				
Comp. Hard/Software	1.0414 ***	0.0398	1.1936 ***	0.0403
Computer Peripherals	0.4806 ***	0.0345	0.5443 ***	0.0339
Information Storage	0.3778 ***	0.0324	0.4296 ***	0.0317
Other Comp. & Comm.	2.3707 ***	0.0652	2.7084 ***	0.0674
Electrical Devices	-0.8256 ***	0.0209	-0.9188 ***	0.0192
Semiconductors	-0.6657 ***	0.0199	-0.7863 ***	0.0186
Other				
Citing From Japan	-0.3078 ***	0.0313	-0.6298 ***	0.0059
Cited Software Patent	9.4217 ***	0.2573	n/a	n/a
Citing From Japan X Cited Software	-6.2592 ***	0.1981	n/a	n/a
Obsolescence	0.3252 ***	0.0095	0.3398 ***	0.0087
Diffusion	3.61e-06 ***	4.79e-07	3.56e-04 ***	4.27e-06
Adj R-Squared	0.9232		0.9674	
Number of Obs.	2940		1470	

The data for regression estimations presented in this table are drawn from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. Regression specifications are estimated in STATA using the nonlinear least squares algorithm. The dependent variable is an empirical measure of the probability a citing patent of a given type cites a cited patent of a given type. All presented coefficients are relative to base categories. They are the following: citing patent grant year = 1990, cited patent grant year = 1989, citing patent type = “Communications”, cited patent category = “non-software” (only applicable to column I), citing patent geography = “Japan”. Patent origin is defined using all inventors listed on the patent document.

Table II: Firm-Level Software Intensity by Sector, 1983-2004

	Share of Software Patents			Share of Citations to Software Patents		
Industry	No. of Obs.	Mean	St. Dev.	No. of Obs.	Mean	St. Dev.
Electronics	65	0.0387 (***/***)	0.0808	65	0.0544 (*/***)	0.0654
Semiconductors	53	0.1069 (***/***)	0.1246	53	0.0768 (*/***)	0.0837
IT Hardware	92	0.1974 (***/***)	0.1681	92	0.1428 (***/***)	0.1109

This table compares measures of software intensity of firms in our sample that belong to different subsectors. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The unit of observation for descriptive statistics and statistical tests presented in this table is *a firm*. The share of software patents for each firm is computed as the number of software patents granted to a firm in the sample period divided by the total number of patents granted to that firm in the sample period. The share of citations to software patents for each firm is calculated as the number of citations directed to software patents generated by the firm's non-software IT patent portfolio divided by the total number of citations generated by the firm's non-software IT patent portfolio. The tests for differences in means across sectors are performed using one-sided t-tests and are reported in the brackets next to the value of the mean. (***) represents the difference being significant at the 0.01 level, (**) at 0.05, and (*) at 0.1. The first series of asterisks in any given bracket represent the results of a one-sided t-test for differences of means using the sector in question and the sector listed in the row above, while the second series of asterisks represents the results of a one-sided t-test using the sector in question and the sector listed in the row below. For sectors listed in the first row, the first series of asterisks refer to a comparison with the sector listed in row immediately below, while the second series of asterisks refer to a comparison with the sector listed in the final row. An identical system applies to the interpretation of asterisks for sectors listed in the final row.

Table II-2: Patent-Level Software Intensity by Sector, 1983-2004

	Share of Software Patents			Share of Citations to Software Patents		
Industry	No. of Obs.	Mean	St. Dev.	No. of Obs.	Mean	St. Dev.
Electronics	67775	0.0476 (***/***)	0.2130	23452	0.0532 (***/***)	0.1429
Semiconductors	83609	0.0995 (***/***)	0.2994	48214	0.0742 (***/***)	0.1678
IT Hardware	251422	0.1439 (***/***)	0.3510	126339	0.1127 (***/***)	0.2092

This table compares measures of software intensity of firms in our sample that belong to different subsectors. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The unit of observation for descriptive statistics and statistical tests presented in this table is *a patent*. The share of software patents for each sector is computed as the number of software patents granted to all firms belonging to that sector in the sample period divided by the total number of patents granted to firms in that sector in the sample period. The share of citations to software patents for each sector is calculated as the number of citations directed to software patents generated by all firms' non-software IT patent portfolios divided by the total number of citations generated all firms' non-software IT patent portfolio. The tests for differences in means across sectors are performed using one-sided t-tests and are reported in the brackets next to the value of the mean. (***) represents the difference being significant at the 0.01 level, (**) at 0.05, and (*) at 0.1. The first series of asterisks in any given bracket represent the results of a one-sided t-test for differences of means using the sector in question and the sector listed in the row above, while the second series of asterisks represents the results of a one-sided t-test using the sector in question and the sector listed in the row below. For sectors listed in the first row, the first series of asterisks refer to a comparison with the sector listed in row immediately below, while the second series of asterisks refer to a comparison with the sector listed in the final row. An identical system applies to the interpretation of asterisks for sectors listed in the final row.

Table III: Software Patent Shares by Sector and Firm Origin, 1983-2004

	U.S. Firms			Japanese Firms		
Industry	No. of Obs.	Mean	St. Dev.	No. of Obs.	Mean	St. Dev.
Electronics	22	0.0806 (*/***)	0.1425	43	0.0173 (/***)	0.0195
Semiconductors	41	0.1341 (*/***)	0.1292	12	0.0138 (/***)	0.0213
IT Hardware	70	0.2411 (***/***)	0.1699	22	0.0585 (***/***)	0.0329

Unit of observation is a firm

	U.S. Firms			Japanese Firms		
Industry	No. of Obs.	Mean	St. Dev.	No. of Obs.	Mean	St. Dev.
Electronics	38902	0.0647 (***/***)	0.2460	28873	0.0247 (***/***)	0.1551
Semiconductors	56833	0.1324 (***/***)	0.3389	26776	0.0298 (***/***)	0.1700
IT Hardware	104998	0.2337 (***/***)	0.4232	146424	0.0795 (***/***)	0.2705

Unit of observation is a patent

This table compares measures of software intensity of firms in our sample that belong to different subsectors, separately for those firms based in Japan and those based in the United States. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The unit of observation for descriptive statistics and statistical tests presented in the upper panel is *a firm*, while it is *a patent* in the lower panel. For details about the construction of software intensity measures please consult Table II. The tests for differences in means across sectors are performed using one-sided t-tests and are reported in the brackets next to the value of the mean. (***) represents the difference being significant at the 0.01 level, (**) at 0.05, and (*) at 0.1. The first series of asterisks in any given bracket represent the results of a one-sided t-test for differences of means using the sector in question and the sector listed in the row above, while the second series of asterisks represents the results of a one-sided t-test using the sector in question and the sector listed in the row below. For sectors listed in the first row, the first series of asterisks refer to a comparison with the sector listed in row immediately below, while the second series of asterisks refer to a comparison with the sector listed in the final row. An identical system applies to the interpretation of asterisks for sectors listed in the final row.

Table III-2: Share of Citations to Software by Non-Software IT Patents by Sector and Firm Origin, 1983-2004

	U.S. Firms			Japanese Firms		
Industry	No. of Obs.	Mean	St. Dev.	No. of Obs.	Mean	St. Dev.
Electronics	22	0.0761 (/***)	0.0921	43	0.0435 (/***)	0.0452
Semiconductors	41	0.0895 (/***)	0.0884	12	0.0286 (/***)	0.0334
IT Hardware	70	0.1647 (***/***)	0.1173	22	0.0738 (***/***)	0.0384

Unit of observation is a firm

	U.S. Firms			Japanese Firms		
Industry	No. of Obs.	Mean	St. Dev.	No. of Obs.	Mean	St. Dev.
Electronics	12915	0.0617 (***/***)	0.1504	10537	0.0430 (***/***)	0.1325
Semiconductors	36389	0.0797 (***/***)	0.1726	11825	0.0572 (***/***)	0.1507
IT Hardware	53706	0.1466 (***/***)	0.2326	72633	0.0877 (***/***)	0.1862

Unit of observation is a patent

This table compares measures of software intensity of firms in our sample that belong to different subsectors, separately for those firms based in Japan and those based in the United States. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The unit of observation for descriptive statistics and statistical tests presented in the upper panel is *a firm*, while it is *a patent* in the lower panel. For details about the construction of software intensity measures please consult Table II-2. The tests for differences in means across sectors are performed using one-sided t-tests and are reported in the brackets next to the value of the mean. (***) represents the difference being significant at the 0.01 level, (**) at 0.05, and (*) at 0.1. The first series of asterisks in any given bracket represent the results of a one-sided t-test for differences of means using the sector in question and the sector listed in the row above, while the second series of asterisks represents the results of a one-sided t-test using the sector in question and the sector listed in the row below. For sectors listed in the first row, the first series of asterisks refer to a comparison with the sector listed in row immediately below, while the second series of asterisks refer to a comparison with the sector listed in the final row. An identical system applies to the interpretation of asterisks for sectors listed in the final row.

Table IV: Tobin's Q Regressions by Period, 1983-2004

lnQ	Entire Sample NLS		1983-1988 NLS		1989-1993 NLS		1994-1999 NLS		2000-2004 NLS	
RD/Assets	**0.1087**		**0.0158**		**-0.0564**		**0.2196**		**-0.0579**	
	(0.0415)	***	(0.1451)		(0.0812)		(0.0897)	**	(0.0495)	
RD/Assets * Japan	**-0.1327**		**0.0008**		**0.0250**		**-0.2844**		**-0.2916**	
	(0.0556)	**	(0.1516)		(0.1129)		(0.1310)	**	(0.1408)	**
lnSales	0.0356		0.0198		0.0309		0.0995		0.0966	
	(0.0039)	***	(0.0069)	***	(0.0062)	***	(0.0059)	***	(0.0050)	***
No. of Obs.	3571		825		833		1082		831	
R-squared	0.2986		0.2763		0.2429		0.4414		0.4049	

The data for regression estimations presented in this table were obtained from Compustat and the Development Bank of Japan for U.S. and Japanese firms, respectively. R&D expenditure data for Japanese firms comes from annual volumes of the Kaisha Shiki Ho survey. The data represent an unbalanced panel of large publicly traded U.S. and Japanese IT firms active in the sample period, 1983-2004. As a consequence of using an unbalanced panel, total number of observations used in regression estimations can vary between time periods. Regression specifications are estimated in STATA using the nonlinear least squares algorithm. The dependent value is the log of Tobin's Q, which is calculated as the ratio of the firm's market value to the replacement value of its total assets. RD/Assets are calculated as the ratio of the stock of firm's accumulated R&D expenditures, calculated using the perpetual inventory method, to the replacement value of the firm's total assets. The Japan dummy equals 1 if the firm is based in Japan. Standard errors are reported in brackets. For detailed information about the specification, sample selection, and variable construction, please consult the main body of the essay. The asterisks that are listed next to coefficients reported in the table denote statistical significance in the following manner: (***) represents significance at the 0.01 level, (**) at 0.05, and (*) at 0.1. For brevity, only coefficients on variables of interest are reported, while coefficients on some of the control variables may be omitted. Detailed estimation results are available from the authors by request.

Table IV-2: Tobin's Q Regressions by Period, Including Firm-Level Software Intensity, 1983-2004

lnQ	Entire Sample NLS		1983-1988 NLS		1989-1993 NLS		1994-1999 NLS		2000-2004 NLS	
RD/Assets	**-0.2342**		**-0.2302**		**-0.2020**		**-0.1580**		**-0.2412**	
	(0.0553)	***	(0.1554)		(0.0945)	**	(0.1189)		(0.0820)	***
RD/Assets * Japan	**0.1992**		**0.2227**		**0.1615**		**0.0779**		**-0.1365**	
	(0.0651)	***	(0.1593)		(0.1208)		(0.1483)		(0.1478)	
RD/Assets * Sof. Int.	**0.9752**		**2.4214**		**0.7938**		**0.9375**		**0.7052**	
	(0.1844)	***	(0.6740)	***	(0.3688)	**	(0.3365)	***	(0.2968)	**
lnSales	0.0419		0.0135		0.0305		0.1093		0.0995	
	(0.0039)	***	(0.0070)	*	(0.0062)	***	(0.0061)	***	(0.0049)	***
No. of Obs.	3571		825		833		1082		831	
R-squared	0.3052		0.2884		0.2465		0.4452		0.4089	

The data for regression estimations presented in this table were obtained from Compustat and the Development Bank of Japan for U.S. and Japanese firms, respectively. R&D expenditure data for Japanese firms comes from annual volumes of the Kaisha Shiki Ho survey. Firm-level software intensity measures were calculated using data from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The data represent an unbalanced panel of large publicly traded U.S. and Japanese IT firms active in the sample period, 1983-2004. As a consequence of using an unbalanced panel, total number of observations used in regression estimations can vary between time periods. Regression specifications are estimated in STATA using the nonlinear least squares algorithm. The dependent value is the log of Tobin's Q, which is calculated as the ratio of the firm's market value to the replacement value of its total assets. RD/Assets are calculated as the ratio of the stock of firm's accumulated R&D expenditures, calculated using the perpetual inventory method, to the replacement value of the firm's total assets. The Japan dummy equals 1 if the firm is based in Japan. Standard errors are reported in brackets. For detailed information about specification, sample selection, and variable construction, please consult the main body of the essay. Regression analysis presented in this table is identical to that presented in Table IV above, except that a measure of firm-level software intensity has been added to the specification. The asterisks that are listed next to coefficients reported in the table denote statistical significance in the following manner: (***) represents significance at the 0.01 level, (**) at 0.05, and (*) at 0.1. For brevity, only coefficients on variables of interest are reported, while coefficients on some of the control variables may be omitted. Detailed estimation results are available from the authors by request.

Table V: Patent Production Function Regressions, Japanese Indicator and Time Trends, Entire Sample and By Sector, 1983-2004

	Entire Sample			Electronics			Semiconductors			IT Hardware		
	OLS	RE	FE	OLS	RE	FE	OLS	RE	FE	OLS	RE	FE
Log R&D	**0.9814**	**0.7429**	**0.6682**	**0.9456**	**0.6944**	**0.6208**	**0.9725**	**0.8241**	**0.6761**	**0.9541**	**0.6865**	**0.6186**
	(0.0392)	(0.0463)	(0.0542)	(0.0762)	(0.0465)	(0.0672)	(0.0907)	(0.1019)	(0.1205)	(0.0582)	(0.0718)	(0.0817)
1989-1993	**0.0066**	**0.1056**	**0.1237**	**0.1132**	**0.2701**	**0.3049**	**0.1310**	**0.1312**	**0.1378**	**0.0029**	**0.0983**	**0.1136**
	(0.0765)	(0.0668)	(0.0680)	(0.1771)	(0.0982)	(0.0995)	(0.1660)	(0.1411)	(0.1420)	(0.0954)	(0.0937)	(0.0969)
1994-1999	**0.1151**	**0.4168**	**0.4942**	**-0.2141**	**0.0723**	**0.1328**	**0.2525**	**0.6259**	**0.8167**	**0.2313**	**0.4461**	**0.5067**
	(0.1269)	(0.1142)	(0.1174)	(0.3336)	(0.3504)	(0.3598)	(0.2278)	(0.1931)	(0.2002)	(0.1677)	(0.1380)	(0.1414)
2000-2004	**0.5053**	**1.0171**	**1.1456**	**-0.1647**	**0.3258**	**0.4280**	**0.3877**	**1.0983**	**1.4642**	**0.9636**	**1.1928**	**1.2684**
	(0.1381)	(0.1230)	(0.1294)	(0.2629)	(0.2137)	(0.2235)	(0.2581)	(0.2317)	(0.2553)	(0.1954)	(0.1718)	(0.1752)
Japan Dummy	**0.7363**	**0.8482**	**n.a**	**-0.0607**	**-0.1600**	**n.a**	**0.5806**	**0.7832**	**n.a.**	**1.2059**	**1.5392**	**n.a.**
	(0.1796)	(0.1922)		(0.2692)	(0.3053)		(0.3523)	(0.3951)		(0.2835)	(0.2843)	
Japan * 1989-1993	**-0.3033**	**-0.1823**	**-0.1584**	**-0.5258**	**-0.4881**	**-0.4850**	**-0.1639**	**0.0697**	**0.1415**	**-0.1511**	**-0.0052**	**0.0230**
	(0.1116)	(0.0984)	(0.0994)	(0.2069)	(0.1341)	(0.1345)	(0.2761)	(0.2772)	(0.2795)	(0.1702)	(0.1451)	(0.1456)
Japan * 1994-1999	**-0.5294**	**-0.5037**	**-0.5111**	**-0.3492**	**-0.2176**	**-0.2118**	**-0.4814**	**-0.5691**	**-0.5924**	**-0.3786**	**-0.4228**	**-0.4283**
	(0.1713)	(0.1435)	(0.1451)	(0.3706)	(0.3584)	(0.3666)	(0.4434)	(0.4132)	(0.4172)	(0.2414)	(0.2086)	(0.2100)
Japan * 2000-2004	**-0.8835**	**-1.0319**	**-1.0758**	**-0.3181**	**-0.4322**	**-0.4551**	**-0.6613**	**-1.0342**	**-1.1847**	**-1.0342**	**-0.9954**	**-1.0056**
	(0.1884)	(0.1740)	(0.1759)	(0.3145)	(0.2392)	(0.2407)	(0.5045)	(0.5781)	(0.6008)	(0.2905)	(0.2771)	(0.2781)

The firm-level R&D expenditure data for regression estimations presented in this table were obtained from Compustat and annual volumes of the Kaisha Shiki Ho survey for U.S. and Japanese firms, respectively. Patent data come from the CASSIS patent database maintained by the United States Patent and Trademark office and from the NBER Patent Data Project database. The data represent an unbalanced panel of large publicly traded U.S. and Japanese IT firms active in the sample period, 1983-2004. The dependent variable is the log of the number of total patents granted in a given year. The Japan dummy equals 1 when a firm is based in Japan. Regression specifications are estimated in STATA using ordinary least squares, random effects, and fixed effects algorithms. Robust and cluster-corrected standard errors are reported in brackets. For detailed information about the specification, sample selection, and variable construction, please consult the main body of the essay. For brevity, only coefficients on variables of interest are reported, while coefficients on some of the control variables may be omitted. Detailed estimation results are available from the authors by request.

Table VI: Tobin's Q Regressions, Comparing Time Trends, By Country, 1983-2004

	Entire Sample				US				Japan		
lnQ	FE		NLLS		FE		NLLS		FE		NLLS
RD/Assets	**-0.0814**		**-0.0167**		**-1.1304**		**-0.5120**		**-0.0273**		**0.0115**
	(0.1257)		(0.0442)		(0.2753)	***	(0.1310)	***	(0.0497)		(0.0352)
RD/Assets * 1989-93	**-0.3011**		**-0.1369**		**0.6919**		**0.1800**		**-0.1295**		**-0.0209**
	(0.1016)	***	(0.0552)	**	(0.2890)	**	(0.1447)		(0.0421)	***	(0.0768)
RD/Assets * 1994-99	**0.1375**		**0.1309**		**1.1809**		**0.5798**		**-0.1191**		**-0.0086**
	(0.1262)		(0.0700)	*	(0.2753)	***	(0.1390)	***	(0.0563)	**	(0.0795)
RD/Assets * 2000-04	**0.0611**		**-0.0396**		**0.9727**		**0.3475**		**-0.1678**		**-0.0897**
	(0.1460)		(0.0663)		(0.2932)	***	(0.1366)	**	(0.2461)		(0.1303)
No. of Obs.	3571		3571		1978		1978		1593		1593

The data for regression estimations presented in this table were obtained from Compustat and the Development Bank of Japan for U.S. and Japanese firms, respectively. R&D expenditure data for Japanese firms comes from annual volumes of the Kaisha Shiki Ho survey. The data represent an unbalanced panel of large publicly traded U.S. and Japanese IT firms active in the sample period, 1983-2004. The regression estimation results presented in this table are analogous to those presented in Tables IV and IV-2, except that they include a direct estimation of the time trends. Regression specifications are estimated in STATA. A linearized version of the specification is estimated using the fixed effects algorithm, while a nonlinear version of the specification is estimated using the nonlinear least squares algorithm. The dependent value is the log of Tobin's Q, which is calculated as the ratio of the firm's market value to the replacement value of its total assets. RD/Assets are calculated as the ratio of the stock of firm's accumulated R&D expenditures, calculated using the perpetual inventory method, to the replacement value of the firm's total assets. Standard errors are reported in brackets. Robust and cluster-corrected standard errors are reported for specifications estimated using the fixed effects algorithm. For detailed information about the specification, sample selection, and variable construction, please consult the main body of the paper. The asterisks that are listed next to coefficients reported in the table denote statistical significance in the following manner: (***) represents significance at the 0.01 level, (**) at 0.05, and (*) at 0.1. For brevity, only coefficients on variables of interest are reported, while coefficients on some of the control variables may be omitted. Detailed estimation results are available from the authors by request.

Table VII: Tobin's Q Regressions, By Industry and Time Period, Fixed Effects, 1983-2004

	Electronics				Semiconductors				IT Hardware	
lnQ	1983-1993		1994-2004		1983-1993		1994-2004		1983-1993	1994-2004
RD/Assets	**-0.3464**		**-1.1880**		**-0.7058**		**0.0609**		**-0.3933**	**-0.2278**
	(0.3059)		(0.3865)	***	(0.1752)	***	(0.0017)	***	(0.3095)	(0.1496)
RD/Assets	**0.2789**		**1.1019**		**0.6043**		**-0.6449**		**-0.0335**	**-0.3502**
* Japan	(0.3040)		(0.4283)	**	(0.1966)	***	(0.9356)		(0.5447)	(0.4091)
No. of Obs.	603		638		349		530		706	745
R-squared	0.1158		0.1030		0.0286		0.0796		0.0966	0.1089

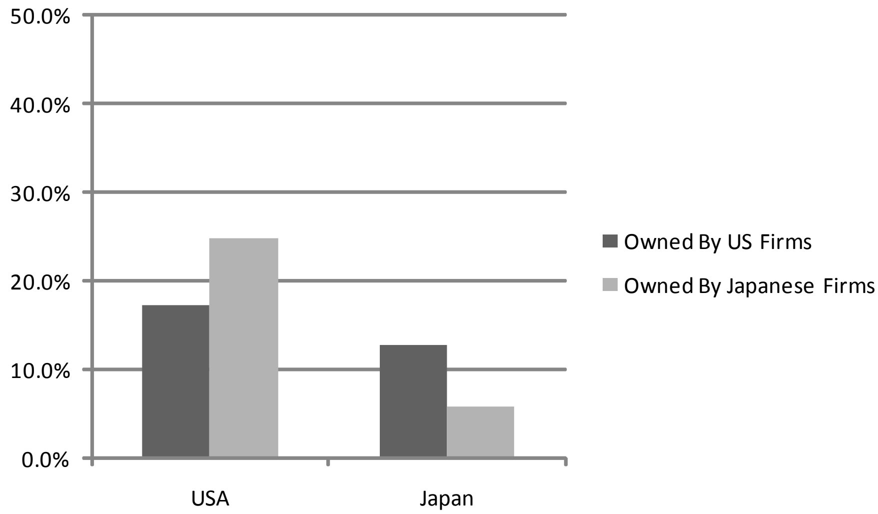
The data for regression estimations presented in this table were obtained from Compustat and the Development Bank of Japan for U.S. and Japanese firms, respectively. R&D expenditure data for Japanese firms comes from annual volumes of the Kaisha Shiki Ho survey. The data represent an unbalanced panel of large publicly traded U.S. and Japanese IT firms active in the sample period, 1983-2004. As a consequence of using an unbalanced panel, total number of observations used in regression estimations can vary between time periods. Regression specifications are estimated in STATA using the fixed effects algorithm. The dependent value is the log of Tobin's Q, which is calculated as the ratio of the firm's market value to the replacement value of its total assets. RD/Assets are calculated as the ratio of the stock of firm's accumulated R&D expenditures, calculated using the perpetual inventory method, to the replacement value of the firm's total assets. The Japan dummy equals 1 if the firm is based in Japan. Robust and cluster-corrected standard errors are reported in brackets. For detailed information about the specification, sample selection, and variable construction, please consult the main body of the essay. The asterisks that are listed next to coefficients reported in the table denote statistical significance in the following manner: (***) represents significance at the 0.01 level, (**) at 0.05, and (*) at 0.1. For brevity, only coefficients on variables of interest are reported, while coefficients on some of the control variables may be omitted. Detailed estimation results are available from the authors by request.

Table VIII: Tobin's Q Regressions, By Industry and Time Period, NLS, 1983-2004

	Electronics			Semiconductors				IT Hardware		
lnQ	1983-1993	1994-2004		1983-1993		1994-2004		1983-1993	1994-2004	
RD/Assets	**-0.0804**	**0.3760**		**-0.2752**		**0.2919**		**-0.1399**	**-0.1412**	
	(0.1216)	(0.1995)	*	(0.0904)	***	(0.1098)	***	(0.1019)	(0.0429)	***
RD/Assets	**0.1070**	**-0.3838**		**0.1239**		**-1.5693**		**-0.3292**	**-0.3107**	
* Japan	(0.1271)	(0.2147)	*	(0.1287)		(0.2756)	***	(0.3255)	(0.2500)	
No. of Obs.	603	638		349		530		706	745	
R-squared	0.4826	0.2414		0.2416		0.6240		0.1431	0.3760	

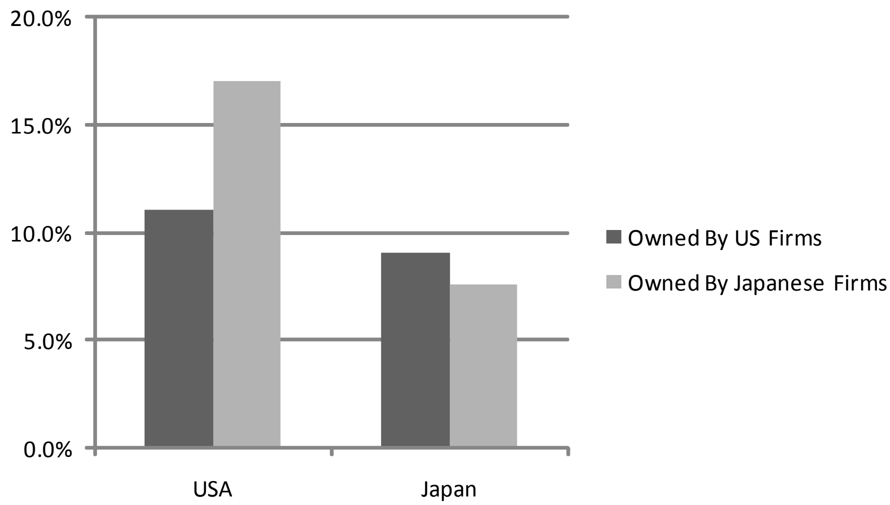
The data for regression estimations presented in this table were obtained from Compustat and the Development Bank of Japan for U.S. and Japanese firms, respectively. R&D expenditure data for Japanese firms comes from annual volumes of the Kaisha Shiki Ho survey. The data represent an unbalanced panel of large publicly traded U.S. and Japanese IT firms active in the sample period, 1983-2004. As a consequence of using an unbalanced panel, total number of observations used in regression estimations can vary between time periods. Regression specifications are estimated in STATA using the nonlinear least squares algorithm. The dependent value is the log of Tobin's Q, which is calculated as the ratio of the firm's market value to the replacement value of its total assets. RD/Assets are calculated as the ratio of the stock of firm's accumulated R&D expenditures, calculated using the perpetual inventory method, to the replacement value of the firm's total assets. The Japan dummy equals 1 if the firm is based in Japan. Standard errors are reported in brackets. For detailed information about the specification, sample selection, and variable construction, please consult the main body of the essay. The asterisks that are listed next to coefficients reported in the table denote statistical significance in the following manner: (***) represents significance at the 0.01 level, (**) at 0.05, and (*) at 0.1. For brevity, only coefficients on variables of interest are reported, while coefficients on some of the control variables may be omitted. Detailed estimation results are available from the authors by request.

Figure 9-1: Software Intensity of Patenting (Share of Software Patents), by Geography of Invention and Country of Ownership, 1983-2004



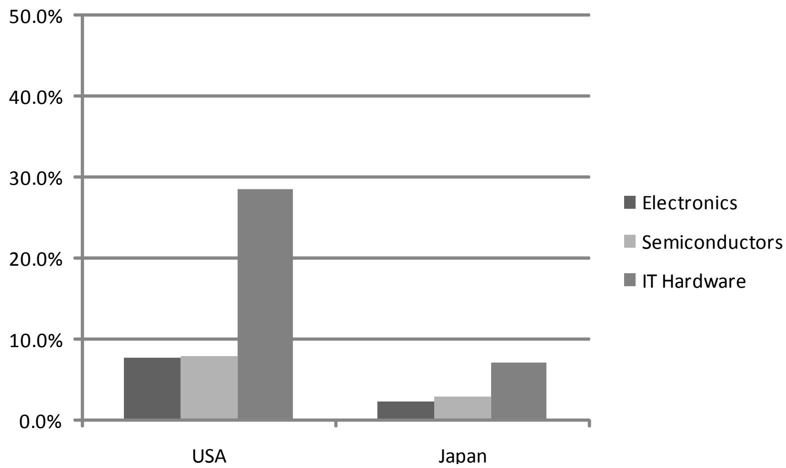
This table compares a measure of firm-level software intensity of patenting for the firms in our sample by the geographical region of their origin and the geographical region of invention. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The software intensity variable is calculated as the share of software patents in total patents granted in the sample period, 1983-2004, averaged across all firms belonging to a given region of origin - region of invention combination. Geography of invention is determined using geographical locations of all inventors listed on the patent document. T-tests for differences in means across geographical groups show that differences are statistically significant at the 0.01 level in the case of all group pairs.

Figure 9-2: Software Intensity of Patenting (Share of Citations Made to Software), by Geography of Invention and Country of Ownership, 1983-2004



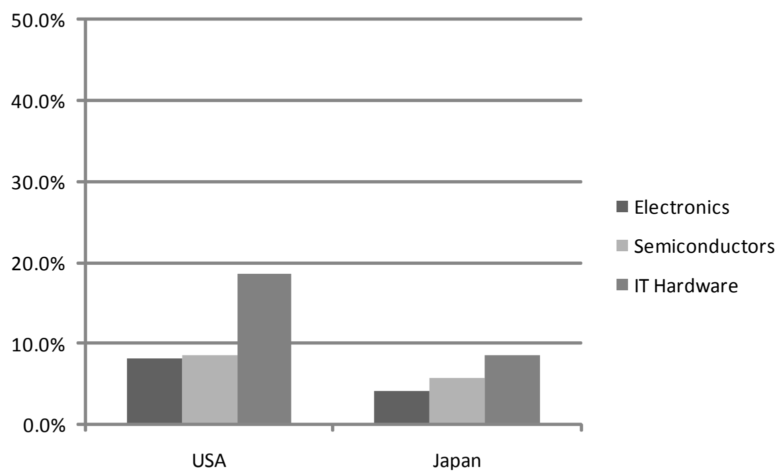
This table compares a measure of firm-level software intensity of patent citations for the firms in our sample by the geographical region of their origin and the geographical region of invention. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The software intensity of citations variable is calculated as the share of citations made to software patents in total citations made by all patents granted to a firm in our sample period, 1983-2004, averaged across all firms belonging to a given region of origin - region of invention combination. Geography of invention is determined using geographical locations of all inventors listed on the patent document. T-tests for differences in means across geographical groups show that differences are statistically significant at the 0.01 level in the case of all group pairs.

Figure 9-3: Software Intensity of Patenting (Share of Software Patents), Japanese Owned Patents, by Industry and Geography of Invention, 1983-2004



This table compares a measure of firm-level software intensity of patenting for the Japanese firms in our sample by the geographical region of invention, separately for three industrial subsectors in Information Technology. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The software intensity variable is calculated as the share of software patents in total patents granted in the sample period, 1983-2004, averaged across all firms belonging to a given region of invention - industrial subsector combination. Geography of invention is determined using geographical locations of all inventors listed on the patent document. T-tests for differences in means across geographical groups show that differences are statistically significant at the 0.01 level in the case of all group pairs, except in the case of "electronics" and "semiconductors" where the region of invention is USA.

Figure 9-4: Software Intensity of Patenting (Share of Citations Made to Software), Japanese Owned Patents, by Industry and Geography of Invention, 1983-2004



This table compares a measure of firm-level software intensity of patent citations for the Japanese firms in our sample by the geographical region of invention, separately for three industrial subsectors in Information Technology. The data used to construct measures of software intensity come from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. The software intensity of citations variable is calculated as the share of citations made to software patents in total citations made by all patents granted to a firm in our sample period, 1983-2004, averaged across all firms belonging to a region of invention - industrial subsector combination. Geography of invention is determined using geographical locations of all inventors listed on the patent document. T-tests for differences in means across geographical groups show that differences are statistically significant at the 0.01 level in the case of all group pairs, except in the case of "electronics" and "semiconductors" where the region of invention is USA.

Essay 2: "Globalization of Software Innovation: Does the U.S. Have an Advantage in Downstream Software Research?"

(with Ashish Arora and Chris Forman)

Introduction

Software is increasingly a global business. Exports of business services and computer and information services have grown at an average annual rate of 27% in India (1995-2003) and at a rate of 46% in Ireland (1995-2004), with similar rapid growth in Brazil, China, and Israel. However, it is important to note that despite the increasingly global nature of the software business, the United States still overwhelmingly dominates both software production and use. Arora and Gambardella (2005) report that software employment in the United States in the years 2000-01 contributed to more than a third of total world software employment, while in 2004 more than $300 billion, or 40-50% of world software sales, were generated in the United States.

Even though the production of software and procurement of software services are being conducted in increasingly many places around the globe, it has been shown that software activities conducted offshore and in the United States tend to differ significantly in their nature. Specifically, prior research has pointed out that the growth in offshore software activities has largely been driven by outsourcing of routine, standardized, and well-specified activities such as software customization and maintenance, mass software customization, and business process outsourcing services. As an illustration, a breakdown of India's software sector for 2006 shows that IT services and business process outsourcing represented 80-85% of total industry revenue, while high value-added engineering and R&D services represented less than 20% of the total (Nasscom, 2006).

Thus, while software production takes place in many countries, innovative software is not created everywhere. As shown in Figure 1, which measures software innovation by counts of U.S.-granted software patents[1], inventive activities in software are disproportionally concentrated in the United States, followed by a small set of other countries, most notably Japan. While there are many issues associated with using software patents to measure innovative activity in software, and we discuss these in detail later on in this essay, the broad implications of the graph are clear: the United States holds an absolute advantage in software research, and despite the recent globalization of the software industry, the U.S. advantage in software innovation seems to have remained largely undiminished.[2]

While the United States holds an absolute advantage in software research, there is anecdotal evidence that the United States enjoys a particular advantage in downstream research related to specific software applications. This advantage may arise from the presence of many more lead users of software in the US than in other regions of the world, particularly in business software. For example, Bloom, Sadun, and Van Reenen (2007) argue that American businesses are more effective at using IT than those from other countries. This may have facilitated the U.S. advantage in downstream software research as it has been argued that the presence of and interaction with lead users can be a crucial input into the innovation process for some types of inventions (e.g. Von Hippel, 1986). Bhidé (2008) provides many case studies showing how software innovation related to downstream applications may be easier to conduct successfully in

[1] Using U.S.-granted patents to measure global patenting activity in software is standard practice in the literature. For a discussion of this topic, see, for example, Bessen and Hunt (2007).

[2] For further evidence on the absolute advantage the US holds in software research, see Arora, Forman, and Yoon (2008).

the U.S., due to the importance of feedback from users. Our own interviews with R&D managers of software research laboratories located in India also confirmed this view.[3]

Apart from this indirect anecdotal evidence, at present there is little direct empirical evidence on whether and how the U.S. enjoys an advantage in the production of application-specific downstream software inventions. This absence of evidence might be an important gap to fill in for several reasons. First, the market for software applications is large. Private US domestic investment in software was $227.3 billion in 2007, according to the National Income and Product Accounts. Second, a comparison of patent characteristics of downstream and upstream software invented in the United States and offshore shows that quality differences between downstream software inventions performed at home and offshore might be more pronounced than those for upstream inventions[4]. Third, answers to this question will further our broader understanding of what factors determine where software research is conducted. Fourth, to the extent that firms with a U.S. presence can use internal knowledge flows and other mechanisms to mitigate the challenges of conducting some types of software research offshore, this would suggest that firms with such a presence will have an advantage in global development of software applications, and will aid in furthering our understanding of knowledge flows within multinational companies (e.g., Branstetter 2006; MacGarvie 2006; Singh 2005).

This essay contributes to the literature by providing empirical evidence for the existence of a U.S. comparative advantage in downstream software research. Using data on patents applied

[3] In December 2007, Arora and Forman conducted a series of interviews in India with managers of U.S. IT firms' Indian R&D centers.

[4] See Table 1d in the Appendix for evidence. Table 1d provides summary statistics for our sample, which is described in detail later on in this paper.

for at the United States Patent and Trademark Office (USPTO), we develop a novel classification algorithm, apply it to a large sample of software patents over the period 1990-2006, and classify them in downstream and upstream categories.[5] We label patents as related to downstream research when they are directed toward a specific business or consumer application, such as business applications software or scientific software, and label the remainder - which includes artificial intelligence software - as upstream.[6] We then explore how geographic location varies with the nature of software innovation in order to empirically address our research question.

We first provide descriptive evidence that the likelihood of producing downstream software inventions is higher in the U.S. than in other countries. A prelude to these results is shown in Figure 2. It shows a graph of the percentage of downstream patents invented in the U.S. and outside of the U.S. over the period 1990-2006.[7] The percentage of downstream software patents is significantly higher in the U.S. than elsewhere. The mean difference in the percentage of downstream software research is 5.4 percentage points and this difference has not changed significantly over time, suggesting that the type of software innovation conducted in the United States and non-U.S. locations differs significantly.

While these basic statistics are informative, we continue by assessing whether the U.S. appears to have a comparative advantage in downstream software research once firm, patent, and time effects are accounted for. We estimate a model of how firms chose to allocate research

[5] We follow the definition of Hall and MacGarvie (2006) in defining software patents. Please see the data section of this paper for a more detailed discussion.

[6] In this paper we use invention and innovation interchangeably, adopting the position of Mokyr (2002) that in the long run invention is a necessary precursor to innovation.

[7] The data points are a moving average with a window of two years before and two years after to smooth year-to-year variations in the data.

projects across geographical regions and show that U.S.-based firms are less likely to conduct downstream research offshore. We observe this result even though, as we detail later on in the essay, our research design is likely to work against us finding any relationship between the nature of invention and the choice of geographical location. While we cannot assert causality, we are able to demonstrate that there appears to be a robust pattern in the data. We then study one possible strategy multinational firms might employ in order to mitigate the challenges associated with conducting inventive activities offshore. Specifically, we ask whether firms use increased internal knowledge flows between the home base and the offshore location to substitute for the possible lack of local capabilities. Using a set of self-citation and co-invention models, we explore how knowledge flows vary by the nature of invention located offshore and find some evidence suggesting that U.S. firms may indeed use internal knowledge flows to mitigate the challenges associated with conducting inventive activities offshore.

Finally, we provide an initial empirical exploration of the suggested possible sources of the U.S. advantage in downstream software innovation. This is worthwhile because uncovering the primary sources for the U.S. advantage in downstream software innovation might help design policy prescriptions on how to preserve and further this advantage. It has been suggested that the U.S. advantage is built on two main pillars. The first suggested pillar is the excellent innovation infrastructure, especially the large stock of software developers in the United States, including those attracted from other countries. The second is the strong presence of lead users in the United States, i.e. the high willingness of U.S. firms to try new software for all types of productive uses in an effective way. Exploiting variation in firm location across industries and geographical regions, we provide suggestive evidence that this assertion might indeed be correct.

Prior research

Our research contributes to four research areas: work on the types of tradable services; research on the mediation of technical knowledge through multinational companies; recent research that has explored the location decisions of multinational firms, and the nascent literature that uses software patents to empirically study innovation in the software industry.

First, our research is related to recent work that has examined the tradability of services. Jensen and Kletzer (2006) identify tradable services industries by examining the extent of their geographic concentration, and then show that tradable services industries do not exhibit slower employment growth. Arora and Forman (2007) show that markets for some IT outsourcing services are local, providing evidence that some of these services have some irreducible “local” component to them. However, none of this research examines the extent to which research activities can be done remotely, nor do they examine how knowledge transfer within an enterprise can mitigate the disadvantages of a conducting research in an offshore location.

Second, this essay is related to a set of papers that have explored how multinational firms transfer knowledge between its subunits. Broadly speaking, this research examines the hypothesis that tacit knowledge is more efficiently transferred within the boundaries of a firm (Kogut and Zander 1993). In particular, itexamines whether knowledge transfer is mediated through exporting and foreign direct investment (Branstetter 2006; MacGarvie 2006; Singh 2005). We add to this literature by examining the link between knowledge flows and different types of research.

Third, this essay pertains to a recent literature that has examined how multinational firms choose to where to locate their activities (e.g., Alcacer and Zhao 2007; Zhao 2006; Chung and

Alcacer 2002). Of this literature, this essay is perhaps closest to Zhao (2006), who examines whether firms modularize the innovation process to manage their intellectual property in countries with weak intellectual property regimes. However, while Zhao studies how citation patterns vary among patents invented in US and foreign countries with weaker IPR regimes, we use the example of software innovation to study how location decisions and citation patterns differ by type of innovation.

Last, this essay adds to a very recent literature that has begun to use patents to empirically explore how innovation is conducted in the software industry. This literature has leveraged an increasing patentability of software inventions and a decreasing importance of copyrights as a tool for intellectual property protection in software (Graham and Mowery 2003; Lerner and Zhu 2007), in order to use patents to study innovation in software. These papers have established that software patents are valuable for firms (Bessen and Hunt 2007; Hall and MacGarvie 2006) and can create entry barriers for new entrants into various segments of the software industry. (Cockburn and MacGarvie 2006, 2007). This essay contributes to this expanding literature by looking at a different question: the decision of where to locate different types of inventive activity in software.

Does the U.S. have a comparative advantage in downstream software innovation?

To study the geography of innovation for downstream applications and upstream software, we proceed in several steps as we seek to demonstrate the existence of a comparative advantage of the U.S. in downstream software innovation. First, we study the likelihood of downstream/upstream software research conditional on region of invention. Second, we model how firm choices on where to locate their research activities differ for downstream and upstream research projects. Last, we examine citation and co-invention patterns of patents resulting from offshore software research projects to examine whether U.S. multinational firms might use internal knowledge transfer as a strategy to mitigate the challenges associated with conducting innovation remotely.

Approach

Step 1: Geographic location and the nature of innovation

Our first step is to provide descriptive evidence for whether the U.S. has a comparative advantage in software research related to downstream software innovation. We follow prior work and use patents as a measure of inventive activity. The limitations of using patents as a measure of inventive activity are well known and there are particular issues associated with using software patent data, which we detail in the data section below.[8] In particular, for our purposes the use of patent data implies that we attempt to identify differences in the geographic pattern of downstream software innovation versus upstream innovation, within the class of innovation that is both patentable and indeed patented.

While Figure 1 provides an initial description of the geographic distribution of software innovation and Figure 2 provides suggestive evidence for the geographic location of downstream and upstream software innovation, we wish to refine the analysis of Figure 2 by examining the differences in geographic distribution between downstream and upstream innovation after time effects, firm-level effects, and differences in patent characteristics have been controlled for. In particular, for patent i, belonging to firm p, and filed for in year t, we estimate the following:

[8] Possible uses of patent data and the limitations to their use in economic research have been well documented. An excellent overview of this topic can be found in Jaffe and Trajtenberg (2002, 2005).

$$Downstream_{ipt} = \alpha US_{ip} + \beta X_{ip} + \mu_i + \tau_t + \varepsilon_{ipt} \tag{1}$$

Here τ_t represents a year dummy, μ_i represents a time-invariant firm fixed effect to control for cross-firm differences in the type of innovation, and X_{ip} represents a vector of patent characteristics controlling for patent originality, importance, and reliance on science as derived from the literature (see The Appendix for details). Our interest is in examining whether $\alpha > 0$; that is, whether a patent is more likely to be associated with downstream software if it is invented in the U.S. We estimate equation (1) with a linear probability model, and use robust standard errors to account for the presence of heteroskedasticity.

Step 2: Exploring the decisions of U.S. firms where to locate inventive activities

Next, we move on to examining the simple choice of US firms whether to locate an invention within the US or in another country. In particular, we examine the choice of US firm i to locate patent p during time t in a country outside of the US.

$$NonUS_{ipt} = \alpha Downstream_{ip} + \beta X_{ip} + \mu_i + \tau_t + \varepsilon_{ipt} \tag{2}$$

Here, τ_t represents a year dummy to control for changes in the propensity to locate innovation outside the US over time, μ_i represents a firm fixed effect to control for cross-firm differences in the propensity to locate innovation globally, and X_{ip} represents the same controls for patent characteristics derived from the literature as before. We estimate this model using robust standard errors.

We are interested in examining whether $\alpha < 0$; if the U.S. as a location does indeed have an advantage in downstream software research downstream from the point of view of U.S. firms, then we should observe that downstream software patents are less likely to be invented outside of the U.S. than upstream patents, other things equal. Note that when using firm-level time-invariant fixed effects, our identification strategy relies on variance within firms in the type of patent and the region of inventor location (inside the U.S. or offshore).

Notice also that by using patents of U.S. firms we are biasing against finding in favor of our hypothesis for two reasons. First, by focusing on the set of U.S. firms for our analysis, we examine the decisions of a set of firms for which the advantage of locating inventive activities in the U.S. is likely to be least important as these firms can presumably use within-firm knowledge transfers and other strategies to mitigate the challenges of "poor" location. The ability of such multinational firms to mediate knowledge transfers has been well documented in the literature (e.g. Branstetter 2006; MacGarvie 2006).

A second bias against a finding of $\alpha < 0$ is that upstream innovation, while it may draw less on the sources of comparative advantage of the US invention location, may also be more technically or scientifically challenging. If such research is more challenging, it may also rely more heavily on local scientific capabilities. For example, prior work in pharmaceuticals has shown that firms frequently draw upon local resources from both public and private sources when completing research (Cockburn and Henderson 1998; Furman, Kyle, Cockburn, and Henderson Forthcoming). It is widely assumed that such concerns have motivated firms in information technology (IT) hardware, software, and pharmaceuticals to cluster together. Thus, if upstream research is more technically challenging and relies on local external resources, then we may observe a positive correlation between downstream software research and Non-US inventor location in our data.

To explore this possibility, we use a multiple choice model of the decision to locate inventions across 4 regions: US, Japan, OECD countries, and developing non-OECD countries.[9] We examine the choice of US firm i to locate patent p during time t in country/region j as follows:

$$U_{ijpt} = \alpha_j Downstream_{ip} + \beta_j X_{ip} + \tau_t + \varepsilon_{ipt} \tag{3}$$

Because our regressors are individual patent specific and do not vary by region, we can use a multinomial logit model to estimate the above equation. As before, we include a vector of patent characteristics and a set of alternative-year dummies in our regression. Note that in this particular specification we are not able to control for time-invariant differences in firm propensities to conduct software innovation across the four regions.[10] Notice, also, that in estimating this model we implicitly assume that firms have a portfolio of innovation projects, all of which can be successfully completed with probability 1, to allocate in the U.S. or offshore. This might not be an innocuous assumption. In addition, note that in estimating the above equation using a multinomial logit we implicitly adhere to the validity of the independence of irrelevant alternatives (IIA) assumption. Specifically, we assume that when determining the probability of locating research in any pair of regions, all other regions are "irrelevant" in determining that probability. In our case, this might not be a realistic assumption to make about substitution patterns between regions. To address this issue, we would have to use a nested

[9] We have also estimated an expanded, 11- and 12-alternative country-choice model, and the results are qualitatively similar. In the second part of the paper, we also estimate a 3-choice version of the model, where we lump together non-OECD and OECD countries into a single region.

[10] We have also attempted to estimate a mixed-effects multinomial logit version of (3) in which we accounted for firm-level "fixed" effects, but we ran into problems with algorithm convergence.

region/sub-region location choice model, which is something we plan to incorporate in future versions of this essay.

It is instructive to look at a comparison of the availability of technical capabilities across the four regions included in the multiple choice model. At the risk of oversimplifying, we expect technical capabilities to be strong in the US; technical capabilities to also be strong in developed OECD countries and Japan; and technical capabilities to be weak in developing non-OECD countries. As a result, if upstream is more difficult to produce, then we expect the net benefits of such research to be relatively greater in Japan and other developed OECD countries than in developing OECD countries. Thus, if upstream software research is more difficult, while we expect $\alpha_{OECD} < 0$, $\alpha_{JP} < 0$, and $\alpha_{OTHER} < 0$, we also expect $\alpha_{OECD} > \alpha_{OTHER}$ and $\alpha_{JP} > \alpha_{OTHER}$ because the presence of greater technical capabilities in Japan and developed OECD countries than in developing non-OECD countries will make upstream research relatively more attractive (and applied software research relatively less attractive) for US firms in those offshore locations.

As already mentioned, in motivating equations (2) and (3) we assumed that firms can choose to source new research projects in a location and each research project will be converted into a patent with probability one. Of course, in practice new research projects may fail to be converted into patents for a variety of reasons. Research projects may, and often do, fail. Further, as is well known, not all inventions are patentable (Jaffe and Trajtenberg 2005). The primary identification assumption for our work is that unobserved differences in the propensity to convert research projects into patents across countries or regions are uncorrelated with whether the research is upstream or downstream. While this assumption may be violated if, for example, foreign patent attorneys are less successful in identifying novelty and nonobviousness in downstream software innovation, we view this possibility as being consistent with the notion that downstream software innovation is more difficult to do in foreign countries.

Step 3: Investigation of strategies for mitigating the challenges of offshore innovation

While US-based multinational firms might find it more difficult to conduct downstream software research remotely, multinational firms may be able to use specific strategies to cope with that challenge. As noted above, prior work has demonstrated the role of multinational firms as a conduit for knowledge flows (e.g., Branstetter 2006; MacGarvie 2006; Singh 2005). In our own interviews of software research labs of U.S. multinationals in India we found that these firms used a number of mechanisms to transfer knowledge to their offshore research labs. For example, these firms would hold product fairs to match research projects to application needs; would direct research groups to work on particular needs of application groups; or, in some cases, would ask research groups to work on needs of particular clients.

We look at self-citations and international co-invention patterns to examine whether U.S. multinational firms use intra-firm knowledge flows from the home country to mediate the challenges associated with conducting downstream software innovation in offshore locations. Trajtenberg et al. (1997) proposed that self-citations can be used to measure the fraction of benefits from a patent accruing to the original inventor, while Hall et al (2005) suggest that self-citations represent internal knowledge transfers that can lead to competitive advantage. Similarly, co-invention resulting from collaboration between home-country inventors and those located in offshore locations can also be viewed as a related strategy multinational firms may employ to substitute internal knowledge flows for inferior local capabilities in offshore locations.

To measure whether internal knowledge flows are particularly important for offshore downstream software innovation of U.S. firms, we first estimate a Poisson count data model of the number of self-citations, motivated by work of Zhao (2006). Specifically, we assume that the number of self-citations follows a Poisson process and is conditional on a set of time-varying covariates. The conditional distribution of the number of self-citations and the conditional mean are thus given as follows:

$$\Pr(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$
$$\xi_{ipt} = \alpha_1 NonUS_{ipt} + \alpha_2 Downstream_{ipt} + \alpha_3 NonUS_{ipt} * Downstream_{ipt} + \beta X_{ipt} + \tau_t \qquad (4)$$
$$f(y) = \frac{e^{-\exp(\xi_t)} e^{\xi_t * y_t}}{y!}$$

We expect the parameter α_3 to be positive, indicating that within-firm knowledge flows will be particularly important for offshore downstream software research. As a robustness check, we also estimate (4) as a negative binomial model, as well as a linear regression version of (4).

A different approach to measuring whether internal knowledge flows are particularly important for offshore downstream software innovation is to look at how co-invention patterns differ between downstream and upstream research projects located offshore. Specifically, we estimate how the likelihood of co-invention varies with type of invention, conditional on in being invented offshore.

$$Coinvented_{ipt} = \alpha Downstream_{ipt} + \beta X_{ipt} + \mu_i + \tau_t + \varepsilon_{ipt} \qquad (5)$$

Here, τ_t represents a year dummy to control for changes in the propensity to co-invent over time, μ_i represents a time-invariant fixed effect to control for cross-firm differences in the propensity to co-invent, and X_{ip} represents the same controls for patent characteristics derived from the literature as before. A patent is assumed to be co-invented if inventors from both the home country (the US) and other countries are jointly listed on the patent document. We expect the parameter α to be positive, indicating that co-within-firm knowledge flows through co-invention will be particularly important for offshore downstream software research.

Data and variables

Software patents

To measure the location of inventive activity and examine knowledge flows within multinational firms, we use data on software patents issued by the USPTO. There are, clearly, significant limitations to the use of software patents as a measure of inventive activity. As Jaffe and Trajtenberg (2002) note, not all inventions meet the US Patent and Trade Office (USPTO) criteria for patentability,[11] and inventors must make an explicit decision to patent an invention, as opposed to relying on some other method of intellectual property protection. Both of these issues are particularly acute in the patenting of software. Historically, inventions in software were not patentable[12] and for a time copyright was the predominant form of formal intellectual property protection in software. However, a series of court decisions widened the scope of software patents. Eventually, this culminated in the Commissioner of Patents issuing guidelines for the patenting of software that allowed inventors to patent any software embodied in physical media (Hall and MacGarvie 2006). In contrast, over the same period a series of cases, including several

[11] Note that not all inventions also meet the criteria for patentability at the European Patent Office (EPO) and/or the Japanese Patent Office (JPO).

[12] The following provides what is necessarily a brief overview of the history of intellectual property protection in software. For a more detailed overview, see Graham and Mowery (2003) and Hall and MacGarvie (2006).

copyright infringement cases brought by Lotus Development weakened the intellectual property protection offered by copyrights. Graham and Mowery (2003) show that over this period the number of granted software patents has increased dramatically while the propensity of firms to copyright has declined substantially.[13] Recent research has shown that the stock of patents is correlated with firm success in the software industry (Merges 2006), suggesting that patents may be a potentially useful metric of the inventive output of firms.

Another challenge in using software patents to measure inventive activity in software is identifying exactly which patents are software patents.[14] Software patents are not assigned to a particular class or subclass in either the USPTO or International Patent Classification (IPC) schemes. Moreover, there is no unique field in patents identifying them as software patents. Graham and Mowery (2003) were the first to attempt to overcome this obstacle by using the patent classification system to systematically identify software patents for research purposes. The Graham-Mowery approach to identify software patents has been used and revised by others. Graham and Mowery (2005) identify software patents using USPTO classifications. Hall and MacGarvie (2006) identify software patents by finding the USPTO class-subclass combinations in which fifteen large software firms patent. To identify their final sample, they intersect the resulting set of patents with another keyword definition used by Bessen and Hunt (2007). We follow the Hall and MacGarvie approach in order to identify the population of software patents.

---

[13] The set of patentable inventions is narrower in Europe than in the US. To be patentable, then European Patent Convention requires that inventions address a particular technical problem and suggest a technical means to solve this problem (Thoma and Torrisi 2006). The implication of this requirement is that "inventions having a technical character that are or may be implemented by computer programs may well be patentable" (EPO 2005).

[14] This section provides an overview of the issues in identifying software patents. For a more complete discussion, see Layne-Farrar (2005) and Hall and MacGarvie (2006).

Classifying software patents

A major data challenge for our work is to find a systematic way of identifying whether a patent is related to upstream research or is related to research that serves a specific user application or need. Cockburn and MacGarvie (2006, 2007) face a similar challenge of mapping software patents to industries in their recent work exploring whether software patents create barriers to entry in software markets. They develop an innovative citation pattern based method for classifying software patents into a predefined set of industries as defined in the Corptech database of high technology companies. A major limitation of their method is that it can only be used to classify a subset of the software patent population. However, we extend their method in a way that allows us to classify all software patents.

We first use the Hall-MacGarvie (2006) method to identify the software patent population. Next, we develop a training dataset by selecting software patents applied for by firms that belong to a single Corptech software category, the method used by Cockburn and MacGarvie (2006). Following them we assume that patents filed by firms that belong to a single Corptech class map to that class. We then use a machine learning algorithm, which mines the text of the patents in the training set to learn the characteristics of patents belonging to each software class. We apply this machine learning algorithm to classify the entire software patent population into these software categories. Finally, we aggregate categories into upstream and downstream software.

Our method uses text mining techniques to classify software patents into industries on the basis of the text of those patents. For example, a patent with the word "MRP" in the patent abstract would be relatively more likely to belong to the patent class "Manufacturing Software" and relatively less likely to belong to the class "Artificial Intelligence." Our classification methodology is discussed in greater detail in the Appendix.

Based on this approach, we were able to map software patents into the industries listed in Appendix, Table A. As the industries used in Cockburn and MacGarvie, these categories are also derived from product codes in the Corptech directory of technology companies. To identify downstream software patents, we used patents that mapped into Corptech industries that extensively referenced applied software: these included categories such as business applications software, educational software, and scientific/technical software. The complement—upstream software—includes artificial intelligence software.

We then used a structured manual approach, the details of which are described in the Appendix, to classify a stratified random sample of 199 software patents over our sample period, where we stratified the sample to include patents from all 15 of our disaggregated software categories, and compared the results of our manual classification effort with those produced by the machine learning algorithm described above. We opted to use only the following categories for which both classification classification methods produced very similar results in our final sample: business application software (downstream), artificial intelligence software (upstream), educational/training software (downstream), manufacturing software (downstream), and science/technical software (downstream). Our system classified these patents correctly between 65.4% and 87.5% of the time.

Our final estimation sample thus includes software patents identified using the Hall-MacGarvie strategy, classified as belonging to one of the classes above, and filed over the period

1990-2006. Finally, we exclude software patents for which we were unable to identify the assignee.

Patent characteristics

To control for differences in patent importance and distance to the technical frontier, we compute measures of patent originality, importance, and percent of citations going to scientific publications as suggested by the literature (Trajtenberg et al, 1997). A detailed description of calculated measures can be found in the Appendix. These measures are based on backward citations rather than forward citations; this is due to the relative newness of our sample and the difficulty of computing reliable measures of forward citations over the years 2004-2007. However, we believe this might not present a major detriment to our analysis as forward-citations based and backward-citations based measures for the patents in are sample are highly correlated.

As has been suggested by prior work (e.g., Zhao 2006), patents may be assigned to a parent company or one of its subsidiaries for unobservable reasons. In order to deal with this issue, we group multiunit firms into single, integrated strategic agents. To this end, we use preliminary (alpha) version of NBER Patent Data Project's 2008 matching dataset. [15]

Descriptive statistics

Table 1a provides descriptive statistics of our estimation sample; as indicated by Figure 1b a little under half (42.8%) of our sample is comprised of downstream software patents, and only 29.0% of patents in our sample were invented outside of the United States. Table 1b shows

[15] For a description of the preliminary (alpha) version of this matching dataset, see http://www.nber.org/~jbessen/matchdoc.pdf

how our sample differs for patents invented in the U.S. versus those invented in foreign countries. Patents invented in the United States exhibit significantly higher mean measured originality (0.526 versus 0.486) and importance (207.1 compared to 118.6). They have also been referenced by a significantly higher number of other patents (14.1 versus 9.4), as well as include a higher number of references to other patents (15.7 versus 10.0). Summary statistics in Figure 1b further suggest that a larger percentage of U.S.-invented than non U.S.-invented patents are downstream (44.9% compared to 37.8%). Finally, Table 1c summarizes the sample by nature of invention. We see that downstream and upstream patents have comparable measured patent characteristics, but a higher share of downstream patents than upstream patents are invented in the United States (74.5% vs. 68.6%).

Results

Step 1: Geographic location and the nature of innovation

We begin by plotting a smoothened ratio of downstream-to-upstream software patents for patents invented in the U.S. and those invented outside the U.S. trough time, which is shown in Figure 2. We observe that the share of downstream patents among U.S.-invented software patents consistently exceeds the share of downstream patents among non-U.S.-invented patents by an average margin of 5.4 percentage points. The difference has not diminished in the period 1990-2006, even though software patent production outside the U.S. has been drastically increased during this period. This descriptive result suggests that downstream software innovation seems to be significantly more concentrated in the U.S. than upstream innovation.

Moving beyond the descriptive analysis, we examine how the likelihood of a software patent being downstream varies with location of invention. Table 2 reports estimates of equation

(1) by employing both ordinary least squares and the fixed-effect estimator, for a set of subsamples and the entire sample. Our baseline model in column (3), which includes a set of firm fixed effects, time fixed effects, and a set of controls for patent characteristics, suggests that software patents invented in the US are 4.3 percentage points more likely to be downstream related than an otherwise identical patent invented outside the US. Columns (1)-(7) suggest this result is robust to a variety of model specifications and estimation routines. In other words, even after extensive controls for time variation, firm-specific differences in the structure of their patent portfolio, and controls for patent characteristics, downstream software innovation seems to be concentrated in the U.S., alluding to the existence of a U.S. relative advantage in downstream software research.

Step 2: Exploring the decisions of U.S. firms where to locate inventive activities

Table 3 reports results of estimating equation (2), exploring how the likelihood of inventing a patent offshore varies based on the extent to which the invention is separate from downstream use. Column (1) shows that without including firm-level fixed effects, downstream patents are 1.6 percentage points less likely to be invented outside of the U.S., and this result is statistically significant at the 1% level. This is consistent with what we would expect if the U.S. was indeed a preferred location for downstream software research from the perspective of U.S. firms, as described above. Column (2) shows that once fixed effects are added to the model the marginal effect remains almost intact at 1.3 percentage points and remains statistically significant at the 1% level. In addition, column (3) reports that even once measures of patent characteristics are accounted for, downstream software patents are still 1.35 percentage points (or 13.4%) less likely to be invented outside of the U.S. We note that these results are strong and statistically significant despite the aforementioned biases against us finding a relationship between downstream-related nature of invention and offshore location.

In columns (4)-(7) we conduct several additional robustness checks. We first estimate versions of equation (2) separately for the post-1995 time period and for patents belonging to firms owning more than 50 patents. We observe that the effect of downstream on the offshoring decision remains strong and statistically significant as before. We also estimate separate regressions for patents in the top and bottom half of the distribution in terms of importance as calculated using backward citations. By running these regressions on subsamples, we can observe if our results hold even among those patents with the highest values for importance. We find that they do. (In unreported regressions, we also show that our results hold among those patents that are highest in importance using forward citations measures.) The import of our findings is to make sure our downstream/upstream classification is not simply a proxy for technical difficulty, thereby providing us reassurance that from the perspective of U.S. firms, offshoring downstream software innovation apparently seems to be genuinely more difficult compared to upstream software innovation.

Results in Table 4 are consistent with the above result, namely that measures of patent importance (difficulty) do not seem to mediate the relationship between the nature of innovation and offshore location choice for U.S. firms. Table 4 presents results of a multinomial logit model of the US firms' decisions of which region to locate a patentable software invention in. It represents a generalization of Table 3, where we break the alternative "offshore" into Japan, developed OECD, and developing non-OECD regions. These results tell us how the marginal effect of the nature of innovation varies across locations. Table 4 shows that U.S. firms are significantly less likely to locate downstream software patents in any of the three offshore regions than in the U.S.. In other words, both among regions with technical capabilities similar to those of the U.S. and those with (as we conjectured) significantly worse technical capabilities, the marginal effect of downstream is significantly negative. In addition to α_{OECD}, α_{JP}, and α_{OTHER} all being statistically significantly negative, which is in line with what we would expect to see if U.S. firms considered the U.S. to be a relatively more advantageous location for downstream research than upstream research (or in other words, if the U.S. firms considered downstream research to be more costly to be conducted offshore than upstream research), Wald tests for $\alpha_{OECD} > \alpha_{OTHER}$ and $\alpha_{JP} > \alpha_{OTHER}$ also fail to reject the null hypotheses that α_{OECD} equals α_{OTHER} , and that α_{OECD} equals α_{OTHER} . Although this is an imperfect test at best, it does seem to suggest that technical capabilities of offshore regions do not seem to mediate the effect of the nature of innovation the software offshoring decisions of US-based multinational firms.

Step 3: Investigation of strategies for mitigating the challenges of innovating in "poor" location

In this section we examine whether, for U.S. firms, within-firm knowledge transfers are especially important to offshore downstream software innovation.

Table 5 reports estimates of equation (5), the self-citation model. Our baseline model is in column (4), which uses conditional firm fixed effects (Hausman, Hall, and Griliches 1984) and a set of patent controls. We find that while the effect of being invented offshore on the propensity to self-cite is larger for downstream than for upstream patents, it is relatively small and statistically indistinguishable from zero, providing little support for the notion that within-firm knowledge transfers are more important for these downstream patents belonging to foreign subsidiaries of U.S. multinational firms. As a robustness check, we estimated a variety of different specifications, including a zero-inflated Poisson model, in part because some self-citations may not yet have been observed. We have also conducted additional robustness checks, the results of some of which are presented in Table 5, but the non-result we find seems to be robust.

Finally, in Table 6 we report the results from estimating a co-invention model in which we explore how the likelihood of international co-invention varies with the nature of innovation for offshore patents belonging to US multinational firms. The results presented in columns (1) - (3), in which we do not use firm-level fixed effects to control for time-invariant firm-level variation in the propensity to co-invent, provide suggestive evidence that the likelihood of co-invention is higher for offshore downstream software innovation projects than offshore upstream software innovation projects. In addition, we see that the effect of downstream in column (2), which estimates the likelihood of co-invention for U.S. assigned offshore patents, when co-invention is defined as the primary inventor being located in the home country, while one or more secondary inventors are located offshore, is positive and statistically significant at the 5% level. In contrast, the effect of downstream in column (3), where co-invention is defined as the primary inventor being located offshore, while at least one secondary inventor is located in the home country, is negative and not statistically significant. This is an interesting result and suggests that U.S. multinational firms may use a domestic primary inventor as a conduit for within-firm knowledge flows to mitigate the challenges of inventing downstream software research offshore.

However, when we add firm fixed effects to this model, while the effects are qualitatively preserved, the coefficients lose their statistical significance. Columns (4)-(6) report these results. The loss of statistical significance in these models might be at least partially due to the fact that co-invention is a relatively rare occurrence. Only 2-3% of the patent population exhibits international co-invention. Thus there might be too little within-firm variation in downstream/upstream co-invention patterns to facilitate identification of the effect of the nature of invention on the likelihood of co-invention.

Exploration of the sources of U.S. advantage in downstream software innovation

The empirical part of this essay thus far has focused on establishing the existence of a comparative advantage of United States in the production of downstream software research. First, we show that a typical software patent, controlling for its characteristics, is more likely to be downstream related if it is invented in the U.S. Second, U.S. assigned patents are less likely to originate offshore if they are downstream, even after firm fixed effects have been accounted for, and even if they are high quality patents. Third, we find some evidence that offshore patents assigned to U.S. firms are more likely to be co-invented with a domestic primary inventor if they are downstream, compared to upstream software patents. We also find some (weak) evidence that such patents are more likely to self-cite prior internal research, pointing to the role multinational firms might play in mitigating the challenges associated with inventing remotely in "poor" locations.

However, while we have alluded to some possible explanations for the observed comparative advantage of the U.S. in downstream related innovation in the text thus far, the question remains if can we discern the primary source(s) for this apparent advantage of the United States in downstream software innovation. There are at least three explanations, not mutually exclusive. The first is that our results are driven by the "firm effect": U.S. firms are simply more likely to engage in downstream research projects, and because U.S. firms "happen to be" disproportionally located in the U.S., we observe a relatively large portion of downstream research being located in the United States. At its extreme, this explanation would force us to implicitly assume that there are no differences in downstream-specific "supply-side factors" between the United States and other regions. Notice the implication: this setup allows us to use an empirical approach that explores how the geographical distribution of the type of patent innovation varies with the origin of the patenting firm, allowing us to test the validity of this first explanation.

A competing explanation argues that the main source for the observed U.S. advantage lies in the relative abundance of downstream-savvy U.S. based inventors and software developers. In other words, the United States is relatively more abundant with managerially and entrepreneurially gifted software R&D labor, which are relatively more efficient at producing downstream software innovations and are available for hire. It is important to note that while this explanation acknowledges that the observed differences in geographical distribution of downstream software research are indeed caused by the "invention location effect", this effect is assumed to be "general" i.e. not industry specific. Yet a third explanation, put forward by anecdotal evidence in prior literature and our own prior work, suggests the observed patterns in our data are driven by the fact that U.S.-located research is blessed with the proximity to and ability to interact with "lead users" of the end products, into which this software innovation is embedded. Put differently, this explanation suggests that the observed comparative advantage of the U.S. in downstream software research is due to the relative abundance of lead users of downstream (applied, business-related) software products (embodying software inventions), and that the proximity to and interaction with these lead users represents a critical input for the production of some types of software innovation. It is assumed that lead users are important because they provide valuable information about needs and uses of a product or service (Von

Hippel, 1986). Proximity to and interaction with lead users are thus important because the knowledge of needs and uses is often tacit or difficult to communicate, may require experimentation and iteration, and requires "rich" interaction, all of which is facilitated by physical proximity.

Do all inventions require input from lead users? Perhaps, but some arguably more so than others, and lead users might differ for different types of inventions. For some software innovations, for example, the lead users of the products they are embodied in might be other software developers (as is the case for software tools), whereas for other types of software innovation, the lead users are located outside the software industry, and in particular, are in leading business organizations. This distinction points to the potentially different uses of software product and services. For example, while software is used to power technical devices (embedded software in electronic goods of various sorts), software is also used to power communication devices of various kinds, including the Internet, and finally, software is used to facilitate the operation of large organizations themselves.

Notice that we have made two important assumptions here: first, that each class of software products is associated with a distinct set of lead users, and second, that the relative abundance of lead users (relative to the endowment of software inventors and developers) is higher in the US than in other regions. Simply put, lead business software users are relatively more concentrated in the United States than are software developers. Bloom, Sadun, and Van Reenen (2007) argue that American businesses are more effective at using IT than those from other countries, while Bhidé (2008) provides many case studies showing how software innovation related to specific business applications software products may be easier to conduct in the U.S., due to importance of feedback from users. In addition, while we remain agnostic on the distribution of lead users for more products into which more technical software is incorporated, we conjecture that they are, relative to the endowment of software developers, relatively more abundant in Japan and other developed countries.

It follows that this skewness of the relative geographical distribution of lead users can help us to observe the rest of the world to have a comparative advantage in the production of more (technical) upstream software innovation and the U.S. in the production of more (applied, business-related) downstream software. However, it is important to note that this observation is also consistent with the explanation that the U.S. is simply relatively more abundant with business- and application-savvy inventors, even if the distribution of lead business users and lead technical users (relative to the endowment of software developers) were uniform across the globe.

In order to distinguish between these two related but not identical explanations for our observation that the U.S. produces relatively more downstream software inventions than upstream software inventions, we can exploit the variation of firms across software-producing IT industries, non software-producing IT industries to examine how the share of upstream patents in U.S. invented patents relative to the share of upstream patents in non U.S. invented patents differs across firms in different industry categories. Conceptually, we expect firms whose primary business is to produce software for external clients to require more critical input from outside (lead) users than firms whose software inventions are primarily used internally to enhance and power other products these firms produce. Thus we expect, for example, software producing firms to rely more heavily on input from external lead users when conducting inventive activities than non-software producing IT firms. We can then exploit the variance in the industry of the patenting firm explore which explanation, the one based on lead-user supply-side factors or the explanation based on non-industry specific supply-side factors resonates most with the patterns in our data.

Evidence for distinguishing between possible hypotheses

While our current data does not enable us to rule out any of the proposed explanations, we can obtain an initial insight by exploring data on patenting behavior of firms across geographical locations by their country of origin and industry affiliation. The first identification strategy we follow is based on the fact that the two possible explanations yield different predictions regarding what how U.S.-firms should allocate different types of innovative activities across geographical locations, as well as how U.S. and non-U.S. firms should undertake software innovation at home and abroad. If the observed comparative advantage of the United States in downstream software innovation production, as presented in Figure 2 and estimations in Step 1, is primarily due to cross-national differences in downstream intensity of firms, then we should not observe invention location choices by firms to depend on the type of innovation.

The following 2x2 matrix summarizes our basic approach:

location of invention

firm location		US	non-US
	US	Pr(Down)(us,us)	Pr(Down)(us,non-US)
	non-US	Pr(Down)(non-US,us)	Pr(Down)(non-US,non-US)

*while keeping firm industry fixed

If we find that the likelihood of a patent being downstream, conditional on location of firm origin, does not vary by location of invention (if Pr(Down)(us,us) = Pr(Down)(us,non-us) & Pr(Down)(non-us,us) = Pr(Down)(non-us,non-us)), then this would suggest that the observed comparative advantage of the US in our data is driven by firm-patenting composition effects rather than supply-side differences in favorability of the U.S. and non-U.S. locations for downstream software innovation.

First note that we have effectively already partially established that inventor location does influence the likelihood of a patent being downstream for U.S. assigned patents in our estimations of equation (1) above, as reported in Table 2. Addition to that, estimates in Tables 7a-7i provide additional evidence in support of this finding. We begin by approximating the likelihood of a patent being downstream conditional on invention location and firm origin using a simple ratio of downstream to all software patents, where we restricted our sample to include only U.S. assigned and Japanese assigned patents invented in either the U.S. or in Japan in order to avoid the possible confounding effect of differences in technical capabilities on the likelihood to conduct various types of innovation offshore. Our approach allows us to directly relate the computed shares to the probabilities from the 2x2 matrix above. Examining the ratios in 7a, we find that Pr(Down)(us,us) > Pr(Down)(us,jp) and that Pr(Down)(jp,us) > Pr(Down)(jp,jp) when we compute the shares across all industries, and when we condition by industry, where we assign firms into two industry categories (software producing firm, non-software producing IT firm) based on their assigned 4-digit NAICS codes in Compustat.[16] Next, tables 7b and 7c report estimates of this likelihood using various specifications of equation (1). We find that across all

[16] We merge unique parent firm codes with Compustat data using the alpha version of NBER's matching data, which we previously also used to identify unique assignee codes and assignee – parent firm relationships above.

firm industry categories, the likelihood of a U.S. assigned patent being downstream is significantly positively associated with it being invented in the U.S. We also find that this is also the case for Japanese assigned patents, although in some model specifications the result is not statistically significant at the typical level. Finally, Tables 7d-7i report estimates of models, where we instead estimate the probability of being invented in the U.S. conditional on downstream and firm origin (US and Japan), and where we estimate this probability using various versions of equations (2) and (3). Again, we find that both U.S. and Japanese firms are more likely to locate their downstream innovation in the US, compares to Japan and other regions. While these effects differ considerably in size as well as in the level of their statistical significance across firm industry categories, they remain qualitatively unperturbed. Our evidence thus overwhelmingly suggests that the observed comparative advantage of the U.S. in downstream software innovation is not primarily driven by the fact that U.S. firms might simply be more efficient at conducting downstream software innovation, even when conditioning on firm industry. Comparison of our results for Japanese and U.S. assigned patents also suggests that both U.S. and Japanese firms recognize this comparative advantage and attempt take advantage of it by disproportionally locating their downstream innovation within the U.S.

Next we move on to attempting to distinguish between two other types of sources for the observed inventor location effect on downstream: relative downstream-savvy of U.S. inventors and relative abundance of lead users in the U.S. relative to other countries. The following 2x2 matrix summarizes our approach:

location of invention

firm industry		US	non-US
	"software-producing firms"	Pr(Down)(sof,us)	Pr(Down)(sof,non-us)
	"non software-producing IT firms"	Pr(Down)(nsof,us)	Pr(Down)(nsof,non-us)

** while keeping country of firm origin fixed*

Here, we investigate how the inventor location effect for downstream varies by industry category between firms that we expect to be more reliant on input from lead software users and those less so.

If the lead user effect is not present and the observed U.S. comparative advantage is primarily driven by non industry specific supply-side factors, then we would expect, using the syntax from the matrix above, Pr(Down)(sof,us) > Pr(Down)(sof,non-us), Pr(Down)(nsof,us) > Pr(Down)(nsof,non-us), but ((Pr(Down)(sof,us) - Pr(Down)(sof,non-us)) = (Pr(Down)(nsof,us) - Pr(Down)(nsof,non-us)). Put simply, while we in this case expect to find that downstream is be associated with higher likelihood of being invented in the U.S. relative to other regions, we do not expect to see this effect to differ for firms across industries. In contrast, if we find that ((Pr(Down)(sof,us) - Pr(Down)(sof,non-us)) >> (Pr(Down)(nsof,us) - Pr(Down)(nsof,non-us)), then this might suggest that the presence of lead users is important, particularly for firms operating in industries for which we assumed proximity to lead users matters most.

Table 7a presents nonparametric estimates of probabilities from the matrix above using a restricted sample of software patents, including U.S. and Japanese assigned patents invented in the U.S. or in Japan, by industry. These descriptive results show that the difference in share of downstream patents between U.S. and Japan locations varies greatly by industry, both for U.S. firms and Japanese firms. Specifically, while the difference in the share of downstream patents between U.S. invented and Japanese invented patents belonging to U.S. software producing firms equals 16 percentage points, this difference in shares is indistinguishable from zero in the case of non-software producing IT firms. We also see this pattern remains quantitatively and qualitatively unperturbed when we look at patents assigned to Japanese firms. In addition, in unreported additional calculations where we also look at software patents assigned to non-IT firms, we find that non-IT firms' differences in shares of downstream patents between U.S. invented and Japanese invented patents resemble those for software-producing firms. This initial evidence is thus consistent with the lead-user based explanation for the observed downstream comparative advantage on the United States.

Next, 7c provides parametric estimates of how the likelihood of downstream differs by invention location (U.S. and Japan) and industry, while keeping firm origin fixed. The results of this analysis are broadly consistent with the evidence presented above. Once time, firm fixed-effects, and patent characteristics are accounted for the invention location effect of the U.S. is largest and highly statistically significant for U.S. software-producing firms, followed by other U.S. firms, where it is moderate and statistically significant at the 10% level, while the likelihood of a patent being downstream is not higher in the U.S. than Japan for non-software producing IT firms. Results are analogous in size when we look at patents of Japanese firms, but the statistical significance of the U.S. invention location coefficient is diminished considerably once firm-fixed effects and patent characteristics are introduced into the model. These results suggest that the effect of being invented outside the U.S. on the likelihood of being downstream is negative and strong for those firms which we ex ante expected rely most on proximity to lead users.

Table 7h reports estimates of a binary choice model to locate innovation offshore, separately by industry and firm origin. We observe that U.S. software-producing firms are less likely to conduct downstream software research offshore, and this effect seems to be stronger than for non-software IT firms. However, the results are blurred once firm fixed effects and patent characteristics are controlled for. We also examine the offshoring decision for Japanese firms in an identical fashion and fail to find a significant relationship between downstream and the probability to locate an invention offshore. When we expand our investigation to a multiple choice model of innovation allocation across three regions – US, Japan, and Other – we find little evidence of industry-differences in the relationship between downstream innovation and the Japanese firms' decisions to locate research in offshore locations. In contrast, when we look at patents assigned to U.S. firms, we do again find relatively strong evidence that U.S. software-producing firms and other firms are much less likely to conduct downstream research in offshore locations than non-software producing IT firms. This effect is particularly strong in the decision to locate offshore innovation in Japan.

In sum, under the assumption that innovation production of software-producing firms requires significant proximity to and interaction with lead users, empirical evidence seems to consistently support the view that the strong presence of lead users, particularly for some software applications, may be an important source of the U.S. comparative advantage in the production of downstream software research.

Conclusions, implications, and next steps

In this essay, we documented the existence of a U.S. comparative advantage in downstream software innovation, and provided an initial exploration of the possible sources of this advantage. Using data on patents applied for at the United States Patent and Trademark Office during the period 1990-2006, we identified a large sample of software patents and used a novel classification algorithm to identify patents related to downstream and upstream software. We then provided evidence that the production of downstream software innovation is significantly more concentrated in the United States, while U.S. firms are significantly less likely to offshore their downstream research than their upstream software research. In addition, we provided some evidence that U.S. multinational firms may use internal knowledge flows to substitute for the lack of capabilities offshore, thereby mitigating challenges we posited are associated with conducting innovation remotely. Finally, exploiting variation in firm origin, industry, and location of invention in software, we presented a set of initial results that suggest patterns in our data are consistent with the hypothesis that this U.S. comparative advantage in downstream innovation in the 1990s and 2000s has been associated with a relative abundance of lead software users in the United States, as well as a greater presence of downstream-savvy inventors and software developers in the U.S.

The empirical results in this essay validate existing case-study evidence about the existence of a relative U.S. advantage in the production of downstream innovation in software (e.g. Bhidé, 2008). In addition, our results contribute, at an aggregate level, to a recent literature on location choices of multinationals (e.g. Zhao, 2006) as well as add to a growing literature on internal knowledge flows within multinationals (e.g. Branstetter, 2006; MacGarvie, 2006; Singh, 2005). However, while providing suggestive evidence, our results fall short of either fully explaining the observed advantage of United States in downstream software innovation, or linking it to firm--level performance. Exploring these questions further would require a model linking firm capabilities, country endowments, nature of innovation, and firm behavior, which is a direction in which we would like to take this research next.

References

Alcacer, Juan and Minyuan Zhao. 2007. Global Competitors as Next-Door Neighbors: Competition and Geographic Co-location in the Semiconductor Industry. Working Paper, Stern School of Business, New York University.

Arora, Ashish and Chris Forman. 2007. Proximity and Information Technology Outsourcing: How Local are IT Services Markets? Journal of Management Information Systems 24(2): 73-102.

Arora, Ashish, Chris Forman, and Jiwoong Yoon. 2008. Software. In Innovation in Global Industries: U.S. Firms Competing in a New World, (eds.) Jeffrey Macher and David Mowery. Washington: National Academies Press.

Bessen, James and Robert M. Hunt. 2007. An Empirical Look at Software Patents. Journal of Economics and Management Strategy 16(1): 157-189.

Bhide, Amar. 2008. The Venturesome Economy. Princeton: Princeton University Press.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2007. Americans do IT Better: US Multinationals and the Productivity Miracle. CEP Discussion Paper No 788.

Branstetter, Lee. 2006. Is foreign direct investment a channel of knowledge spillovers? Evidence from Japan's FDI in the United States. Journal of International Economics 68: 325-344.

Bresnahan, Timothy F. and Alfonso Gambardella. 2004. Building High-Tech Clusters: Silicon Valley and Beyond. Cambridge: Cambridge University Press.

Cockburn, Iain M. and Rebecca Henderson. 1998. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. Journal of Industrial Economics 46(2): 157-182.

Cockburn, Iain M. and Megan MacGarvie. 2006. Entry, Exit, and Patenting in the Software Industry. NBER Working Paper #12563.

Cockburn, Iain M. and Megan MacGarvie. 2007. Patents, Thickets, and the Financing of Early Stage Firms: Evidence from the Software Industry. NBER Working Paper #13644.

Forman, Chris, Avi Goldfarb, and Chris Forman. 2008. Understanding the Inputs into Innovation: Do Cities Substitute for Internal Firm Resources? Journal of Economics and Management Strategy 17(2): 295-316.

Furman, Jeff, Margaret Kyle, Iain Cockburn, and Rebecca Henderson. 2008. Public and Private Spillovers, Location, and the Productivity of Pharmaceutical Research. Annales d'Economie et de Statistique, Forthcoming.

Gambardella, Alfonso and Salvatore Torrisi. 1998. Does Technological Convergence Imply Convergence in Markets? Evidence from the Electronics Industry. Research Policy 17(5): 445-463.

Graham, Stuart J.H. and David C. Mowery. 2003. Intellectual Property Protection in the U.S. Software Industry. In Patents in the Knowledge-Based Economy (eds.) W.M. Cohen and S.A. Merrill, p. 219-258. Washington, DC: National Academies Press.

Graham, Stuart J.H. and David C. Mowery. 2005. Software Patents: Good News or Bad News. In Intellectual Property Rights in Frontier Industries: Software and Biotechnology (ed.) Robert W. Hahn, p.45-80. Washington, DC: AEI-Brookings Joint Center for Regulatory Studies.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2005. Market value and patent citations. RAND Journal of Economics 36(1): 16-38.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. The NBER Patent Citation Data File: Lessons, Insights, and Methodological Tools. NBER Working Paper #8489.

Hall, Bronwyn H. and Megan MacGarvie. 2006. The Private Value of Software Patents. NBER Working Paper #12195.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52: 909-938.

Jaffe, Adam and Manuel Trajtenberg. 2005. Patents, Citations, and Innovations: A Window on the Knowledge Economy. Cambridge: MIT Press.

Jensen, J. Bradford and Lori Kletzer. 2006. Tradable Services: Understanding the Scope and Impact of Services Offshoring. In Offshoring White Collar Work—Issues and Implications, eds. Lael Brainard and Susan M. Collins. Washington, DC: Brookings Trade Forum.

Layne-Farrar, Anne. 2005. Defining Software Patents: A Research Field Guide. Working Paper 05-14, AEI-Brookings Joint Center for Regulatory Studies.

Lerner, Josh and Feng Zhu. 2007. What is the impact of software patent shifts? Evidence from Lotus v. Borland. International Journal of Industrial Organization 25(3): 511-529.

Kogut, B. and U. Zander. 1993. Knowledge of the firm and the evolutionary theory of the multinational corporation. Journal of International Business Studies 24(4): 625-645.

Macher, Jeffrey Y. and David C. Mowery. 2008. Innovation in Global Industries: U.S. Firms Competing in a New World. Washington: National Academies Press.

MacGarvie, Megan. 2006. Do Firms Learn From International Trade? Review of Economics and Statistics 88(1): 46-60.

Merges, Robert P. 2006. Patents, Entry, and Growth in the Software Industry. Working Paper, University of California, Berkeley.

Mitchell, Thomas. 1997. Machine Learning. New York: McGraw Hill.

Nasscom. Indian IT Industry - Fact Sheet, http://www.nasscom.in (accessed Apr 26th, 2009)

Rosenberg, Nathan. 1963. Technological Change in the Machine Tool Industry, 1840-1910. Journal of Economic History 23: 414-443.

Rosenberg, Nathan. 1983. Inside the Black Box: Technology and Economics. Cambridge: Cambridge University Press.

Saxenian, A. 1996. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Cambridge: Harvard University Press.

Singh, Jasjit. 2005. Collaborative Networks as Determinants of Knowledge Diffusion Patterns. Management Science 51(5): 756-770.

Szulanski, Gabriel. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. Strategic Management Journal 17(Winter Special Issue): 27-43.

Teece, David J. 1977. Technology transfer by multinational firms: The resource cost of transferring technological know-how. Economic Journal 87: 242-261.

Thoma, Grid and Salvatore Torrisi. 2006. The Evolution of the Software Industry in Europe. Working Paper, CESPRI, Bocconi University.

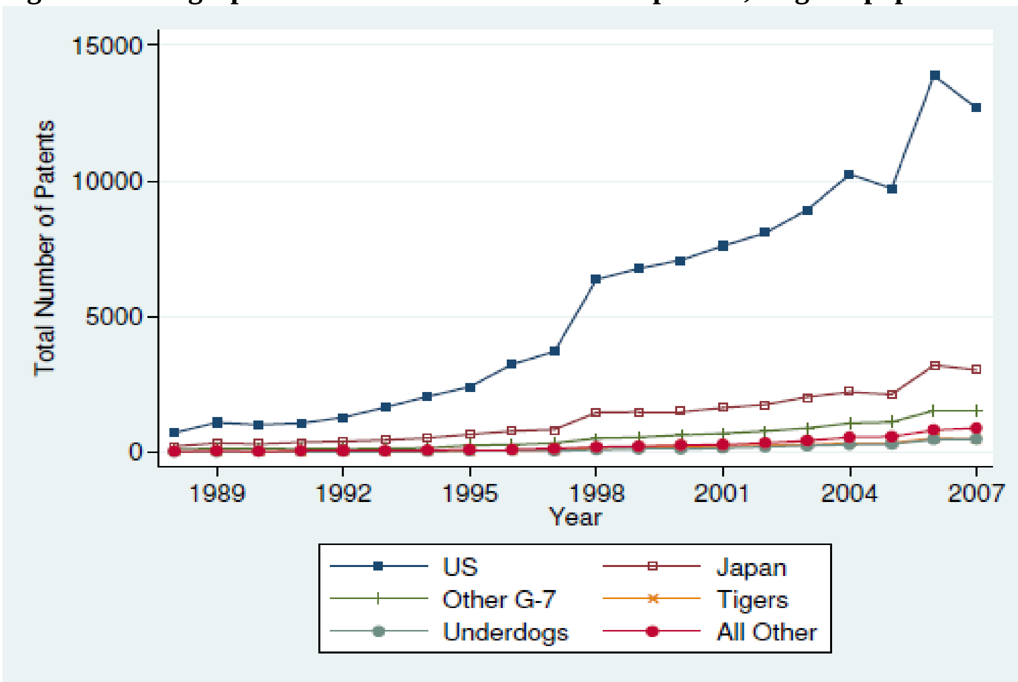
Trajtenberg, Manuel, Rebecca Henderson, and Adam Jaffe. 1997. University Versus Corporate Patents: A Window on the Basicness of Invention. Economics of Innovation and New Technology 5: 19-50.

Von Hippel, Eric. 1986. Lead Users: A Source of Novel Product Concepts. Management Science 32(7): 791-805.

Zhao, Minyuan. 2006. Conducting R&D in Countries with Weak Intellectual Property Rights Protection. Management Science 52(8): 1185-1199.

Tables and Figures

Figure 1a: Geographic distribution of U.S. software patents, original population



Note: "Underdogs" include India, China, Israel, Ireland, Brazil, and Russia, while "Tigers" include high-growth East and South East Asian countries such as South Korea and Taiwan.

Figure 1b: Geographic distribution of U.S. software patents, final sample

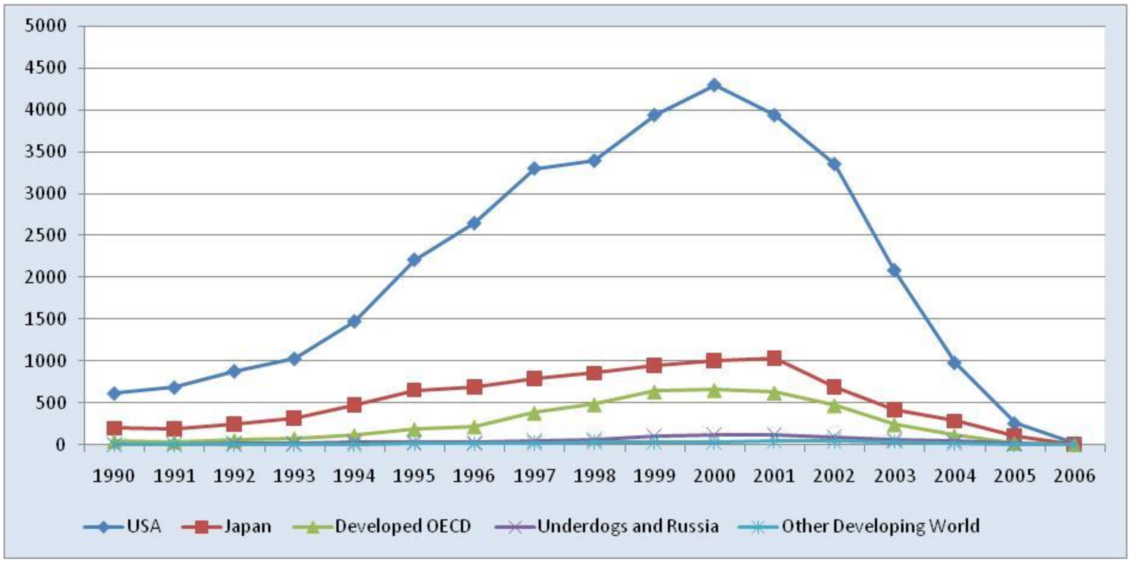


Figure 2: Percentage of downstream software patents, 3-year moving average, initial sample

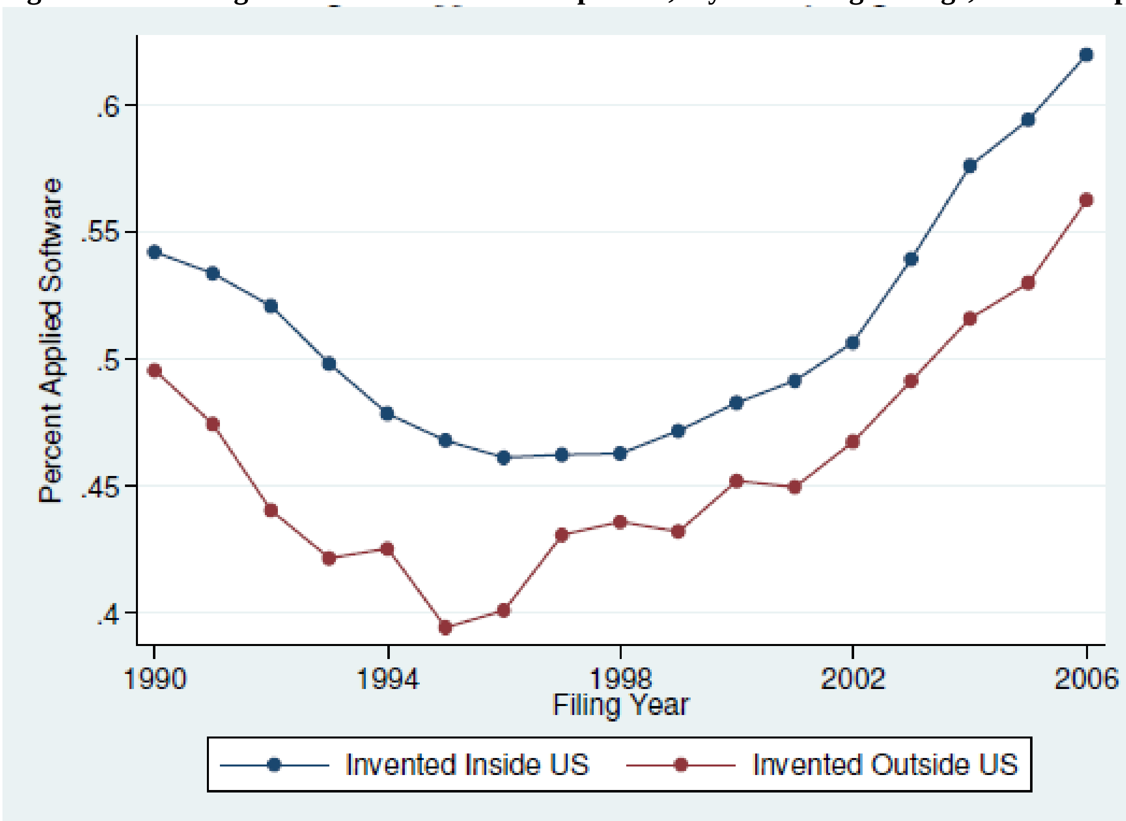


Table 1a: Summary statistics, entire sample

	Observations	Mean	Std. Dev	Minimum	Maximum
US	49376	0.7109	0.4533	0	1
Downstream	49376	0.4282	0.4948	0	1
Filing Year	49376	1998.454	3.3303	1990	2006
Importance	49084	181.5798	471.3922	1	12470.5
Originality	49084	0.5146	0.2539	0	0.9529
Percent Science References	49302	0.1559	0.2097	0	1

Table 1b: Summary statistics, by location of invention

	Observations	Mean	Std. Dev	Minimum	Maximum
Patents Invented in US					
Downstream	35102	0.4485	0.4974	0	1
Filing Year	35102	1998.42	3.3416	1990	2006
Importance	34928	207.0886	530.0314	1	12470.5
Originality	34928	0.5262	0.2503	0	0.9529
Percent Science References	35070	0.1591	0.2104	0	1
Patents Invented outside US					
Downstream	14274	0.3782	0.4849	0	1
Filing Year	14274	1998.537	3.3012	1990	2006
Importance	14156	118.6405	267.8857	1	4922.75
Originality	14156	0.4858	0.2606	0	0.9394
Percent Science References	14232	0.1479	0.2076	0	1

Table 1c: Summary statistics, by nature of invention

	Observations	Mean	Std. Dev	Minimum	Maximum
Downstream Patents					
US	21143	0.7446	0.4361	0	1
Filing Year	21143	1998.59	3.4377	1990	2006
Importance	20995	190.1018	526.0042	1	12470.5
Originality	20995	0.5278	0.2547	0	0.9529
Percent Science References	21108	0.1623	0.2186	0	1
Upstream Patents					
US	28233	0.6857	0.4643	0	1
Filing Year	28233	1998.352	3.2439	1990	2006
Importance	28089	175.2101	425.9236	1	11589
Originality	28089	0.5046	0.2529	0	0.9425
Percent Science References	28194	0.1511	0.2025	0	1

Table 1d: Patent characteristics, by nature of invention and location of invention, US-assigned patents

	Mean	St. Dev.	Mean	St. Dev.
U.S. Invented	Upstream		Downstream	
Importance	202.06	490.11	214.53	581.12
Originality	0.5138	0.2504	0.5434	0.2491
Percent Science References	0.1516	0.2007	0.1668	0.2201
Offshore Invented	Upstream		Downstream	
Importance	120.45	212.29	108.09	213.42
Originality	0.4831	0.2598	0.5151	0.2550
Percent Science References	0.1539	0.2101	0.1817	0.2339

Table 2: Likelihood of downstream software research conditional on region of invention, all software patents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	Fixed Effects	With patent controls	Post-1995	Companies with >50 patents	Above median in importance	Below median in importance
US	0.0709	0.0413	0.0430	0.0405	0.0448	0.0484	0.0469
	(0.0048)**	(0.0123)**	(0.0125)**	(0.0127)**	(0.0127)**	(0.0128)**	(0.0157)**
Importance			-0.0000	-0.0000	-0.0000		
			(0.0000)+	(0.0000)+	(0.0000)+		
Originality			0.0264	0.0206	0.0264		
			(0.0216)	(0.0199)	(0.0223)		
Science			0.0361	0.0234	0.0384		
			(0.0211)+	(0.0226)	(0.0217)+		
Constant	0.4353	0.4401	0.4232	0.4266	0.4188	0.3915	0.4676
	(0.0174)**	(0.0199)**	(0.0227)**	(0.0332)**	(0.0232)**	(0.0266)**	(0.0223)**
Year dummies	YES	YES	YES	YES	YES	YES	YES
Firm dummies	NO	YES	YES	YES	YES	YES	YES
Observations	49376	49376	49376	42629	45989	24847	24529
Number of assignee groups		1262	1262	1141	298	1014	821

Dependent variable is applied software. Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 3: Binary choice model of the invention location decision, US-assigned patents only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	Fixed Effects	With patent controls	Post-1995	Companies with >50 patents	Above median in importance	Below median in importance
Downstream Software	-0.0161	-0.0131	-0.0135	-0.0123	-0.0142	-0.0124	-0.0173
	(0.0028)**	(0.0049)**	(0.0049)**	(0.0048)**	(0.0049)**	(0.0042)**	(0.0081)*
Importance			-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
			(0.0000)*	(0.0000)*	(0.0000)*	(0.0000)**	(0.0002)
Originality			-0.0301	-0.0269	-0.0312	-0.0257	-0.0027
			(0.0194)	(0.0179)	(0.0201)	(0.0164)	(0.0111)
Science			0.0124	0.0191	0.0152	0.0034	0.0156
			(0.0128)	(0.0134)	(0.0133)	(0.0167)	(0.0151)
Constant	0.0702	0.0502	0.0663	0.0428	0.0669	0.0673	0.0648
	(0.0096)**	(0.0085)**	(0.0084)**	(0.0185)*	(0.0086)**	(0.0134)**	(0.0119)**
Year dummies	YES	YES	YES	YES	YES	YES	YES
Firm dummies	NO	YES	YES	YES	YES	YES	YES
Observations	36398	36398	36204	31303	33440	18126	18078
Number of firm groups		1145	1145	1027	272	900	751

Dependent variable is non-US. Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 4: Multinomial logit of choice of invention location across 4 regions, US-assigned patents

	(1)	(2)	(3)
	Japan	Developed OECD	Non-OECD
Downstream Software	-0.0019	-0.0082	-0.0039
	(0.0007)**	(0.0021)**	(0.0013)**
Importance	-0.0000	-0.0000	-0.0000
	(0.0000)**	(0.0000)**	(0.0000)**
Originality	-0.0030	-0.0024	-0.0079
	(0.0013)*	(0.0044)	(0.0026)**
Science	-0.0034	0.0065	0.0066
	(0.0016)*	(0.0049)	(0.0030)*
Pr(Region)	0.0059	0.0437	0.0185
Observations	36204		

Marginal effects at the mean reported. Base category is "US". Robust standard errors in parentheses. Regressions include alternative-year dummies + significant at 10%; * significant at 5%; ** significant at 1% (compared to the base category). Results are robust to inclusion or omission of a variety of control variables.

Table 5: Model of patent propensity to self-cite, linear regression and count data models, US-assigned patents only

	(1)	(2)	(3)	(4)	(5)
	Linear regression (no firm fixed effects)	Linear regression with fixed effects	Poisson model	Panel Poisson model with conditional firm fixed effects	Panel negative binomial with conditional firm fixed effects
Dep.var.	Log(Number of self citations)	Log(Number of self citations)	Number of self citations	Number of self citations	Number of self citations
Invented offshore	-0.0804	-0.1329	-0.3849	-0.4474	-0.3398
	(0.0253)**	(0.0528)**	(0.0509)**	(0.1046)**	(0.0374)**
Downstream software	0.0111	0.0043	-0.0243	-0.0403	-0.0539
	(0.0115)	(0.0291)	(0.0311)	(0.0685)	(0.0153)**
Invented offshore x downstream software	0.0095	0.0264	-0.0279	0.0048	0.0373
	(0.0401)	(0.0515)	(0.0802)	(0.1044)	(0.0588)
Importance	0.0001	0.0001	-0.0000	-0.0002	-0.0004
	(0.0000)+	(0.0001)+	(0.0000)**	(0.0002)	(0.0000)**
Originality	-0.3036	-0.2824	0.5043	0.4643	0.3919
	(0.0243)**	(0.0468)**	(0.0541)**	(0.1237)**	(0.0308)**
Science	-0.0318	0.0889	-0.3624	-0.1345	-0.1035
	(0.0267)	(0.0606)	(0.0606)**	(0.1118)**	(0.0366)**
Number of backward citations	0.4629	0.4803	0.0095	0.0160	0.0171
	(0.0099)**	(0.0436)**	(0.0011)**	(0.0033)**	(0.0003)
Year dummies	YES	YES	YES	YES	YES
Firm dummies	NO	YES	NO	YES	YES
Observations	17686	17686	36205	35023	35023

Lower number of observations in columns (1) and (2) are when self-citations=0. Results are qualitatively similar if log(1+self citations) are used. Robust standard errors in parentheses in columns (1)-(4), classical standard errors in column 5. + significant at 10%; * significant at 5%; ** significant at 1%

Table 6: Likelihood of a patent being co-invented by a multinational team with a domestic inventor conditional on it being invented offshore, US-assigned patents

	(1)	(2)	(3)	(4)	(5)	(6)
	Without Firm-Level Fixed Effects			With Firm-Level Fixed Effects		
Dep.var.	Co-invented	Co-invented and domestic inv. being primary inv.	Co-invented and offshore inv. being primary inv.	Co-invented	Co-invented and domestic inv. being primary inv.	Co-invented and offshore inv. being primary inv.
Downstream Software	0.0217	0.0317	-0.0099	0.0023	0.0237	-0.0214
	(0.0162)	(0.0145)*	(0.0121)	(0.0256)	(0.0164)	(0.0092)
Importance	0.0002	0.0002	0.0000	0.0001	0.0002	0.0000
	(0.0000)**	(0.0000)**	(0.0000)	(0.0001)	(0.0000)*	(0.0001)
Originality	0.1151	0.0587	0.0564	0.0847	0.0457	0.0389
	(0.0314)**	(0.0277)*	(0.0229)*	(0.0291)	(0.0613)	(0.0321)
Science	0.1779	0.1545	0.0234	0.1596	0.1582	0.0014
	(0.0369)**	(0.0341)**	(0.0274)	(0.0982)	(0.1306)	(0.0324)
Constant	0.1187	0.0989	0.0198	0.2681	0.1952	0.0729
	(0.0599)*	(0.0547)+	(0.0317)	(0.2473)	(0.2874)	(0.0401)
Year dummies	YES	YES	YES	YES	YES	YES
Firm dummies	NO	NO	NO	YES	YES	YES
Observations	3832	3832	3832	3832	3832	3832

Robust standard errors in parentheses (1)-(3). Robust and inventor region-clustered standard errors are in parentheses in (4)-(6).. + significant at 10%; * significant at 5%; ** significant at 1%

Table 7a: Share of downstream software research conditional on region of invention, US and JP invented patents, restricted to US and Japan inventor locations, nonparametric estimates

All Industries

firm location	location of invention: US	location of invention: JP
US	0.453	0.371
JP	0.422	0.382

Software-Producing Firms

firm location	location of invention: US	location of invention: JP
US	0.486	0.326
JP	0.529	0.359

Non Software-Producing Firms

firm location	location of invention: US	location of invention: JP
US	0.524	0.382
JP	0.426	0.403

Table 7b: Likelihood of downstream software research conditional on region of invention, US and JP invented patents, restricted to US and Japan inventor locations

	(1)	(2)	(3)	(4)	(5)	(6)
	US-Assigned Patents			JP-Assigned Patents		
	OLS	Fixed Effects	With patent controls	OLS	Fixed Effects	With patent controls
US	0.0779	0.0556	0.0569	0.0459	0.0459	0.0456
	(0.0250)**	(0.0244)*	(0.0209)**	(0.0202)*	(0.0373)	(0.0363)
Importance			-0.0000			0.0000
			(0.0000)**			(0.0000)
Originality			0.0523			-0.0656
			(0.0224)*			(0.0233)**
Science			0.0545			-0.0256
			(0.0219)*			(0.0367)
Constant	0.4311	0.4324	0.3993	0.4325	0.4414	0.4756
	(0.0320)**	(0.0265)**	(0.0304)**	(0.0354)**	(0.0335)**	(0.0326)**
Year dummies	YES	YES	YES	YES	YES	YES
Firm dummies	NO	YES	YES	NO	YES	YES
Observations	33914	33914	33744	9134	9134	9058
Number of assignee groups		1119	1119		29	29

Dependent variable is downstream software. Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 7c: Likelihood of downstream software research conditional on region of invention, US and JP invented patents, restricted to US and Japan inventor locations, by industry

	(1)	(2)	(3)	(4)	(5)	(6)
	US-Assigned Patents			JP-Assigned Patents		
	Software Firm	IT Manuf. Firm	Other Firm	Software Firm	IT Manuf. Firm	Other Firm
US	0.0742	0.0477	0.0455	0.1140	0.0428	0.0493
	(0.0042)**	(0.0516)	(0.0240)+	(0.1398)	(0.0341)	(0.0896)
Importance	-0.0000	-0.0000	-0.0000	0.0001	0.0001	-0.0001
	(0.0000)*	(0.0000)+	(0.0000)	(0.0001)	(0.0000)+	(0.0001)
Originality	0.0751	0.0347	0.0898	-0.0473	-0.0549	-0.0893
	(0.0275)**	(0.0185)+	(0.0402)*	(0.0322)	(0.0276)+	(0.0510)+
Science	0.1046	0.0230	0.0960	0.1959	-0.0284	-0.0684
	(0.0191)**	(0.0253)	(0.0255)**	(0.3033)	(0.0449)	(0.0505)
Constant	0.4277	0.3924	0.4332	0.1259	0.4921	0.4755
	(0.0138)**	(0.0601)**	(0.0491)**	(0.0989)	(0.0554)**	(0.0369)**
Year dummies	YES	YES	YES	YES	YES	YES
Firm dummies	YES	YES	YES	YES	YES	YES
Observations	5676	20768	7300	513	6268	2277
Number of assignee groups	178	491	481	7	21	23

Dependent variable is downstream software. Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 7d: Binary choice model of the invention location decision, JP-assigned patents only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	Fixed Effects	With patent controls	Post-1995	Companies with >50 patents	Above median in importance	Below median in importance
Downstream Software	-0.0129	-0.0127	-0.0125	-0.0164	-0.0125	-0.0165	-0.0089
	(0.0055)*	(0.0120)	(0.0112)	(0.0127)	(0.0112)	(0.0168)	(0.0081)
Importance			-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
			(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Originality			-0.0412	-0.0428	-0.0412	-0.0358	-0.0123
			(0.0146)**	(0.0148)**	(0.0146)**	(0.0263)	(0.0162)
Science			-0.1086	-0.1186	-0.1086	-0.1798	-0.0749
			(0.0303)**	(0.0293)**	(0.0303)**	(0.0559)**	(0.0199)**
Constant	0.9802	0.9587	1.0047	1.0666	1.0047	0.9885	1.0012
	(0.0114)**	(0.0139)**	(0.0190)**	(0.0578)**	(0.0191)**	(0.0128)**	(0.0229)**
Year dummies	YES	YES	YES	YES	YES	YES	YES
Firm dummies	NO	YES	YES	YES	YES	YES	YES
Observations	9358	9358	9282	7851	9277	4653	4629
Number of firm groups		29	29	27	26	28	23

Dependent variable is non-US. Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 7e: Multinomial logit of choice of invention location across 4 regions, JP-assigned patents

	(1)	(2)	(3)
	US	Developed OECD	Non-OECD
Downstream Software	0.0103	-0.0052	/
	(0.0048)*	(0.0022)*	/
Importance	0.0000	0.0000	/
	(0.0000)**	(0.0000)**	/
Originality	0.0554	0.0128	/
	(0.0095)**	(0.0044)**	/
Science	0.0951	0.0382	/
	(0.0105)**	(0.0047)**	/
Pr(Region)	0.0574	0.0153	/
Observations		9282	

Marginal effects at the mean reported. Marginal effects for “Non-OECD” could not be computed due to an insufficient number of observations. Base category is “JP”. Robust standard errors in parentheses. Regressions include alternative-year dummies + significant at 10%; * significant at 5%; ** significant at 1% (compared to the base category). Results are robust to inclusion or omission of a variety of control variables.

Table 7f: Multinomial logit of choice of invention location across 3 regions, US- and JP-assigned patents jointly.

	(1)	(2)
	US	Japan
Downstream Software	0.0385	-0.0041
	(0.0134)**	(0.0028)
US Firm	0.8559	-0.8936
	(0.0048)**	(0.0053)**
US Firm x Downstream Software	-0.0145	-0.0062
	(0.0154)	(0.0046)
Importance	0.0000	-0.0000
	(0.0000)**	(0.000)**
Originality	0.0476	-0.0348
	(0.0076)**	(0.0043)**
Science	0.0309	-0.0591
	(0.0091)**	(0.0057)**
Observations	45486	

Marginal effect at the mean reported. Base category is “Other”. Robust standard errors in parentheses. Regressions include alternative-year dummies + significant at 10%; * significant at 5%; ** significant at 1% (compared to the base category).Results are robust to inclusion or omission of a variety of control variables.

Table 7g: Invention location binary choice linear panel data model, by industry, US and JP assigned patents

	US-Assigned Patents				JP-Assigned Patents			
	Software-Producing Firm		Non-Sof. IT Firm		Software-Producing Firm		Non-Sof. IT Firm	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	Fixed Effects	OLS	Fixed Effects	OLS	Fixed Effects	OLS	Fixed Effects
Downstream Software	-0.0268	-0.0175	-0.0134	-0.0151	0.0287	-0.0149	0.0064	-0.0135
	(0.0076)**	(0.0087)*	(0.0035)**	(0.0068)*	(0.0188)	(0.0217)	(0.0080)	(0.0127)
Importance		-0.0000		-0.0000		-0.0000		-0.0001
		(0.0000)		(0.0000)**		(0.0000)		(0.0001)
Originality		-0.1023		-0.0071		0.0136		-0.0607
		(0.0433)*		(0.0136)		(0.0337)		(0.0134)**
Science		0.0090		0.0109		-0.0858		-0.1158
		(0.0300)		(0.0201)		(0.0614)		(0.0309)**
Constant	0.0764	0.0898	0.0528	0.0583	0.0000	1.0602	0.0232	1.0219
	(0.0197)**	(0.0234)**	(0.0129)**	(0.0135)**	(0.0000)	(0.0220)**	(0.0152)	(0.0219)**
Year dummies	YES	YES	YES	YES	YES	YES	YES	YES
Firm dummies	NO	YES	NO	YES	NO	YES	NO	YES
Observations	6207	6151	22400	22303	524	516	6522	6473
Number of firm groups		185		500		7		21

Dependent variable is non-US for US-firms, and non-Japan for JP-firms. Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 7h: Multinomial logit of choice of invention location across 3 regions, by industry, US-assigned patents

	(1)	(2)
	Japan	Other
Downstream Software	-0.0002	-0.0118
	(0.0009)	(0.0032)**
Software Firm	0.0185	0.0118
	(0.0031)**	(0.0047)*
Other Firm	0.0089	-0.0107
	(0.0018)**	(0.0041)**
Software Firm X Downstream Software	-0.0022	-0.0036
	(0.0008)**	(0.0063)
Other Firm X Downstream Software	-0.0019	0.0031
	(0.0009)*	(0.0068)
Importance	-0.0000	-0.0000
	(0.0000)**	(0.0000)**
Originality	-0.0026	-0.0103
	(0.0011)*	(0.0051)*
Science	-0.00375	0.0111
	(0.0013)**	(0.0058)+
Pr(Region)	0.0046	0.063
Observations	36204	

Marginal effects at the mean reported. Base category is "US". Robust standard errors in parentheses. Regressions include alternative-year dummies + significant at 10%; * significant at 5%; ** significant at 1% (compared to the base category). Results are robust to inclusion or omission of a variety of control variables.

Table 7i: Multinomial logit of choice of invention location across 3 regions, by industry, JP-assigned patents

	(1)	(2)
	US	Other
Downstream Software	0.0096	-0.0038
	(0.0054)+	(0.0019)*
Software Firm	-0.0419	-0.0133
	(0.0072)**	(0.0024)**
Other Firm	-0.0228	-0.0137
	(0.0060)**	(0.0023)**
Software Firm X Downstream Software	0.0301	0.0475
	(0.0403)	(0.0685)
Other Firm X Downstream Software	-0.0015	-0.0064
	(0.0111)	(0.0051)
Importance	0.0000	0.0000
	(0.0000)**	(0.0000)**
Originality	0.0542	0.0104
	(0.0093)**	(0.0037)**
Science	0.0944	0.0345
	(0.0103)**	(0.0043)**
Pr(Region)	0.0560	0.0129
Observations	9282	

Same as above. Base category is "Japan". Robust standard errors in parentheses.

Table 8a: Share of downstream software research conditional on region of invention, US and JP invented patents, restricted to US and Japan inventor locations, nonparametric estimates, by firm origin.

Both U.S. and Japanese firms

firm industry	location of invention: US	JP
"software-producing firms"	0.487	0.354
"non software-producing IT firms"	0.414	0.377

U.S. Firms

firm industry	location of invention: US	JP
"software-producing firms"	0.486	0.326
"non software-producing IT firms"	0.419	0.413

Japanese Firms

firm industry	location of invention: US	JP
"software-producing firms"	0.529	0.359
"non software-producing IT firms"	0.419	0.376

Appendix: Classification Algorithm

Following Cockburn and MacGarvie (2006), we seek to map software patents to software industries as defined in the Corptech directory of technology companies. To identify the universe of software patents, we utilize the Hall and MacGarvie (2006) software patent definition that defines software patents as the intersection of the Bessen and Hunt (2007) keyword approach with an approach based on USPTO classes.

To develop a training data set, we use patents from firms that appeared in only one Corptech SOF sub-industry during all years when they appeared in the Corptech database.[17] Using this method some Corptech SOF industries have a very small number of patents in our training data set. As a result, we were forced to consolidate some Corptech SOF industries. Table 1 lists the resulting Corptech industries with their descriptions.

Our classification approach is as follows: we first use the Hall-MacGarvie (2006) method to obtain the set of all software patents. Next, we develop a training dataset by selecting software patents applied for by firms that belong to a single Corptech software category as described above. Third, we classify all software patents, using machine learning algorithms, into these software categories. Finally, we aggregate categories into upstream and downstream software.

In order to classify patents into software categories using machine learning algorithms, we had to select a set of features that would be used to classify patents into Corptech SOF classes. To this end, we conducted two things. First, we turned words in patent abstracts and

[17] Thus, we assume that single-industry patents will patent in only their "home" industry. Of course, it is well known that firms do sometimes invent and patent outside of their industry (Gambardella and Torrisi 1998), so that will introduce some noise into our classification procedure.

titles into word vectors. Due to the resource constraints, only single word vectors were feasible, but our method could easily be extended to allow for multi-word vectors. The word vector initially returned approximately 13000 words. To make it easier to run algorithms, the number of features was reduced to 1000 using WEKA feature selection feature with ChiSquared and Ranker. Feature selection was done on training data. Word features were used as binary features.[18] Second, we used primary USPTO and IPC class codes as additional features.

Once we selected the features, we tested a variety of classification algorithms: naïve Bayes, LWL naïve Bayes, complement naïve Bayes, J48, nearest neighbor, and SMO, among others. Cross-validation showed that Naïve Bayes with kernel estimator correctly classified 72.0% of patents in the training data set with a Kappa statistic of 0.6348. This is the classification algorithm we finally used to construct the dataset used in this essay. Because we knew it was likely to occur that single Corptech industry firms also patented outside their industry, which would make one of our assumptions underlying the training dataset construction weak, we tested for the severity of this problem by examining whether the distribution of IPC classes for the class that we classified a patent into was similar to that of the most common incorrect class (i.e., the class into which Science/Technical software is most commonly misclassified is Artificial Intelligence, so we compared the distribution of IPC classes of Science/Technical software to that of Artificial Intelligence). If some of our classification error is due to noise in our procedure from our assumption about single-industry firms, these distributions would broadly have the same shape. We found that was indeed the case.

[18] We also experimented with using word counts rather than binary measures of the appearance of a word in patent title or abstract. However, the resulting kappa statistic was consistently lower when word count was used instead of binary measures.

To reassure ourselves about the validity of our classification algorithm further, we examined a stratified random sample of 199 software patents over our sample period, where we stratified the sample to include patents from all 15 of our disaggregated software categories. We then manually classified these as belonging to either downstream or upstream software using a structured approach involving a set of binary yes/no questions. Some of the questions included whether the need for user needs is important to do the innovation, whether the innovation is application oriented towards a business industry, and whether the innovation is oriented towards a particular application or is a general purpose invention. We then compared the results of our manual classification effort with that produced by the machine learning algorithm described above. We opted to use the following classes for which the two classification methods produced very similar results in our sample: business application software (downstream), artificial intelligence software (upstream), educational/training software (downstream), manufacturing software (downstream), and science/technical software (downstream). Our system classified these patents correctly between 65.4% and 87.5% of the time.

Appendix Table 1: List of Corptech Classes Used

A	Other Business Applications Software (operating on text)
B	Artificial intel software
C	Communications mgmt
D	Database/file mgmt sof
E	Educational/training sof
F	Manufacturing software
G	Media communications sof
H	Office automation sof
I	Program Development Software
J	Sales/Marketing Software
L	Transportation Software
M	Technical/Scienfitic Software
N	Utility System Software
O	Warehousing/distribution
P	All Other

Appendix: Patent Characteristics Measures

We followed Trajtenberg, Henderson, Jaffe (1997) and Hall, Jaffe, Trajtenberg (2001) in constructing the measures below.

"Science": an indicator for the patent's reliance on non-patent (scientific) research. The variable is constructed as a ratio of non-patent references made to all references made. It ranges from 0 to 1, a higher value indicates greater reliance on scientific research.

$$SCIENCE = \frac{NPCITESi}{NPCITESi + NCITEDi}$$

Here, "ncitedi" is the number of patent citations made, "ncitingi" is the number of US patent citations received prior to Jan 1^{st} 2008, and "npcitesi" is the number of non-patent citations made.

"Importb": an indicator for the patent's reliance on prior patented research (a backward looking measure). The variable is constructed as the sum of all citations made by a patent, and the discounted sum of citations received by the patents cited. The idea is to capture both the breadth and weight of citations made by a patent. The value ranges from 0 to infinity, a higher value indicating greater reliance on prior patented research. Trajtenberg, Henderson, and Jaffe (1997) find that greater reliance on important patented research is positively correlated with other measures of patent importance.

$$IMPORTB = NCITEDi + \omega \sum_{j=1}^{Ni} NCITING_{i-1,j}$$

Here, w = 0.25 and Ni is the number of patents cited by the focal patent.

"Importf": an indicator of the research impact of the focal patent (a forward looking measure). The variable is constructed as the sum of all citations received by a patent, and the discounted sum of citations received by the patents that cite the focal patent. The idea is to capture the number and importance of citations received by the patent. The value ranges from 0 to infinity, a higher value indicating greater impact for future patented research. Trajtenberg, Henderson, and Jaffe (1997) find this indicator to be significantly positively correlated to "importb", the backward looking measure of importance, the correlation coefficient being around 0.25. We find the same sign and magnitude of correlation for patents in our dataset, meaning that forward and backward looking measures give broadly the same information. This is important since we have to rely predominantly on backward looking measures in our dataset due to the fact we are not able to observe forward citations for more recent patents in our data.

$$IMPORTF = NCITINGi + \omega \sum_{j=1}^{Ni} NCITING_{i+1,j}$$

Here, w = 0.25 and Ni is the number of patents citing the focal patent.

"Originality": an indicator of the breadth of patented research the patent built on (a backward looking measure). Hence, a measure of the breadth of the research process. The underlying idea is that synthesis of divergent ideas is a characteristic of originality, hence if a patent cites patents coming from a wide range of technological categories, it is deemed more original. The value ranges from 0 to 1, a higher value representing more original research. The variable is constructed as a normalized Hirschman-Herfindahl index of concentration of citations in technology space.

$$ORIGINALITY = 1 - \sum_{j=1}^{Nj} s_{ij}^2 = 1 - \sum_{j=1}^{Nj} \left(\frac{NCITESij}{NCITEDi}\right)^2$$

Here, Nj is the number of 3-digit USPTO classes that exist (app. 400).

"Generality": an indicator of the breadth of patented research that cites the focal patent. Hence, a measure of the generality of the outcome of research represented by the focal patent. The underlying idea is that the more divergent technological areas the citations come from, the more general purpose the patent should be. The value ranges from 0 to 1, a higher value representing higher generality. The variable is constructed as a normalized Hirschman-Herfindahl index of concentration of citations in technology space. Trajtenberg, Henderson, and Jaffe (1997) find originality and generality to be significantly positively correlated, indicating that the backward and forward looking measures convey similar information. We find this to be the case in our dataset as well, giving us some confidence in the backward-looking measures we employ.

$$GENERALITY = 1 - \sum_{j=1}^{Nj} s_{ij}^2 = 1 - \sum_{j=1}^{Nj} \left(\frac{NCITINGij}{NCITINGi}\right)^2$$

Here, Nj is the number of 3-digit USPTO classes that exist (app. 400).

Essay 3: "Who's Your Daddy? Foreign Investor Origin, Multi-Product Firms, and the Benefit of Foreign Investment"

Introduction

Researchers have studied the impact of foreign direct investment for decades and there now exists an impressive body of work exploring the effects of foreign investment on the overall performance of local firms. This literature has shown that foreign subsidiaries often exhibit higher productivity, larger exports, and higher survival rates than their domestically owned peers, that firms receiving foreign investment subsequently pay higher wages and that they exhibit increased R&D performance. However, our knowledge of how and when these improvements take place remains very limited.[19] This essay addresses this gap and contributes to the literature by exploring the mechanisms that underlie the overall performance effects of foreign investment that have been previously reported in the literature. Specifically, I find that firms respond to receiving foreign investment by altering the scale of their activities, modifying the scope of their product mix, changing the scope and composition of the export markets they serve, and by ex-ante increasing the level of capital goods imports from the geographical region of the investor.

In order to achieve this, I build on a theoretical framework of Bernard, Redding, and Schott (2006, 2011) in which firms endogenously choose their product mix and geographical scope. The model yields theoretical predictions about how receiving foreign investment affects the scope and scale of target firms in terms of product space and geographical coverage. These changes occur because target firms' managerial and technological abilities increase through

[19] For a detailed discussion of this literature, please see Section II of the paper.

foreign investment. Exploiting the unusually rich panel data on behavior and performance of the universe of Slovenian manufacturing firms in the period 1994-2010, I empirically test the model's predictions and show that they align very closely with my empirical results. I find strong evidence supporting the notion that recipient firms respond to foreign investment by expanding the scope of their activities, broadening their geographical scope, and expanding the scope of their product mix, especially when foreign investors take on a large share of the recipient firm's capital. They also dramatically increase imports of capital goods from the geographical region of investor origin in the years immediately before and after the investment, consistent with the notion that foreign investors transfer their superior management and technological practices to local firms via production re-tooling.

These findings provide an insight into the mechanisms and strategies that underlie the overall performance impacts observed in this essay and in the literature as a whole, and provide an alternative explanation for recent empirical findings presented by Guadalupe, Kuzmina, and Thomas (forthcoming, American Economic Review) that observe local firms engaging in self-reported product and process innovation after receiving foreign investment. My findings suggest what they describe as "innovation" could be better understood as the transfer of already developed product and process knowledge from foreign firms to their local subsidiaries. These findings thus also relate this essay to the literature exploring the mechanisms by which foreign multinationals embed local firms into their supply chains, and transfer their organizational practices and technological capabilities to their subsidiaries.[20] To the best of my knowledge, this

[20] Prominent example of the former is Baldwin (2011), while examples of the latter include Caves (1996), Branstetter, Fisman, and Foley (2006), Bloom and Van Reenen (2010), and Bloom, Sadun, and Van Reenen (2012).

is the first paper that empirically explores the product mix and export scope dimensions of the effects of foreign investment.

In addition, this essay also adds to the literature by examining whether heterogeneity in investor characteristics affects the ability of local firms to benefit from receiving foreign investment, which is a question that has not received much attention in the literature.[21] I contribute to existing research by examining the way one important source of investor heterogeneity, geographical origin, impacts the ability of target firms to benefit from receiving foreign investment. Using my dataset of Slovenian manufacturing firms, I empirically show that all foreign investments are not created equal and that investor heterogeneity matters for how target firms benefit from foreign investment. Consistent with a hypothesis that investor origin proxies for differences in average managerial and technological ability of investing firms, I find that firms receiving foreign investment from advanced country investors (which are likely to be higher-ability investors) outperform their domestically owned peers to a larger degree than those who receive investment from developing country investors (likely lower-ability investors). Further, building on my previous results I show that firms receiving investment of advanced country origin exhibit a greater degree of expansion in their product and geographical scope, and a larger drop in product prices, than firms receiving investment from investors of developing country origin.

While the empirical focus of this essay is on examining the effects of foreign investment on domestic firms, my analysis also allows me to answer questions about the mechanisms that drive FDI decisions in the context of my data, specifically how local firms are selected for

[21] A notable exception is the work of Chen (2011), who empirically examines differences in overall ex-post performance of acquired U.S. firms, depending on the geographical origin of the investor.

investment. I find strong evidence that foreign investors select the largest, most productive local firms (i.e. "cherries") for investment, which confirms the results of several recent studies.[22] I also find some preliminary evidence in support of the notion that foreign investors are choosing to invest in local firms in order to exploit their existing export networks, which is consistent with very recent theoretical and empirical findings in the literature.[23]

This essay is organized as follows. Section 2 briefly reviews the literature on foreign direct investment and how this essay relates to existing research, Section 3 introduces the theoretical model and its predictions, and Section 4 describes the data and provides a brief overview of Slovenia's economic context. Section 5 describes my empirical approach, while Section 6 presents the results and discusses how they align with the model. Finally, Section 7 discusses the implications and limitations of the essay and concludes.

The Impact of Foreign Direct Investment: A Brief Review of the Literature

A large body of work in economics and management strategy has explored the effects of foreign investment on the overall performance of local firms. Researchers have examined the impact of foreign investment on the productivity and survival of local firms and found that foreign subsidiaries often exhibit higher productivity, larger exports, and higher survival rates than their domestically owned peers, and further that this seems to be at least partially a causal effect of receiving foreign investment. Kronborg and Thomsen (2009), Criscuolo and Martin (2009), Ramondo (2009), and Arnold and Javorcik (2009) are recent examples of this work. Researchers have also studied the effects of foreign ownership on wages and employment of

[22] These include Guadalupe Kuzmina, and Thomas (forthcoming) and Blonigen et al (2012).

[23] See Blonigen et al (2012) for details.

target firms and have found mixed results (Aitken, Harrison, and Lipsey, 1996; Heynman, Sjoholm, and Tinvall, 2007; Huttunen, 2007). In addition, they have explored the effect of foreign investment on target firms' R&D investment and innovation and found evidence of a positive effect of foreign ownership on target firms' subsequent R&D performance (Falk, 2008; Guadalupe, Kuzmina, and Thomas, forthcoming), as well as mixed evidence for the presence of knowledge spillovers from foreign multinationals to local firms (Haddad and Harrison, 1993; Branstetter, 2000; Javorcik, 2004).

However, much is still unknown about the mechanisms that underlie the overall effects that have been observed in the literature. For example, we know little about how firms respond to receiving foreign investment by shifting the scope and geographical focus of the foreign markets they serve, altering the scope and quality of their product mix, and adjusting the prices they charge for their products. Exploiting the unusually rich panel data on behavior and performance of the universe of Slovenian manufacturing firms in the period 1994-2010, I contribute to this literature by examining how receiving foreign investment affects the product market and export market choices of local firms, which is novel in the empirical literature on foreign investment.

Another focus of academic literature on foreign investment has been the study of firm-level determinants of why foreign investors engage in FDI and how domestic firms are selected for foreign investment. One important stream of recent literature in economics, management strategy, and industrial organization has tried to understand the decision for engaging in FDI, as opposed to choosing another way to serve a foreign market, from the perspective of the foreign investor.[24] Researchers have emphasized a variety of motives, from difficulties related to contracting with foreign firms (e.g. the "hold-up" problem)[25] and the exploitation of complementarities between firm-specific and country-specific assets in the spirit of the "resource-based theory of the firm"[26] to issues related to the interplay of firms' strategic decision-making about gaining and retaining market power.[27] Studies conducted in a variety of geographical and industry contexts have found empirical evidence supporting all of the motivations listed above. Helpman, Melitz, and Yeaple (2004), Nocke and Yeaple (2007), and others have, for example, introduced models that attempt to explain how firms choose the mode of serving foreign customers – either through exports or via FDI, or between greenfield and brownfield investment – as the result of trade-offs between variable trade costs and fixed costs of setting up foreign subsidiaries, or conversely, the trade-offs between mobile and immobile capabilities of firms. They find that characteristics of the focal firm, such as its productivity, determine whether or not it will engage in FDI, and similarly whether or not it will engage in greenfield investment or foreign acquisition. While this stream of research, along with the majority of the managerial literature, has focused primarily on exploring heterogeneity in the mode of foreign market entry and on examining how value from foreign direct investment is realized and transferred to the investor, my work in contrast focuses on examining exactly how subsidiary is transformed after receiving foreign investment.

[24] See, for example, the papers by Blonigen (1997), Shaver (1998), Chung and Alcacer (2002), Luo and Tung (2007), and Seth, Song, and Pettit (2009), among others.

[25] See, for example, Hennart (1991), Shane (1994), Grossman and Helpman (2002), Antras (2003), and Feenstra and Hanson (2005),

[26] Examples include Helpman, Melitz and Yeaple (2004), Jovanovic and Braguinsky (2004), Nocke and Yeaple (2007), and Meyer, Estrin, Bhaumik, and Peng (2009).

[27] Examples of this research include Kamien and Zhang (1990), Horn and Persson (2001), and Neary (2003).

In addition, recent research has focused on a different type of heterogeneity – that between local firms that investors target with investment. Researchers have studied the process by which domestic firms are selected for investment, and they have been particularly interested in resolving the old debate in economics, finance, and management literatures about whether foreign investors select underperforming ("lemons") or high performing local firms ("cherries"). Traditional literature has emphasized the view in which the merger and acquisition activity is a consequence of natural selection in which winners absorb losers, which would imply that foreign investment is a process in which high-performing foreign firms take over the assets of poorly performing local firms.[28] Some recent work, however, has presented evidence supporting the opposite view. Guadalupe, Kuzmina, and Thomas (forthcoming, American Economic Review), for example, present a model in which foreign investors earn higher payoffs when investing in high-performing local firms and provide supporting empirical evidence that corroborates the story that foreign investors engage in "cherry-picking" in the context of Spanish manufacturing firms. Similarly, Blonigen et al. (2012) provide a model in which foreign investors tend to select high-performing local firms that have experienced a recent period of poor performance. Using French data, they present empirical evidence that foreign investors tend to target "cherries that are on sale." While the empirical focus of this essay is on examining the effects of foreign investment on domestic firms, my analysis allows me to also answer questions about how local firms are selected for investment in the context of my data, thereby validating the results of these recent studies.

[28] Some examples of this research include Lichtenberg and Siegel (1987) and Neary (2007).

While literature has explored heterogeneity in the mode of investor entry and heterogeneity between local firms that investors target with investment, research to date has paid only little attention to another source of heterogeneity that might be important, especially from the perspective of policymakers in domestic markets – namely, that of the heterogeneity of investors that do engage in direct foreign investment. This essay contributes to this literature by examining the way an important source of investor heterogeneity, geographical origin, impacts the ability of target firms to benefit from receiving foreign investment. Using my dataset of Slovenian manufacturing firms, I empirically show that investments associated with different investor origins exhibit differential effects on the ex-post behavior and performance of domestic firms, consistent with a hypothesis that investor origin proxies for differences in average managerial and technological ability of investing firms. With a notable exception of Chen (2011), who empirically examined differences in overall ex-post performance of acquired U.S. firms as a function of the geographical origin of the investor, this is to the best of my knowledge the first paper that empirically explores the effect of heterogeneity in investor origin on the impact of foreign direct investment on local firms.

Theoretical Framework

To inform my empirical analysis, it is useful to first think about the effects of foreign investment on their recipient in the theoretical framework of multi-product firms developed by Bernard, Redding, and Schott (2010b). I borrow the model these authors used to describe the effects of trade liberalization on multi-product firms to consider the effects of receiving foreign investment on local firms. The description bellow characterizes a portion of that model that is useful for the purpose of this essay and develops a prediction that I then take to the data. A full description of the general equilibrium framework and properties of the model can be found in Bernard, Redding, and Schott (2010b).

The model is a natural generalization of Melitz's (2003) single-product, heterogeneous-firms model of trade in horizontally differentiated products. As in the Melitz model, there are a continuum of countries and products, and firm entry involves a sunk cost that reveals the entrant's productivity. But in this model, firms can then endogenously choose to produce any number of products and serve any number of markets in order to maximize their profits. Firm profitability depends on a measure of the firm's overall productivity dubbed "ability", as well as on a set of product attributes which vary among products and possibly across export markets, but are common across firms.[29]

When firms export, they face fixed costs of entering each market and fixed costs of supplying each product to that market. Thus, because higher ability firms are able to generate sufficient variable profits to cover these fixed costs, they in equilibrium supply a wider range of products to a wider range of export markets. It also follows that firms with sufficiently low productivity exit production altogether, firms with somewhat higher productivity produce only domestically, and only firms above a certain productivity cutoff export.

To see this formally, we have to introduce the model in some more detail. Suppose the world consists of many countries, indexed by $i \in \{1, \ldots, J\}$, and firms that produce many

[29] Bernard, Redding, and Schott (2010b) also derive a model where product attributes may vary across firms. But it turns out that under the assumption of constant elasticity of substitution preferences and monopolistic competition in the spirit of Dixit and Stiglitz, the model's predictions are very similar to those in the simplified version of the model.

products, indexed by k, and within each product, many varieties of that product. Each firm is assumed to produce at most one unique variety of any given product.

There is an unbounded continuum of potential firms prior to entry, but in order to enter, each firm has to incur a sunk cost of entry in the home market, $f_i > 0$. The overall ability of a firm only gets revealed after entry and after the sunk cost is incurred. There are two components of production technology and product characteristics that influence firm profitability: a firm-specific ability captured by the scalar φ and an idiosyncratic measure of product characteristics, captured by the k-dimensional vector λ, which we assume is independent of firm ability and is common across firms and countries. We can think of φ as firm productivity and λ as closeness to consumer preferences for various varieties of products.

Once the firm enters, it observes its ability, φ, and the set of product attributes for each product *k*, λ_k. Firm ability, $\varphi \in [0, \infty]$ is drawn from a continuous distribution $g_i(\varphi)$ that may vary across countries. We thus allow firm ability to be differentially distributed in different countries, consistent with empirical observations in the literature that firms in highly advanced countries possess superior managerial and technological expertise than firms in less advanced countries. A firm then decides whether to stay in the market and what products and markets to supply. Firms in country *i* face a fixed cost of entering country *j* of $F_{ij} > 0$ as well as a fixed cost of supplying product k to that country, $f_{ikj} > 0$. The first fixed cost component is intended to capture the initial costs of building a distribution network in a new export market, while the second is intended to capture the product-specific costs of market research, advertising, and conforming to regulatory standards for each product. In addition to fixed costs, firms also face a constant marginal cost of production for each product that is negatively related to firm ability (thus more productive firms can produce more cheaply), as well as variable costs of trade, capturing transportation costs, which take the standard "iceberg" form.[30]

Under constant-elasticity-of-substitution (CES) demand structure and the Dixit-Stiglitz type monopolistic competition market structure, the demand for each variety of a product will solely depend on its price, the prices of other varieties and other products, and on aggregate expenditure. As we assume there to be a continuum of varieties of any given product, each firm is therefore unable to influence the price index for any product. Its profit maximization problem reduces to choosing the price for each product variety separately to maximize its profits. The solution to this optimization problems leads to the typical result that the equilibrium price of a product variety is a constant mark-up over marginal cost. Since production technology and demand elasticity of substitution do not vary across varieties of the same product, we can derive the equilibrium profits that a firm from country i receives from selling a particular product to country j, which are as follows:[31]

$$\pi_{ij}(\varphi, \lambda) = \frac{r_{ij}(\varphi, \lambda)}{\sigma} - \omega_i f_{ij} \qquad (1)$$

[30] We can also allow the exporting costs to vary by firm or by firm-destination pair, consistent with the notion that some firms in a given country might have better access to export markets than others. The predictions of the model would not be qualitatively affected by this modification.

[31] For a detailed derivation, see Bernard, Redding, and Schott (2010b), pp.12-13.

Where $r_{ij}(\varphi,\lambda)$ are revenues the firm generates and are a function of its ability and its product characteristics, while σ is an elasticity-of-substitution parameter from the demand function. It follows from the equation above that there exists, for each firm ability level φ, a zero-profit cutoff for product attributes, $\lambda^*_{ij}(\varphi)$, such that a firm from country *i* will only sell a product to country *j* if its ability draw is above the threshold value. This cutoff is lower for higher ability firms, which thus have the ability to generate sufficient variable profits to cover fixed costs at lower values for product attributes. Since product attributes are independently distributed across the continuum of products, the share of products supplied by a firm with a given ability from source country *i* to destination *j* is just the probability of drawing a value of product attributes above the threshold, $\lambda^*_{ij}(\varphi)$: $\left[1 - Z(\lambda^*_{ij}(\varphi))\right]$, where Z is the cumulative distribution of product characteristics. We can now derive the total profits a firm will generate in each market. They equal the (expected) profits from each product minus the market fixed costs:

$$\pi_{ij}(\varphi) = \int_{\lambda^*_{ij}(\varphi)}^{\infty} \left(\frac{r_{ij}(\varphi,\lambda)}{\sigma} - \omega_i f_{ij}\right) z(\lambda)\, d\lambda - \omega_i F_{ij} \qquad (2)$$

It follows from the above expression that as lower ability firms face a higher zero-profit product cutoff, they will, all else equal, supply a smaller fraction of products to a given market and, combined with expression (1), have lower expected profits from each product. For sufficiently low ability levels, overall profits from supplying products to a country may fall below the level necessary to cover the fixed costs of market entry, and such firms would exit that market.

This result allows us to think about what would happen to a firm if it got taken over by another firm from a different country. Suppose the investing firm is of a high-ability, which is consistent with much of the findings in the empirical and theoretical FDI literature that finds high-productivity firms to be the firms most likely to engage in FDI[32]. Further suppose that this investor firm implants the local firm with its superior management and technology practices, effectively raising the ability level of the recipient firm to match its own (Bloom and Van Reenen, 2007, 2010; Bloom, Sadun, and Van Reenen, 2012).[33] In the model, this translates into the local firm exhibiting a positive shock to its ability, φ. The model's prediction would be that this firm would, in equilibrium:[34]

(a) *Increase the scale of its operations*. As the firm's ability increases, it is able to sell larger quantities of its existing products, and find it profitable to introduce new products to its product mix.

(b) *Increase the scope of its product mix.* With a higher level of ability, the firm is able to export a larger share of products to any given market.

(c) *Increase the scope if its geographical export presence.* With a higher level of ability, the firm finds it profitable to export products to new markets with less favorable draws of product variety tastes for its products.

---

[32] Please refer to the introduction to this paper for some examples of this research.

[33] This assumption is common in the theoretical literature. See, for example, Burstein and Monge-Naranjo (2009), Ramondo and Rodriguez-Clare (2009), and McGrattan and Prescott (2010). Empirical evidence for cross-country differences in management practices abounds (see Bloom and Van Reenen (2007, 2010), as does evidence for the transfer of technology and management expertise from foreign firms to their subsidiaries (e.g. Bloom, Sadun, and Van Reenen (2012). For additional examples, please refer to the introduction to this paper.

[34] For a formal proof of sections of the above proposition, please see Web Appendix to Bernard, Redding, and Schott (2010b)

(d) *Lower the prices of its existing products.* Given the construction of the model, the equilibrium price a firm charges is a constant mark-up over its marginal cost. As marginal costs decline with higher firm ability, its equilibrium price drops as well.

To see that part (a) holds, one must just examine the structure of the product-country specific profit expression from equation (1). As $r_{ij}(\varphi, \lambda)$ is monotonically increasing in firm ability, φ, a firm that exhibits a positive shock to its ability, $\varphi_s > \varphi$, will in equilibrium exhibit increased scale of operations across all of the products it chooses to supply. To see that part (b) holds, refer back to the share of products supplied by a firm with a given ability (above the minimum ability cutoff) from source country *i* to destination *j*, which equals $[1 - Z(\lambda^*_{ij}(\varphi))]$, where Z is the cumulative distribution of product characteristics. It can be shown that $\lambda^*_{ij}(\varphi)$ is monotonically decreasing in φ, while we know that by construction Z is monotonically increasing in λ. It then follows that $[1 - Z(\lambda^*_{ij}(\varphi))]$ is monotonically increasing in firm ability and that increasing firm ability increases the share of products it chooses to supply to any market in equilibrium. For part (c), once again refer to $[1 - Z(\lambda^*_{ij}(\varphi))]$ and note that this expression also tells us the probability that a product is exported to any country *j* in equilibrium. Since we have established that it is increasing in firm ability, a positive shock to φ will increase the probability that a firm in equilibrium chooses to supply any product to any country, increasing the equilibrium geographical scope of that firm. Finally, part (d) holds as by construction the equilibrium pricing rule takes on the structure of a constant mark-up over its marginal cost. Also by construction, marginal costs are monotonically decreasing in firm ability. Thus, as a firm experiences a positive shock to φ it in equilibrium charges less for its existing products.

This proposition gives us a simple set of predictions of how foreign investment would affect the ex-post behavior and performance of local firms that we can directly take to the data. In addition, it also allows us to consider how heterogeneity in investor ability would result in differential effects on the target firms. It is trivial to show that in the model receiving investment (and ability levels) from higher-ability investors leads to larger increases in target firm scale, larger changes in the scope of the firm's product mix and geographical presence, and a larger decline in prices charged for existing products, all else equal. While I do not directly observe investor ability in the data, I can exploit the fact that firm productivity, management practices, and technological prowess, have all been shown to differ across countries in a way that is closely related to the countries' levels of economic development. In the empirical analysis that follows, I thus attempt to use investor origin as broad proxy for investor ability, and evaluate whether the data support the hypothesis that heterogeneity in investor origin leads to effects in the performance and behavior of target firms that are consistent with what we would expect if the source of heterogeneity was in fact investor ability.

Data Description and Historical Context

Description of Data

In my empirical analysis, I use a rich panel dataset containing a wealth of information on the universe of Slovenian firms during the period 1994-2010. The data were made available for this project by the Statistical Office of the Republic of Slovenia and Bank of Slovenia, and contain detailed information on financial accounts of Slovenian firms, detailed transaction-level data on all their import and export activities at the product-destination level, annual firm-level information on all foreign direct investments received by Slovenian firms, and additional descriptive information that allow me to observe their primary industry, number of employees, and geographical location.

For every firm-year combination, I thus observe detailed information from their balance sheets and income statements, allowing me to compute a variety of firm performance indicators, including revenue-based measures and total factor productivity (TFP)[35]. For every export and import transaction a firm reports, I observe product information at the 8-digit level of the Slovenian version of the Combined Nomenclature, the transaction value, a measure of product quantity and weight (if available), and information about export destination or import origin. This allows me to compute measures of a firm's product mix scope and scale in its export activities, as well as to observe how firms behave at the level of specific products and export destinations. In particular, I compute measures of a firm's geographical scope (number of export destinations served) and geographical focus (intensity and representation of exports in various geographical regions), as well as track prices (unit values) that a firm charges for its products.

Furthermore, investment data allow me to observe all inward foreign non-portfolio investments that surpass at least 10% of the local firm's outstanding capital in a given year. As Slovenian firms are required by law to annually report this information to the central bank, I can be confident I observe the universe of qualified foreign direct investment flows into the country. Aside from observing the fact that a firm has received foreign direct investment, I also know the

[35] I compute TFP using the Levinsohn-Petrin method and use materials expenditure as the proxy for unobservable firm-level productivity shocks.

investment amount and the origin of the investor, and am able to track this information at the firm level on an annual basis. I use investor origin to group investors into two main geographical groups: “advanced country investors” and “developing country investors”. The former group contains investors originating from high-income member states of the Organization of Economic Cooperation and Development (OECD)[36], while the latter group is defined as a complement to the former, excluding countries that are typically deemed to be offshore tax havens[37].

Table 1 provides summary statistics for key variables in the dataset, both for the entire dataset and for three subgroups: firms that remain domestically owned during their entire spell in the dataset, firms that were initially domestically owned but then received foreign investment from an advanced country investor, and those that were initially domestically owned but then received foreign investment from a developing country investor. I thus drop all firms that reported having received foreign investors during their entire spell in the dataset, which include subsidiaries of foreign firms spawned by greenfield investment and domestic firms that received foreign investment prior to the first year of our sample period. In addition, I drop all firms which did not report positive revenue or variables needed for TFP calculations before and after receiving foreign investment. This insures that I am able to observe a firm for at least 5 continuous years in my data. In order to achieve comparability across periods, I first denominate values of all financial variables in terms of a common currency.[38] I then employ a series of price

[36] These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States.

[37] These include countries such as Luxembourg, Lichtenstein, Cyprus, and the Cayman Islands.

[38] As Slovenia adopted the European common currency, the euro, in 2007, I use average annual exchange rates as published by the Bank of Slovenia to convert all values prior to 2007 from the Slovenian Tolar into Euros.

deflators in order to remove the effect of temporal price changes from financial variables used in the analysis.

The resulting dataset is an unbalanced panel that contains 8,171 firms, 7,970 of which are domestic and 194 of which are targets of foreign investment. Out of the latter group, 163 firms are targeted by investors originating from advanced countries, while 31 are targeted by investors from developing countries. The three most common advanced investor origins include Austria, Germany, and Italy, while developing country investor origins include, among others Croatia, Bosnia and Herzegovina, the Czech Republic, Belarus, and Hungary.

Comparing descriptive statistics for these groups of firms in Table 1, we observe that domestically owned firms differ substantially from firms that are targeted by foreign investors. They are, on average, significantly smaller, both in terms of their revenue and the number of employees; they export and import less, source imports from a more geographically narrow set of countries and export to a narrower set of market destinations. They also, on average, pay their employees substantially less than their peers who receive foreign investment, indicating that they might have lower levels of human capital.

If we compare the firms that receive foreign investment from advanced and developing country investors, we see that they are quite similar, at least compared to domestically owned firms. While developing country investor targets are somewhat larger, advanced country investor targets are more profitable, and pay their employees a higher wage. We should note however, that Table 1 provides descriptive statistics for variables for the entire duration of the sample and thus includes information on these firms both for the period before and after receiving foreign investment. To observe the effects foreign investment has on the ex-post performance of target firms, a more sophisticated analysis is required. While I turn to describing and implementing such an analysis in the next section, it might be useful to first provide the reader some context about Slovenia's economic history in the period from which our data originate, and to briefly discuss a small set of illustrative examples of foreign investment in Slovenia.

Brief Summary of Slovenia's Economic History during the Sample Period

A former republic in the Socialist Federal Republic of Yugoslavia (henceforth, Yugoslavia), Slovenia declared its independence from Yugoslavia in 1991, and after a brief period of ethnic strife charted a new chapter in its political and economic history, embarking on a process of economic transition from socialism to capitalism and beginning a process of alignment with Western Europe that included membership in the WTO in 1995, entry into the European Union and NATO in 2004, the adoption of the European common currency in 2007, and admission into the OECD in 2010.

Slovenia, being the northernmost state of Yugoslavia, bordering Austria to the North and Italy to the West, enjoyed the status of the federation's most economically advanced region. While it represented only one thirteenth of the population of Yugoslavia, it accounted for more than a fifth of the federation's GDP, and its firms represented an estimated one third of Yugoslav exports. Unlike in other countries of the Soviet Bloc, where collective ownership of productive assets meant state ownership in gigantic production complexes, Slovenian firms benefited from a policy of a decentralized system of self-management by the workers themselves, with moderate levels of interference from local councils and party organs (Pogatsa, 2012). As a consequence, Slovenian firms were well-positioned, compared to their peers from other Eastern European economies, to successfully manage the transition from a socialist to a capitalist system.

Nevertheless, macroeconomic instability in the early years of economic transition had left the Slovenia of the early 1990s in a difficult economic position with high rates of inflation and negative economic growth. This was a period of mass economic restructuring and privatization in which the majority of large and medium-sized manufacturing firms received private ownership, went bankrupt, and/or split into smaller independent units. In this period, a number of large and medium sized manufacturing firms were acquired by foreign investors, while others, especially in what were deemed "strategic sectors" such as banking, insurance, telecommunications, and steel production, retained at least partial state ownership.

This period also marked the start of a radical period of trade liberalization, characterized by falling tariffs and non-tariff barriers to trade. Slovenia removed restrictions on foreign investment and expanded its network of bilateral and multilateral trade and investment agreements, which included the 1997 Interim Free-Trade Agreement with the European Union, liberalizing cross-border capital movements and reducing tariffs and some non-tariff barriers to trade for Slovenian exporters to EU member states, as well as bilateral trade and investment agreements with most former Yugoslavian republics, which eased cross-border business and investment.

From the mid-1990s until approximately the end of the time period studied in this essay and the onset of the global economic slowdown as a consequence of the 2008 financial crisis, Slovenia experienced a favorable pace of economic development with real economic growth in the 3-5% range. The country's traditional export-oriented manufacturing industries[39] expanded, fueled by exports to both Western European markets and the rebuilding of trade ties with

[39] These included chemicals, electrical equipment and electronics, food processing, metal products and industrial equipment, motor vehicles and components, lumber and paper products, pharmaceuticals, and textiles.

traditional markets in Eastern Europe and former Yugoslavia. Business investment levels rose, unemployment and real wage trends trended favorably, and the overall export-intensity of the economy, while already significantly higher than in most other post-socialist economies, inched further upward (Lorber, 1999; SURS, 2012). These trends continued throughout the early and mid 2000s and were coupled with both a gradual structural transition of Slovenia's economy from manufacturing towards services, as well as a slow transition in manufacturing itself toward sectors characterized by higher technological sophistication of production. Slovenia's transition towards closer resemblance of the economic structure of the world's advanced economies was aided by the country's integration within the European Union.

As proved to be the case with many economies on the "periphery" of the European Union, the country's integration into the European economic and financial system brought along a boom in capital inflows, investment, and wage inflation that lasted until the financial crisis of 2008. Convergence of interest rates with the rest of the European Union fueled business and public investment alike, especially in residential construction and infrastructure, and Slovenian wages and standards of living converged toward the European Union averages. In terms of GDP per capita, Slovenia passed some existing European Union member states such as Portugal and Greece, and moved on a path towards OECD membership, which was officially granted in 2010. The financial crisis of 2008 and its lasting aftermath have significantly affected Slovenia's economic dynamism and the country's economic woes mirror those of many European economies. As a small open economy with a particularly export-oriented private sector, the economic slowdown in its main European trading partners negatively affected Slovenia's exporters and a slowdown in business investment resulted in stagnation of Slovenia's economy in the recent years.

Illustrative Examples of Foreign Investments in Slovenia

In order to help elucidate the effects receiving foreign capital via foreign direct investment may have on recipient firms, as well as to illustrate how investor origin might moderate its effects, it is useful to provide a brief discussion of some prominent cases of Slovenian firms that received foreign capital of either developed or developing country origin during the sample period. While privacy protection policies that were a part of the data licensing agreement prohibit me from determining if these cases are actually featured in my data, they nonetheless provide useful insights into the dynamic effects receiving foreign investment has had on targeted Slovenian firms.

From its declaration of independence from Yugoslavia in 1991 and the subsequent turn toward capitalism in the early 1990s, Slovenia's manufacturing firms received a steady stream of foreign direct investments. A large majority of investments of developed country origin came from Slovenia's main Western European trading partners, though some notable investments also originated from the United States, but very few from Asia and elsewhere. Similarly, direct investments of developing country origin most frequently originated from the country's main trading partners in South-Eastern Europe, especially Croatia and (later) Serbia, but also from other former Soviet bloc countries such as Hungary, the Czech Republic, and Russia. The manufacturing industries that developed country investors targeted most frequently included chemical manufacturing, pharmaceuticals, industrial machinery and products, automobile components, and the manufacture of electrical and electronic equipment. Conversely, industries prominently featured among investments of developing country origin included paper and paper products manufacture, packaging manufacturers, food processing, and, in service-oriented industries, tourism, wholesale merchants, and retail chains.

The companies that were frequently targeted were large manufacturing firms with an already established tradition of export-oriented production, albeit one that was often primarily oriented toward markets in ex-Yugoslavia and the former Soviet Bloc. Foreign investors were typically large Western multinational concerns with existing operations in industries in which target firms were primarily engaged. While there is no shortage of cases to choose from, here is one illustrative example of developed country FDI in Slovenia during the 1990s and 2000s:[40]

- *Bosch and Siemens Home Appliances Group Nazarje:* Starting in 1993, Germany's Bosch and Siemens Hausgeräte GMBH (BSH Group), one of world's largest manufacturers of home appliances, began fostering an equity relationship with Slovenia's largest manufacturer of small motor-based home appliances, Tovarna malih hisnih aparatov Nazarje. Previously a division of Slovenia's Gorenje, one of Central Europe's largest manufacturers of (predominantly large) home appliances, the company had a relatively successful 30-year history of producing small home appliances, and it mainly focused on serving the Slovenian, ex-Yugoslavian, and Eastern European markets. After BSH Group's acquisition in the 1990s, however, the company expanded to become a prominent regional production and R&D hub for its parent company, as well as serve as BSH Group's sales and marketing headquarters for a large chunk of Central and South-Eastern Europe. By 2002, a decade after initial acquisition, BSH Nazarje's revenue had increased more than fourfold, with similar increases in R&D investment and a significant expansion in production capacity. The firm subsequently became fully integrated into BSH Group's global supply chain and presently produces approximately 5.5 million high-end home appliances of various types which are marketed globally under Bosch and Siemens brands.

[40] For additional examples, please refer to the Web Appendix.

Until the most recent past, foreign direct investments of developing country origin did not feature as prominently in the media and public consciousness as did developed country investments. While this was primarily due to their lower frequency, it was also related to the fact that prior to the most recent period of Slovenia's economic history, there were few examples of investors from developing countries buying majority stakes in Slovenian manufacturing firms that were considered to be national champions. This changed when in 2007 Slovenian Steel Group, the country's largest steel manufacturer, was taken over by Russia's KOKS Group, one of the world's largest metallurgical conglomerates. Soon thereafter, Droga Kolinska, Slovenia's largest processed food producer, was acquired by its peer from Croatia, Atlantic Group, while Fructal, Slovenia's largest fruit processing company, was acquired by its Serbian peer, Nectar. Nevertheless, the years prior to 2007 have seen some major manufacturing investments with developing-country origins, one prominent example of which is the following:

- *Valkarton:* in 2002, Belisce, Croatia's largest paper and packing products manufacturer, acquired a majority stake in Valkarton, Slovenia's largest producer and exporter of corrugated cardboard products and packaging, laminated packaging, and folding boxes. Following acquisition by Belisce, which is Valkarton's main supplier of raw materials, the company continued on its existing path of incremental upgrades to its technology and equipment, and on its strategy of growth by acquisition of smaller competitors in the former Yugoslavian republics, but it did not expand its product mix or international footprint dramatically. Today, Valkarton sells the majority of its products in Slovenia, while its subsidiaries predominantly serve their local ex-Yugoslavian markets. Other exports represent approximately 20% of the firm's revenue base and mainly include Italy, Hungary, Austria, and the Netherlands.

The distinctions between advanced and developing country investments highlighted in these particular cases are borne out in the complete analysis that follows. The next section systematically investigates these differences and embeds them in a formal econometric framework.

Estimation Approach

In order to estimate the effect of foreign investment on ex-post measures of target firm performance and behavior, I write a simple empirical model linking foreign investment and subsequent firm-level outcomes of interest as follows:

$$Y_{it} = \alpha + \beta F_{it-1} + d_t + d_i + \varepsilon_{it}, \qquad (3)$$

where Y_{it} is an outcome of interest for firm *i* in year *t*, F_{it-1} is an indicator of whether the firm had received foreign investment in the prior year and equals one in every year thereafter,[41] and d_t and d_i represent year- and firm-level fixed effects. I include firm-level fixed effects to control for the effect of time-invariant firm-level characteristics that might affect firm behavior and performance over the sample period and time effects to account for secular factors that might impact all firms operating in year *t*.

Recent literature tells us that it is very unlikely that assignment of foreign investment is random across firms. If foreign investors select their targets based on characteristics of these

[41] Please note that the coding of the foreign investment indicator implies we are not identifying a one-year effect of foreign investment on the firm-level variable of choice, but rather a (weighted) average effect of receiving foreign investment on the firm-level variable of interest over the entire post-investment horizon.

firms that vary over time, estimates of expression (3) will be biased and inconsistent. In order to alleviate this problem, I follow the approach of Chen (2011) and Guadalupe, Kuzmina, and Thomas (forthcoming) and propose a selection mechanism for foreign investment that depends on observable characteristics of target firms. If this selection mechanism, as described by the ex-ante trajectory of firm characteristics, is a sufficiently exhaustive description of the process by which foreign investors select their targets in my data, then by purging the selection effect in the equation above I may obtain consistent estimates of the effect of foreign investment on ex-post measures of target firm performance and behavior.

In order to implement this approach, I draw on the literature that discusses the use of propensity score estimation techniques in order to identify average effects of treatment.[42] This literature uses observed characteristics of participants and non-participants in a particular treatment program to estimate a single-dimensional propensity score that summarizes the relationship between participant characteristics and treatment and serves as an estimate of the probability that a participant will be treated. The propensity score is then used to adjust for selection into treatment on the basis of observable characteristics, allowing for consistent estimates of the average treatment effect.

The effectiveness of these methods depends on the validity of two assumptions: (1) whether observed pre-treatment characteristics do indeed predict participation in the program to the extent that treatment can be thought of being random, conditional on observed pre-treatment characteristics (this is often referred to as "unconfoundedness" or the "conditional independence" assumption), and (2) whether we can observe a sufficient number of similar

[42] See Busso, DiNardo, and McCrary, 2011 for a recent survey of the literature.

participants and non-participants to successfully build an empirical counterfactual for treatment by comparing the two groups (this is often referred to as the "overlap" assumption). Provided they both hold, the researcher can use these methods to consistently estimate the treatment effect of the program, and under some circumstances these estimators might even have desirable finite-sample efficiency properties (Busso, DiNardo, and McCrary, 2011).

I employ two variations of the above approach to estimating the average effect of foreign investment ("treatment") on target firms using propensity scores. First, I follow the method proposed Dehejia and Wahba (1999) and Busso, DiNardo, and McCrary (2011) and implemented by Guadalupe, Kuzmina, and Thomas (forthcoming) and employ a reweighting estimator, which uses estimated propensity scores to calculate re-weighted observations in equation (3), then estimate that equation using weighted least squares. Secondly, I employ a semi-parametric matching estimator that uses kernel regression matching to associate treated firms with an appropriate weighted set of untreated firms, then calculates the average treatment effect on the treated non-parametrically as the average difference in means of the outcomes of interest between the treated and control firms, conditional on the differences for the treated and control firms in the pre-treatment time period. This is the so-called difference-in-difference matching estimator used by Chen (2011) and others to study the effects of foreign investment on target firms.[43] The advantage of the first approach is its ease of implementation and possibly desirable efficiency properties, while the second approach requires fewer parametric assumptions and explicitly allows me to purge any systematic differences between target firms and matched firms that may be unobservable, as long as they are time-invariant. In addition, the difference-in-

[43] See Heckman, Ichimura, Smith, and Todd (1998) or Todd (2006) for a discussion.

difference matching estimator also lends itself directly to the estimation of the average effects of treatment with varying lags. For robustness, I employ both approaches in parallel and check that they produce qualitatively similar results.

In order to build a propensity score measure, however, it is necessary to specify an empirical model for the decision of a foreign firm to acquire a domestic firm. I follow the selection process as proposed by Guadalupe, Kuzmina, and Thomas (forthcoming) and assume that, in the presence of positive or negative selection, there is a threshold value of an underlying latent variable that measures future growth prospects of the domestic firm at any point in time, so that the firm will be acquired only if the threshold value is surpassed in the presence of positive selection or the firm will be acquired only if the threshold value is below some value in the presence of negative selection. Assuming that the observable underlying future growth prospects of the domestic firm, from the perspective of the foreign acquirer, can be proxied by observable characteristics of the domestic firm captured in our data, then I can write an empirical model for the acquisition decision in terms of variables observed in the data as follows:

$$F_{it} = \alpha + \beta\varphi_{it-1} + d_t + d_s + v_{it}, \qquad (4)$$

where F_{it} is a dummy variable indicating if firm i received foreign investment in year t, φ_{it} is a vector of a set of proxy variables for lagged underlying growth ability of firm i, and d_t and d_s are dummy variables representing year- and industry-specific fixed effects.

Estimating equation (4) gives me a set of propensity scores that I use to obtain consistent estimated of the parameter of interest in equation (3) and to estimate the difference-in-difference matching estimator. Equation (4) also allows me to empirically examine the presence and form of selection on observable characteristics of domestic firms in my data and, as a consequence, determine whether foreign acquirers target the most productive domestic firms (i.e. they "cherry-pick") or the least productive domestic firms (i.e. they target "lemons"), a question that has recently attracted renewed attention in the literature.[44]

I follow recent empirical literature, particularly Chen (2011) and Guadalupe, Kuzmina, and Thomas (forthcoming), in my selection of firm-level observable characteristics and use a broad set of proxy variables for the underlying growth ability of the domestic firm, including lagged export status, lagged total factor productivity, lagged labor productivity, lagged capital intensity of the firm as measured by fixed assets per worker and the share of fixed assets in total revenue, lagged productivity relative to the industry mean, lagged firm size measured in terms of revenue and employment, lagged skill intensity of the firm as measured by wages per worker, and lagged profitability measured as the share of net profit in total revenue. I also investigate a variety of functional forms and lag structures on the relationship between the set of proxy variables and foreign investment, and estimate propensity scores separately by industry to account for any inter-industry differences in the targeting behavior of foreign investors.

My propensity score estimation results provide clear evidence that election into foreign investment is strongly correlated with observable firm-level characteristics, and my industry-specific probit propensity score estimates allow me to achieve covariate balance for virtually all industry-variable combinations. In addition, I explore the robustness of my empirical results to various propensity score specifications, and find that the findings I report below are not

[44] Please refer to the introduction to this paper for a brief discussion of this literature.

qualitatively sensitive to the particular choice of functional form and lag structure. For additional details, please refer to the Web Appendix to this essay.

Results

The Effect on Target Firm Performance

My empirical results indicate that receiving foreign investment has significant positive effects on the ex-post performance of target firms. As the fixed effects ordinary least squares estimate from the first column of Table 2 tells us, target firms more than double their revenues after receiving foreign investment, controlling for time-invariant differences between firms. Even after correcting for the selection process using the re-weighting estimator as presented in the fourth column of the same table, I still find that receiving foreign investment causes target firms to increase their revenues by more than 30%. This result is corroborated by the difference-in-difference matching estimator approach presented in Figure 1, which shows that targeted firms, relative to their domestically owned "matches", increase their revenues by 13% in the first year after investment, and this difference increases to almost 25% by the end of the fourth year.

I observe similar results when looking at the scale of the firms' export and import activities. As the simple fixed effects estimator in Table 3 tells us, target firms more than double their exports after receiving foreign investment, and this effect remains even after I control for selection, even though it becomes marginally statistically insignificant. Turning to the results from the difference-in-difference matching estimator, we observe a similar story: target firms' exports increase, relative to their peers, after receiving investment, but this effect takes several years to become statistically significant. The results from Figure 1 indicate that target firms' exports increase by 37%, relative to their peers, by the fourth year after having received foreign investment.

These results are consistent with what the theoretical model would predict if foreign investment indeed led to an increase in the managerial and technological abilities of targeted local firms. In the second panel of Table 2, I attempt to measure this increase directly using target firms' total factor productivity. Looking at the simple fixed effects estimate, I find that target firms exhibit a 30% increase in their TFP after receiving foreign investment, but this effect goes away once we impose the propensity score re-weighting structure on the estimate. However, the difference-in-difference matching estimator finds that target firms do indeed exhibit a modest relative TFP increase over their peers. By the end of the fourth year after investment, target firms increase their TFP by an average of 12%. Given that empirical literature has shown it is very difficult to accurately measure total factor productivity for multi-product firms with aggregate financial data, these results are all the more striking.

If the observed increases in the performance and the scale of operations of target firms are indeed due to the effects of receiving foreign investment, we would expect the intensity of foreign investment to be positively associated with the observed measures of ex-post firm performance. This is exactly what the results in Figure 5 suggest. Firms that are targeted with investment that takes on an above-median share of the recipient firms' capital (i.e. high intensity investment) outperform their domestically owned peers to a much larger extent than those firms that are targeted with foreign investment that takes a below-median share of the recipient firms' capital (i.e. low intensity investments). By the end of the fourth year, high intensity investment targets' relative improvement is strong and statistically significant along all measures of firm performance and scale.

Do targets of investors of different geographical origins exhibit different ex-post scale and performance effects? My results suggest that this is the case. Comparing the estimates in Tables 4 and 5, I find that while receiving investments originating from both advanced and developing countries leads to increases in target firms' revenues, the effect is significantly larger for those firms that receive investment from advanced country investors. My results indicate that the revenue increase for firms receiving advanced country investment is 13 percentage points larger than for those receiving developing country investment. Similarly, estimates from Tables 6 and 7 tell us that this result also holds for the target firms' increase in the scale of their export and import activities. While simple fixed effects estimates show that investment from both origins leads to significant increases in the scale of exports, the coefficient on developing country investments goes away after we control for the selection process. These findings are qualitatively confirmed by the difference-in-difference matching estimator results. Firms receiving investment from advanced country investors exhibit sustained increases in exports that become statistically significant by the fourth year after investment. As Figure 10 suggests, this is especially true for targets of high intensity developed country investment. On the other hand, the estimates for firms receiving investment from developing country investors are very unstable and statistically indistinguishable from zero in most years, though this might be at least partially due to a smaller number of observed investments originating from developing countries.

The Effect on Target Firm Scope

My results on the effects of foreign investment on target firms' scale and performance are thus far largely consistent with what the model would predict if investor origin signified heterogeneity in average investor ability across the two origin groups. However, I can test this notion further by empirically investigating the model's prediction that receiving investment from higher ability investors would lead to larger increases in target firms' scope as well.

As estimates from Tables 8 and 9 reveal, overall foreign investment leads to target firms' increasing the scope of their product mix, consistent with what the model would predict if foreign investment led to improvement in their overall managerial and technological ability. However, results from Table 9 suggest that these results are entirely driven by increases in product mix scope of firms receiving investment from advanced country investors. Difference-in-difference matching estimators largely confirm this view and actually paint an even starker picture: as Figure 4 suggests, while advanced country investor targets exhibit moderate increases in the scope of their export product mix, developing country investor targets seem to actually decrease their scope in the product space.

The overall results align closely with the empirical findings from Guadalupe, Kuzmina, and Thomas (forthcoming), who find that local firms exhibit sustained increases in self-reported rates of product innovation after receiving foreign investment. My findings, however, suggest that their treatment of all investors as essentially homogenous may be obscuring differential effects of investors of different abilities that might be underlying their estimates, provided investor heterogeneity dynamics in the context of Slovenian firms translate into their context of Spanish firms as well.

Similar results as in the case of product choices are found when examining the scope of destinations to which target firms export after receiving foreign investment. Results presented in Tables 10 and 11 show that recipient firms significantly expand the number of export destinations they service as a result of foreign investment. Simple fixed effects estimates from the first column of Table 10 suggest that target firms add in excess of 6 new export destinations after receiving foreign investment, and even after we control for selection, the coefficient in the fourth column of Table 10 still shows that target firms exhibit a statistically significant increase in the scope of their geographical presence. Table 11 suggests that the increase in scope is large and statistically significant for firms that receive investment from advanced country investors, but after controlling for selection, the effect becomes statistically insignificant for firms that receive investment from developing country investors. The difference-in-difference matching estimator paints a similar picture, but again suggests that firms targeted by developing country investors might actually reduce their geographical scope. While the standard errors are large, the estimated average treatment effects are positive and marginally statistically significant for high intensity advanced investor targets and negative but largely statistically insignificant for developing investor targets.

The Effect on Product Prices and Capital Goods Imports

While my empirical results seem largely consistent with the hypothesis that heterogeneity in investor ability, as proxied by their origin, leads to differential effects on target firms' ex-post scope and scale, the model also yields a prediction that firms exhibiting increases in their ability would, on average, lower the prices they charge on their existing products. Tables 12 and 13 empirically examine this notion and find modest evidence in support of this hypothesis. While the difference-in-difference matching estimator results in Table 12 are very noisy, the point estimates do seem to suggest that firms receiving foreign investment might lower the average price of their products, relative to their peers, in the years following investment, and that this effect, although at most marginally statistically significant, is entirely driven by decreases in average prices by firms receiving investment from advanced country investors. Table 13 takes a step further and examines the post-investment change in price for existing and new products. The results suggest that products introduced after receiving foreign investment exhibit significantly higher prices relative to existing products of the firm, and this effect is much stronger for firms receiving investment from advanced country investors. Combining this insight with results from Table 12 suggests that advanced country investor targets indeed lower the prices of their existing product portfolio after being targeted with foreign investment.[45]

If we assume foreign investors indeed transfer their superior managerial and technical abilities to their investment targets, we would expect to observe that these firms undergo extensive retooling of their production processes after receiving investment. While this is something I cannot directly observe in the data, I can observe whether target firms increase imports of manufacturing equipment and related capital goods before and after the entrance of the foreign investor. Table 14 estimates a simple fixed effects regression that suggests this is the case. Target firms experience a 26% increase in imports of capital goods after receiving foreign investment, and this effect is entirely driven by firms that receive investment from advanced country investors. Further, the results suggest that a majority of capital goods imports come from advanced OECD countries, consistent with the view that advanced country investors retool their local targets using superior production technology. These findings again mirror those presented in Guadalupe, Kuzmina, and Thomas (forthcoming), who observe that local firms report significant increases rates of process innovation and assimilation of foreign technologies after receiving foreign investment. While they suggest this reflects actual increases in indigenous rates of innovation, my findings would suggest they rather reflect evidence of production re-tooling

[45] Results presented in the Web Appendix suggest the lower prices are not a consequence of a shift toward lower quality products.

and technology transfer from the foreign investor to the local firm via imports of superior machinery and equipment.

When controlling for selection using the difference-in-difference matching estimator, however, the results paint a more nuanced picture. As the estimates from Figures 7 and 8 suggest, while targeted firms do exhibit a relative increase in the imports of capital goods in the year immediately following foreign investment, the vast majority of the relative increase in capital goods imports actually comes in the years immediately leading up to foreign investment. These findings are particularly interesting in that they suggest target firms might be undergoing pre-emptive upgrading of their production in anticipation of foreign investment. These results seem to be consistent with recent findings of the literature on "learning to export", which has shown in a variety of contexts that firms can be induced to invest in productivity improvements by being presented with improved exporting opportunities (Lileeva nad Trefler, 2010; Bustos (2011). While this literature has mostly focused on exploring the effects of trade liberalization on investment decisions of local firms, my preliminary findings could suggest that foreign investment might be an alternative mechanism for inducing local firms to invest in productivity improvements.

Additional Findings

My analysis also allows me to investigate the validity of certain findings emphasized by recent papers, specifically the proposition put forth in Guadalupe, Kuzmina, and Thomas (forthcoming) that foreign investors "cherry-pick" when selecting acquisition targets, as well as the proposition laid out in Blonigen et al (2012) that foreign investors acquire local firms in order to exploit their export distribution networks. My data provide results consistent with both of these propositions.

Figures A1-A4 and Tables A15-A20 reported in the Web Appendix[46] clearly show that foreign investors do not randomly select their targets. Instead, they target the largest and most productive local firms, i.e. they invest in the local "cherries". Firms that receive investment are significantly larger and exhibit higher initial productivity than firms that do not receive foreign investment. They are also much more likely to be already active in export markets, and I find that the selection mechanism exhibits similar properties in the case of advanced country investors as in the case of developing country investors. My results thus clearly support the notion from Guadalupe, Kuzmina, and Thomas (forthcoming) that foreign investors engage in "cherry-picking" on observable characteristics of target firms and underscore the need to control for ex-ante differences in the characteristics of target firms when attempting to estimate causal effects of receiving foreign investment on their ex-post performance.

My data also allow me to engage in an initial exploration of the validity of the notion put forth in Blonigen et al (2012) that foreign investors might seek to acquire local firms for their proprietary export distribution networks. As I have discussed in my description of the country's historical context, Slovenian firms have enjoyed a long history of economic ties to markets in former Yugoslavia, Eastern and South-Eastern Europe. If Blonigen et al (2012) are correct, we should see target firms disproportionally increasing exports to these markets after receiving foreign investment, especially from advanced country investors. While the results I discussed above tell us that target firms expand the scope of their export presence overall as well as their export presence in high-income OECD countries after receiving foreign investment, I find mixed evidence that they do indeed disproportionally increase the volume of ex-post exports to

[46] The Web Appendix is available online at http://www.andrew.cmu.edu/user/mdrev/

countries where their ex-ante export ties were strongest. As Tables 15 and 16 report, export volume increases are strongest in non-OECD destinations, and are particularly strong in ex-Yugoslavia and in former Eastern Bloc markets. Interestingly, this is especially true for firms receiving investment from advanced country investors, which is consistent with the story that advanced country investors might partially target local firms in order to exploit their regional export networks and which confirms anecdotal evidence from the illustrative examples. Difference-in-difference matching estimator results, however, provide little evidence in support of this notion.

Conclusions

In this essay, I have used panel data on Slovenian firms to measure the effects of receiving foreign investment on subsequent behavior and performance of targeted local firms. Consistent with several recent studies, I find evidence that firms receiving foreign investment improve their ex-post performance. I take a step beyond existing literature by exploring the importance of investor ability, as proxied by their origin, on the ability of local firms to benefit from foreign investment in the context of a developing country.[47] I find evidence that firms receiving investment from advanced country (i.e. higher-ability) investors experience a larger performance boost ex-post than do firms receiving investment from developing country (i.e. lower-ability) investors. This suggests heterogeneity in investor ability might be important, which is something most recent studies on foreign investment have not focused on.

[47] Chen (2011) has explored overall target firm performance effects of foreign investor origin using data on foreign acquisitions of firms in the United States.

Building on a theoretical framework developed by Bernard, Redding, and Schott (2010b), I show that foreign investment, if accompanied by a transfer of superior managerial and technological abilities from the foreign investor to the local firm, results in an expansion in the local firm's product mix and export destination scope, as well as in a decrease in the prices the firm charges for its existing product portfolio. I present empirical evidence supporting the assertion that local firms endogenously shift their scope, in addition to the scale of their operations, as a result of increases in their ability following foreign investment. My empirical results also provide evidence that local firms modify the scope of their operations in a way consistent with the view that advanced country (i.e. high-ability) investments result in larger increases in target firm ability than developing country (i.e. low-ability) investments, especially when foreign investment is of high intensity. While one needs to exercise caution in drawing general policy implications from these findings, my results do suggest local policymakers in developing countries might maximize the outcomes for local firms offered for investment by targeting high-ability foreign investors and engaging in investor "cherry-picking."

These findings suggest several fruitful avenues for future research. As my data currently only allow me to observe a small set of investor characteristics, most notably their origin, future work using richer data on foreign investors should focus on understanding how investor and target firm characteristics jointly determine the ability of local firms to benefit from foreign investment. Drawing on existing theoretical literature in management strategy and economics that investigates the determinants and effects of cross-border mergers and acquisitions and multinational activity, research using data on a universe of firms in a particular country or set of countries to empirically examine what synergies between investor and target firm abilities are required for a local firm to benefit from foreign investment could yield important insights for future researchers and policy decision-makers.

My empirical results also provide suggestive evidence in support of several notions recently reported in the literature: I find that foreign investors "cherry-pick" when deciding which local firms to target, validating a notion that was put forth in several recent papers. Similarly, I find some preliminary evidence consistent with the view that foreign investors might target local firms in order to exploit their regional export networks. These results suggest it would be useful to extend my theoretical framework to formally include the investment decision, which is something this essay currently abstracts from. Embedding the above stylistic facts, alongside investor heterogeneity and the multi-product multi-destination nature of firms into an internally consistent theoretical framework holds the promise to give us a new depth of understanding of the mechanisms that underlie the results reported in recent literature and in this essay.

References

Aitken, Brian, Ann E. Harrison, and Robert Lipsey, "Wages and Foreign Ownership: A Comparative Study of Mexico, Venezuela, and the United States," Journal of International Economics, 40:3-4 (1996), 345-371.

Antras, Pol, "Firms, Contracts, and Trade Structure," The Quarterly Journal of Economics, 118:4 (2003), 1375-1418.

Arnold, Jens Matthias, and Beata S. Javorcik, "Gifted Kids or Pushy Parents? Foreign Direct Investment and Plant Productivity in Indonesia," Journal of International Economics, 79:1 (2009), 42-53.

Baldwin, Richard, "Trade and Industrialization After Globalization's 2nd Unbundling: How Building and Joining a Supply Chain are Different and Why It Matters," NBER Working Paper #17716 (2011).

Baldwin, John, and Wulong Gu, "Global Links: Multinationals, Foreign Ownership and Productivity Growth in Canadian Manufacturing," The Canadian Economy in Transition Series Research Paper (2005).

Barba Navaretti, Giorgio, and Anthony Venables, "Multinational Firms in the World Economy," Princeton, NJ: Princeton University Press, 2004.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multi-Product Firms and Product Switching," American Economic Review, 100:1 (2010), 70-97.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multi-Product Firms and Trade Liberalization," Working Paper (2010).

Blonigen, Bruce A., "Firm-Specific Assets and the Link between Exchange Rates and Foreign Direct Investment," American Economic Review, 87:3 (1997), 447-465.

Blonigen, Bruce A., Lionel Fontagne, Nicholas Sly, and Farid Toubal, “Cherries for Sale: Export Networks and the Incidence or Cross-Border M&A,” NBER Working Paper #18414 (2012).

Bloom, Nicholas, and John Van Reenen, “Measuring and Explaining Management Practices Across Firms and Countries,” The Quarterly Journal of Economics, 122:4 (2007), 1351-1408.

Bloom, Nicholas, and John Van Reenen, “Why Do Management Practices Differ across Firms and Countries?” Journal of Economic Perspectives, 24:1 (2010), 203-224.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, “Americans do IT Better: US Multinationals and the Productivity Miracle,” American Economic Review, 102:1 (2012), 167-201.

Branstetter, Lee G., Raymond Fisman, and Fritz Foley, “Do Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from the U.S. Firm-Level Panel Data,” The Quarterly Journal of Economics, 121:1 (2006), 321-349.

Burstein, Ariel, and Alexander Monge-Naranjo, “Foreign Know-How, Firm Control, and the Income of Developing Countries,” The Quarterly Journal of Economics, 124:1 (2009), 149-195.

Busso, Matias, John DiNardo, and Justin McCrary, “New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators,” Working Paper (2011).

Bustos, Paula, “Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms,” American Economic Review, 101:1 (2011), 304-340.

Caves, Richard, "Multinational Enterprise and Economic Analysis (2nd Edition)," Cambridge, MA: Cambridge University Press (1996).

Chen, Wenjie, "The Effect of Investor Origin on Firm Performance: Domestic and Foreign Direct Investment in the United States," Journal of International Economics 83 (2011), 219-228.

Chung, Wilbur, and Juan Alcacer, "Knowledge Seeking and Location Choice of Foreign Direct Investment in the United States," Management Science, 38:12 (2002), 1534-1554.

Criscuolo, Chiara, and Ralf Martin, "Multinationals and U.S. Productivity Leadership: Evidence from Great Britain," Review of Economics and Statistics, 92:2 (2009), 263-281.

Dehejia, Rajeev H. and Sadek Wahba, "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," Journal of the American Statistical Association, 94:448 (1999), 1053-1062.

Dixit, Avinash, and Joseph Stiglitz, "Monopolistic Competition and Optimum Product Diversity," American Economic Review, 67:3 (1977), 297-308.

Estrin, Saul, Jan Hanousek, Evzen Kocenda, and Jan Svejnar, "The Effects of Privatization and Ownership in Transition Economies", Journal of Economic Literature, 47:3 (2009), 699-728.

Falk, Martin, "Effects of Foreign Ownership on Innovation Activities: Empirical Evidence for Twelve European Countries," National Institute Economic Review, 204:1 (2008), 85-97.

Feenstra, Robert C., and Gordon H. Hanson, "Ownership and Control in Outsourcing to China: Estimating the Property-Rights Theory of the Firm," The Quarterly Journal of Economics, 120:2 (2005), 729-761.

Guadalupe, Maria, Olga Kuzmina, and Catherine Thomas. "Innovation and Foreign Ownership," American Economic Review (forthcoming).

Grossman, Gene M., and Elhanan Helpman, “Integrations versus Outsourcing in Industry Equilibrium,” The Quarterly Journal of Economics, 117:1 (2002), 85-120.

Haddad, Mona, and Ann E. Harrison, “Are There Positive Spillovers from Direct Foreign Investment? Evidence from Panel Data for Morocco,” Journal of Development Economics, 42:1 (1993), 51-74.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, “Characterizing Selection Bias Using Experimental Data,” Econometrica, 66:5 (1998), 1017-1098.

Hennart, Francois, “The Transaction Costs Theory of Joint Ventures: An Empirical Study of Japanese Subsidiaries in the United States,” Management Science, 37:4 (1991), 483-497.

Heynman, Fredrik, Fredrik Sjoholm, and Patrik Gustavsson Tinvall, “Is There Really a Foreign Ownership Premium? Evidence from Employer-Employee Matched Data,” Journal of International Economics, 73:2 (2007), 355-376.

Helpman, Elhanan, Marc Melitz, and Stephen Ross Yeaple, “Export Versus FDI with Heterogeneous Firms,” American Economic Review, 94:1 (2004), 300-316.

Horn, Henrik, and Lars Persson, “The Equilibrium Ownership of an International Oligopoly,” Journal of International Economics, 53:2 (2001), 307-333.

Huttunen, Kristiina, “The Effect of Foreign Acquisition on Employment and Wages: Evidence from Finnish Establishments,” The Review of Economics and Statistics, 89:3 (2007), 497-509.

Javorcik, Beata, “Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages,” American Economic Review, 94:3 (2004), 605-627.

Jovanovic, Boyan, and Serguey Braguinsky, “Bidder Discounts and Target Premia in Takeovers,” American Economic Review, 94:1 (2004), 46-56.

Kamien, Morton I., and Israel Zhang, “The Limits of Monopolization Through Acquisition,” The Quarterly Journal of Economics, 105:2 (1990), 465-499.

Kronborg, Dorte, and Steen Thomsen, “Foreign Ownership and Long-Term Survival,” Strategic Management Journal, 30 (2009), 207-219.

Lichtenberg, Frank R., and Donald Siegel, “Productivity and Changes of Ownership in Manufacturing Firms,” Brookings Papers on Economic Activity, 1987:3 (1987), 643-83.

Lileeva, Alla, and Daniel Trefler, "Improved Access to Foreign Markets Raises Plant-Level Productivity…For Some Plants," The Quarterly Journal of Economics, 125:3 (2010), 1051-1099.

Lorber, Lucka, “The Economic Transition of Slovenia in the Process of Globalization”, Geografski zbornik, 39 (1999), 134-166.

Luo, Yadong, and Rosalie L. Tung, “International Expansion of Emerging Market Enterprises: A Springboard Perspective,” Journal of International Business Studies, 38:1 (2007), 481-498.

McGrattan, Ellen, and Edward Prescott, “Technology Capital and the U.S. Current Account,” American Economic Review, 100:4 (2010), 1493-1522.

Melitz, Mark, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” Econometrica 71:6 (2003), 1695-1725.

Meyer, Klaus E., Saul Estrin, Sumon Kumar Bhaumik, and Mike W. Peng, ”Institutions, Resources, and Entry Strategies in Emerging Economies,” Strategic Management Journal, 30 (2009), 61-80.

Neary, Peter J., “Cross-Border Mergers as Instruments of Comparative Advantage, ”Review of Economics Studies, 74:4 (2007), 1229-57.

Nenadic, Tina, “Prestrukturiranje predelovalnih dejavnosti v Sloveniji” (Structural Changes in Slovenian Manufacturing), UMAR Working Paper, 21:2 (2012).

Nocke, Volker, and Stephen R. Yeaple, “Cross-Border Mergers and Acquisitions Versus Greenfield Foreign Direct Investment: The Role of Firm Heterogeneity,” Journal of International Economics, 72:2 (2007), 336-365.

Nocke, Volker, and Stephen R. Yeaple, „An Assignment Theory of Foreign Direct Investment,” Review of Economic Studies, 72:2 (2008), 529-557.

Pogatsa, Zoltan, “Slovenia: The Only Successful Case of Economic Transition”, Hungarian Review, 3:4 (2012).

Ramondo, Natalia, “Foreign Plants and Industry Productivity: Evidence from Chile,” The Scandinavian Journal of Economics, 111:4 (2009), 789-809.

Ramondo, Natalia, and Andres Rodriguez-Clare, “Trade, Multinational Production, and the Gains from Openness,” NBER Working Paper #15604 (2009).

Seth, Anju, Kean P. Song, and Richardson Pettit, “Synergy, Managerialism, or Hubris? An Empirical Examination of Motives for Foreign Acquisitions of U.S. Firms,” Journal of International Business Studies, 31:3 (2000), 387-405.

Shane, Scott, “The Effect of National Culture on the Choice between Licensing and Direct Foreign Investment,” Strategic Management Journal, 15 (1994), 627-642.

Shaver, J. Myles, “Accounting for Endogeneity When Assessing Strategy Performance: Does Entry Mode Choice Affect FDI Survival?” Management Science, 44:4 (1998), 571-585.

Simoneti, Marko, Joze P. Damijan, Matija Rojec, and Boris Majcen, “Case-by-Case Versus Mass Privatization in Transition Economies: Initial Owner and Final Seller Effects on Performance of Firms in Slovenia”, World Development, 33:10 (2005), 1603-1625.

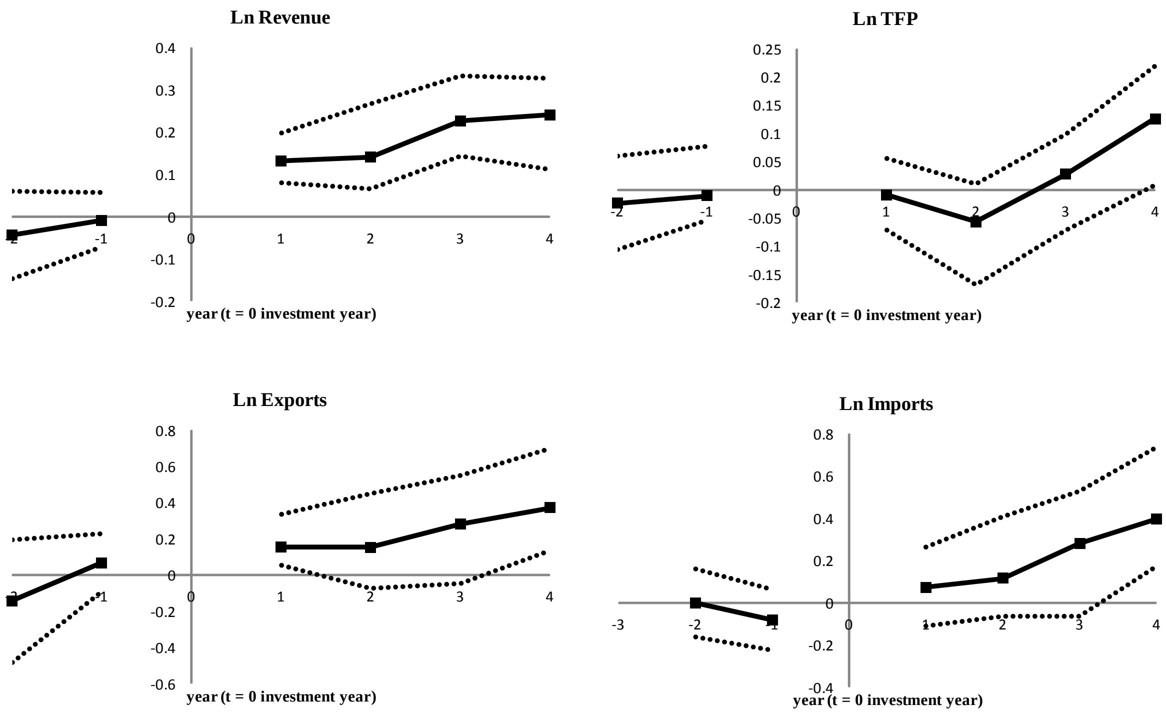
Statistical Office of the Republic of Slovenia (SURS), “Statisticni letopis”, Multiple Vintages (1995-2012), (http://www.stat.si/publikacije/pub_letopis_prva.asp).

Todd, Petra. E., “Matching Estimators,” Working Paper (2006).

Yeaple, Stephen Ross, “Offshoring, Foreign Direct Investment, and the Structure of U.S. Trade,” Journal of the European Economic Association, 4:2-3 (2006), 602-611.

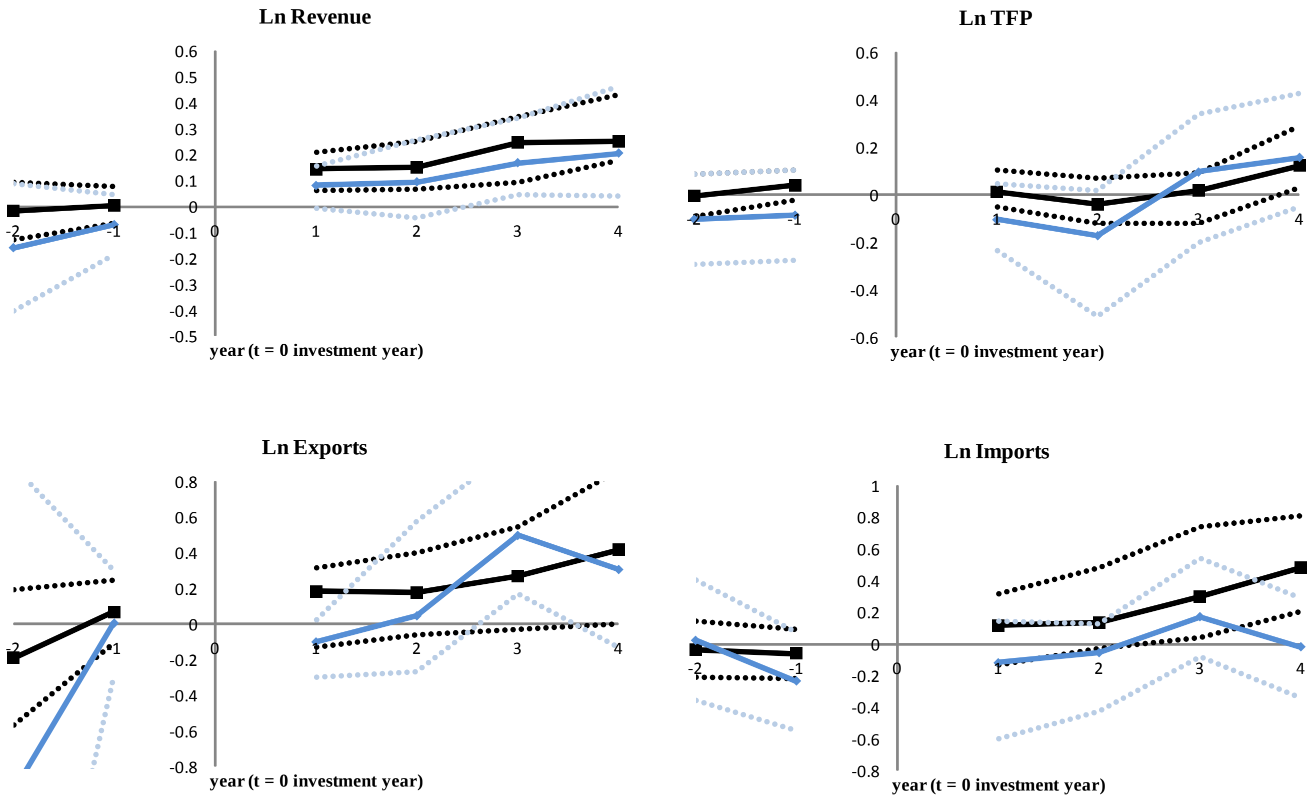
Tables and Figures

Figure 1: Effect of Foreign Investment on Firm Performance and International Trade Dynamics, Difference-in-Difference Matching Estimator



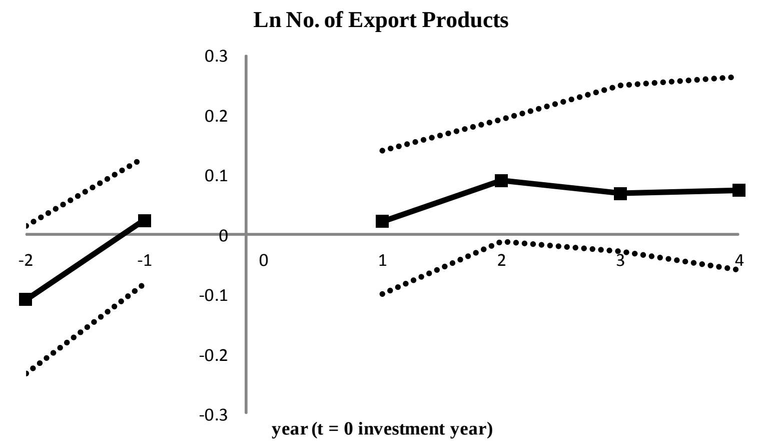
This figure documents difference-in-difference matching estimator results for the post-acquisition performance between firms who received foreign investment and "matched" firms who stayed domestically owned. Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Confidence intervals are calculated using bias-adjusted bootstrapped standard errors.

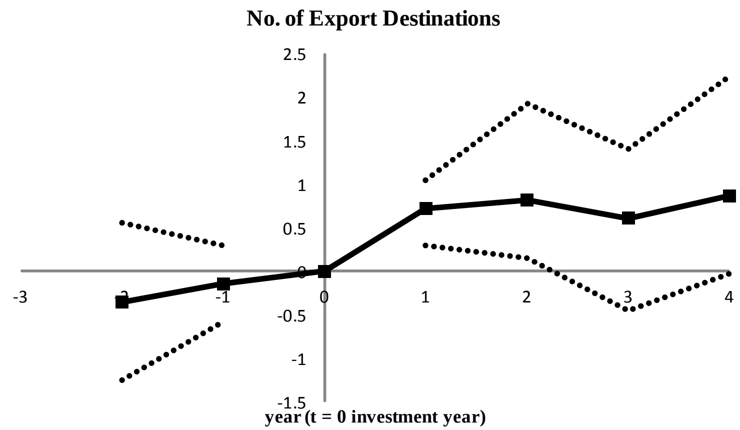
Figure 2: Investor Origin, Firm Performance, and International Trade Dynamics, Difference-in-Difference Matching Estimator



This figure documents difference-in-difference matching estimator results for the post-acquisition performance between firms who received foreign investment from a certain geographical origin and matched firms who remained domestically owned. Black line denotes advanced country investors, while blue line denotes developing country investors Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Confidence intervals are calculated using bias-adjusted bootstrapped standard errors.

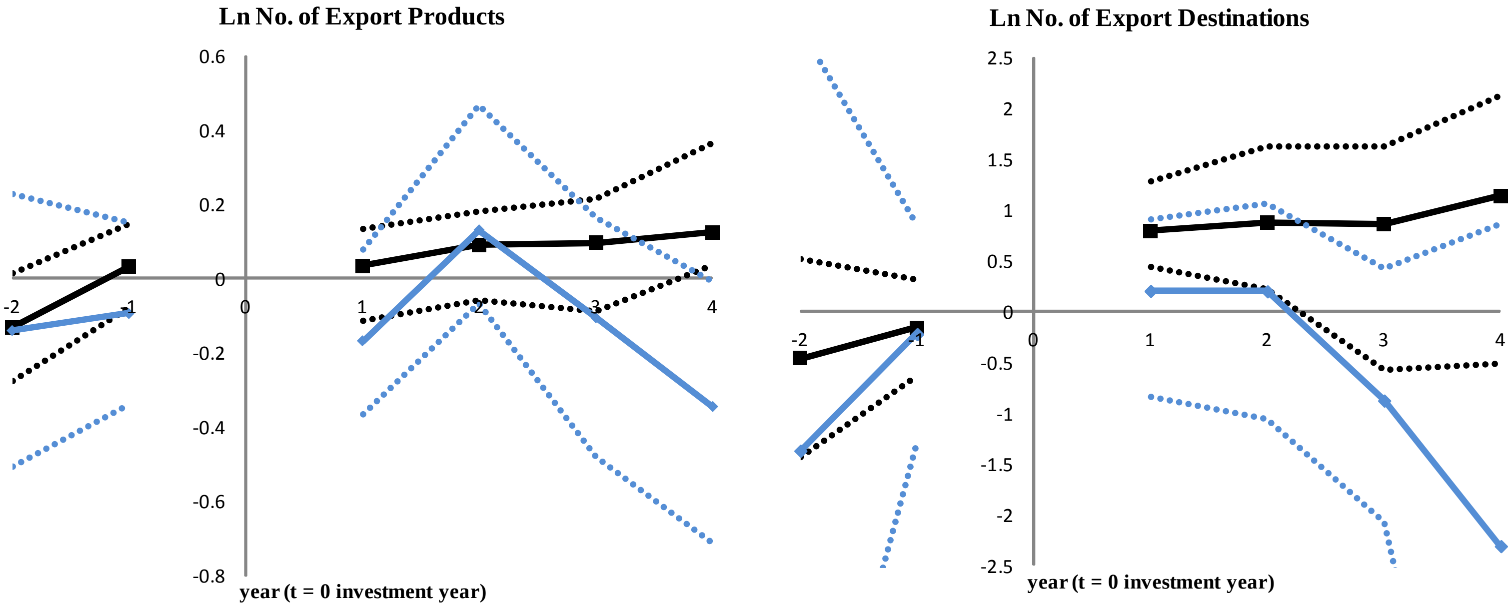
Figure 3: Effect of Foreign Investment on Firm Product Mix and Export Destination Scope, Difference-in-Difference Matching Estimator





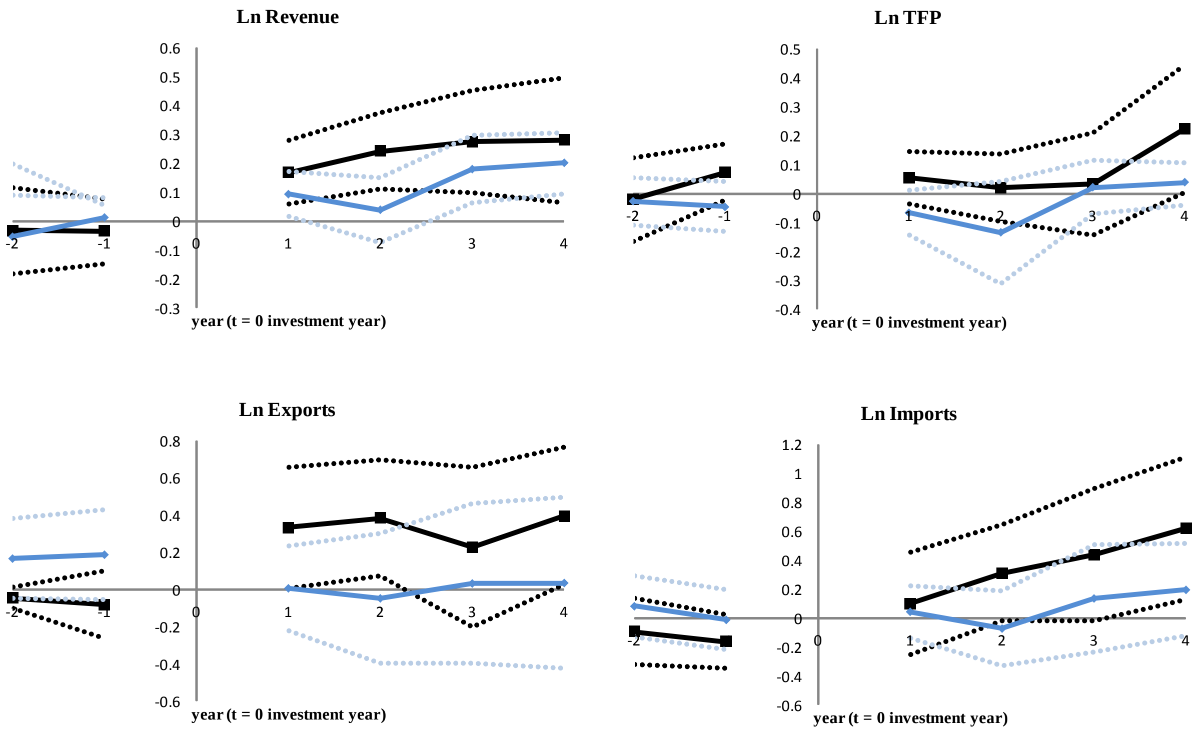
This figure documents difference-in-difference matching estimator results for the post-acquisition product mix and export destination scope between firms who received foreign investment and "matched" firms who stayed domestically owned. Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Confidence intervals are calculated using bias-adjusted bootstrapped standard errors.

Figure 4: Investor Origin, Firm Product Mix, and Export Destination Scope, Difference-in-Difference Matching Estimator



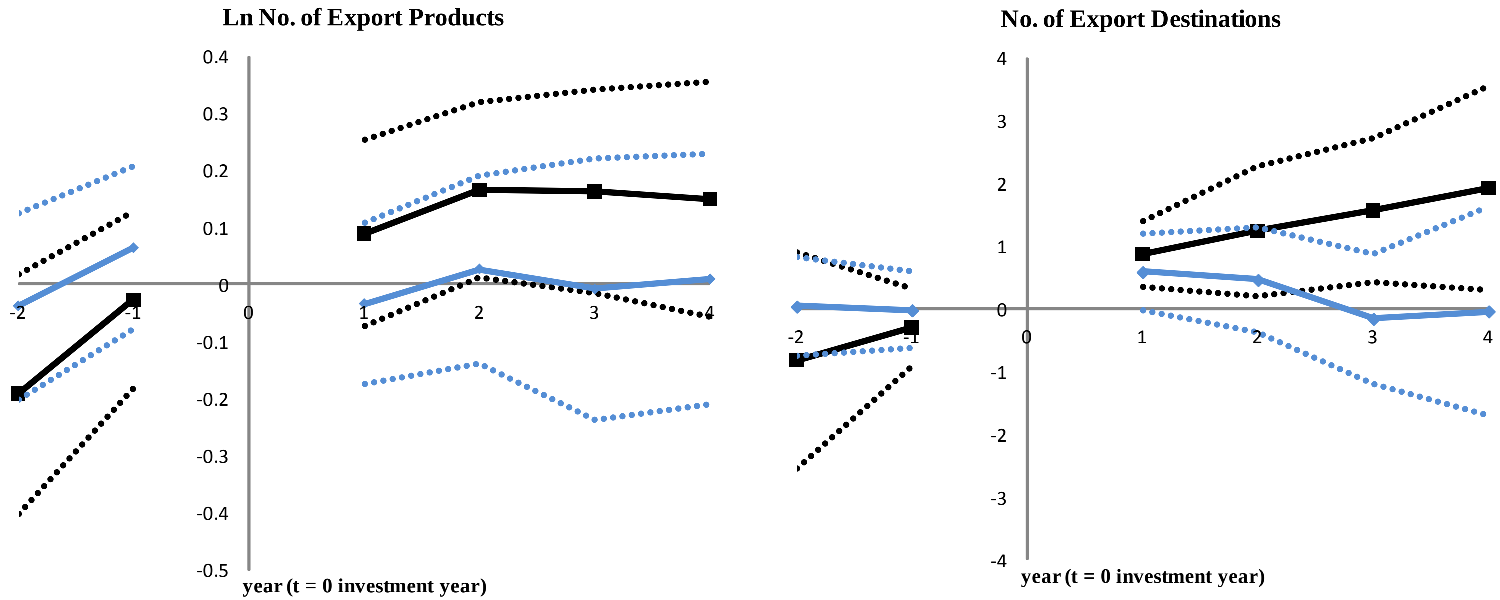
This figure documents difference-in-difference matching estimator results for the post-acquisition product mix and export destination scope between firms who received foreign investment from a certain geographical origin and matched firms who remained domestically owned. Black line denotes advanced country investors, while blue line denotes developing country investors. Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Product mix was calculated at the 8-digit level of the Slovenian version of the Combined Nomenclature. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Confidence intervals are calculated using bias-adjusted bootstrapped standard errors.

Figure 5: Investment Intensity, Firm Performance, and International Trade Dynamics, Difference-in-Difference Matching Estimator



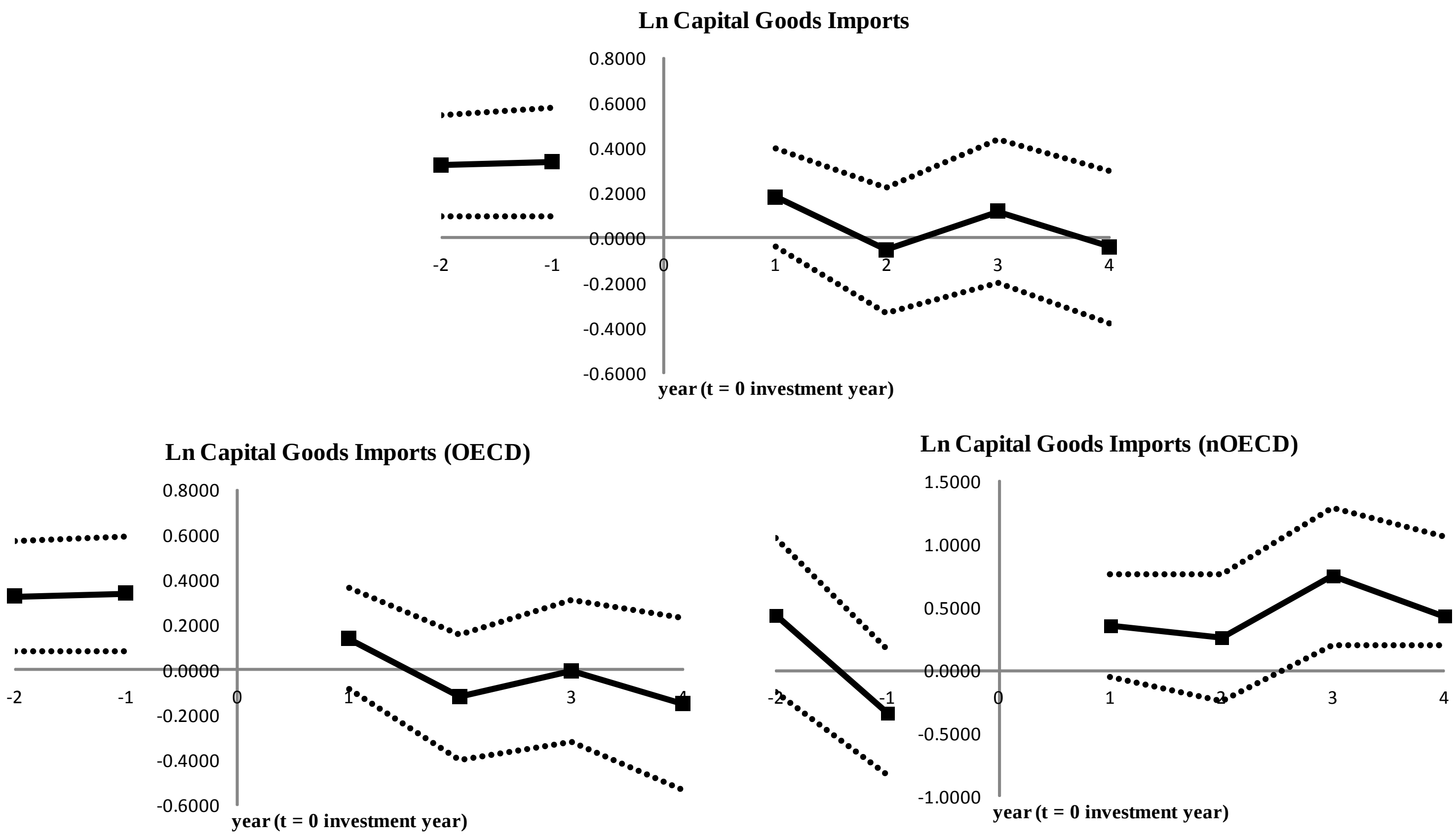
This figure documents difference-in-difference matching estimator results for the post-acquisition performance between firms who received foreign investment and matched firms who remained domestically owned. Black line denotes investment target for which the scaled initial investment amount was above the median of the sample (i.e. high intensity), while blue line denotes investment target for which the scaled initial investment amount was below the median of the sample (i.e. low intensity). Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Confidence intervals are calculated using bias-adjusted bootstrapped standard errors.

Figure 6: Investment Intensity, Firm Product Mix, and Export Destination Scope, Difference-in-Difference Matching Estimator



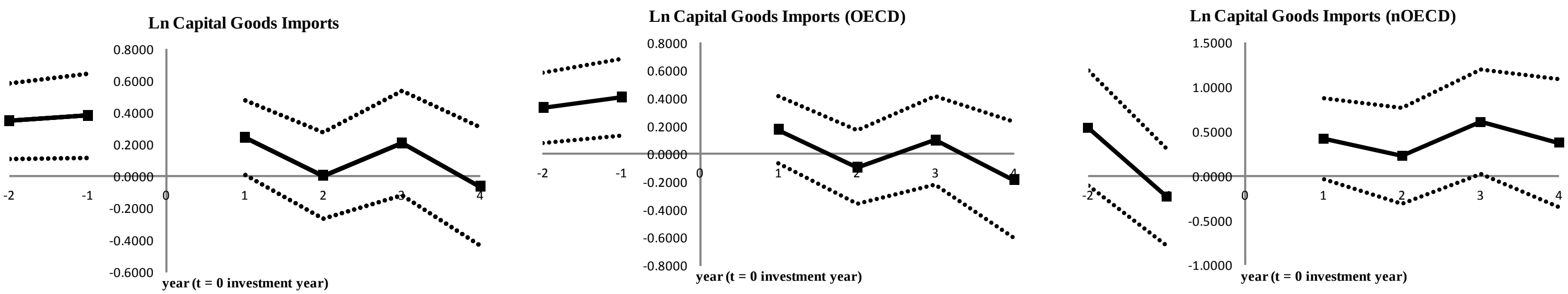
This figure documents difference-in-difference matching estimator results for the post-acquisition product mix and export destination scope between firms who received foreign investment and matched firms who remained domestically owned. Black line denotes investment target for which the scaled initial investment amount was above the median of the sample (i.e. high intensity), while blue line denotes investment target for which the scaled initial investment amount was below the median of the sample (i.e. low intensity). Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Product mix was calculated at the 8-digit level of the Slovenian version of the Combined Nomenclature. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups..

Figure 7: Foreign Investment and Imports of Capital Goods by Target Firms, Difference-in-Difference Matching Estimator

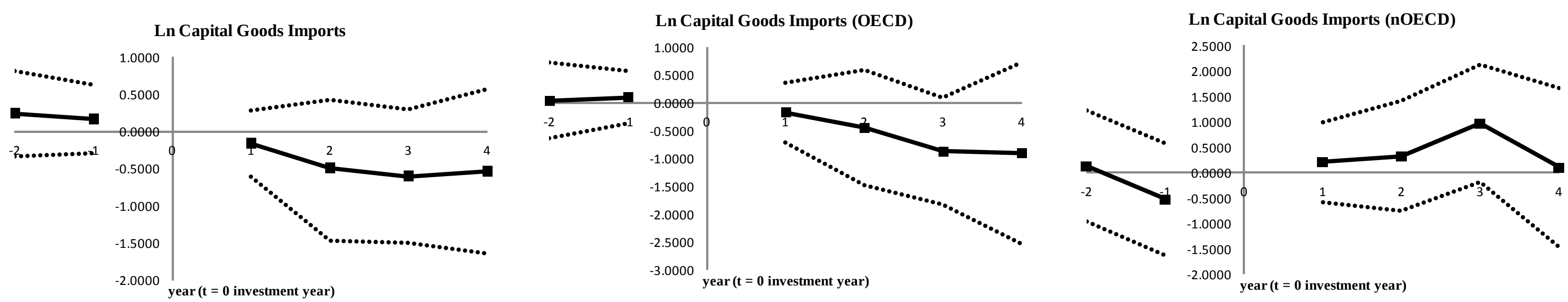


This figure documents difference-in-difference matching estimator results for the imports of capital goods between firms who received foreign investment and matched firms who remained domestically owned. Capital goods are defined using the Slovenian vintages of the Combined Nomenclature at the 4-digit level (codes 8201-9033). Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups.

Figure 8: Investor Origin and Imports of Capital Goods by Target Firms, Difference-in-Difference Matching Estimator



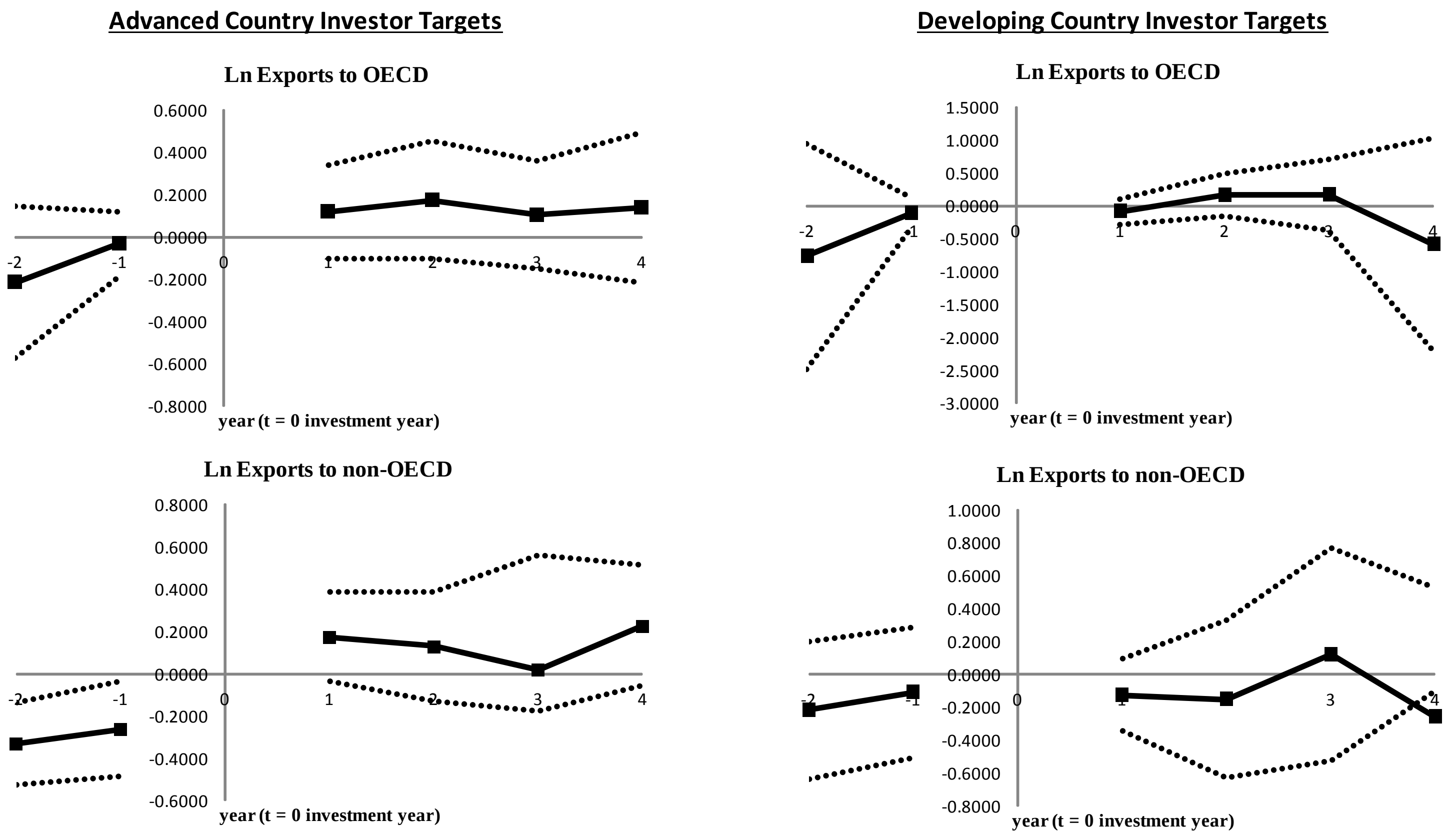
Advanced Country Investor Targets



Developing Country Investor Targets

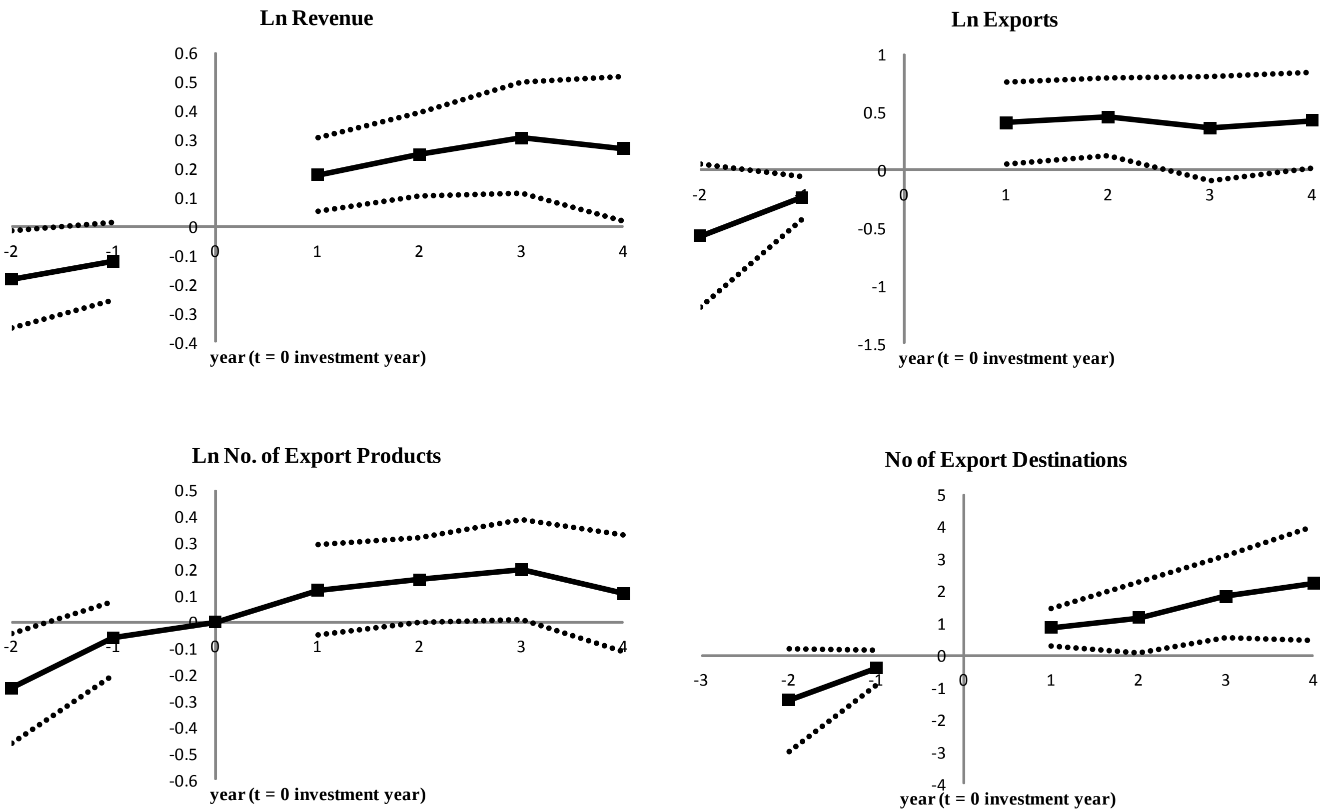
This figure documents difference-in-difference matching estimator results for the imports of capital goods between firms who received foreign investment and matched firms who remained domestically owned. The top panel contains targets of advanced country investors, while the bottom panel contains targets of developing country investors. Capital goods are defined using the Slovenian vintages of the Combined Nomenclature at the 4-digit level (codes 8201-9033). Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups.

Figure 9: Investor Origin and Geography of Export Destinations, Difference-in-Difference Matching Estimator



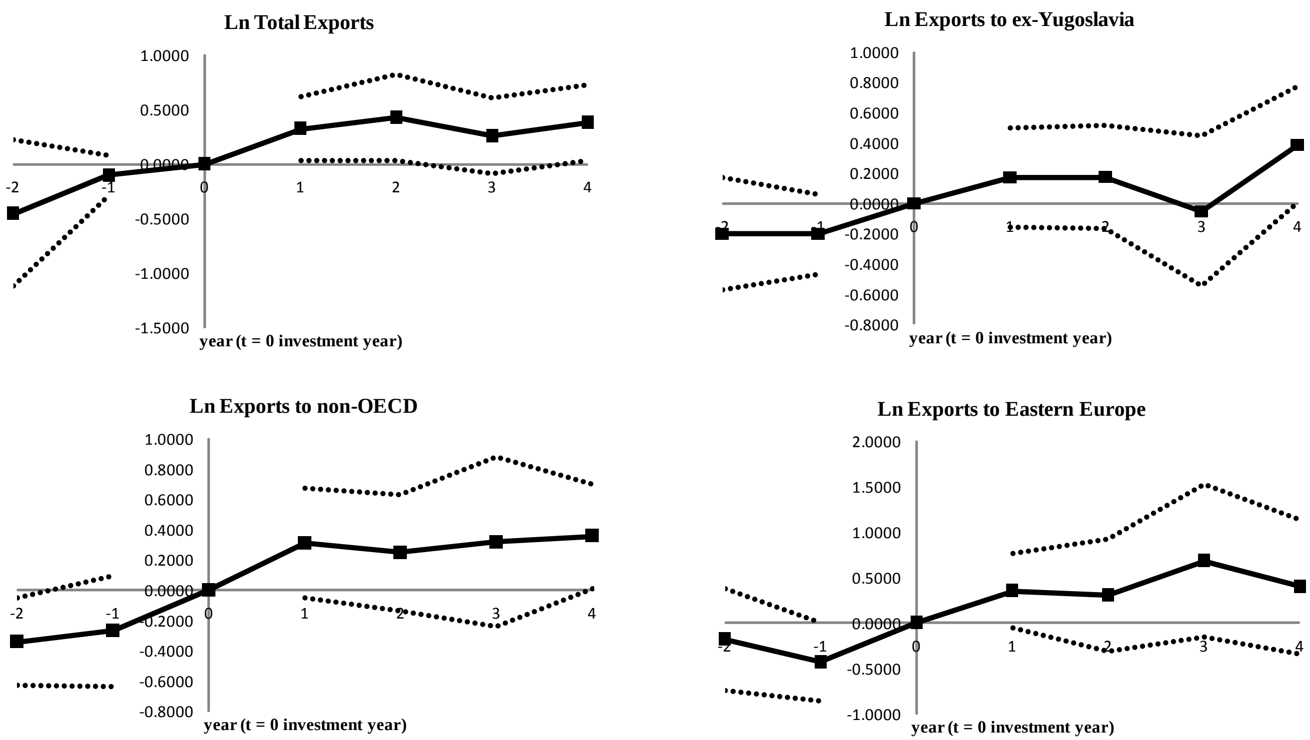
This figure documents difference-in-difference matching estimator results for the geography of export destinations between firms who received foreign investment and matched firms who remained domestically owned. The left-hand panel contains targets of advanced country investors, while the right-hand panel contains targets of developing country investors. Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups.

Figure 10: Effect of High-Intensity Foreign Investment of Developed Country Investor Origin on Firm Performance and International Trade Dynamics, Difference-in-Difference Matching Estimator



This figure documents difference-in-difference matching estimator results for the post-acquisition performance between firms who received high-intensity foreign investment from developed country investors and "matched" firms who stayed domestically owned. Only those targets of developed country investors for which the scaled initial investment amount was above the median of the sample (i.e. high intensity) were used in the estimation. Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Confidence intervals are calculated using bias-adjusted bootstrapped standard errors.

Figure 11: High-Intensity Foreign Investment of Developed Country Origin and Geography of Export Destinations, Difference-in-Difference Matching Estimator



This figure documents difference-in-difference matching estimator results for the for the geography of export destinations between firms who received high-intensity foreign investment from developed country investors and "matched" firms who stayed domestically owned. Only those targets of developed country investors for which the scaled initial investment amount was above the median of the sample (i.e. high intensity) were used in the estimation. Bold line indicates the point estimate, while dotted lines indicate boundary of the 95% confidence interval. Kernel matching with a bandwidth of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Confidence intervals are calculated using bias-adjusted bootstrapped standard errors.

Table 1: Summary Statistics

Variable	Total Sample			Domestic Firms			Advanced Investor Targets			Developing Investor Targets		
	Mean	St. Dev.	# of Obs.	Mean	St. Dev.	# of Obs.	Mean	St. Dev.	# of Obs.	Mean	St. Dev.	# of Obs.
Revenue	1873.30	11867.20	96940	1581.10	10697.50	94074	10419.64	25234.47	2301	16159.57	48813.66	466
Net Income	29.31	1132.91	96800	21.50	1023.94	93943	326.55	3251.81	2290	150.54	1647.76	468
Fixed Assets	905.96	6268.96	96672	774.18	5927.59	93812	5271.31	13352.27	2295	4761.25	8210.42	466
Materials Expenditure	835.67	6063.94	96564	709.79	5776.96	93704	4584.48	10300.96	2297	6281.79	15558.43	464
Labor Expenditure	404.20	2294.64	94611	344.11	2066.28	91755	2324.67	5818.47	2290	2242.87	5850.83	467
Value Added	561.69	4018.58	96911	472.33	3593.56	94045	3521.91	11358.63	2301	3268.57	7487.84	466
Total Exports	2515.41	15129.40	34856	2060.78	14034.28	32462	7342.36	20672.23	1964	15736.38	42261.90	340
Total Imports	1268.30	7158.38	42524	1048.39	6565.98	39922	3943.16	10216.37	2110	7734.42	21677.08	393
Number of Exp. Destinations	5.69	8.59	35579	5.20	7.93	33151	12.18	12.99	1992	13.28	14.78	343
Number of Imp. Destinations	4.68	5.24	43269	4.41	4.98	40632	8.73	6.91	2139	8.67	7.31	396
Number of Exp. Products (8-Digit)	14.35	28.08	35579	13.01	25.73	33151	31.50	44.68	1992	31.06	43.65	343
Number of Imp. Product (8-Digit)	32.10	56.15	43269	29.17	52.29	40632	78.92	87.52	2139	57.08	66.50	396
Number of Exp. Product Lines (4-Digit)	8.54	13.98	35579	7.82	12.93	33151	18.07	21.63	1992	18.43	23.08	343
Number of Imp. Product Lines (4-Digit)	18.64	26.47	43269	17.10	24.78	40632	43.77	38.66	2139	32.61	32.73	396
Number of Employees	29.96	146.15	94075	25.85	136.54	91228	160.62	283.73	2284	149.44	364.79	464
Average Wage	7.85	4.20	80700	7.77	4.14	77934	10.18	5.60	2217	9.59	3.46	450

This table provides summary statistics of key variables available in the dataset used in this essay. The first column provides summary statistics for the entire dataset, while the 2nd, 3rd, and 4rd columns split the dataset into three categories: domestic firms, firms targeted by advanced country investors, and firms targeted by developing country investors. Domestic firms are those that remain domestically owned in the entire period they appear in the data set. Advanced country investor targets are those firms that are initially domestically owned, but then report receiving investment from an advanced country investor. Developing country investor targets are those firms that are initially domestically owned, but then report receiving investment from an investor from a developing country. All financial accounts and trade data values are in thousands of real Euros, with 2000 set as the base year. Number of export/import destinations is the number of distinct countries a firm reports exporting/importing to/from in a given year. Number of exported/imported products is the number of exported/imported products a firm reports in a given year. The product identification was conducted using the Slovenian version of the Combined Nomenclature, either at the 8-digit level (products) or at the 4-digit level (product lines)

Table 2: Foreign Investment and Measures of Firm Performance, Linear Regression Approach

Variable / Performance Measure				
Panel A	Ln Revenue			
Lag Foreign	2.8388***	0.3443***	0.1238***	0.2780***
	(0.1443)	(0.0645)	(0.0392)	(0.0712)
No. of Observations	88768	88768	63511	59586
R-Squared	0.8614	0.8501	0.9238	0.9580
Panel B	Ln TFP			
Lag Foreign	0.3096***	0.0971***	0.0445*	0.0097
	(0.0316)	(0.0305)	(0.0270)	(0.0338)
No. of Observations	73571	73571	60007	56820
R-Squared	0.7562	0.4508	0.5114	0.7824
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the natural log of deflated sales revenue of a firm in year t, while the dependent variable in Panel B is the natural log of firm-level Levinsohn-Petrin estimated TFP in year t. Lag Foreign is an indicator variable that equals one if firm had reported at least 10% foreign ownership in year t-1. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Number of observations can differ between columns due to the fact that we are using an unbalanced panel and not all variables were available for all firm-year observations. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 3: Foreign Investment and Measures of International Trade Dynamics, Linear Regression Approach

Variable / Performance Measure				
Panel A	Ln Exports			
Lag Foreign	2.5529***	0.3206***	0.1965**	0.1837
	(0.2118)	(0.0986)	(0.0915)	(0.1223)
No. of Observations	33051	33051	27316	26164
R-Squared	0.7398	0.8263	0.8564	0.9415
Panel B	Ln Imports			
Lag Foreign	2.4988***	0.3866***	0.1550**	0.1663
	(0.1622)	(0.0771)	(0.0651)	(0.1057)
No. of Observations	39851	39851	31871	30462
R-Squared	0.7281	0.7985	0.8320	0.9296
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the natural log of deflated exports of a firm in year t, while the dependent variable in Panel B is the natural log of deflated imports of a firm in year t. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Export intensity was top-censored at one. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 4: Advanced Country Foreign Investment and Measures of Firm Performance, Linear Regression Approach

Variable / Performance Measure				
Panel A	Ln Revenue			
Lag Foreign	2.9011***	0.3670***	0.1371***	0.3158***
	(0.1482)	(0.0726)	(0.0439)	(0.0833)
No. of Observations	88336	88336	63132	59377
R-Squared	0.8613	0.8488	0.9230	0.9475
Panel B	Ln TFP			
Lag Foreign	0.3148***	0.0974***	0.0457*	0.0376
	(0.0337)	(0.0309)	(0.0264)	(0.0348)
No. of Observations	73158	73158	59638	56615
R-Squared	0.7558	0.4508	0.5114	0.7054
Industry FEs	Yes			
Time Fes	Yes	Yes	Yes	Yes
Firm Fes		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the natural log of deflated sales revenue of a firm in year t, while the dependent variable in Panel B is the natural log of firm-level Levinsohn-Petrin estimated TFP in year t. Lag Foreign is an indicator variable that equals one if firm had reported at least 10% foreign ownership in year t-1. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Advanced country investor is defined as foreign investment originating from high-income OECD countries. These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States. .Number of observations can differ between columns due to the fact that we are using an unbalanced panel and not all variables were available for all firm-year observations. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 5: Developing Country Foreign Investment and Measures of Firm Performance, Linear Regression Approach

Variable / Performance Measure				
Panel A	Ln Revenue			
Lag Foreign	2.6220***	0.2636**	0.0905	0.1845*
	(0.4154)	(0.1146)	(0.0704)	(0.1023)
No. of Observations	86613	88336	61665	58663
R-Squared	0.8596	0.8437	0.9198	0.9297
Panel B	Ln TFP			
Lag Foreign	0.2800***	0.0805	0.0292	-0.0914
	(0.0797)	(0.0853)	(0.0811)	(0.0771)
No. of Observations	71526	71526	58202	55914
R-Squared	0.7521	0.4473	0.5070	0.6857
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the natural log of deflated sales revenue of a firm in year t, while the dependent variable in Panel B is the natural log of firm-level Levinsohn-Petrin estimated TFP in year t. Lag Foreign is an indicator variable that equals one if firm had reported at least 10% foreign ownership in year t-1. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Developing country investor is defined as investment originating from a complement of the set of advanced investor countries, except countries that are offshore tax haven countries, which are excluded from this part of the empirical analysis. Number of observations can differ between columns due to the fact that we are using an unbalanced panel and not all variables were available for all firm-year observations. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 6: Advanced Country Foreign Investment and Measures of International Trade Dynamics, Linear Regression Approach

Variable / Performance Measure				
Panel A	Ln Exports			
Lag Foreign	2.6339***	0.3488***	0.2154**	0.2310*
	(0.2186)	(0.1083)	(0.1011)	(0.1407)
No. of Observations	32731	32731	27072	26002
R-Squared	0.7393	0.8243	0.8546	0.9362
Panel B	Ln Imports			
Lag Foreign	2.5696***	0.4213***	0.1642**	0.2306**
	(0.1583)	(0.0853)	(0.0712)	(0.1081)
No. of Observations	39481	39481	31538	30279
R-Squared	0.7268	0.7965	0.8303	0.9183
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the natural log of deflated exports of a firm in year t, while the dependent variable in Panel B is the natural log of deflated imports of a firm in year t. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Export intensity was top-censored at one. Advanced country investor is defined as foreign investment originating from high-income OECD countries. These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 7: Developing Country Foreign Investment and Measures of International Trade Dynamics, Linear Regression Approach

Variable / Performance Measure				
Panel A	Ln Exports			
Lag Foreign	2.1120***	0.2919	0.2682	0.0696
	(0.6169)	(0.2152)	(0.1964)	(0.2286)
No. of Observations	31182	31182	25666	25352
R-Squared	0.7288	0.8181	0.8493	0.8741
Panel B	Ln Imports			
Lag Foreign	2.2063***	0.2356	0.1427	0.0118
	(0.5367)	(0.1554)	(0.1368)	(0.2581)
No. of Observations	37850	37850	30124	29596
R-Squared	0.7146	0.7910	0.8253	0.8683
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the natural log of deflated exports of a firm in year t, while the dependent variable in Panel B is the natural log of deflated imports of a firm in year t. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Export intensity was top-censored at one. Developing country investor is defined as investment originating from a complement of the set of advanced investor countries, except countries that are offshore tax haven countries, which are excluded from this part of the empirical analysis. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 8: Foreign Investment and Firm Export Product Mix Scope, Linear Regression Approach

Variable / Performance Measure				
Panel A	Ln Number of Exported Products (8-digit CN level)			
Lag Foreign	0.9504***	0.2117***	0.0875*	0.0996*
	(0.1027)	(0.0561)	(0.0504)	(0.0593)
No. of Observations	33730	33730	27778	26592
R-Squared	0.6638	0.7625	0.7996	0.9142
Panel B	Ln Number of Exported Products (4-digit CN level)			
Lag Foreign	0.8823***	0.1953***	0.0804*	0.0780
	(0.0917)	(0.0493)	(0.0444)	(0.0551)
No. of Observations	33730	33730	27778	26592
R-Squared	0.6291	0.7499	0.7872	0.9098
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the natural log of the number of exported products measured at the 8-digit Combined Nomenclature level for a firm in year t, while the dependent variable in Panel B is the natural log of the number of exported products measured at the 4-digit Combined Nomenclature level for a firm in year t. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Only firms that exported at least one product in a given year were included in the regressions. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 9: Investor Origin and Firm Product Mix Scope, Linear Regression Approach

Variable / Performance Measure				
Panel A	Advanced Country Investor			
Lag Foreign	1.013***	0.2744***	0.1285**	0.1250**
	(0.1054)	(0.0586)	(0.0521)	(0.0605)
No. of Observations	33407	33407	27486	26430
R-Squared	0.6632	0.7611	0.7984	0.9065
Panel B	Developing Country Investor			
Lag Foreign	0.6611**	-0.1342	-0.1170	-0.0776
	(0.3004)	(0.1096)	(0.1168)	(0.1648)
No. of Observations	31833	31833	26104	25780
R-Squared	0.6514	0.7556	0.7929	0.8316
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable is the natural log of the number of exported products measured at the 8-digit Combined Nomenclature level for a firm in year t. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Only firms that exported at least one product in a given year were included in the regressions. Advanced country investor is defined as foreign investment originating from high-income OECD countries. These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States. Developing country investor is defined as investment originating from a complement of the set of advanced investor countries, except countries that are offshore tax haven countries, which are excluded from this part of the empirical analysis. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 10: Foreign Investment and Firm Export Geographical Scope, Linear Regression Approach

Variable / Performance Measure				
Panel A	Number of Export Destinations			
Lag Foreign	6.6794***	1.5621***	0.8652**	1.4830***
	(1.0419)	(0.4314)	(0.3756)	(0.4170)
No. of Observations	33730	33730	27778	26592
R-Squared	0.3578	0.8828	0.9085	0.9594
Panel B	Number of OECD Export Destinations			
Lag Foreign	4.2756***	0.9283***	0.4888**	0.7457***
	(0.6159)	(0.2193)	(0.1990)	(0.2368)
No. of Observations	33730	33730	27778	26592
R-Squared	0.3366	0.8714	0.8957	0.9553
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable in Panel A is the number of export destination countries for a firm in year t, while the dependent variable in Panel B is the number of OECD-member export destination countries for a firm in year t. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Only firms that exported to at least one country in a given year were included in the regressions. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 11: Advanced Country Foreign Investment and Firm Export Destination Scope, Linear Regression Approach

Variable / Performance Measure				
Panel A	Advanced Country Investor			
Lag Foreign	6.8906***	1.8776***	1.1033***	1.6211***
	(1.1194)	(0.4659)	(0.4058)	(0.4493)
No. of Observations	33407	33407	27486	26430
R-Squared	0.3574	0.8810	0.9069	0.9562
Panel B	Developing Country Investor			
Lag Foreign	5.5620**	0.3043	0.2329	1.7513
	(2.5254)	(0.9376)	(0.9188)	(1.3676)
No. of Observations	31833	31833	26104	25780
R-Squared	0.3436	0.8865	0.9072	0.9216
Industry FEs	Yes			
Time FEs	Yes	Yes	Yes	Yes
Firm FEs		Yes	Yes	Yes
Selection Controls			Yes	
Propensity Score Weights				Yes

The dependent variable is the number of export destination countries for a firm in year t. Industry fixed effects were inserted at the 2-digit industry classification level where applicable, while selection controls include covariates reported in the firm selection decision and propensity score estimation specifications, lagged one year relative to the foreign investment decision. Only firms that exported to at least one country in a given year were included in the regressions. Advanced country investor is defined as foreign investment originating from high-income OECD countries. These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States. Developing country investor is defined as investment originating from a complement of the set of advanced investor countries, except countries that are offshore tax haven countries, which are excluded from this part of the empirical analysis. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm.

* indicates 10% significance; ** 5% significance; *** 1% significance

Table 12: Foreign Investment and Average Export Price, Difference-in-Difference Matching Estimator

Panel A	Combined FDI			
Measure / Time Lag	Foreign+1	Foreign +2	Foreign +3	Foreign +4
Ln	-0.2541	-0.1649	-0.2010	-0.2584
Average Export Price	(0.0998)	(0.1847)	(0.2050)	(0.2246)
Panel B	Developed Country Investor			
Measure / Time Lag	Foreign+1	Foreign +2	Foreign +3	Foreign +4
Ln	-0.2580	-0.2605	-0.3466	-0.2816
Average Export Price	(0.1345)	(0.2016)	(0.2779)	(0.2988)
Panel C	Developing Country Investor			
Measure / Time Lag	Foreign+1	Foreign +2	Foreign +3	Foreign +4
Ln	-0.47961	0.2460	0.4311	0.0731
Average Export Price	(0.3577)	(0.5782)	(0.8664)	(0.7230)

This table documents difference-in-difference matching estimates for the post-acquisition export price dynamics between firms who received foreign investment and "matched" firms who stayed domestically owned. "Foreign" denotes the foreign investment year. Average export price is weighted average of deflated value of exported products price per kilogram by a firm in given year. Advanced country investor is defined as foreign investment originating from high-income OECD countries. These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States. Developing country investor is defined as investment originating from the complement of the above list of countries, except countries that are offshore tax haven countries, which are excluded from this part of the empirical analysis. Kernel matching with a bandwith of 0.005 and caliper of 0.005 was employed. Post-matching balancing tests reveal covariate balance in the treated and matched control groups. Reported are bootstrapped standard errors. Reported are bootstrapped standard errors.

Table 13: Foreign Investment and Price of New Versus Continuing Products, Linear Regression Approach

Variable			
Subset	*Total Sample*	*Developed Country*	*Developing Country*
Panel A	Difference in Average Price of Whole Vs. Continuing Product Mix		
Lag Foreign	1.4114*	1.5578*	0.5353**
	(0.7869)	(0.9179)	(0.2052)
No. of Observations	1918	1650	268
R-Squared	0.6655	0.6657	0.1493
Panel B	Difference in Share of Exports to OECD of Whole Vs. Continuing Product Mix		
Lag Foreign	0.0052	0.0002	0.0349
	(0.0058)	(0.0054)	(0.0239)
No. of Observations	1918	1650	268
R-Squared	0.3078	0.3361	0.1953
Firm FEs	Yes	Yes	Yes

The dependent variable in Panel A is % difference in the average price of the entire product mix a firm is exporting in year t and the average price of the "continuing" product mix a firm is exporting in year t. "Continuing" product mix is defined as the set of products a firm was exporting before receiving FDI. The dependent variable in Panel A is 0 before a firm receives FDI and can then deviate from 0 after FDI was received, provided the firm exports new products, and that the average price of these products differ from that of the "continuing" products. The dependent variable in Panel B is the difference in the share of the entire product mix a firm is exporting to the OECD in year t and the share of the "continuing" product mix a firm is exporting in year t. Estimations include only firms that received FDI. Advanced country investor is defined as foreign investment originating from high-income OECD countries. These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States. Developing country investor is defined as investment originating from the complement of the above list of countries, except countries that are offshore tax haven countries, which are excluded from this part of the empirical analysis. Yearly time effects and firm fixed effects were included in the estimations. "Lag Foreign" is an indicator that foreign investment was received in previous year, whereas All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 14: Foreign Investment and Imports of Capital Goods by Target Firms, Linear Regression Approach

Variable						
Subset	*Total FDI*		*Developed Country*		*Developing Country*	
Panel A	Imports of Capital Goods					
Lag Foreign	0.2414**	0.2605**	0.3336***	0.3764***	-0.3173	-0.3924
	(0.1135)	(0.1198)	(0.1188)	(0.1249)	(0.3341)	(0.3446)
Lag Foreign _2	-0.1502			-0.1681		-0.0509
	(0.1198)			(0.1015)		(0.3446)
No. of Observations	2169	1999	1870	1722	299	277
R-Squared	0.7346	0.7499	0.7284	0.7466	0.7671	0.7688
Panel B	Imports of Capital Goods from OECD					
Lag Foreign	0.1707	0.2316*	0.2427**	0.2844**	-0.3303	-0.0798
	(0.1164)	(0.1004)	(0.1225)	(0.1326)	(0.3457)	(0.3500)
Lag Foreign _2		-0.2144**		-0.1888*		-0.4182
		(0.1004)		(0.0993)		(0.4158)
No. of Observations	2119	1954	1845	1698	274	256
R-Squared	0.7202	0.7367	0.7144	0.7339	0.7498	0.7497
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variable in Panel A is natural log of imports of capital goods by a firm in year t, while the dependent variable in Panel B is natural log of imports of capital goods originating in OECD countries by a firm in year t. Capital goods are defined using the Slovenian vintages of the Combined Nomenclature at the 4-digit level (codes 8201-9033) Estimations include only firms that received FDI. Advanced country investor is defined as foreign investment originating from high-income OECD countries. These include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and United States. Developing country investor is defined as investment originating from the complement of the above list of countries, except countries that are offshore tax haven countries, which are excluded from this part of the empirical analysis. Yearly time effects and firm fixed effects were included in the estimations. “Lag Foreign” is an indicator that foreign investment was received in previous year, whereas “Lag Foreign _2 is lagged two years. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 15: Foreign Investment and Geography of Export Destinations, Linear Regression Approach

Variable						
Subset	*Total FDI*		*Developed Country*		*Developing Country*	
Panel A	Total Exports					
Lag Foreign	0.4094***	0.2213*	0.4412***	0.2682*	0.3613*	0.1122
	(0.1032)	(0.1224)	(0.1137)	(0.1409)	(0.2118)	(0.2291)
No. of Observations	33729	26592	33406	26430	31832	25780
R-Squared	0.8108	0.9370	0.8085	0.9312	0.8023	0.8646
Panel B	Exports to “High-Income” OECD					
Lag Foreign	0.1860*	-0.0259	0.2225*	-0.0381	0.0365	-0.0887
	(0.1081)	(0.1539)	(0.1184)	(0.1733)	(0.2636)	(0.3935)
No. of Observations	21878	17325	21643	17188	20205	16572
R-Squared	0.8147	0.9404	0.8127	0.9384	0.8071	0.8470
Panel C	Exports to Non- “High-Income” OECD					
Lag Foreign	0.4296***	0.3176***	0.4473***	0.3542***	0.3478*	0.2024
	(0.1057)	(0.1061)	(0.1162)	(0.1195)	(0.2059)	(0.1810)
No. of Observations	27206	21786	26893	21631	25562	21070
R-Squared	0.7826	0.9225	0.7796	0.9150	0.7761	0.8433
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Prop. Score Weights		Yes		Yes		Yes

The dependent variable in Panel A is the natural log of total exports for a firm in year t, the dependent variable in Panel B is the natural log of total exports to “high-income” OECD-member export destinations for a firm in year t, while the dependent variable in Panel C is the natural log of total exports to non “high-income” OECD-member export destinations. High-income OECD member destinations are defined as current OECD member states minus Chile, Czech Republic, Estonia, Greece, Hungary, Mexico, Poland, Portugal, Slovakia, Slovenia, and Turkey. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance

Table 16: Foreign Investment and Geography of non-OECD Exports, Linear Regression Approach

Variable						
Subset	*Total FDI*		*Developed Country*		*Developing Country*	
Panel A	Exports to ex-Yugoslavia					
Lag Foreign	0.2645***	0.2579***	0.2732**	0.2480**	0.1769	0.2884
	(0.1000)	(0.0977)	(0.1073)	(0.1057)	(0.2281)	(0.2163)
No. of Observations	25340	20398	25037	20244	23799	19721
R-Squared	0.7739	0.9212	0.7721	0.9142	0.7691	0.8355
Panel B	Exports to “post-Communist” Eastern Europe					
Lag Foreign	0.3012**	0.1678	0.3072**	0.2481*	0.2443	-0.0347
	(0.1211)	(0.1302)	(0.1346)	(0.1426)	(0.2031)	(0.2230)
No. of Observations	9807	7853	9618	7752	8726	7356
R-Squared	0.7510	0.9281	0.7445	0.9241	0.7473	0.8053
Panel C	Exports to Other Non- “High-Income” OECD Markets					
Lag Foreign	0.6021	0.6194	0.7682*	0.5512	-0.5257	0.3209
	(0.3947)	(0.4614)	(0.4452)	(0.5221)	(0.5027)	(0.6598)
No. of Observations	9200	7481	9033	7392	8181	7043
R-Squared	0.6513	0.8731	0.6516	0.8707	0.6591	0.7243
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Prop. Score Weights		Yes		Yes		Yes

The dependent variable in Panel A is the natural log of total exports to ex-Yugoslavia for a firm in year t, the dependent variable in Panel B is the natural log of total exports to “post-Communist” Eastern Europe export destinations for a firm in year t, while the dependent variable in Panel C is the natural log of total exports to non “high-income” OECD-member, non ex-Yugoslavia, non “post-Communist” Eastern Europe export destinations. Ex-Yugoslavia member destinations are defined as Croatia, Bosnia and Herzegovina, Serbia, Montenegro, and FYR Macedonia. Post-Communist Eastern Europe member destinations are defined as Albania, Belarus, Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Moldova, Poland, Romania, Russia, Slovakia, and Ukraine. All reported standard errors are calculated using heterogeneity-robust estimators and are clustered at the level of the firm. * indicates 10% significance; ** 5% significance; *** 1% significance