Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 788

December 2003

How Does the Border Affect Productivity?
Evidence from American and Canadian Manufacturing Industries

Robert J. Vigfusson

How Does the Border Affect Productivity? Evidence from American and Canadian Manufacturing Industries*

Robert Vigfusson

December 8, 2003

Abstract

This paper studies how much of productivity fluctuations are industry specific versus how much are country specific. Using data on manufacturing industries in Canada and the United States, the paper shows that the correlation between cross-border pairings of the same industry are more often highly correlated than previously thought. In addition, the paper confirms earlier findings that the similarity of input use can help describe the comovement of productivity fluctuations across industries.

Keywords: productivity, comovement, border effects
JEL D24 F4

1 Introduction

A national border often appears to reduce the comovement between economic variables that are on opposite sides of the border. Knowing whether this border effect extends to productivity fluctuations would be of interest for a number of reasons. For example, evidence either for or against a border effect would help us understand the true nature of these productivity shocks. Evidence that productivity shocks are industry-specific shocks that affect industries equally on either side of a national border would be evidence in favor of interpreting productivity shocks as being caused by changes in technology. Such a finding would, therefore, let us take seriously the notion of using these industry-level measures to examine technology-driven explanations of economic fluctuations (as in Basu Fernald and Kimball 1999). Evidence that productivity shocks do not depend on industry characteristics but rather are country-specific phenomena would be evidence against technology-driven models (Stadler 1994). Evidence of a large country-specific factor would emphasize the need to understand further the role of government, culture and other country-specific factors in determining industry-level productivity.

The current paper studies productivity growth of manufacturing industries in Canada and the United States. Because these two countries are so similar, one might expect to find no border effect.[1] Studies of both trade and prices, however, find found such border effects. McCallum

*Thanks for comments from David Bowman, Tim Conley, Brian Doyle, Joe Gagnon, Maral Kichian, John Rogers, and Jonathan Wright. Thanks also to the Bureau of Labor Statistics and Statistics Canada for their assistance with data. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any person associated with the Federal Reserve System.

[1] Although the countries are similar, Baldwin and Sabourin (1998) report survey evidence that Canadian and American firms differ on their views on technology adoption.

(1995) and Helliwell(1996) find border effects for trade. Recent work by Anderson and Van Wincoop (2003) suggests that the border effect is smaller; but that the border still reduces trade by 44 percent. For prices, Engel and Rogers (1996) find a border effect for Canada and the United States. Furthermore, Costello (1993) studies the cross-country industry-level productivity correlations for some OECD countries. In her study, none of the correlations between Canadian and American versions of the same industry are statistically significant. Because several pairs of American industries and several pairs of Canadian industries had significant correlations, Costello's finding of no significant cross border correlations can be interpreted as evidence of a considerable border effect.[2]

In this paper, productivity is measured using the KLEMS data sets from Statistic Canada and the U.S. Bureau of Labor Statistics.[3] These data sets are what the respective statistical agencies use to measure multifactor productivity in the manufacturing industries. The acronym KLEMS stands for capital, labor, energy, materials, and business services. For each industry, the data set reports both an index of the amount used (the volume) and also the nominal expenditure (the value) on any given input. The data set also reports the value and volume of each industry's gross output.

Productivity studies often use data sets that lack information on intermediate inputs. Because productivity is measured as a residual, any omitted variables can potentially result in an inaccurate measure. In particular, Basu and Fernald (1995) argue that measuring productivity using data only on capital and labor can result in an inaccurate measure that over states the degree of cross-industry comovement.

An important contribution of this paper is that it constructs productivity using information on intermediate inputs, in contrast to other work, such as Costello (1993).[4] In addition, because there are more industries in the KLEMS dataset, the current paper reports on correlations on 10 additional industries beyond the 5 studied in Costello. The increased coverage of the KLEMS dataset results in finding more positive correlations than is found for a dataset restricted to Costello's original five.

Another advantage of calculating productivity with data on intermediate inputs is that it allows a reassessment of Conley and Dupor (2003). They study the industry-level comovement of productivity growth rates in the United States. They construct their productivity series using information on the changes of output, labor, and capital services (measured using electricity usage). Because they study data at the quarterly frequency, they however do not have data

[2] The current paper focuses on exogenous fluctuations in technology and hence is of most relevance to macroeconomics. Depending on the difference between exogenous and endogenous fluctuations, this paper's results may have implications for trade theory. One might hold the view that the channels through which endogenous improvements in an industry's productivity transfers to other industries are the same as for exogenous improvements. If there is a large border effect and hence knowledge spill-overs are country specific then comparative advantage in research and development can become endogenous (Grossman and Helpman 1990). Branstetter (2001) uses firm level data on patents to examine spill-overs in knowledge between U.S. and Japanese firms. He found evidence that the knowledge spillovers are more intra-national than international in scope. Related to the current paper, he too uses a measure of technological proximity to weight how much other firms' research spills over to an individual firm. Finding an important role for within country spill-over effects has some important policy implications. In particular, if there are within country spill-overs, a country might want to subsidize research to gain comparative advantage

[3] One desirable extension of this research would be to compare industry-level productivity fluctuations across regions in the United States with each other and Canadian industries. Unfortunately, the data to do such a comparison are not available.

[4] Other papers, such as Basu Fernald and Kimball (1999), that have constructed productivity measures using intermediate inputs have not looked for border effects.

on quarterly changes in intermediate inputs. Conley and Dupor's finding of a high degree of comovement could be subject to the criticisms of Basu and Fernald (1995). In particular, Basu and Fernald report that the earlier comovement results of Caballero and Lyons (1992) were highly dependent on constructing productivity without using data on material usage.

Conley and Dupor's comovement results are important because they provide an organizing principle for the observed empirical comovement patterns. In particular, Conley and Dupor provide evidence that comovement between industries depends on the similarity of the inputs that each industry uses. (While they do not have quarterly data on intermediate input use, Conley and Dupor can classify industries by their input use from the input-output tables. These benchmark input-output tables, however, are available only every five years.) Comovement has often been viewed as a defining characteristic of business cycles. Insights into comovement, therefore, should lead to a better understanding of business cycles. Confirmation of their results would emphasize the need for further investigation into how input usage leads to comovement.

The empirical evidence presented here confirms Conley and Dupor's results in two crucial ways. First, constructing the productivity series using data on material usage does not change qualitatively the observed comovement patterns in the United States. Second, the dependence on input similarity is not restricted to the United States. In Canada, comovement also depends on input similarity.

The rest of the paper has the following structure. Section 2 lays out the modelling framework. Section 3 describes the data and some of the measurement issues in constructing productivity series. Section 4 reports the empirical work. The section first reports simple correlations between industry pairs to provide evidence against earlier claims of a large border effect and then reports on the correlations between all industry pairs. As in Conley and Dupor (2001), using information on input usage can concisely summarize the comovement patterns found in the data. Here also, the border effect is found to be small. In particular, variation in input usage has a larger effect on correlations than does being separated by a border. Section 5 offers conclusions and provides suggestions for future research.

2 Modelling Framework

In studying the covariance between productivity growth rates, this paper examines the connection between the similarity of input use and the covariance between industries. Conley and Dupor present an econometric model where the covariance between industries is a function of the similarity of input usage. However, they do not explicitly model why input usage might matter.[5] Although not a full structural model for estimation, the following discussion should provide some intuition.

2.1 Economic Modelling

To provide some intuition for why input usage might matter, consider the following production function.[6] Industry j produces output Y_j by combining inputs x_{ji} in the following production

[5] The closest model is that of Dupor (1996) where productivity comovement should depend on where industries sell their output. The evidence in Conley and Dupor, however, is more supportive of productivity comovement depending on where industries buy their inputs.

[6] This version of the model does not take into account the international aspect of productivity comparisons. Appendix B presents a version of the model with two countries.

function

$$Y_j = \theta_A \theta_j \prod_{i=1}^{N} (x_{ji} v_i)^{\gamma_{ji}}$$

where θ_A is an aggregate productivity term, θ_j is an industry-specific productivity term, and v_i is an input-specific productivity term. For intuition, the reader could think of θ_A as arising from gains due to national infrastructure, θ_j from industry-specific production improvements that are not readily transferable to other industries, such as for the steel industry finding an improved way to make steel, and v_i a general technology that makes better use of a particular input i. Examples of input-specific improvements include: vintage capital, human capital[7], labor-augmenting technological progress, or even something as simple as fewer defects per box of metal fasteners.[8] The log of each of these productivity terms is assumed to be a random walk with the disturbances having a constant variance and being independent of each other. Suppressing time subscripts for notational simplicity, these assumptions can be expressed as:

$$\begin{aligned} \Delta \ln v_i &= u_i \\ \Delta \ln \theta_j &= \varepsilon_j \\ \Delta \ln \theta_A &= \varepsilon_A \end{aligned}$$

where $[u_i, \varepsilon_j, \varepsilon_A]$ are independent both across time and each other with variances $\left\{\sigma_{vi}^2, \sigma_{\theta j}^2, \sigma_A^2\right\}$. The change in the log of industry-level productivity Δz_j would be the following

$$\Delta z_j = \Delta \ln \theta_A + \Delta \ln \theta_j + \sum \gamma_{ji} \Delta \ln v_i$$

Hence the covariance between any two industries can be written as follows.

$$\begin{aligned} E\Delta z_j \Delta z_k &= E\left(\Delta \ln \theta_A\right)^2 + E\left(\Delta \ln \theta_j \Delta \ln \theta_k\right)^2 + \sum \left(\gamma_{ji}\gamma_{ki}\right) E\left(\Delta \ln v_i\right)^2 \\ &= \sigma_A^2 + \sum \left(\gamma_{ji}\gamma_{ki}\right) \sigma_{vi}^2 \quad \text{if } j \neq k \\ &= \sigma_A^2 + \sigma_{\theta j}^2 + \sum \left(\gamma_{ji}\right)^2 \sigma_{vi}^2 \text{ if } j = k \end{aligned}$$

Assuming constant-returns-to-scale and industry-level cost minimization, the above Cobb-Douglas production function implies that γ_{ji} equals the industry's expenditure share s_{ji}. With a further assumption that all of the individual input-productivity terms (σ_{vi}^2) have the same variance (σ_v^2), the resulting covariance function would be the following.

$$\begin{aligned} E\Delta z_j \Delta z_k &= \sigma_A^2 + \sigma_v^2 \sum s_{ji} s_{ki} \quad \text{if } j \neq k \\ &= \sigma_A^2 + \sigma_{\theta j}^2 + \sigma_v^2 \sum \left(s_{ji}\right)^2 \text{ if } j = k \end{aligned}$$

Therefore, the covariance between two industries is larger for those industries that have more similar input shares, because more similar shares implies a larger value for the product of input shares $\sum s_{ji} s_{ki}$.

[7]Although the BLS does account for changes in skill composition for private (nonfarm) business, the BLS does not correct for changes in skill composition at the industry-level. (BLS Handbook of Methods, 1997).

[8]An industry that reduces the fraction of defective output would likely have a measured increase in productivity only if the defective output had not been previously counted as output. Furthermore, if improved inspections reduced the amount of output sold but not produced, then measured productivity would actually fall. In such a case, a lower rate of defective inputs would then show up as an improvement in productivity for the using industry.

2.2 Econometric Spatial Model

Conley and Dupor estimate a similar but more general function for covariance between industries.[9] Their assumption is that the covariance between any two industries can be written as a function of input shares,

$$E\Delta z_j \Delta z_k = c\left(\sqrt{\sum (s_{ji} - s_{ki})^2}\right)$$

where $c()$ is a smooth function. The estimator becomes a 'spatial estimator' by viewing the similarity of input usage as a measure of economic distance, where distance is defined as:

$$d_{jk} = \sqrt{\sum (s_{ji} - s_{ki})^2}$$

This specification implies that each industry has a location that is described by the vector s_j that describes its input shares $\{s_{ji}\}_{i=1}^{N}$. The covariance between two industries then depends on the Euclidian distance d_{jk} between these industries' locations. Note, that this distance is an economic distance not a geographic distance.

Although Conley and Dupor estimated $c()$ using non-parametric kernel regression, I will estimate $c()$ using a version of the sieve estimator describe in Chen and Conley (2001). The function form of $c()$ is the following

$$c(d) = \sum b_k H_k(d)$$

where H_k are the set of functions, described in the Appendix, that approximates the covariance function. With the restrictions that b_1 is positive and $b_{k+1} > b_k$, the sieve estimator has the advantage of guaranteeing a valid covariance matrix, which is important for the hypothesis testing reported here.

3 Discussion of the Data

Productivity measures how much output can be produced for a given amount of inputs. The standard approach is to linearize a production function and thereby express the growth in productivity dz as the growth in output dy minus the growth in inputs. Several measures of the inputs are considered here. Beginning with the most general, output Y is produced by function that has as inputs: technology z, capital services k, labor services l, and intermediate inputs (energy e, materials m, and services s).

$$Y = F(K, L, E, M, S, Z)$$

Linearizing this production function results in

$$dz = dy - (\gamma_k dk + \gamma_l dl + \gamma_e de + \gamma_m dm + \gamma_s ds) \tag{1}$$

The output elasticities γ are calculated in several different ways. The empirical part of this paper reports results where the values of γ are estimated by an instrumental variables regression. An

[9] Appendix B reports on the similarities and differences between the production function described here and Conley and Dupor's covariance function. In particular, the Conley and Dupor econometric model is estimated on the simulated data from the economic model described above. While the two models are not identical, the econometric model does approximate the simulated covariances.

alternative approach is that taken by Statistics Canada and the Bureau of Labor Statistics to calculate their productivity measure, multi-factor productivity (MFP). Under their assumptions of constant returns to scale, the values of γ_j are the nominal expenditure on an input j as a share of the total nominal value of gross output.

$$dz^* = dy - (s_k dk + s_l dl + s_e de + s_m dm + s_s ds) \tag{2}$$

More restricted measures are often constructed due to data limitations. For many countries, the available data sets, (such as the OECD STAN data set) do not have data on intermediate input use. In the United States, no statistical agency publishes data on intermediate input use at higher frequencies than annual. Therefore, to do higher frequency work, one has to follow Conley and Dupor and estimate the following equation that measures productivity using gross output as a function of just capital and labor services.

$$dz^y = dy - (\gamma_k dk + \gamma_l dl) \tag{3}$$

Rather than measuring output as the total value of production, one could measure output using value added (for example, industry-level GDP). Productivity measured in value-added terms can be written as.

$$dz^v = dv - (\gamma_k^v dk + \gamma_l^v dl) \tag{4}$$

where dv is the growth rate of value added. In both of these cases, the values of γ will be estimated. An alternative approach is to ignore the data on capital usage and focus entirely the amount of gross output produced using labor. Labor productivity is defined as

$$dz^l = dy - dl \tag{5}$$

The data set studied here is the KLEMS data set for Canada and the United States for 1960-1997. Using this data set rather than other options has several advantages. First, for any given industry, the statistical agencies construct output and inputs with a goal of being consistent. Other studies that have examined productivity often combine output data reported by one statistical agency with input data reported by another. For example, industrial production from the Federal Reserve is often matched with hours data from the BLS Current Employment Statistics survey.[10] As such, one should be concerned that the firms used in measuring output are not the same firms as those used in measuring inputs. An additional advantage of using the official KLEMS data sets from the two countries is that others have already studied the comparability of the two data sets. Eldridge and Sherwood (2001) found that the differences for the two data sets are minor and do not contribute substantially to any differences in average productivity growth.

The industries studied here are the two-digit Standard Industrial Classification (SIC) manufacturing industries. These industries can be divided into two groups: the durable goods industries and the nondurable goods industries. Table 1 lists the industries studied in this paper and their corresponding U.S. 1987 SIC code. The Canadian industries were mapped into their U.S. counterpart.

[10]Choosing to combine different data sources is one of the many trade-offs a researcher must make. For example, choosing to use the KLEMS dataset implies that the researcher also chooses to work with annual data.

3.1 The Effect of Missing Data on Materials Inputs

Materials usage data are often not available. Because productivity growth is measured as the growth of output minus the growth of inputs, missing inputs should therefore lead to a contaminated productivity series. Two different strategies can attempt to deal with this problem. First, production functions can be assumed to be separable: observed inputs such as labor and capital move together with the missing inputs. The problem with the separability assumption is that counter examples are easy to find. For example, suppose that the workers vary their effort. When they have work to do, they work and when there is nothing to do, they still are counted as working. In that case, to produce more output, a firm may increase the amount of intermediate inputs without changing the amount of employed labor. Hence, the production function is not separable. The other strategy is to use output that has be constructed with the contribution of intermediate inputs removed, value-added data.

Basu and Fernald (1995), however, showed that these intermediate inputs can still contaminate productivity measures constructed using value-added data. There are several different ways to measure value added using data on intermediate inputs and gross output. One approach measures value added as the growth rate of output minus the expenditure share weighted growth rate of intermediate inputs. Basu and Fernald discuss other approaches and show that these other methods have more biases. Abstracting away from energy and services, one can express valued added as follows. Assuming constant returns to scale, real gross output can be written as a combination of materials and value added:

$$dy = (1 - s_m)\, dv + s_m dm$$

where s_m denotes the expenditure on materials as a fraction of total input expenditure. This equation can be rewritten as an expression for value added

$$dv = \frac{1}{1 - s_m}\,(dy - s_m dm)$$

This equation is sensible with perfect competition since it implies that if gross output and materials both increase by one percent, value added will also increase by the same amount.

Notice however the implication for the measure of productivity when the elasticity of output with respect to materials γ_m differs from the expenditure share, such as when the industry has increasing returns to scale.

$$\begin{aligned} dv &= \frac{1}{1 - s_m}\,((\gamma_k dk + \gamma_l dl + \ \gamma_m dm\ ) - s_m dm) + \frac{1}{1 - s_m} dz \\ &= \frac{\gamma_k}{1 - s_m} dk + \frac{\gamma_l}{1 - s_m} dl \ + \ \frac{\gamma_m - s_m}{1 - s_m} dm + \frac{1}{1 - s_m} dz \end{aligned}$$

But in estimating this equation, one will actually estimate the following equation

$$dv = \gamma_k^v dk + \gamma_l^v dl \ + dz^v \tag{6}$$

where

$$dz^v = \frac{\gamma_m - s_m}{1 - s_m} dm + \frac{1}{1 - s_m} dz$$

Hence as long as γ_m does not equal the expenditure share s_m, then the technology measure will be contaminated with material usage data. Material usage is likely correlated across industries within a country because of demand linkages and aggregate shocks such as monetary policy. Therefore one might expect to find larger border effects when examining productivity calculated using value-added data than using gross output data.[11]

3.2 Input Output Tables

To construct distances between sectors requires a measure of industry similarity. One source of such data is the Input-Output table that describes how much of each input an industry uses in production. The inputs reported are materials and labor. Under the assumption that industries that use similar inputs are more similar than industries that do not, the information on inputs is used to calculate the economic distance between sectors.

In order to reduce the dimension of the space, the Input-Output table is condensed into a table with 18 rows: fifteen rows for the manufacturing industries studied here, one row for petroleum, one row for labor, and one for all other industries. Likewise the aggregated table has 15 columns representing each industry's use of the inputs. Each element of this table is divided by its associated column sum to make a matrix S of input shares. The element s_{ij} denotes the share of sector j's inputs that come from sector i. The distance between sectors j and k, d_{jk} is defined as the Euclidean distance between the two sectors input shares.

$$d_{jk} = \left(\sum (s_{ij} - s_{ik})^2\right)^{1/2}$$

The input-output matrices have several important properties. First the matrices are sparse. At the two digit level, most of the entries are near zero.[12] A subsection of the input output table is reported in Table 2. For most industries, one of the largest sources of inputs is itself. The Chemicals industry is one of the few industries that is an input into all of the others. Also the durable goods industries that are between SIC 33 and SIC 37 are the most interconnected manufacturing industries.

3.2.1 Distribution of Distances

Figure 1 reports the distribution of distances. The distribution has a mass at zero because the distance between a sector and itself is zero. About eighty percent of the distribution is located between 0.25 and 0.6. The number of zero distances is smaller for the cross-border distances because the Canadian and U.S. sectors have different input requirements.

3.3 Capital Services and Utilization

The discussion of measuring productivity referred to the inputs as capital services and labor services rather than capital and labor. This distinction is to emphasize the need to differentiate between changes in output due to changes in inputs versus changes in outputs due to better use of the same inputs. For example, a fixed amount of capital (a particular machine) can be run at a greater utilization rate and thereby provide additional capital services. Therefore, in this paper, the estimation procedure will not measure capital services using the published series for

[11] Estimating the coefficients in equation (6) may mitigate the problem of using value added data if the instruments are correlated with the omitted variable (dm). With the instruments correlated with material usage, the estimated coefficients should be biased in such a way to reduce the amount of material usage in the residual.

[12] The sparseness of these matrices has implications for economic modelling. As noted in Horvath (2000), sparseness postpones the law of large numbers result that idiosyncratic sector-specific shocks would nullify each other in aggregate output.

capital.[13] Rather, as in Burnside Eichenbaum and Rebelo (1993), Costello (1993), and Conley and Dupor (2003), changes in capital services are approximated by changes in energy usage.[14] Dropping capital may be controversial. As such, results are also reported for the published multifactor productivity series that includes the measured capital stock as an input. As will be clear from the results reported in the following sections, the paper's conclusions are not overly sensitive to the treatment of capital.

4 Empirical Results

The empirical results are presented in four parts. The first part describes the estimated coefficients of the production function that are used to construct the estimated productivity series. The second part reports pair-wise correlations for both the estimated productivity series and also the published labor and multifactor productivity series. The pair-wise correlations make clear that the border effect is smaller than had been previously reported. The third part characterizes the dependence of the observed productivity comovement on the similarity of input use. The final part presents results regarding weak instruments and hypothesis testing.

Two different approaches are used to measures the dependence on input use. The first approach is a linear regression of the sample correlation between industries on the distance measure of input use similarity. The second approach uses the analogous covariance function of Conley and Dupor. Both approaches support the conclusion that the observed comovement is dependent on the similarity of input use. To strengthen these conclusions, several alternative hypotheses are tested. The hypothesis testing confirms that input use is a useful organizing tool. There, however, needs to be further refinement.

4.1 Estimation Based Productivity Series

To compute the correlations between American and Canadian productivity series, I need to estimate the most general production function and the resulting technology series. For each industry j the technology series is given by the residual from the following regression.

$$\triangle \ln y_j = \gamma_l \triangle \ln l_j + \gamma_e \triangle \ln e_j + \gamma_m \triangle \ln m_j + \gamma_s \triangle \ln_s + z_j \tag{7}$$

The technology shock is defined as z. The observable variables in the equation are output y, labor l, energy e, materials m, and services s.

The values of the observed inputs may depend on the value of the technology shock. Therefore, a consistent estimate of the equation 7 requires using instrumental variables. For both countries, the estimation process uses five instrumental variables to estimate the four coefficients. For the American industries, the instruments are current and lagged values of the changes in real U.S. defense expenditure, and the changes in the IMF spot oil price, and the current value of a monetary policy shock taken from the VAR estimated in Christiano Eichen-

[13]Estimating the production function with the reported change in Δk results in the estimate of the coefficient on Δk being strongly negative with an average value of -0.3. This may reflect some kind of adjustment cost.

[14]Basu and Fernald (2000) discuss other approaches that one can use to approximate for changes in utilization. Their preferred approach, however, requires data on average hours worked per employee which are not available in the datasets used here.

baum and Evans (1999).[15] For Canada, the instruments include current and lagged changes in the oil price and the U.S. monetary policy shock. An additional instrument is the current change of non-oil commodity prices. That the Canadian economy responds to commodity price fluctuations is well established. For example, Amano and van Norden (1995) report on the strong relationship between the Canadian exchange rate and commodity prices. Since the technology shocks may be correlated across industries, the equations for all of the industries in a particular country are estimated jointly using three-stage least squares (two-step GMM). The weighting matrix is constructed using the standard sample covariance matrix.

Table 3 reports the coefficient estimates. The coefficient estimates do vary between durables and nondurables and between Canada and the United States. First the nondurable sectors in the United States places almost all of its weight on labor and business services, the most labor-like of the intermediate inputs. The U.S. durable sector places much more weight on material usage.[16] The Canadian estimates are quite different. Energy and material usage are much more important. These factors are the ones most likely to correct for utilization. Standard asymptotic confidence intervals are also reported.[17]

4.1.1 Returns to Scale

Table 3 also reports a measure of the returns-to-scale, the sum of the estimates γ_i. The statistical agencies assume constant returns-to-scale when computing multi-factor productivity. The estimation approach used here allows for returns-to-scale different from one. For the United States, the nondurable industries have increasing returns to scale while the durable industries have decreasing returns to scale. Neither of these estimates is statistically significant.[18] In Canada, both industries have increasing returns to scale but only the estimate for the Non-durable industries is statistically significant.

4.2 Pair-wise Correlations

Table 4 reports the sample correlations between industries pairs across borders and also for different industries within a country. This exercise is analogous to Costello (1993).[19] For Canada and the United States, Costello did not find any statistically significant cross-border own-industry correlations. There, however, were a number of significant correlations for industries that were located in the same country.

These correlations are reported for several different productivity measures. In all cases, the measures are constructed using the BLS and Statistics Canada KLEMS data sets. The first

[15]In the Christiano Eichenbaum and Evans paper, the sample period is only from 1964 to 1995. I extend the sample back to 1959 and forward to 1997 in order to match the span of the KLEMS data. The shocks do result in very similar impulse responses. Due to data revisions, the shocks identified here, however, are not identical to the shocks identified in the CEE paper. The correlation between their monetary policy shocks and the ones used here is 0.62.

[16]A large weight on materials is compatible with Basu's (1993) use of materials to approximate for changes in utilization.

[17]Section 5 discusses the possibility of weak instruments and how it might affect inference concerning the size of the border effect.

[18]For the United States, economists have found different estimates of the relative size of the returns-to-scale in the durable and non-durable good producing industries. Basu Fernald and Kimball (1999) found that nondurables have lower returns to scale than durables, but Conley and Dupor found the opposite result. Given the wide confidence intervals of the two estimates, the results reported here are insufficient evidence to decide the issue.

[19]While the sample period is longer than Costello's, using the shorter sample is not responsible for the results.

column reports the correlation between industries using labor productivity. The second column reports calculations using the multi-factor productivity measures calculated by the BLS and Statistics Canada. The third column reports the results from the estimated productivity series.

Comparing the results for the MFP measure and the estimated measure suggests that the estimation procedure has removed a common term from all of Canada's measures. The average correlation between solely Canadian pairs has gone down dramatically. The average correlation between a Canadian and American pair has not seen the same reduction. In terms of a border effect, one could apply Costello's criterion of statistical significance. Out of the 15 industries studied here, labor productivity correlations are statistically significant for nine industries, multifactor correlations are significant for 12 industries, and the estimated productivity series has significant correlations for 7 industries.

4.2.1 Comparing Results With Costello

The large percentage of statistically significant industries suggests that the border effect is less important than Costello had claimed. A major reason behind the difference in results is the industries studied. The five industries that Costello studied are some of the least likely to have significant cross border correlations. For the results reported here, both Food and Textiles are never statistically significant. For all three measures, the Chemicals industry is statistically significant. Primary Metals and Fabricated Metals both have statistically significant correlations for the MFP measure. For the estimated productivity series, although the points estimates reported in Table 4 are above 0.25, the correlations are not statistically significant. Overall, these results suggest that Costello's conclusion of a large border effect was highly dependent on the industries that she studied. Studying a greater number of industries reverses her conclusion and allows us to conclude that the border effect is smaller than previously claimed.

4.2.2 Further Diagnostics

Table 5 reports further diagnostics of the productivity series. For each industry, the table reports the standard deviation of the MFP measure and the ratio of that measure to two others: the labor productivity series and the estimated productivity series. One of the standard reasons for mismeasured productivity growth is labor hoarding, where workers vary their effort over the business cycle. If labor hoarding were a serious issue at the annual frequency then the labor productivity series would have a much higher standard deviation than the MFP measure. If the estimation procedure actually corrects for the roles of market power and the changes in utilization, then the estimated productivity series should have a much lower standard deviation than the multifactor productivity series.[20]

As reported in Table 5, labor productivity often has a much larger standard deviation than does the MFP. A larger standard deviation is suggestive of a labor hoarding hypothesis. Labor hoarding is most likely in industries with highly skilled workers that are difficult to replace. In both countries, the two industries that had the largest ratios were Apparel [SIC 23] and Transportation Equipment [SIC 37]. That the industry that produces airplanes and

[20] The claim may be false if utilization changes and technology changes were negatively correlated. Many macroeconomic models would have a positive correlation. Of course, counter examples do exist. For example, suppose that an increase in utilization results in a greater depreciation rate for capital. In a model with investment adjustment costs, utilization and technology could then be negatively correlated. A reduction in utilization would be way to 'avoid' the investment adjustment costs and to preserve capital for the next period.

automobiles [SIC 37] is a high skill industry seems sensible. The high ratio for Apparel is perhaps more puzzling as it contradicts conventional stereotypes.

For Canada, the estimated productivity series has, on average, a much lower standard deviation than the MFP measure. The lower value further suggests that the estimation process did correct for the role of utilization. In particular, the standard deviation of durable good producing industries are much lower. It is commonly thought that durable goods industries would be the industries most likely to have large time-varying utilization rates.

4.3 Distance Based Correlation Functions

The following section reports two sets of calculations describing how industry comovement depends on the similarity of input usage. The first results are from OLS regressions on the sample correlations. The second set of results are from methods similar to those used in Conley and Dupor, where the correlation between industries is a function input-use similarity.

4.3.1 Linear Regressions

To understand the relationship between distance and the correlation between industries, I first estimate the following simple regression.

$$\rho\left(\Delta z_i, \Delta z_j\right) = \beta_0 + \beta_1 d_{ij} + \beta_2 d_{ij}^2 + u_{ij}$$

where $\rho\left(\Delta z_i, \Delta z_j\right)$ is the correlation between productivity growth in industry i and industry j. Table 6 also reports the R^2 of the regression for each possible set of pairings (i.e. a regression for all Canadian pairs, another for all American pairs, and another for cross-border pairs). The table also reports the F-test of the distance coefficients both equal to zero. For both countries, input similarity helps explain the pattern of productivity correlations. In particular, the F-test reject the hypothesis that input similarity does not help. As can be seen by the R^2 statistics, even though the input-usage information is useful, there remains a great deal of unexplained variation. Whether the unexplained variation comes from omitted explanatory variables or mismeasured productivity growth is an open question. From Figure 2, the decline in correlation as a function of distance is somewhat greater for the U.S. estimates than the Canadian estimates. The labor productivity results are flatter for the cross-border pairs. The evidence for cross border pairs depending on distance is strongest for the estimated productivity series.

To investigate further the cross-border effects, consider the following direct test of the border effect for all pairs both within each country and also across countries

$$\rho\left(\Delta z_i, \Delta z_j\right) = \beta_0 + \beta_1 d_{ij} + \beta_2 d_{ij}^2 + \phi_{cross}\delta + u_{ij}$$

where δ equals one if it is a cross border pairing and otherwise equals zero. The border effect measured as ϕ_{cross} is largest for the constructed productivity series. With the average correlation between U.S. pairs is 0.30, a decline of 0.06 is somewhat small. The result for the estimated productivity series is statistically significant. The values of ϕ_{cross} found for the other productivity measures are smaller and are not statistically significant. The size of these coefficients are similar to the results reported in the next section.

4.4 Explicit Distance Function Based Estimates

The correlation regressions reported in the previous section may be sufficient proof that (1) the border effect is relatively small and that (2) the correlation between industries is dependent on what inputs the industries use. One, however, may want to consider the more general distance based measure used in Conley and Dupor. One advantage of their approach is that these measures allow for changes in distances between industries over time. Over time, the subcomponents of the industries might change. These changes will be reflected in their input usage and cause changes in the distance measures. For example, SIC 35 includes the computer industry. As the importance of that industry grows relative to other industries that are also included in SIC 35, the types of inputs SIC 35 buys should change. They may use more Electrical Machinery SIC 36 and less Fabricated Metal SIC 34. Therefore, the distance from other sectors that continue to use a lot of metals would grow over time.

This estimation implements the semi-parametric methods of Chen and Conley in estimating the size of the border effect. Although their methods are proposed for a distance-dependent covariance matrix, the approach taken here will be to first normalize each growth rate by dividing by its sample standard deviation. The covariance matrix on the resulting normalized growth rates corresponds to the correlation matrix of growth rates. The advantage of this approach is that it abstracts away from heterogeneity in variances across industries and countries already described in Table 5. This abstraction is not because the heterogeneity is uninteresting but rather because the heterogeneity is of only secondary importance to the question of border effects.

Figure 3 plots the correlation function for all three measures. For each productivity measure, three estimates of $C(d)$ are reported. The thick solid line is the estimate of C(d) found by estimating the following equation

$$\Delta z_i \Delta z_j = \sum b_k H_k \left(d_{ij}\right)$$

for all pairings of i and j and with the restriction that $b_{k+1} > b_k$. The estimation is a straightforward application of the methods in Chen and Conley, with no allowance made for border effects. The sign restriction on b_k insures that the correlation matrix is positive definite, which is important for constructing the confidence intervals. The gray interval is the bootstrap confidence interval generated as in Chen and Conley. The covariance matrix Σ_t of the productivity shocks z_t is a time-varying matrix that depends on the time-varying distances between industries. As such, a vector of i.i.d errors u_t with variance one can be constructed as

$$u_t = \text{chol}\left(\Sigma_t\right)^{-1} z_t$$

where chol denotes the matrix operator that maps between a positive definite matrix Σ and its upper triangular Cholesky factorization. A simulation run is constructed by drawing with replacement from the set of the errors $\{u_s\}_{s=1}^{T}$ for each time period a value u_t^* and then multiplying it by the Cholesky factorization of the corresponding variance covariance matrix to get a simulated productivity series

$$z_t^* = \text{chol}\left(\Sigma_t\right) u_t^*$$

For each of five hundred simulation runs, the coefficient estimates of C(d) are stored. These simulations are then used to construct a ninety percent confidence interval for the covariance function.[21]

The second set of results allow for two border effects. The first controls for cross-border pairs of the same industry and the second controls for all cross-border pairs of different industries.

$$\Delta z_i \Delta z_j = \phi_1 \delta_1 + \phi_2 \delta_2 + \sum b_k H_k \left(d_{ij}\right)$$

where δ_1 equals one when i and j indicate the equivalent industries in the two different countries and δ_2 equals one when i and j indicates non-equivalent industries that are located in different countries. In addition, the restriction of the b_k's is not imposed because these off-diagonal constants imply that the resulting correlation matrix is not necessarily positive-definite. The thin line reports the value of $\sum b_k H_k (d)$ and the line with circles reports the value of $\phi_2 + \sum b_k H_k (d_{ij})$. Hence the difference between the two lines is the value of ϕ_2, the effect of the border.

The figure reports $C(d)$ over the entire range of observed distances. In addition, the dashed lines indicate the interval between 0.23 and 0.70 that contains 90 percent of the non-zero distances. The estimates of C(d) are all fairly steep. For example, within-country industries that use similar inputs (a distance of 0.25) have an estimated correlation for labor productivity of 0.36 that decline to 0.10 when the distance increase to 0.70.

Table 8 reports the cross-border coefficients and the associated sampling uncertainty. The first result is a test of the null hypothesis of no border effect between the same industry in the two countries. These results are generated by the same bootstrap simulations of the data generating process described above that made no allowance for border effects. As an additional check, a 90 percent confidence interval for the cross border coefficient is estimated by a alternative bootstrap where the vector of growth rates is sampled with replacement. This process is typically less efficient (Horowitz 2002) but is valid under the assumption of no serial correlation.

For both labor productivity and the estimated productivity series, one would fail to reject the null hypothesis that ϕ_1 equals zero at the 10 percent significance level. For the MFP measure, one would reject the null hypothesis that ϕ_1 equals zero at the 1 percent significance level. For all three productivity measures, one would reject the null hypothesis of no border effect for non-identical industries. Similar results are found using the confidence intervals generated by the alternative bootstrap procedure. One can conclude that the statistical evidence favors non-identical cross-border industries being less correlated than similar industries located in the same country. The evidence is weaker for cross-border pairs of industries in the same country. For these cross-border pairs of the same industry, stronger evidence may be found if we imposed the sign restrictions on b_k. Part of the relative fall in $C(d)$ for small distances may be attributed to controlling for cross-border pairs of the same industry. As mentioned in the discussion of Figure 1, the cross-border pairs are close together because they use very similar inputs.

Having established statistical significance, the next question concerns the border's economic significance. As mentioned above, the dependence on distance shows a fairly strong decline. For example for labor productivity, industries that use similar inputs (a distance of 0.23) have an estimated correlation 0.26 more than industries that are quite dissimilar (a distance of 0.70). Crossing the border results in a 0.14 decline in correlation. Hence, for labor productivity, cross-border pairs that use similar inputs are more correlated than within country pairs that

[21] The confidence interval is constructed as follows. First, generate 500 simulations of the statistic of interest x_i where x^* is the empirical estimate of the same statistic. Calculate the differences between the simulated statistics and the empirical statistic. Denote the 5^{th} percentile of these differences as ζ_{low} and the 95^{th} percentile as ζ_{high}. The resulting confidence interval is $[x^* - \zeta_{high}, x^* + \zeta_{low}]$

use dissimilar inputs. The border effect, however, is smaller for the multifactor and estimated productivity series. Therefore we can conclude that, although the border effect does still exist, it is much smaller than earlier evidence suggested.

5 Allowing for Weak Instruments

Conventional confidence intervals often over state the precision of instrumental variables estimates. This section describes how adopting the weak instruments approach affects inference of the importance of the border effect. As in Stock and Wright (2001), an S set describes all those parameters whose GMM objective function is less than the critical value from a chi-square distribution with degrees of freedom equal to the number of moment functions

$$S = \left\{ \gamma : J(\gamma) < \chi^2 \right\}$$

where $J(\gamma)$ is the continuous-updated version of the objective function[22] estimated in Section 4.1.[23] Each γ in S results in a productivity growth series for each industry $\Delta z_i(\gamma)$ that can then be used to estimate the border effect coefficients $\phi_1(\gamma)$ and $\phi_2(\gamma)$. One could therefore construct, confidence intervals for $\phi_1(\gamma)$ and $\phi_2(\gamma)$.[24] Theoretically, it is known that these confidence intervals will be conservative (in other words, wider than optimal). With degrees of freedom equal to 180 and a resulting critical value of 237, a 90 percent confidence interval for the size of the border effect is quite large. For ϕ_1 the interval is from -1.13 to 0.10. For ϕ_2 the interval is from -0.97 to 0.04. Therefore, one can conclude that the instruments used here are not strong enough to impose a tight confidence interval on these border effects.[25]

Although the wide interval suggests a problem with the estimated technology series, the weak instrument criticisms are not applicable to the results for multi-factor productivity or labor productivity. Because these results confirm each other, overall conclusions are more robust.

5.1 Further Hypothesis Testing

One might be interested in knowing how well the estimated model matches the unconditional covariance matrix. Based on a classical likelihood ratio statistic, the evidence suggests that unexplained influences on the comovement remain. Overall, however, the input-based measure of similarity appears to be a useful organizing principle.

If the distance matrix is fixed at a constant value, then the distance based covariance matrix could be nested by the sample covariance matrix.[26] Furthermore, assuming that the

[22]For ease of computation, I modify the objective function in two respects. First, I allow for covariance between Canadian and American industries. Second, I impose additional moment conditions that the Canadian industries productivity growth rates are orthogonal to the defense expenditure shock and that the American industries productivity growth rates are orthogonal to the commodity price shock.

[23]The estimates reported in Table 3 are from a two-step GMM estimator, where the efficient weight matrix is evaluated at the one-step estimator. For the continuous updated estimator, the efficient weight matrix is evaluated at the same value as the moment conditions.

[24]Because the objective function is invariant to reparameterization, having the objective function depend on the γ's rather than the ϕ's does not affect the S-set.

[25]In spite of the large intervals, the objective function, however, is not entirely flat with respect to ϕ_1 and ϕ_2. A much smaller critical value like 190 would produce much tighter intervals with ϕ_1 being between -0.39 and -0.02 and ϕ_2 being between and -0.23 and -0.10.

[26]I continue to normalized the productivity growth rates by dividing by the sample standard deviations.

vector of productivity shocks are normally distributed implies the following likelihood ratio test (Hamilton, 1994).

$$L(\Omega_U) - L(\Omega_R) = \frac{T}{2}\log\left|\Omega_U^{-1}\right| - \frac{1}{2}\sum_{t=1}^{T} z_t\Omega_U^{-1}z_t - \frac{T}{2}\log\left|\Omega_R^{-1}\right| + \frac{1}{2}\sum_{t=1}^{T} z_t\Omega_R^{-1}z_t$$

With 15 industries in each of the two countries, the number of degrees of freedom is large, with 465 coefficients in the sample covariance Ω_U. The distance based measure has only the 30 diagonal terms, and the 4 distance coefficients. Hence in the current application there are 431 degrees of freedom. A 95 percent critical value is 480.40. Fixing the distance matrix at the 1987 values, the estimated likelihood ratio is 536.44 for labor productivity, 504.67 for the multifactor productivity, and 442.93 for the estimated productivity. Therefore, as for the R^2 statistics reported earlier, there is evidence that there remains unexplained variation. Trying to understand what additional factors or refinements could help explain this variation is a topic for future research.

5.2 The Implications of Not Having Data on Materials Inputs

As mentioned earlier, an advantage of the KLEMS data set relative to other data sets is that the KLEMS data set has information on intermediate inputs. This section reports two sets of results to describe the effect of excluding the intermediate inputs data. In the first set of results, the measure of output is the value-added measure expressed as the growth rate of gross output minus the share-weighted growth rates of materials and services. The only inputs are labor and energy. Therefore, for the value-added data, productivity is calculated as the residual from the following regression. :

$$\begin{aligned}(\Delta y - s_m\Delta m - s_s\Delta s) &= \Delta v \\ \Delta v &= \gamma_l\Delta l + \gamma_e\Delta e + z\end{aligned}$$

An alternative is to measure output using gross output but still have as inputs only labor and energy.

$$\Delta y = \gamma_l\Delta l + \gamma_e\Delta e + z$$

This equation is similar to the equations estimated in Conley and Dupor (2003) and Costello (1993).

In terms of simple correlations, the different measures do not make a substantial difference. The first rows of Table 9 report the average correlation between cross-border pairs of the same industry for the estimated productivity and these two new measures. The number of significant correlations was also very similar.

For the border effect, the magnitudes of ϕ_1 and ϕ_2 were qualitatively similar. For the measures without intermediate inputs, non-identical industries have a somewhat larger border effect. For all three measures, the correlation patterns depend on the similarity of input use.

6 Conclusions

The paper has made three contributions to understanding industry-level productivity comovement. First, for Canada and the United States, cross-border productivity fluctuations are more highly correlated than had been reported. Second, the paper confirms that the similarity of input usage can help explain the pattern of covariances between industries. Third, there is a border effect. The correlation in productivity growth between similar industries in Canada and the United States is smaller than the correlations between similar industries within a country. In contrast to some of the border effect literature, the difference between cross-border correlations and within-country correlations of similar industries is not overwhelming. In particular, for a given industry, its correlation with a within-border industry that uses dissimilar inputs is less than its correlation with a cross-border industry that uses similar inputs.

These results suggest directions for further theoretical and empirical research. In particular, understanding what generates the dependence on input-usage could lead to a better understanding of the sources of productivity fluctuations. One possibility is that the dependence on input similarity reflects a measurement problem. Improvements in the quality of intermediate inputs is being measured as improved productivity by the users rather than the producers of the input.

As discussed in Engel and Rogers (2001), the 'first generation' of border-effect papers document the size of the border effect. As Engel and Rogers did for prices, there needs to be a 'second generation' of papers that explore the economic forces behind border effects. For productivity, one could follow the example of Evans (2003), who uses firm-level data to better understand the economics behind border effects for trade. In particular, she examines variations in sales between domestic and foreign multinationals to determine the importance of nationality versus location. A similar study for productivity growth would be very informative. In particular, if one could obtain firm-level information on intermediate input usage, then one might be able to better understand the dependence on input usage.

A Details on the Estimation

The following section describes the calculation behind the estimation procedure used to calculate the covariance between industries as a function of economic distance. The covariance is written as the weighted sum of basis functions $H_k()$:

$$C(d_{ij}) = \sum b_k H_k(d_{ij})$$

where $H_k()$ arises from the spectral representation of the covariance function. The spectral representation is implemented here using the following specification

$$H_k(d_{ij}) = \int h(\omega d_{ij}) B^1_{m,k}(d_{ij}) d\omega$$

where $B^1_{m,k}(d_{ij})$ is the first-derivative of the k-th b-spline[27] of order m. The function $h(x)$ is defined as

$$h(x) = 2^{(l-2)/2}\Gamma(k/2)\frac{J_{(l-2)/2}(x)}{x^{(l-2)/2}}$$

where $J_{(l-2)/2}(x)$ is a Bessel function of the first kind and l is the dimension of the location vector. In the current application, l equals the number of inputs in the input-output table. The covariance matrix is computed using the values of $\{b_k\}$ and the distance between sectors. The off-diagonal elements of the covariance matrix are defined as a function of the distance between the two industries $C(d_{ij})$. The diagonal elements are defined as $\sigma_j^2 + C(0)$ where $C(0)$ is the value of
$sumb_k H_k(0)$ and σ_j^2 is the industry-specific covariance term. The own-industry covariances are computed as

$$\sigma_j^2 = \max\left(\frac{1}{T}\sum_{t=1}^{T}\Delta z_{jt}^2 - C(0), 0\right)$$

The values of b_k are computed as the solution to the following least squares problem subject to the constraint that b_1 is non-negative and that $b_{k+1} \geq b_k$.

$$\min \sum_{t=1}^{T}\sum_{i=1}^{n-1}\sum_{j\neq i}\left(\Delta z_{it}^c \Delta z_{jt}^c - \sum b_k H_k\left(d_{ij}^c\right)\right)^2$$

$$+\sum_{t=1}^{T}\sum_{i=1}^{n-1}\sum_{j\neq i}(\left(\Delta z_{it}^u \Delta z_{jt}^u - \sum b_k H_k\left(d_{ij}^u\right)\right)^2$$

$$+\sum_{t=1}^{T}\sum_{i=1}^{n-1}\sum_{j\neq i}(\Delta z_{it}^c \Delta z_{jt}^u - \phi_2 - \sum b_k H_k\left(d_{ij}^a\right))^2$$

$$+\sum_{t=1}^{T}\sum_{i=1}^{n-1}(\Delta z_{it}^c \Delta z_{it}^u - \phi_1 - \sum b_k H_k(d_{ii}^a))^2$$

where Δz_j^c is the productivity measure from Canadian industry j, Δz_j^u is the productivity measure from the American industry j, d_{ij}^c is the distance between Canadian industries i and j, d_{ij}^u is the distance between American industries i and j, and d_{ij}^a is the distance between Canadian industry i and American industry j. When ϕ_1 and ϕ_2 are set equal to zero, these conditions on the coefficient estimates insure the resulting variance covariance matrix is a valid positive-definite matrix (Yaglom, 1987).

[27] Judd (1998, p 227-8) gives a concise discussion of b-splines.

B Similarity Between the Production Function and the Estimation

B.1 Economic Modelling

A two country version of the economic model described in the text is a straightforward extension. Consider the following production function for industries in two countries: Home and Foreign. Each country has J industries that each produce output Y_j^* by combining inputs x_{ji}^* in the following production function, where $*$ is either H for Home or F for Foreign,

$$\begin{aligned} Y_j^H &= \theta^H \theta_j^H \theta_j \prod_{i=1}^{N} \left(x_{ji}^H v_i \right)^{\alpha_{ji}^H} \\ Y_j^F &= \theta^F \theta_j^F \theta_j \prod_{j=1}^{N} \left(x_{ji}^F v_i \right)^{\alpha_{ji}^F} \end{aligned}$$

where θ^H is a country-specific productivity term, θ_j^H is a country-specific industry-specific productivity term, θ_j is an industry-specific productivity term, and v_i is an input-specific productivity term. The log of each of these productivity terms is assumed to be a random walk with the disturbances having a constant variance and being independent of each other. These assumptions can be expressed as

$$\begin{aligned} \Delta \ln v_i &= u_i \\ \Delta \ln \theta_j &= \varepsilon_j \\ \Delta \ln \theta^H &= \varepsilon^H \\ \Delta \ln \theta^F &= \varepsilon^F \end{aligned}$$

where $\left[u_i, \varepsilon_j, \varepsilon^H, \varepsilon^F\right]$ are independent of each other with variance $\left[\sigma_{vi}^2, \sigma_{\theta j}^2, \sigma_H^2, \sigma_F^2\right]$

The growth rate in industry level productivity would be the following (suppressing time subscripts for notational simplicity):

$$\begin{aligned} \Delta z_j^H &= \Delta \ln \theta^H + \Delta \ln \theta_j + \sum \alpha_{ji}^H \Delta \ln v_i \\ \Delta z_j^F &= \Delta \ln \theta^F + \Delta \ln \theta_j + \sum \alpha_{ji}^F \Delta \ln v_i \end{aligned}$$

Hence the covariance between any two industries can be written as follows.

$$\begin{aligned} E\Delta z_j^H \Delta z_k^H &= E\left(\Delta \ln \theta^H\right)^2 + E\left(\Delta \ln \theta_j \Delta \ln \theta_k\right)^2 + \sum \left(\alpha_{ji}^H \alpha_{ki}^H\right) E\left(\Delta \ln v_i\right)^2 \\ &= \sigma_H^2 + \sum \left(\alpha_{ji}^H \alpha_{ki}^H\right) \sigma_{vi}^2 \qquad \text{if } j \neq k \\ &= \sigma_H^2 + \sum \left(\alpha_{ji}^H\right)^2 \sigma_{vi}^2 + \sigma_{\theta j}^2 \quad \text{if } j = k \\ E\Delta z_j^H \Delta z_k^F &= E\left(\Delta \ln \theta_j \Delta \ln \theta_k\right)^2 + \sum \left(\alpha_{ji}^H \alpha_{ki}^F\right) E\left(\Delta \ln v_i\right)^2 \\ &= \sum \left(\alpha_{ji}^H \alpha_{ki}^F\right) \sigma_{vi}^2 \qquad \text{if } j \neq k \\ &= \sum \left(\alpha_{ji}^H \alpha_{ji}^F\right) \sigma_{vi}^2 \quad + \sigma_{\theta j}^2 \qquad \text{if } j = k \end{aligned}$$

$$
\begin{aligned}
E\Delta z_j^F \Delta z_k^F &= E\left(\Delta \ln \theta^F\right)^2 + E\left(\Delta \ln \theta_j \Delta \ln \theta_k\right)^2 + \sum\left(\alpha_{ji}^F \alpha_{ki}^F\right) E\left(\Delta \ln v_i\right)^2 \\
&= \sigma_F^2 + \sum\left(\alpha_{ji}^F \alpha_{ki}^F\right) \sigma_{vi}^2 \qquad \text{if } j \neq k \\
&= \sigma_F^2 + \sum\left(\alpha_{ji}^F \alpha_{ki}^F\right) \sigma_{vi}^2 + \sigma_{\theta j}^2 \quad \text{if } j = k
\end{aligned}
$$

Assuming industry-level cost minimization, the above Cobb-Douglas production function, implies that α_{ji} equals the industry's expenditure share s_{ji}. With a further assumption that σ_{vi}^2 equals σ_v^2 for all i, the resulting covariance function would be the following.

$$
\begin{aligned}
E\Delta z_j^H \Delta z_k^H &= \sigma_H^2 + \sigma_v^2 \sum\left(s_{ji}^H s_{ki}^H\right) \qquad \text{if } j \neq k \\
&= \sigma_H^2 + \sigma_{\theta j}^2 + \sigma_v^2 \sum\left(s_{ji}^H\right)^2 \quad \text{if } j = k \\
E\Delta z_j^H \Delta z_k^F &= \sigma_v^2 \sum\left(s_{ji}^H s_{ki}^F\right) \qquad \text{if } j \neq k \\
&= \sigma_v^2 \sum\left(s_{ji}^H s_{ji}^F\right) \qquad + \sigma_{\theta j}^2 \quad \text{if } j = k \\
E\Delta z_j^F \Delta z_k^F &= \sigma_F^2 + \sigma_v^2 \sum\left(s_{ji}^F s_{ki}^F\right) \qquad \text{if } j \neq k \\
&= \sigma_F^2 + \sigma_{\theta j}^2 + \sigma_v^2 \sum\left(s_{ji}^F s_{ki}^F\right) \quad \text{if } j = k
\end{aligned}
$$

Hence the covariance between two industries is larger for those industries that have more similar input shares and hence have large values of the product of input shares. There would also be a border effect because of the role of country-specific shocks.

B.1.1 Production Function Based Estimation:

The economic model was useful in order to give some perspective on the relevance of input usage. Besides providing intuition, the model could itself be estimated. To estimate the model, one would estimate the following regression.

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} \Delta z_{jt}^H \Delta z_{kt}^H &= \sigma_H^2 + \sum_{i=1}^{N} s_{ji}^H s_{ki}^H \, \sigma_{vi}^2 \qquad \text{if } j \neq k \\
\frac{1}{T}\sum_{t=1}^{T} \Delta z_{jt}^H \Delta z_{kt}^F &= \sum s_{ji}^H s_{ki}^F \, \sigma_{vi}^2 \qquad \text{if } j \neq k \\
\frac{1}{T}\sum_{t=1}^{T} \Delta z_{jt}^H \Delta z_{kt}^F &= \sigma_F^2 + \sum s_{ji}^F s_{ki}^F \sigma_{vi}^2 \qquad \text{if } j \neq k
\end{aligned}
$$

The only caveat of this regression is a need to reduce the number of inputs to a more manageable number. Because the input-output matrix is sparse, the number of inputs is reduced to the ten most important. The first column of Table A lists these inputs. The coefficient estimates are not too surprising. Both Chemicals (SIC 28) and Primary Metals (SIC 33) are important inputs to many different industries. Users of these inputs tend to have higher covariances. Having a high labor shares in both industries is the most important predictor of a high covariance. The one counter intuitive estimate is for textile users (SIC 24). The estimated coefficient is negative but this value may be a fixed effect for Textiles and Apparel industries.

B.1.2 Simulation Evidence to Compare the Two Models

The production-based model described in Section 2 and the econometric model of Conley and Dupor are similar but not identical. Using a simulation approach, this appendix explores how well the Conley and Dupor model can match the economic model. A panel of productivity fluctuations is simulated using the above production functions as the data generating process [DGP]. This simulated data are then used to estimate the Conley and Dupor spatial model. While obviously not a proof, it helps illustrate the properties of the two models. The simulation was made with the following assumption with the respect to the variances

$$
\begin{aligned}
\sigma_H^2 &= \sigma_F^2 = 0.25^2 \\
\sigma_\theta^2 &= 1 \\
\sigma_v^2 &= 1
\end{aligned}
$$

The input shares used are the actual 1987 input shares for the American and Canadian industries. The panel of industries was simulated for 500 observations. These data were then used in estimating the Chen and Conley spatial model with 5 splines of order 2. The resulting covariance functions and the true theoretical correlations are reported in Figure A1. The estimation method does preserve the general shape of the correlations from the production function model. In addition, the cross-border effect is estimated to be close to its true value.

References

[1] Amano, Robert A and Simon van Norden 1995. 'Terms of Trade and Real Exchange Rates: The Canadian Evidence' *Journal of International Money and Finance*, 14(1) 83-104

[2] Anderson, James and Eric Van Wincoop 2003. 'Gravity with Gravitas: A Solution to the Border Puzzle' *American Economic Review*, 93(1), 170-92

[3] Baldwin, John R. and David Sabourin 1998. 'Technology Adoption: A comparison Between Canada and the United States' *Statistic Canada Working Paper* 119

[4] Basu, Susanto and John G. Fernald 1995. 'Are Apparent Productive Spillovers a figment of Specification Error?' *Journal of Monetary Economics* 36(1), 165-88

[5] Basu, Susanto and John G. Fernald 2000. 'Why is Productivity Procyclical? Why do we care?' *NBER Working Paper* 7940

[6] Basu Susanto, John G. Fernald, and Miles S. Kimball. 1999. 'Are Technology Improvements Contractionary' Manuscript.

[7] Branstetter Lee G. 2001. 'Are knowledge spillovers international or intranational in scope? Microeconometric evidence from the U.S. and Japan' *Journal of International Economics*, 53, 53-79

[8] Burnside, Craig 1996. 'Production Function Regressions, Returns to Scale and Externalities,' *Journal of Monetary Economics*, 37, 177-201.

[9] Burnside, Craig, Martin Eichenbaum and Sergio Rebelo, 1996. 'Sectoral Solow Residuals,' *European Economic Review,* 40, 861-869.

[10] Caballero, Ricardo J. and Richard K. Lyons 1992. 'External Effects in U.S. Procyclical Productivity', *Journal of Monetary Economics*, 29(2), 209-25.

[11] Chen, Xiaoheng and Timothy G. Conley 2001. 'A New Semiparametric Spatial Model for Panel Time Series', *Journal of Econometrics*, 105(1), 59-83

[12] Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans, 2000. 'Monetary Policy Shocks: What Have We Learned and to What End?', *Handbook of Monetary Economics*, ed.s Michael Woodford and John Taylor.

[13] Costello, Donna M. 1993. 'A Cross-Country Cross-Industry Comparison of Productivity Growth' *Journal of Political Economy*, 101(2), 207-222

[14] Conley, Timothy G. and Bill Dupor 2003. 'A Spatial Analysis of Sectoral Complementarity' *Journal of Political Economy*, 111(2), 311-52

[15] Dupor, Bill. 1996 'Aggregate Fluctuations and Production Complementarities.' Manuscript.

[16] Eldridge, Lucy P. and Mark K. Sherwood. 2001. 'A perspective on the U.S.-Canada manufacturing productivity gap,' *BLS Monthly Labor Review*, February 124(2), 31-48

[17] Engel, Charles, and John H. Rogers 1996. 'How Wide is the Border?' *American Economic Review* 86, 1112-1125

[18] Engel, Charles, and John H. Rogers 2001. 'Deviations from Purchasing Power Parity: Causes and Welfare Costs' *Journal of International Economics* 55(1) 29-57

[19] Grossman, Gene M. and Elhanan Helpman 1990 'Comparative Advantage and Long-Run Growth', American Economic Review, 80(4), 796-815.

[20] Helliwell, John F. 1996. 'Do National Borders Matter for Quebec's Trade?' *Canadian Journal of Economics*, 29(3), 507-22

[21] Horowitz, Joel L. 2000 'The Bootstrap' in the *Handbook of Econometrics*, editors James J. Heckman and Edward Leamer, 5(52)

[22] Horvath, Michael 2000. 'Sectoral Shocks and Aggregate Fluctuations' *Journal of Monetary Economics*, 45(1), 69-106

[23] Johnson, Joanne 1997. 'A KLEMS Database: Describing the Input Structure of Canadian Industry' *Aggregate Productivity Measures* Statistics Canada Cat No 15-204E, 19-32

[24] Judd, Kenneth L. 1998. *Numerical Methods in Economics* MIT Press

[25] Lum Sherlene K.S. and Brian C. Moyer 1998. 'Gross Product by Industry 1995-97', *Survey of Current Business* November

[26] McCallum, John 1995. 'National Borders Matter: Canada-U.S. Regional Trade Patterns' *American Economic Review*, 85(3), 615-23

[27] Stadler, George W. 1994. 'Real Business Cycles' *Journal of Economic Literature*, 32(4), 1750-83

[28] Stock, James H. and Jonathan H. Wright 2000. 'GMM With Weak Identification', *Econometrica* 68(5,) 1055-1096

[29] Yaglom A. M. 1986. *Correlation Theory of Stationary and Related Random Functions* Springer Verlag New York

C Tables

Table 1: Durable and NonDurable Industries

Durable	SIC Code	NonDurable	SIC Code
Lumber	24	Food	20
Furniture	25	Textiles	22
Glass Stone & Clay	32	Apparel	23
Primary Metals	33	Paper	26
Fabricated Metals	34	Printing	27
Industrial Machinery	35	Chemicals	28
Electrical Machinery	36	Rubber and Plastics	30
Transportation (Equipment)	37		

Table 2: A subsection of the U.S. Input Output Table 1987

	Industries							
Inputs	22	26	28	29	33	35	36	37
Other	22	22	29	82	34	18	20	17
20 Food	0	0	0	0	0	0	0	0
22 Textiles	**29**	1	0	0	0	0	0	0
26 Paper	0	**29**	2	0	0	0	1	0
28 Chemicals	21	8	**33**	1	3	0	2	1
29 Petroleum	0	0	1	**8**	0	0	0	0
33 Primary Metals	0	0	0	0	**27**	10	7	5
34 Fabricated Metals	0	0	1	0	1	4	4	7
35 Industrial Machinery	0	0	0	0	2	**15**	1	5
36 Electrical Machinery	0	0	0	0	0	7	**16**	4
37 Transportation Equipment	0	0	0	0	0	0	0	**23**
Labor	22	26	25	5	26	39	39	26

Notes: Each table entry reports the percentage of total inputs used by the industry listed in the column produced by the industry listed in the row.
Bold text denotes the input share produced by the own-industry.

Table 3: Production Function Coefficient Estimates

	United States			Canada		
	Nondurables			**Nondurables**		
	Cost Shares	Estimate	95 percent Confidence Interval	Cost Shares	Estimate	95 percent Confidence Interval
Capital γ_k						
Labor γ_l	0.31	0.77	(0.56, 0.97)	0.23	0.17	(0.03, 0.30)
Energy γ_e	0.027	-0.01	(-0.11, 0.08)	0.02	0.04	(-0.01, 0.10)
Materials γ_m	0.45	0.07	(-0.06, 0.20)	0.48	0.67	(0.57, 0.77)
Services γ_s	0.08	0.27	(0.14, 0.39)	0.14	0.25	(0.13, 0.37)
Returns to Scale		1.09	(0.96, 1.22)		1.13	(1.10,1.16)
	Durables			**Durables**		
	Cost Shares	Estimate	95 percent Confidence Interval	Cost Shares	Estimate	95 percent Confidence Interval
Capital γ_k						
Labor γ_l	0.37	0.09	(-0.08, 0.26)	0.30	0.21	(0.06, 0.36)
Energy γ_e	0.03	0.18	(0.06, 0.30)	0.02	0.07	(0.01, 0.14)
Materials γ_m	0.41	0.53	(0.41, 0.64)	0.44	0.68	(0.57, 0.79)
Services γ_s	0.07	0.14	(0.06, 0.21)	0.12	0.05	(-0.10, 0.19)
Returns to Scale		0.93	(0.86, 1.00)		1.02	(0.84, 1.20)
Test of Over-identifying Restrictions		76.96			79.35	

Notes: Costs shares are the average expenditure on the input. Estimated coefficients are from a two-step GMM process. The partition of industries into durable and nondurable are reported in Table 1.

Table 4: Correlations Between Industry Pairs

Industry SIC Code	Measure of Productivity Labor	Multifactor	Estimated
20 Food	**0.26**	**0.15**	**0.21**
22 Textiles	**0.26**	**-0.12**	**-0.03**
23 Apparel	-0.14	0.36*	0.23
24 Lumber	-0.09	0.12	0.26
25 Furniture	-0.04	0.50*	0.58*
26 Paper	0.62*	0.49*	0.61*
27 Printing	0.44*	0.33*	0.24
28 Chemicals	**0.55***	**0.56***	**0.48***
30 Rubber and Plastics	0.29*	0.54*	0.59*
32 Glass Stone & Clay	0.43*	0.64*	0.33*
33 Primary Metals	**0.51***	**0.33***	**0.25**
34 Fabricated Metals	**0.26**	**0.37***	**0.28**
35 Industrial Machinery	0.30*	0.35*	0.16
36 Electrical Machinery	0.58*	0.44*	0.30*
37 Transportation Equipment	0.60*	0.62*	0.30*
	Average Correlation between Industry Pairs		
American Pairs	0.22	0.27	0.33
Canadian Pairs	0.28	0.37	0.19
Cross Border Pairs	0.12	0.23	0.17
	Percentage of Statistically Significant Correlations		
American Pairs	51%	55%	65%
Canadian Pairs	54%	74%	31%
Cross Border Pairs	32%	50 %	30%

* denotes correlations that are statistically significant
at the 90 percent significance level.
Bold text denotes industries examined in Costello (1993).
Sample Period 1961-1997

Table 5: Comparing the Standard Deviation of Various Productivity Measures

Industry	United States			Canada		
	Std Dev of MFP	Ratio of Std Dev to MFP Labor	Estimated	Std Dev of of MFP	Ratio of Std Dev.to MFP Labor	Estimated
20 Food	2.41	0.96	1.00	0.87	2.17	1.03
22**Textiles**	1.98	1.25	1.08	2.44	1.65	0.60
23 Apparel	1.14	3.51	2.81	1.68	1.87	0.88
24 Lumber	3.43	1.30	1.23	2.35	1.56	0.75
25 Furniture	1.79	1.42	1.51	3.13	1.47	0.79
26 Paper	2.96	0.67	1.00	2.87	1.21	0.56
27 Printing	1.68	1.43	1.11	2.32	1.19	0.75
28 **Chemicals**	4.04	1.04	0.97	2.29	1.76	0.72
30 Rubber and Plastics	2.33	1.32	1.15	2.92	1.46	0.68
32 Glass Stone & Clay	2.26	0.87	0.95	3.44	1.19	0.56
33 **Primary Metals**	2.94	1.18	1.06	1.90	2.62	0.71
34 **Fabricated Metals**	1.77	1.34	1.24	1.90	1.85	0.65
35 Industrial Machinery	2.99	1.20	1.15	3.02	1.62	0.53
36 Electrical Machinery	3.02	1.33	1.10	2.64	1.69	0.76
37 Transportation Equipment	2.92	1.63	1.21	2.37	2.56	0.83
Average	2.51	1.36	1.24	2.41	1.72	0.72

Standard Deviation Of Multifactor Productivity Growth Multiplied by 100.
Bold text denotes industries examined in Costello (1993).

Table 6 : OLS Regressions on Distance-Based Correlations

	R^2 of Regression			F-test of No Explanatory Role for Inputs		
	U.S.	Canada	Cross	U.S.	Canada	Cross
Labor	0.11	0.22	0.02	6.13	14.46	2.27
P-values				*<0.01*	*<0.01*	*0.1059*
Multifactor	0.23	0.17	0.07	15.41	10.53	7.60
P-values				*<0.01*	*<0.01*	*<0.01*
Estimated	0.29	0.10	0.09	20.57	5.59	10.67
P-values				*<0.01*	*<0.01*	*<0.01*

Notes: For the U.S. and Canada regressions, there are 2 and 15*7-3 degrees of freedom
For the Cross Border regression, there are 2 and15*14-3 degrees of freedom

Table 7: Estimated Coefficients for Border Effect

	β_0	β_1	β_2	ϕ_{CROSS}	R^2	F-test
Labor Productivity	0.44	-0.40	0.01	-0.04	0.04	4.57
	2.91	*-0.54*	*0.01*	*-1.44*		
MFP	0.79	-1.77	1.62	-0.01	0.06	6.17
	5.36	*-2.42*	*1.83*	*-0.55*		
Estimated Productivity	0.83	-2.16	1.71	-0.06	0.15	18.48
	5.68	*-2.97*	*1.93*	*-2.54*		

Notes: Numbers in italics are t-statistics for the hypothesis that coefficient equals zero. F-test is that β_1 and β_2 equals zero.

Table 8: Testing for A Border Effect

Productivity:	Labor	MFP	Estimated
Cross-Border Decline: Identical Industries	-0.073	-0.164	-0.092
$P\left(\phi_{1,sim} < \phi_{1,est} \mid H_0\ \phi_1 = 0\right)$	0.112	0.008	0.100
90% Confidence Interval	(-0.205 0.012)	(-0.306 -0.036)	(-0.215 0.028)
Cross-Border Decline: Non-identical Industries	-0.147	-0.117	-0.081
$P\left(\phi_{2,sim} < \phi_{2,est} \mid H_0\ \phi_2 = 0\right)$	0	0	0
90% Confidence Interval	(-0.209 -0.088)	(-0.185 -0.043)	(-0.194 0.049)
Confidence Interval for $\phi_{1,sim} - \phi_{2,sim}$	(-0.029 0.141)	(-0.156 0.031)	(-0.112 0.107)

Notes: Hypothesis Tests are generated using 500 Simulations of Distance-Based Correlation Matrix DGP with no Border Effect. Confidence Intervals are generated by drawing with replacement . from vector time series of productivity growths. 500 Simulations

Table 9: Effect of Not Having Intermediate Inputs Data

	Baseline	Value Added	Gross Output No Intermediate Inputs
Average Correlation	0.3192	0.3158	0.3212
Significant Correlations	7	8	9
ϕ_1	-0.092	-0.113	-0.059
ϕ_2	-0.081	-0.113	-0.124
d		C(d)	
0	0.511	0.507	0.461
0.126	0.372	0.404	0.345
0.253	0.354	0.357	0.335
0.379	0.281	0.260	0.266
0.505	0.162	0.147	0.155
0.632	0.087	0.079	0.085
0.758	0.047	0.042	0.047

D Figures

Figure 1: Histogram of Distances

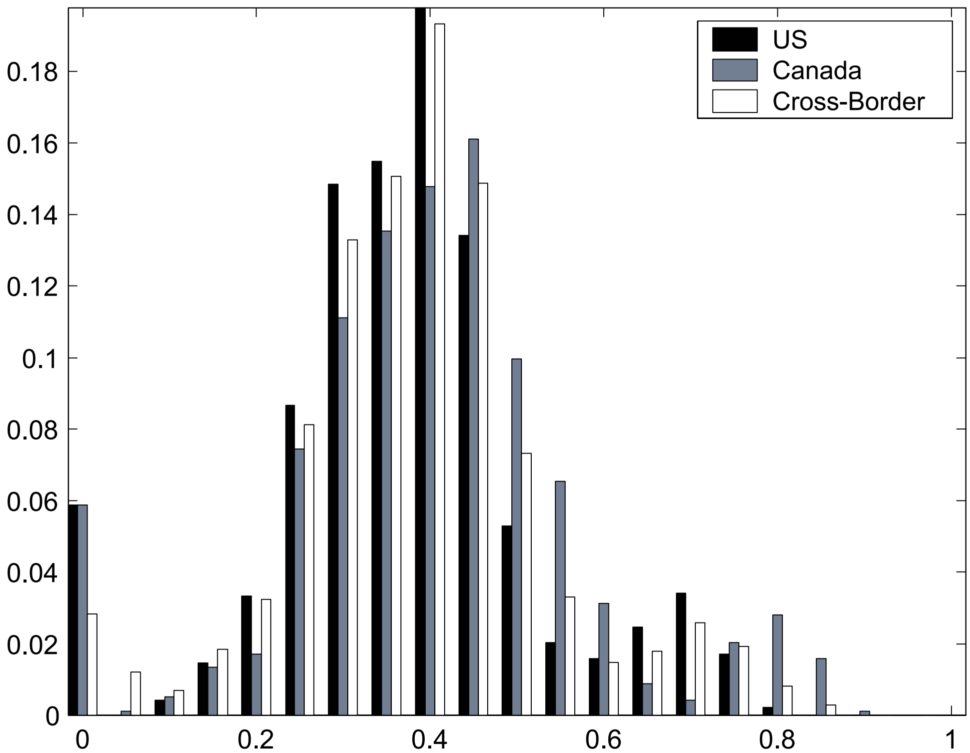


Figure 2: The Correlation of Productivity Growth as Function of Similar Input Use

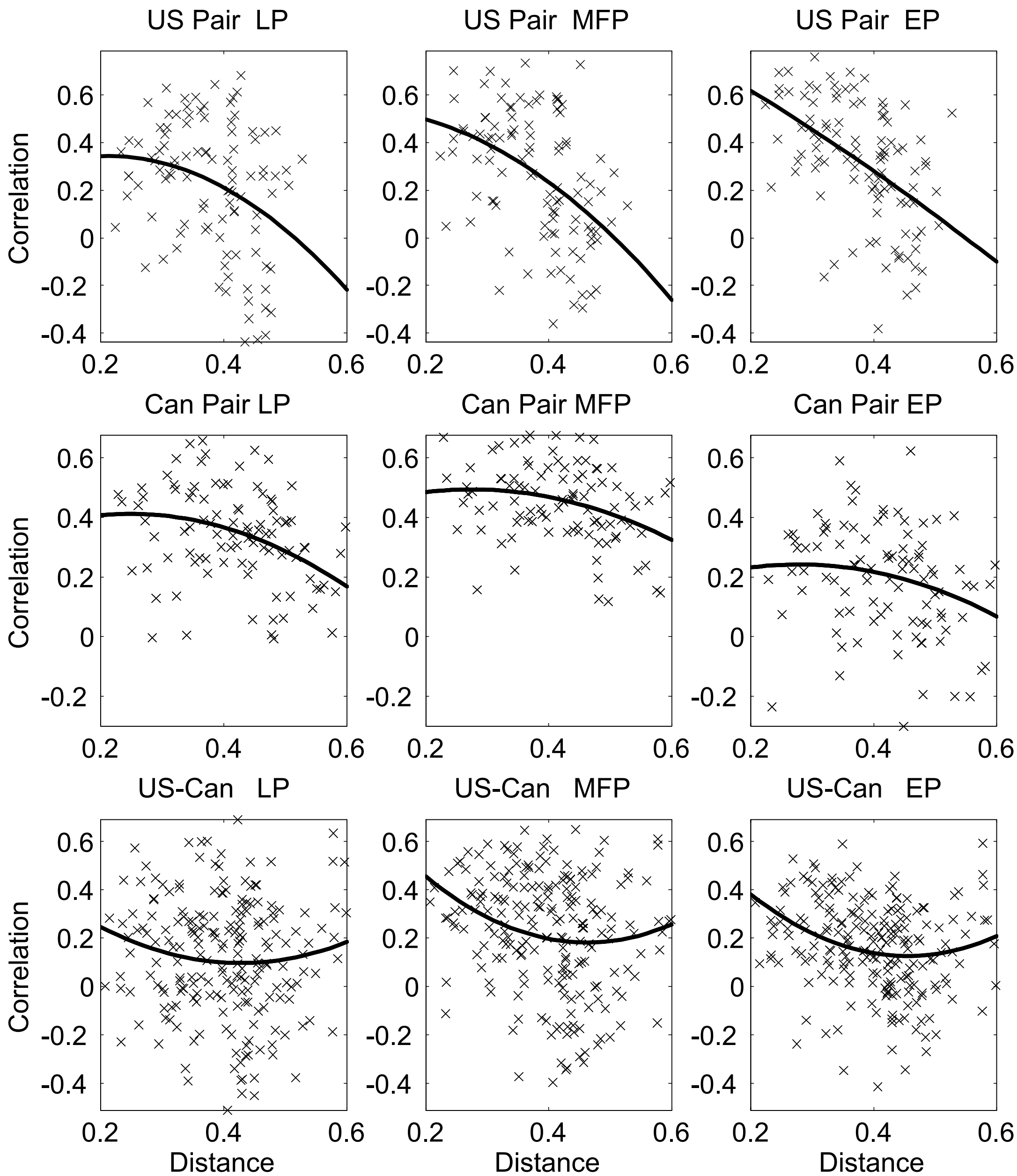


Figure 3 Correlation Between Industries As Function Of Input Similarity

Figure 3a: Labor Productivity

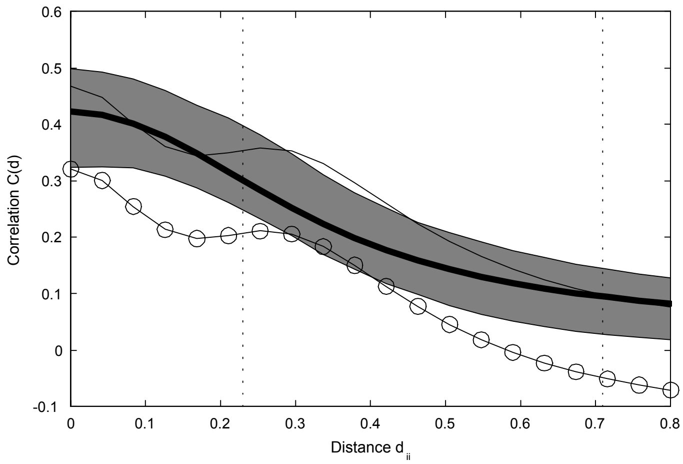


Figure 3b: Multifactor Productivity

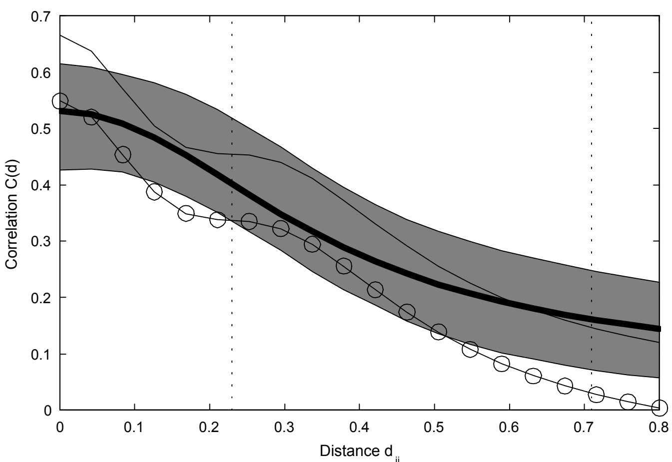
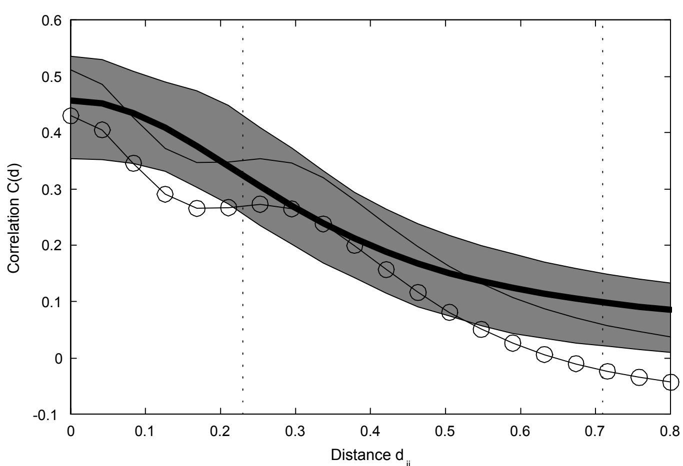


Figure 3c: Estimated Productivity



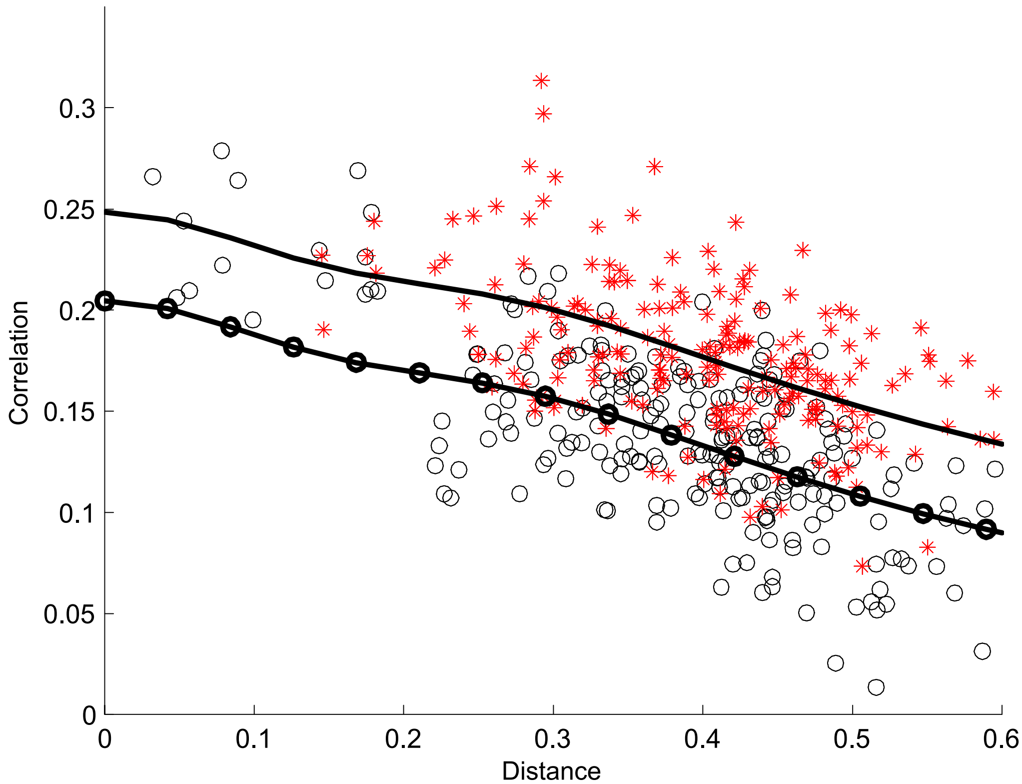
Notes: Thick Line, C(d) estimated under null hypothesis of no border effect

Thin Line: Estimate of C(d) absent a border effect, Circles: Estimate of C(d) $+\phi_2$ (include a border effect)

Grey Area 90 percent confidence interval of C(d) under null of no border effect.

Dashed Lines. Interval Between 0.23 and 0.7 that includes 90 percent of non-zero distances

Figure A: Comparing the Two Models



Solid Lines: C(d)
Line with Circles C(d)+ϕ_2
Stars Correlations Between within Border Pairs
Circles Correlations Between Cross Border Pairs

Appendix Table A

	β	T-stat	Average Product	Median Product
Constant	-0.4	-0.98		
U.S.	1.8	10.99		
Canada	-0.2	-1.51		
Other	0.8	0.39	0.047	0.041
22	–0.1	-0.01	-0.000	-0.000
24	-78.2	-5.32	-0.064	-0.000
26	12.1	0.92	0.015	0.001
28	22.9	4.33	0.087	0.011
33	9.9	1.25	0.026	0.0002
34	63.6	1.08	0.039	0.008
35	-6.4	-0.09	-0.002	0
36	113.9	2.48	0.040	0
37	27.8	0.21	0.007	0
Labor	10.5	3.91	0.790	0.763