H
ARVARD

E
CONOMICS

R
EVIEW

Dear Reader,

In this issue, we delve into the field of international economics and examine important issues related to trade and globalization. The movement of goods, people, and technology between countries has been a topic at the forefront of debate – most recently, the presidential debates. Trade and globalization benefit and hurt different parts of our economy such that it is impossible to claim these phenomenon as pareto improvements. Our increasingly interconnected world is a fascinating place, and we are fortunate to feature the work of very accomplished economists who are helping us understand it.

Ying Zhang and Andre van Stel share their findings on how entrepreneurship stemmed from and now drives economic development in China's contemporary economy. Davin Chor addresses the phenomenon of global trade slowdown and discusses how its impact depends on whether the cause is fundamental or cyclical. Marc Melitz and Daniel Trefler explain how trade is no longer only driven by comparative advantage and differences in natural resources. They propose several novel ways that trade is improving efficiency for producers and variety available to consumers.

Esteban Rossi-Hansberg examines the impact of migration by acknowledging its positive and negative consequences for Americans and by ultimately suggesting that increased economic activity can lead to a net benefit for all. Thomas Prusa questions the efficiency of anti-dumping laws by reevaluating when a trade law is "unfair" and when anti-dumping laws are needed to combat unfairness.

In the context of the post-crisis global economy, Carter Johnson analyzes the consequences of unpegging the Swiss franc and looks broadly at the monetary policies of Europe, Japan, and the United States. Masoud Movahed returns to the pre-crisis economy and suggests that the 2008 crisis resulted from a discrepancy between "what is good for banks" and "what is good for the economy." Finally, Diego Perez reminds us of what we often take for granted: the availability of economic statistics. Perez studies an episode in which the Argentinian government purposefully understated inflation, leading to inefficiency and underproduction.

Thank you to the writers for sharing your work and knowledge.

Thank you to the editorial board for putting together this issue. We hope you enjoy.

Sincerely,

Angela Ma

Angela Ma

Editor-in-Chief
Angela Ma

Business Directors
Mirza Uddin
Osama Arif

Content Director
Justin Lee

Technology Director
Julian Salazar

Design Director
Jordan Berman

Contents

About Harvard Economics Review

The Harvard Economics Review is a biannual periodical addressing issues of national and global importance through economic analyses accessible to professionals and common readers alike. We feature articles and interviews with some of the world's most prominent economists in academia, business, and government, as well as a regular blog from our Harvard staff.

harvardecon.org

The Harvard Economics Review deeply appreciates the support of the Harvard Undergraduate Council, without whom this issue could not have been printed.

Economic Transition and Entrepreneurship Development

One of the most prominent features of China's transition from a centrally planned economy to a market-based economy is the emergence of entrepreneurship, although previous literature discusses this phenomenon descriptively rather than prescriptively. In this article, we consider entrepreneurship developments in China since the end of the 1970s and argue that the role of entrepreneurship in the economy has changed considerably over the last four decades. Our perception is that initially, China's entrepreneurship development stemmed from China's economic transition, and currently, entrepreneurship is both influenced by and influences economic development. We propose a conceptual model of the role of entrepreneurship in China's contemporary economy, which we test using a unique database for 31 Chinese regions during the period from 1997 to 2009. Our analysis shows that two types of entrepreneurial organizations (siyingqiye and getihu) in China play important but distinct roles in stimulating China's economic development.

By Ying Zhang[a] & Andre van Stel[b]
[a]Rotterdam School of Management, Erasmus University Rotterdam, the Netherlands
[b]Trinity College Dublin, Ireland & Kozminski University, Warsaw, Poland

The Chinese economy has gone through major transitions in the last decades, where a substantial part of economic activity has shifted from state-owned sectors to private sectors. China has transitioned from a tightly centrally-planned economy to a market-oriented economy that continues to be shaped by the government's long-term economic development plan and entrepreneurship focus (cf. Huang 2010). Policies have accelerated China's economic development by making important adjustments in the areas of education, national innovation system, economic openness, market function, infrastructure investment, and more. This transition cycle, with a great leap in economic growth, has run for more than three decades and still plays a critical role in China's economic growth. One important characteristic of this economic and institutional transition, we argue, is the attitude transition in acknowledging entrepreneurship while developing from a factor-driven economy to an efficiency-driven economy in the past 30 years, and now towards the innovation-driven stage. These developments are not independent of each other with entrepreneurship undoubtedly contributing to China's fascinating economic growth.

Since the late 1970s, Chinese entrepreneurship has experienced three generations of organizational forms: commune and brigade enterprises, Township and Village enterprises (TVE), and finally, the emergence of private firms such as getihu and siyingqiye. Commune and brigade enterprises, as the first generation, were designed by the central government to deal with the negative economic consequences of China's Cultural Revolution (1966 to 1976), particularly in non-agricultural industries. This organizational form did not function in the Chinese economy longer than a decade. It was then replaced by the second generation, TVEs, which were characterized by shared ownership of local government and collectives. Similarly, TVEs faced tremendous questioning on its ownership and nature as private firms. Though it contributed significantly to China's economy in the late 1980s (20% of China's gross output), TVEs were terminated, and gradually replaced by the third generation of organizational forms in the late 1980s: getihu and siyingqiye.

Getihu (in Chinese) are private businesses that are registered at the Chinese Industry and Commerce Office in the enterprise category with no more than seven people hired as employees. In June 1988, the Chinese central government added a new provision on private enterprise (TSPE), stipulating that a business with privately owned assets and more than seven employees could be registered as another form of private enterprise called siyingqiye (in Chinese). Getihu are restricted to only using individual or household assets for business operations but enjoy the privilege of registering with a shorter procedure and wider cognitive acceptance. In contrast, siyingqiye are given much more relaxed conditions in terms of more allowed sources of registered capital (e.g., shareholders can come from outside of the entrepreneur's family). However, siyingqiye are required to hold a fixed amount of registration capital, which translates to higher start-up costs. As a result of these differences, getihu are often smaller than siyingqiye in terms of organizational size.

Introducing getihu and siyingqiye as new organizational forms of private firms was a landmark for China's economic transition and was very critical in China's entrepreneurship development. Private firms have since started co-existing with State-Owned Enterprises (SOEs).

CONCEPTUAL MODEL: THE ROLE OF ENTREPRENEURSHIP IN MODERN CHINA

Given China's economic development in recent decades, we posit that the fundamental pillar of this growth—entrepreneurship—is powered by its economic transition and its economic & institutional antecedents. Figure 1 displays our view on the bigger picture of the role of entrepreneurship in China's contemporary economy.

Since 1978, the Chinese government has been gradually experimenting with various transitional policies, which have led to the improvement of welfare (the standard of living) for the majority of the population (illustrated by the direct link between the leftmost and rightmost boxes in Figure 1); these transitional policies have also promoted entrepreneurial activities in two ways. First, by allowing and acknowledging the important role of private enterprise in the Chinese economy, the policies directly contributed to an increase in the numbers of private firms (arrow 1 in Figure 1). Second, by improving the conditions (antecedents) for entrepreneurship (e.g., investing in education or upgrading the institutional environment), the policies boosted entrepreneurial activities in a more indirect manner as well (arrow 2 in Figure 1).

It is the link between economic antecedents and entrepreneurship (arrow 2 in Figure 1) that we investigate empirically. In particular, we explain regional rates of getihu and siyingqiye over the period of 1997 to 2009 using a number of economic antecedents and investigate the extent to which these antecedents are in line with the theory of the 'entrepreneurial economy' (Audretsch and Thurik 2000). The theory describes how 'productive' entrepreneurship (Baumol 1990) contributes to macroeconomic growth in the innovation-driven stage of economic development. This theory is par-

Figure 1 ---Economic Mechanism of China's Entrepreneurship and Economic Development

ticularly relevant for entrepreneurship development in China, which is now moving from an efficiency-driven economy towards an innovation-driven one. If economic antecedents of regional private firm rates in China are in line with those described in the theory of the 'entrepreneurial economy', that is, if antecedents are conducive to knowledge production and knowledge spillovers, it may be argued that the economy has the righ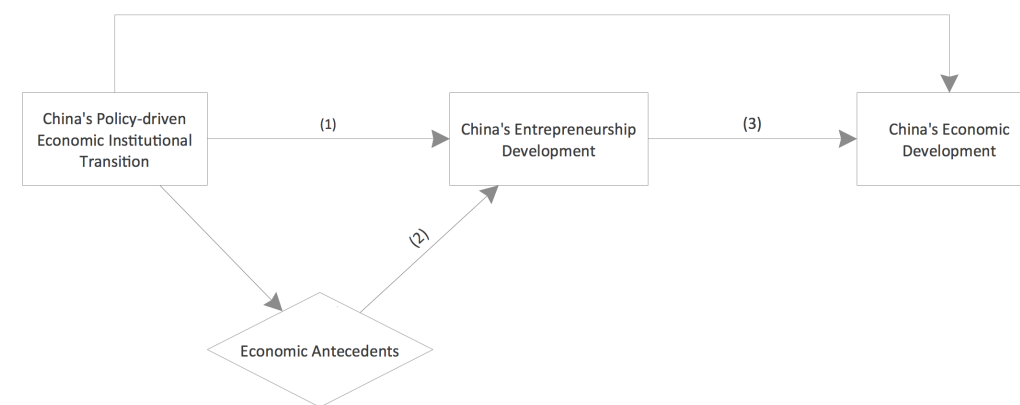t incentive structure in place to produce high quality entrepreneurship, and hence, that the entrepreneurial sector (private firm population) contributes substantially to economic development. The antecedents found in the empirical analysis therefore lead us to reflect on the extent to which entrepreneurship development in China has kept pace with the transition to the innovation-driven stage of economic development, i.e., to what extent entrepreneurship in China may be expected to contribute to economic development (arrow 3 in Figure 1).
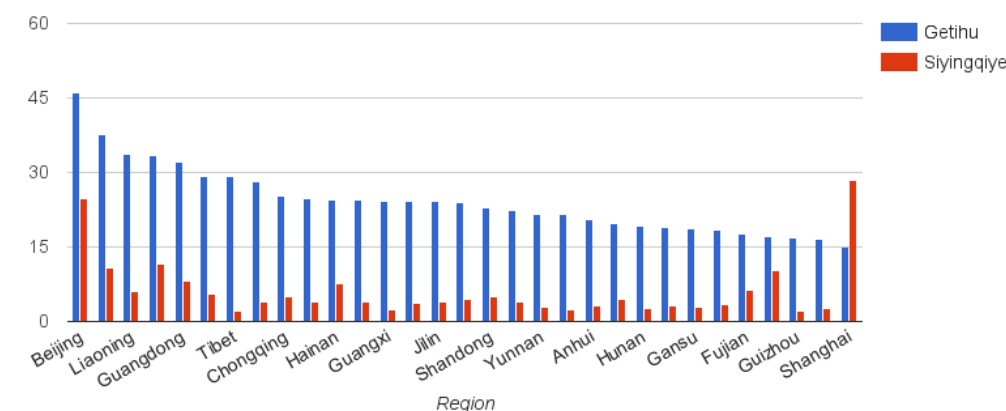
EMPIRICAL MODEL

We select economic antecedents from the literature on regional determinants of entrepreneurship (e.g., Reynolds et al.,

1994), in particular indicators reflecting the remuneration of labor, the stock of human capital, the institutional environment, the degree of agglomeration, and economic openness. Using a two-equation model, we investigate how these variables affect the regional rates of getihu and siyingqiye (scaled on the regional population) over the period of 1997 to 2009. In our full paper we derive hypotheses on how these variables may influence the rates of getihu and siyingqiye differently (Zhang and Van Stel, 2016). We also control for the regional industry structure and the investment intensity of state-owned enterprises. Moreover, our model allows us to investigate the degree of interaction between getihu and siyingqiye firms at the regional level.

We use seemingly unrelated regression (SUR) estimation to take into account the correlation between our two dependent variables, the getihu and siyingqiye rates. Moreover, we estimate two different set-ups of our two-equation model. In the first set-up, we focus on explaining the variation over time of the regional rates of getihu and siyingqiye, using a fixed effects set-up (dynamic approach).

In the second set-up, we include variables for all regions but include only the years 1998, 2003 and 2009. Here, we do not include regional fixed effects but instead estimate a pooled (SUR) model to explain the variation across regions of the regional rates of getihu and siyingqiye (static approach).

With one exception, all variables in our study are derived from the various China Statistical Yearbooks from the National Bureau of Statistics of China (NBSC) database, covering 31 Chinese regions over 13 years (from 1997 to 2009, in total 403 region-year observations). As an illustration of our dependent variables, Figure 2 presents the rates of getihu and siyingqiye in Chinese regions for our most recent year of data, 2009.

Figure 2---Rates of Getihu and Siyingqiye across regions (2009)

RESULTS AND CONCLUSION[5]

We find that, among other findings, a region's prevalence of human capital (measured as the share of college graduates in the local population) is positively related to the rate of siyingqiye but not so to the rate of getihu. Our results also suggest that agglomeration advantages accrue to siyingqiye rather than getihu firms. Thus, siyingqiye firms are relatively more often present in regions where economic antecedents are conducive to knowledge production and knowledge spillovers. As knowledge is the main source of competitive advantages in innovation-driven economies, we may therefore say that regional incentive structures for siyingqiye seem to be in line with a modern competitive economy, where knowledge-based entrepreneurship is key.

Although siyingqiye thus seem to be conducive to China's innovation-driven economic development, our estimation results also show that the presence of large companies, either in the form of inward FDI or in the form of SOEs, still crowds out domestic siyingqiye firms. This suggests that, although China is transitioning towards an innovation-driven economy and away from an efficiency-driven economy, economies of scale still play an important role.

Regarding getihu, our results suggest that antecedents of regional private firm rates are less in line with the 'entrepreneurial economy'. For instance, we do not find evidence of a positive association between the education level of the regional population and the number of getihu firms. This may suggest that a portion of getihu firms are started out of a necessity-motivation. When labor market participants have no other options for work, they may start their own firm, and the getihu form is the easiest (and cheapest) organizational form to choose. However, notwithstanding the presence of necessity-motivated entrepreneurs in the getihu sector, nowadays entrepreneurial activity among getihu is increasingly opportunity-based. In this respect, the getihu form may also offer advantages to ambitious entrepreneurs as the small scale and scope enable them to be flexible and to use a low-profile environment to experiment with new ideas.

To investigate the degree of interaction between getihu and siyingqiye rates, we performed a simulation exercise based on our estimation results, where we found evidence for considerable interaction between regional rates of getihu and siyingqiye, arguably a sign that the getihu sector in China is of considerable quality. We found that the impact of getihu on siyingqiye (estimated elasticity 0.51) is even stronger than vice versa (0.29), predicting that, ceteris paribus, the gap between the number of getihu and siyingqiye will decrease in the near future. (Note that, with the exception of Shanghai, the getihu rate is higher than the siyingqiye rate, see Figure 2). As siyingqiye antecedents have been found to be more in line with the 'entrepreneurial' economy, this predicted increase in the share of siyingqiye firms (relative to getihu) suggests that China is slowly but surely transitioning towards an 'entrepreneurial' economy.

On balance, our results suggest that siyingqiye and getihu each play their own role in the modern Chinese economy. Siyingqiye firms have the possibility to grow and are therefore attractive to ambitious, opportunity-oriented entrepreneurs. Our analysis suggests that siyingqiye entrepreneurship in China is indeed to a large extent opportuni-

ty-driven as its rates are found to be highest in regions where economic antecedents are conducive to running profitable, competitive firms in the modern knowledge-based economy. Nevertheless, siyingqiye rates are lower in regions with a high large-firm presence, marking China's transitional stage between the efficiency-driven and innovation-driven economy where economies of scale still play an important role.

In contrast, antecedents of getihu firms are less clearly linked to reaping the fruits of the 'entrepreneurial' economy, possibly indicating that the getihu sector consists of a mixture between necessity-driven and opportunity-driven entrepreneurs. However, our analysis also suggests that the share of high-quality, opportunity-driven entrepreneurs in the getihu sector may be increasing, as getihu firms were found to play an important role in enabling the number of siyingqiye firms to increase.

We believe that by investigating the economic antecedents of regional private firm rates and by distinguishing between two types of private firms, getihu and siyingqiye, our study contributes to a better understanding of the complex role that entrepreneurship plays in China's contemporary economy.

> "We found that the impact of getihu on siyingqiye (estimated elasticity 0.51) is even stronger than vice versa (0.29), predicting that, ceteris paribus, the gap between the number of getihu and siyingqiye will decrease in the near future. As siyingqiye antecedents have been found to be more in line with the 'entrepreneurial' economy, this predicted increase in the share of siyingqiye firms (relative to getihu) suggests that China is slowly but surely transitioning towards an 'entrepreneurial' economy"

[1] For an extended version of this article we refer to Zhang, Ying, and André van Stel (2016), "Who should be Running Ahead? The Roles of Two Types of Entrepreneurship in China's Contemporary Economy", Harvard Business School Working Paper, No. 16-086. http://www.hbs.edu/faculty/Publication%20Files/16-086_752131c0-5f40-426a-ad68-a8e776d9a7f3.pdf
[2] This policy was based on Marx's theory that if a business is allowed more than seven employees, the result could be the exploitation of labor (from "Das Kapital" by Karl Marx, 1867).
[3] In the early stage of China's entrepreneurship development, entrepreneurs who chose to set up getihu were more often from necessity orientation. Therefore, the getihu form received more legitimate support from government.
[4] See the arrow between China's policy-driven economic institutional transition and economic antecedents in Figure 1.
[5] For a more detailed presentation of our estimation results, including the regression tables, we refer to Zhang and Van Stel (2016).
[6] For instance, China's national team in the Global Entrepreneurship Monitor reports that the proportion of opportunity-driven entrepreneurship increased from 40% in 2002 to 58% in 2011 (http://www.gemconsortium.org/country-profile/51).
[7] Indeed, in the 2010s, in certain sectors such as the internet industry, many ambitious entrepreneurs choose the getihu form as it enables them to experiment in a small-scaled environment while enjoying a relatively easy registration procedure (http://news.xinhuanet.com/fortune/2015-03/26/c_1114778026.htm)
[8] Getihu and siyingqiye firms may interact in different ways. First, siyingqiye firms often act as demanders of specialized, high-quality, intermediate goods and services, while getihu, because of their specialized activity and size, often act as suppliers. Second, a big population of (by definition small) getihu firms will together create an evolutionary process of competition, where the 'winners' may want to grow bigger and hence switch to the siyingqiye organizational form. Importantly, we argue that both types of interaction can only take place if the getihu firms are of sufficient quality.

REFERENCES:
Audretsch, David B., and A. Roy Thurik (2000), "Capitalism and Democracy in the 21st Century: From the Managed to the Entrepreneurial Economy", Journal of Evolutionary Economics 10(1), 17–34.
Baumol, William J. (1990), "Entrepreneurship: Productive, Unproductive and Destructive", Journal of Political Economy 98(5), 893-921.
Huang, Yasheng (2010), "Debating China's Economic Growth: The Beijing Consensus or The Washington Consensus", Academy of Management Perspectives 24(2), 31–47.
Reynolds, Paul, David J. Storey, and Paul Westhead (1994), "Cross-National Comparisons of the Variation in New Firm Formation Rates", Regional Studies 28(4), 443–56.
Zhang, Ying, and André van Stel (2016), "Who should be Running Ahead? The Roles of Two Types of Entrepreneurship in China's Contemporary Economy", Harvard Business School Working Paper, No. 16-086.

Globalization in Retreat?

The Recent Slowdown in International Trade Flows

By Davin Chor
National University of Singapore

After several decades of steady, almost uninterrupted expansion, international trade has run into significant speedbumps in recent years. At the height of the global financial crisis, trade volumes went into a sudden decline as the world plunged into a deep recession. Trade in fact contracted more sharply than world GDP, as the ratio of world merchandize trade (exports plus imports) to GDP decreased from a peak of 52.0% in 2008 to 42.4% in 2009. This "trade collapse" triggered much concern from policy makers, drawing uncomfortable parallels with the freefall in global trade seen during the Great Depression. Meanwhile, trade economists (including myself) sifted through the data to better understand the underlying causes of this sharp decrease.

The worst-case scenario was avoided (thankfully) when world trade rebounded almost immediately in 2010. But as Figure 1 shows, global commerce is not out of the woods yet. Although the merchandize trade-to-GDP ratio has recovered in levels,

the healthy pre-crisis growth trend in this indicator of openness has been replaced by a distinctly flat pattern. In the last few years, this trade ratio

Figure 1: Trade-to-GDP Ratios (World)
Source: World Development Indicators, with author's own calculations.

even crept downward, prompting warnings that the world economy may have hit "peak trade" and that we are in the midst of a "global trade slowdown".

Why is globalization, as measured by international trade flows, apparently in a slow retreat? In this piece, I overview several of the hypotheses that have been put forth in the public domain as potential explanations. Given the recentness of these trends, the discussion here is exploratory (and at times speculative) in nature, and not meant as a definitive anatomy of what has happened to world trade. As a recent VoxEU eBook on this topic makes clear, there is considerable ongoing debate among economists as to what forces provide the best explanation for this trade slowdown (VoxEU 2015). What is nevertheless emerging as a critical issue is the following: Is this trade decline merely the result of cyclical forces, or does it mask m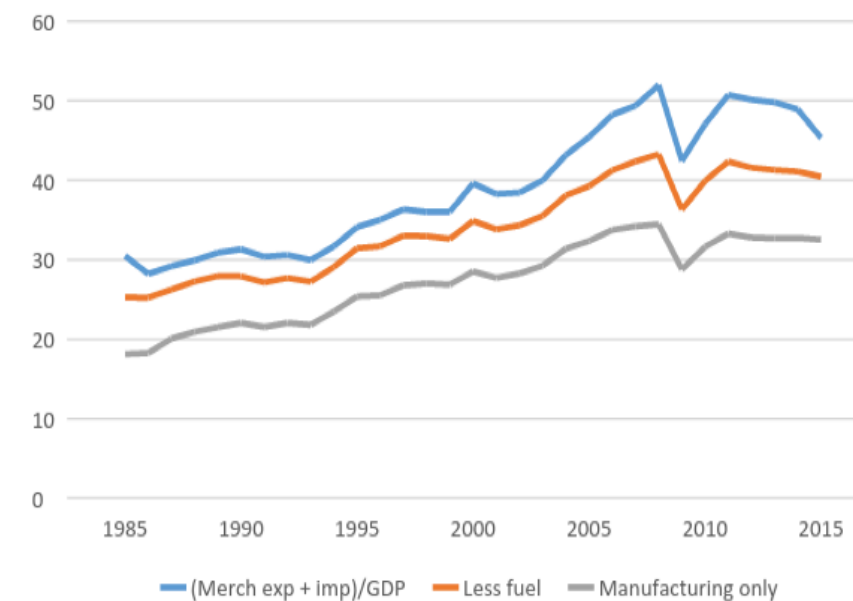ore deep-seated structural changes in the manner in which global production and trade are conducted? The answer to this question will have enormous bearings on the outlook for world trade.

THE TRADE SHOWDOWN IN A (LITTLE) MORE DETAIL

But first, a closer look at the existing data is warranted. A skeptic might point out that in the past few years, crude oil prices have dropped by more than a half, so the apparent decline in world trade might simply be an artefact of lower oil prices. Returning to Figure 1 though, this cannot account for the entire decline, as the world trade-to-GDP ratio still exhibits a decrease after removing trade in fuel-related products from its numerator. When focusing on manufacturing trade (i.e., taking out agricultural and mineral products/commodities), a relatively flat pattern in world trade still remains evident.

Examining the four largest trading countries in the world separately – namely, the USA, China, Germany and Japan – reveals some telling differences. Figure 2 plots the ratio of manufacturing exports plus imports to GDP for each of these countries over time. For the USA, its manufacturing trade ratio mirrors that seen earlier in Figure 1 for the world as a whole: A steady increase leading up to the global financial crisis, a sharp dip followed by a quick recovery, but flat and even falling thereafter. Contrast this on the other hand with the experience of Germany or Japan. In these countries, the ratio of manufacturing trade to GDP has started to show signs of an uptick between 2013 and 2015. The pattern is distinct however for China: Somewhat surprisingly, China's manufacturing trade ratio had already started falling in 2006, with the post-crisis years merely extending what was already a systematic decrease.

The above discussion should therefore caution us against over-generalizing in any search for explanations behind the trade decline. The case of China in particular deserves special attention, given the starkly different pattern it presents. We turn now to these potential explanations.

IS IT CYCLICAL DEMAND?

One view of the current trade slowdown is simply that it is a cyclical phenomenon. The world economy as a whole has struggled to find sustained growth momentum since the global financial crisis. For example, Europe is still sorting its way through the effects of sovereign debt overhang. Elsewhere, China has slowed down as it nears the end of its phase of rapid catch-up growth. World demand thus remains relatively weak when compared against past recoveries from recessions, which could help to account for why growth in world trade has been lackluster. In countries where the short- to medium-term growth outlook remains uncertain, consumers could be adopting a "wait and see" attitude before undertaking purchases of durable goods, while firms could be holding off from fresh investment commitments to expand production and export capacity.

In support of this view, one could argue that there are signs of a possible turnaround in external demand for several key export-oriented countries: As Figure 2 indicates, the manufacturing trade to GDP ratio in both Germany and Japan appears to be on the increase once again. If this cyclical interpretation of events is correct, this current trade slowdown should abate once the world economy gets back on a stronger recovery path. The outlook for world trade will then hinge on how quick and successful countries are in reinvigorating macroeconomic growth and demand.

IS IT A DECLINE IN GLOBAL SUPPLY CHAIN ACTIVITY?

An alternative argument that has gained serious consideration is that the ongoing trade decline is instead being driven by a structural shift in global production arrangements, specifically by an apparent rolling up of some supply chain activity back within country borders.
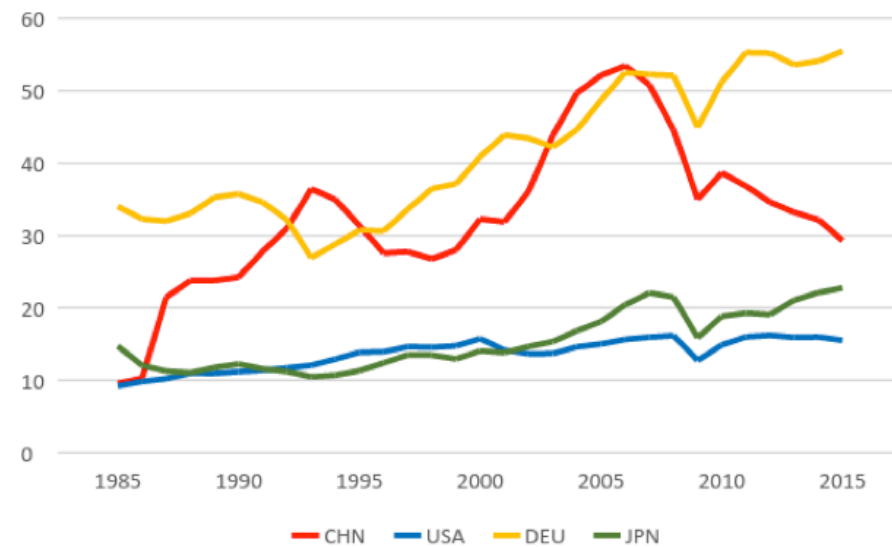
The backdrop to this is of course the rapid rise of global production. Transport and communication costs have fallen to a level where it is now feasible for firms to source for parts and components from around the world, in order to tap into the niche comparative advantage of each country. As a result, trade in intermediate inputs has grown much faster than trade in final goods. In an influential article, Yi (2003) argued that this rise in "vertical specialization" can account for a sizeable portion of the increase in the trade-to-GDP ratio over the later half of the 20th century. There is a straightforward intuition for this: Trade flows are by convention reported in gross values, whereas GDP is a value-added concept. (Recall that GDP is the sum of value-added that is generated in production by domestic entities.) Along a global production process where each country successively adds a little more value to the product in question, the total value of trade in intermediate inputs recorded in gross terms will end up double-counting the value-added that is actually accumulated. An increase in the use of imported intermediates would thus be expected to raise the observed trade-to-GDP ratio.

This trend towards ever more international sourcing appears however to have reached a peak. Figure 3 below provides an illustration of this. Trade economists have developed more sophisticated indicators of the degree of "vertical specialization", but a simple way to capture this phenomenon is through the value of total intermediate imports divided by the value of final goods exports (where the latter refer to goods classified as consumption or capital goods). Broadly speaking, this helps to capture the share of value in final goods exports that was derived from imported inputs. For the world economy, this simple ratio was steadily increasing up until the global financial crisis, consistent with the earlier narrative of a rise in global supply chains. However, the most recent years in this series suggest instead that the use of imported intermediates in global production is now falling. By running the logic in Yi (2003) in reverse, this provides a potential explanation for the decline in the ratio of world trade to GDP.

> For the USA, its manufacturing trade ratio mirrors that seen earlier in Figure 1 for the world as a whole: A steady increase leading up to the global financial crisis, a sharp dip followed by a quick recovery, but flat and even falling thereafter.

Figure 2: Merchandize Trade-to-GDP Ratios (Largest Trading Countries)
Source: World Development Indicators, with author's own calculations.

AS CHINA GOES, SO GOES WORLD TRADE?

The case of China warrants separate consideration, given China's important position as a manufacturing center of the world. Figure 2 earlier showed that the trade-to-GDP ratio for China had started decreasing in the mid-2000s. In Figure 3, we see further that there was a tapering off in the use of imported intermediate inputs by the Chinese economy. This visual evidence is in line with the findings from a careful empirical study conducted by Kee and Tang (2016). Using detailed firm-level and customs data, they document that China's processing export firms have been gradually reducing their use of imported inputs and instead replacing these with domestically-sourced materials and components. This has led to a gradual rise in the domestic value-added content of these firms' exports – from 65% to 70% – between 2000 and 2007.

This should be viewed as a significant development for the Chinese economy. The rapid growth of China in the preceding decade was achieved on the back of the rapid expansion of its manufacturing sector, which established China as a hub for downstream assembly activity: Upstream parts and components were brought in from abroad, Chinese labor inputs were added, and the finished product was then shipped off to consumers around the world. The recent decline in the use of imported intermediates thus points to a shift in the modus operandi of Chinese manufacturing firms and their engagement in global supply chains. What could be behind this shift? Why are Chinese firms apparently reconsidering their optimal production and sourcing practices? One plausible explanation is that China could be attempting to "move up the global value chain". The processing assembly and export activity that has fueled the expansion of Chinese manufacturing is by its nature relatively low in value-added content, since key parts and components – such as the semiconductor chips for computer assembly – are often shipped in from abroad. Instead, China appears to be transitioning into activities that would reduce its dependence on such imported inputs, to perform these high value-added stages domestically; for instance, data on China's exports show that its semiconductor industry has been developing and has in fact gained a small but significant share in the world semiconductor market. Consistent with this view, the share that processing exports occupy in China's total exports has been declining steadily, from 55% in 2004 to 39% in 2013, as reported in Xing (2016).

Whether this is the result of a conscious industrial policy to discourage low value-added manufacturing and encourage other activities remains unclear – policies that govern tax incentives to firms in China tend to be opaque – but this trend of shifting industrial activity within China bears close monitoring. For countries that are reliant on their trade linkages with China as upstream suppliers of intermediate inputs to China's export processing firms, there would be cause for concern if China was systematically rolling back on its use of foreign intermediate inputs. Conversely, other developing countries that continue to have large pools of surplus unskilled labor could stand to benefit, as low value-added manufacturing industries such as textiles and footwear migrate out of China in search of lower labor costs.

CONCLUSION

In this article, we have argued that the outlook for world trade will hinge on the extent to which the slowdown we are witnessing is cyclical in nature, as opposed to being driven by structural shifts in global value chains. With the latter scenario, we could be seeing the start of a significant rearrangement of global production activity, especially if the underlying changes are emanating from the Chinese economy. Apart from these economic considerations, the outlook for world trade is further complicated by the prevailing political climate in many developed countries, where rhetoric and public opinion towards deeper integration (such as through the TPP) has taken a decided turn for the worse in the past few years. Global trade is indeed at a cross-roads.

> For countries that are reliant on their trade linkages with China as upstream suppliers of intermediate inputs to China's export processing firms, there would be cause for concern if China was systematically rolling back on its use of foreign intermediate inputs.

[1] Davin Chor (A.B. 2000, Ph.D. 2007) is an Associate Professor of Economics at the National University of Singapore, with research interests in international trade and political economy. He is a Visiting Scholar at the Department of Economics in Fall 2016.
[2] Based on author's own calculations, with data from the World Bank World Development Indicators; equal to merchandize exports plus imports divided by GDP, all measured in current US dollars.
[3] Several leading explanations emerged to account for this "global trade collapse". These include: the sharp decline in global demand at the height of the crisis (e.g., Eaton, Kortum, Neiman and Romalis (2016)), the tightening of credit conditions affecting the availability of trade finance (e.g., Chor and Manova (2012)), and the decline in trade in intermediate inputs (e.g., Bems, Johnson and Yi (2010), Levchenko, Logan and Tesar (2010)), among others.
[4] As of 2015, the USA, China, Germany and Japan collectively accounted for about one-third of world exports and imports.
[5] For an overview of the rise in global production, see Antràs (2015), Chapter 1.
[6] The UN Broad Economic Categories (BEC) classification system is used here to distinguish between consumption, capital and intermediate goods. The patterns described are qualitatively similarly when excluding products with Harmonized System codes commencing in "27" (mineral fuels and related products). For examples of more sophisticated measures of the degree of cross-border vertical specialization, see Johnson and Noguera (2012), and Koopman, Wang and Wei (2014).

References
Antràs, Pol, (2015), Global Production: Firms, Contracts, and Trade Structure, Princeton University Press.
Bems, Rudolfs, Robert C. Johnson, and Kei-Mu Yi, (2010), "Demand Spillovers and the Collapse of Trade in the Global Recession," IMF Economic Review 58(2): 295-326.
Chor, Davin, and Kalina Manova, (2012), "Off the Cliff and Back? Credit Conditions and International Trade during the Global Financial Crisis," Journal of International Economics 87(1): 117-133.
Eaton, Jonathan, Samuel Kortum, Brent Neiman, and John Romalis, (2016), "Trade and the Global Recession," American Economic Review, forthcoming.
Johnson, Robert C., and Guillermo Noguera, (2012), "Accounting for Intermediates: Production Sharing and Trade in Value Added," Journal of International Economics 86(2): 224-236.
Kee, Hiau Looi, and Heiwai Tang, (2016), "Domestic Value Added in Exports: Theory and Firm Evidence from China," American Economic Review 106(6): 1402-1436.
Koopman, Robert, Zhi Wang, and Shang-Jin Wei, (2014), "Tracing Value-Added and Double Counting in Gross Exports," American Economic Review 104(2): 459-494.
Levchenko, Andrei, Logan Lewis, and Linda Tesar, (2010), "The Collapse of International Trade during the 2008–09 Crisis: In Search of the Smoking Gun," IMF Economic Review 58(2): 214-253.
VoxEU, (2015), The Global Trade Slowdown: A New Normal?, edited by Bernard Hoekman, CEPR Press eBook.
Xing, Yuqing, (2016), "Rising Wages, Yuan Appreciation and China's Processing Exports," GRIPS Discussion Paper 16-01.
Yi, Kei-Mu, (2003), "Can Vertical Specialization Explain the Growth of World Trade?" Journal of Political Economy 111(1): 52-102.
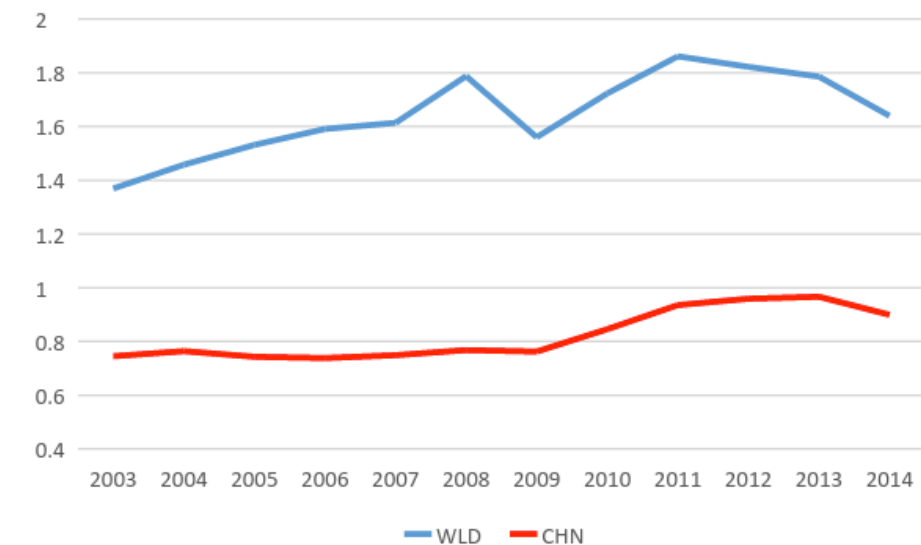
Figure 3: Imports of Intermediates as a Share of Final Goods Exports
Source: BACI trade data, with author's own calculations.

Gains from Trade when Firms Matter

Marc J. Melitz is the David A. Wells Professor of Political Economy, Harvard University, Cambridge, Massachusetts.

Daniel Trefler is the J. Douglas and Ruth Grant Canada Research Chair in Competitiveness and Prosperity, Rotman School of Management, University of Toronto, Toronto, Ontario, Canada

By Marc Melitz and Daniel Trefler[1]

The gains from long-distance international trade have been understood and exploited since prehistoric times. Our pre-urban ancestors were benefitting from long-distance trade in obsidian some 10,000 years ago, Plato's Academy was built on the profits of Athenian silver exports, and Rome was not built in a day partly because goods moved too slowly in the vast Roman trade network. But whereas trade was once dominated by the movement of goods that could only be produced, harvested, or mined regionally, the international trade landscape is now dominated by two striking facts. The first is the rise of intra-industry trade—that is, two-way trade in similar products. Chinese consumers can now buy a midsize car from Toyota (Japan), Kia (Korea), General Motors (United States), and Chery (China). Ditto for consumers in Japan, Korea, and the United States. The second striking fact is that world trade is dominated by huge, innovative and extraordinarily productive firms. For example, Intel is so large that it is the largest industrial employer in both Oregon and New Mexico and accounts for 20% of Costa Rica's exports. China's Foxconn infamously employs 450,000 workers in a single one of its many export-oriented electronics factories. These are big companies … and if you are reading this document on an Apple computer you will know that there are other large companies.

The rising prominence of intra-industry trade and huge multinationals has transformed the way economists think about the gains from trade. In the past, we focused on gains that stemmed either from endowment differences (wheat for iron ore) or inter-industry comparative advantage (David Ricardo's classic example of cloth for port). Today, we focus on three sources of gains from trade: (1) love-of-variety gains associated with intra-industry trade, (2) allocative efficiency gains associated with shifting labor and capital out of small, less productive firms and into large, more productive firms, and (3) productive efficiency gains associated with trade-induced innovation.

Back in the 1980s, a "New Trade Theory" was developed that focused on intra-industry trade in differentiated goods produced subject to increasing returns to scale. This theory centered on an elegant tension: Consumers love variety and are willing to pay a premium for the perfect product, but as the market fragments into niche products, producers struggle to attain the volumes needed to recoup their product development costs. International trade creates a larger market place, which means that each firm can operate at a larger scale and hence more firms can survive. The result reads like an advertisement for free trade: lower prices, more varieties. Paul Krugman earned the Nobel Prize in 2008 in large part for his work highlighting how economies of scale and product differentiation lead to intra-industry trade, just as in our example above of midsize cars. See Krugman (1979, 1980), Helpman and Krugman (1985), and Helpman (2011, Ch. 4) for a review of love-of-variety gains from trade.

More recently, a very different source of gains from trade has emerged from the research of Melitz (2003) and Bernard et al. (2003). This is the firm-level "reallocation" effect that arises when there is firm heterogeneity. By firm heterogeneity we mean that even within narrowly defined industries some firms are much larger and more profitable than others because, for example, they are much more productive. Globalization generates both winners and losers among firms within an industry and these effects are magnified by heterogeneity. Better-performing firms thrive and expand into foreign markets, while worse-performing firms contract and even shut down in the face of foreign competition. This generates a new source of gains from trade: as production is concentrated towards better-performing firms, the overall efficiency of the industry improves. In this way, globalization raises average efficiency within an industry. Why is it that only the better-performing firms grow? Globalization expands markets but also increases competition in those markets. This competition effect dominates for the worse performing firms while the increased market access dominates for the better performing firms. Also, a firm's international expansion – whether by exporting, by offshore outsourcing of intermediate components and assembly, or by building plants abroad (multinationals) – entails some up-front fixed costs; and only the best-performing firms have the sales volumes needed to justify these fixed costs.

Our third source of gains from trade flows from the positive impacts of larger markets on innovation. New productivity-enhancing products and processes require up-front development costs. Trade integration, by expanding the size of the market, encourages firms to pony up these development dollars and this in turn raises productivity. Theories of innovation-based gains from trade with homogeneous firms were developed by Grossman and Helpman (1991) and are supported by country-level evidence (Helpman, 2004, Ch. 5.6). At the firm level, there is a strong relationship between exporting and innovation. For example, Intel and Apple are major patent holders and Foxconn holds 40% of all Chinese patents filed in the United States (Eberhardt et al., 2011). Of course, this correlation between exporting and innovation is not causal and lacks a framing theory featuring heterogeneous firms. Recently, however, there has been a great deal of theoretical and empirical progress. Lileeva and Trefler (2010) show theoretically and empirically how the market-expanding effects of international integration causally encourage firms to innovate. Verhoogen (2008), Bustos (2011) and Aw, Roberts and Xu (2011) assess other interesting channels through which trade promotes firm-level innovation. Note that this third source of gains deals with within-firm efficiency; in contrast, the second source of gains above deals with between-firm or allocative efficiency.

This paper reviews these three sources of gains from trade both theoretically and empirically. Our empirical evidence will be centered on the experience of Canada following its closer economic integration in 1989 with the United States (the largest example of bilateral intra-industry trade in the world); but we will also describe evidence for other countries. The related literature is huge. Here we focus on firms that expand internationally via exporting as in Melitz (2003) and Bernard et al. (2003). Another related research topic analyzes how firm boundaries evolve across borders as multinational firms look abroad to "outsource" key parts of their production chain. The interested reader is directed to surveys by Antras and Rossi-Hansberg (2010) and Helpman (2011, Ch. 6).

Gains from Love of Variety:

Economies of Scale and Product Variety

Our first source of gains from trade is intimately related to intra-industry trade. To measure intra-industry trade, one needs to start with a classification system that assigns trade flows to a particular "industry." One can then categorize trade flows as either intra-industry (two-way trade within the same industry classification code) or inter-industry (imports and exports in separate industry codes). The United Nations uses the Standard International Trade Classification, or SITC, to categorize world trade flows. In its most detailed form, the SITC contains 1,161 separate industry codes (that can be consistently traced back over time), but these industries are often aggregated into a smaller subset of industries.

Figure 1 shows the time trend for the share of intra-industry trade according to this most detailed classification, and a more aggregated version with only 59 industry codes. Mechanically, the share of intra-industry rises with the level of aggregation for the industrial classification system (after all, with a single aggregate industry code, all trade would be "intra" to this aggregated industry). However, the

Figure 1: World Share of Intra-Industry Trade 1962-2006

Source: Data from Bruhlhart (2009). We thank Marius Brulhart for generously sharing his data.
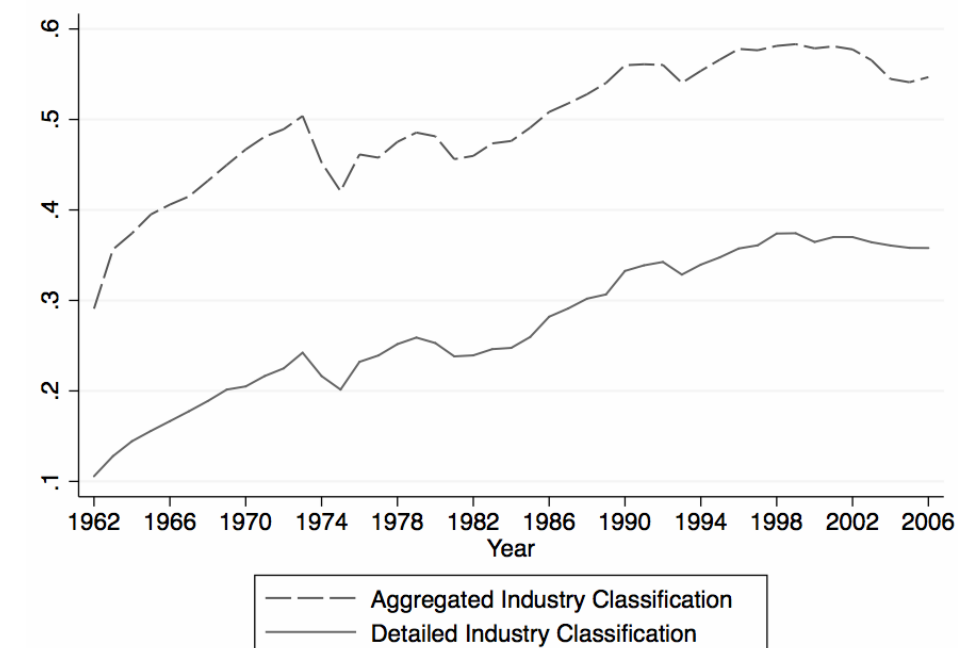
time trends for the two series are very similar: intra-industry trade grew rapidly from 1962 to the mid-1990s, before stabilizing at a substantially higher level. As countries industrialize, they tend to experience a higher share of intra-industry trade, because they tend to produce and export differentiated manufactured goods that are similar to other brands of goods that are imported. However, some of the countries with the highest shares of intra-industry trade in 2000 were newly industrializing nations such as the Czech Republic (77 percent), the Slovak Republic (76 percent), Mexico (73 percent), and Hungary (72 percent). For comparison, the U.S. had a 69 percent share of intra-industry trade in 2000 (OECD Economic Outlook, 2002, Ch. 6; based on the 59-industry level of aggregation). Most recently, China's share of intra-industry trade has risen above the 50 percent mark.

Why might a country both export and import goods that are similar? As a starting point, consider world trade in automobiles. Consumers in a car-producing country are not limited to buying the car models that are produced domestically: many of those consumers choose to buy models that are produced elsewhere and imported. The extent of this product differentiation is then limited by high fixed start-up costs for a new brand

two identical countries, which provides a stark contrast with the gains from inter-industry trade that arise from exploiting differences across countries such as differences in technology (Ricardo) or differences in factor supplies (Heckscher-Ohlin).

In our theoretical example, two identical countries produce differentiated widget varieties subject to the same constant-returns-to-scale technology. Assume that one worker can produce 1 widget, but that production of any new variety of widgets requires 4 workers to cover fixed overhead costs: this implies decreasing average costs of production as the fixed cost is spread over an increasing number of output units (hence the economies of scale). Also to be specific, suppose that both countries have a fixed supply of 12 workers. If they do not trade, then each country can produce: a) 8 units of 1 variety, or b) 2 units each of 2 different varieties.

Allowing countries to trade leads to a new possibility that is better than what either country can achieve on its own. Suppose that each country produces 8 units of 1 variety and exports 4 of these units to the other country. Consumers are now consuming 4 units of the home variety and 4 units of the foreign variety. This is preferred to either of the no-trade production plans above. Compared to choice

product variety. Economic integration allows production of each individual variety to be consolidated for the whole integrated market; given increasing returns to scale, this reduces average production costs. At the same time, product variety increases because consumers can buy varieties produced anywhere in the integrated market.

One of the most salient real-world examples of economic integration between similar countries occurred between the United States and Canada. This integration started with the signing of the North American Auto Pact in 1964. Before then, most car models were produced in the United States for U.S. consumers and in Canada for Canadian consumers. High tariffs on auto trade made it uneconomical to export most car models across the border. Since the Canadian auto market was roughly one-tenth the size of the U.S. market, this implied substantial scale disadvantages for production in the Canadian market: labor productivity there was about 30 percent below the U.S. level. The U.S. market was large enough that assembly lines could be dedicated to one particular car model, while Canadian assembly lines had to switch across models, involving costly down-time and reconfiguration costs, while also holding substantially higher inventory levels. The 1964 Pact established a free trade area

Canada increased as the remaining models produced in Canada supplied the U.S. market as well as the Canadian one. Canadian automotive exports to the United States increased from $16 million in 1962 to $2.4 billion in 1968. That same year, U.S. automotive exports to Canada were valued at $2.9 billion: intra-industry trade in action. Today, $85 billion worth of automotive products cross the U.S.-Canada border each year – roughly half in each direction. The productivity gains associated with this consolidation were also substantial: by the early 1970s, the Canadian auto industry's 30 percent labor productivity shortfall relative to its U.S. counterpart had disappeared.

Later, this transformation of the automotive industry was extended to include Mexico. In 1989, Volkswagen consolidated its North American operations in Mexico, shutting down its plant in Pennsylvania. This process continued with the implementation the North American Free Trade Agreement between the United States, Canada, and Mexico. In 1994 Volkswagen started producing the new Beetle for the entire North American market in that same Mexican plant. This consolidation in response to closer economic integration with the United States was not limited to the auto industry. Following the implementation of the Canada-U.S Free Trade Agreement in 1989, each Canadian manufacturing industry experienced a dramatic reduction in its product offerings, concentrating on a smaller number of products (Baldwin, Beckstead, and Caves, 2002; Baldwin, Caves, and Gu, 2005; Baldwin and Gu, 2006b, Bernard et al., 2011). Baldwin, Caves and Gu (2005) also report that the decrease in product offerings was accompanied by substantial increases in production runs for individual products. This process is even evident in the Canadian wine industry, an industry that exclusively produced low-end wines that could not possibly compete with Californian giants such as Gallo. In response to the Agreement, Canadian manufactures dramatically reduced the number of varietals produced and focused on the varietals used to produce ice wine. The industry is now healthier than ever (Beamish and Celly, 2003).

Another prominent example of economic integration began in 1957, when the major countries of Western Europe es-

tablished a free trade area in manufactured goods: the European Economic Community or EEC. Many politicians evinced an old-fashioned Ricardian prediction that German manufacturers would eradicate all their European competitors. The facts did not treat politicians kindly: trade within the EEC grew twice as fast as world trade

during the 1960s, and intra-industry trade as a share of EEC trade more than doubled from 1960 to 1990. The benefits of the original European Community agreement were about 1% of GDP for the largest economies and about 3% of GDP for mid-sized economies such as Belgium (Harrison et al. 1989). (These numbers capture more than

Figure 2: Performance Differences Across Firms

(a)

(b)
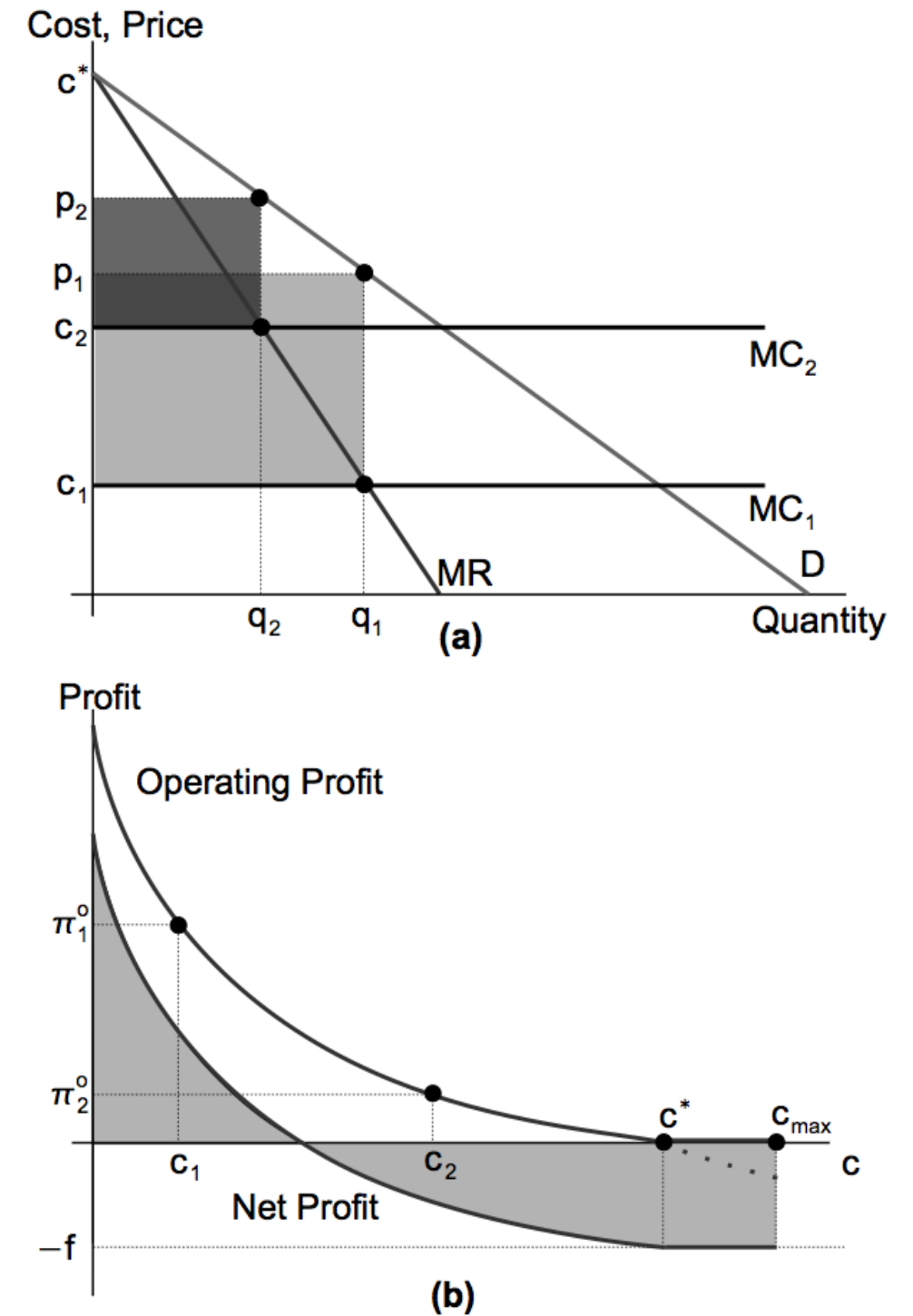
> In 1957, the major countries of Western Europe established a free trade area in manufactured goods (the European Economic Community or EEC). The result was a rapid growth of trade, especially intra-industry trade. Trade within the EEC grew twice as fast as world trade during the 1960s, and intra-industry trade as a share of EEC trade more than doubled from 1960 to 1990.

and by the related economies of scale.

We now highlight how the combination of product differentiation and economies of scale generates intra-industry trade using a theoretical example. Notice that this source of gains from trade provides a rationale for trade between

b, there is the same number of varieties (2 varieties), but more of each variety (4 versus 2). Compared to choice a, there is the same number of units (8 units), but more varieties (2 versus 1). Thus, trade expands the set of consumer choices and eases the tradeoff between consumption units and

for autos that allowed manufacturers to consolidate the production of particular car models in one country, and export that model to consumers in the other country. For example, General Motors cut in half the number of models assembled in Canada. However, total production of autos in

just pure love-of-variety gains.) Economic integration has continued in Europe as more countries have joined the free trade area, which is today called the European Union or EU, and as a subset of EU countries have adopted the Euro as a common currency in 1999. Eurozone members have

tinue to be large for developing countries.

Trade expands product variety both in final goods (which benefits consumers) as well as in specialized production inputs (which benefits the firms that use those inputs). Ethier (1982) showed that there is a close parallel between these two. Instead

port intermediate inputs. Using Hungarian data, Halpern et al. (2005) show that importing many varieties of foreign inputs increases firm productivity by 12 percent. Using Indonesian data, Amiti and Konings (2007) show that a 10 percentage point fall in input tariffs leads to a productivity gain

> " We assume that entrants face some randomness about their future production cost. This randomness disappears only after f is paid and is sunk. Thus, some firms will regret their entry decision as their net profit is negative. This is the case for firm 2 in panel (b). On the other hand, some firms discover that their production cost is very low and earn a high (and positive) net profit. Firms consider all these possible outcomes, captured by the net profit curve in panel (b) of Figure 4 when they make their entry decision. "

experienced strong trade growth, especially intra-industry trade growth, relative to non-EU countries and even relative to EU countries that have not adopted the Euro. A substantial portion of the increased trade that comes with economic integration also delivers increased product variety to consumers. Balistreri et al. (2011) show that the worldwide elimination of all trade barriers would raise the number of varieties available by about 3%, lower manufacturing prices by a similar amount, and raise world welfare by 2%. Most of these gains would accrue to developing countries such as China. Broda and Weinstein (2006) estimate that the number of products available to U.S. consumers through imports tripled between 1972 and 2001, resulting in welfare gains to U.S. consumers equivalent to a 2.6 percent rise in U.S. GDP. Feenstra (2010, table 2.1) examines how worldwide welfare would change if all countries went from autarky to their 1996 levels of trade. He estimates that the welfare gains from increased varieties are comparable to a 12.5 percent rise in world GDP. While the exact magnitudes of the gains from increased variety differ across studies due to differences in what exactly is being modeled, the main message here is that the gains have been very large for developed countries and con-

of the love-of-variety that accrues to consumers, firms benefit from the increased productivity derived from an increased range of available production inputs. Recent firm-level research has confirmed this product variety benefit for firms that im-

of 12 percent for firms that import their inputs. Kasahara and Rodrigue (2008), Kasahara and Lapham (2007), and Goldberg et al. (2010) show similarly large gains for Chile and India. In the context of the Canada-U.S. Free Trade Agreement, Lileeva and

Figure 3: Winners and Losers from Market Integration (a)

(a)
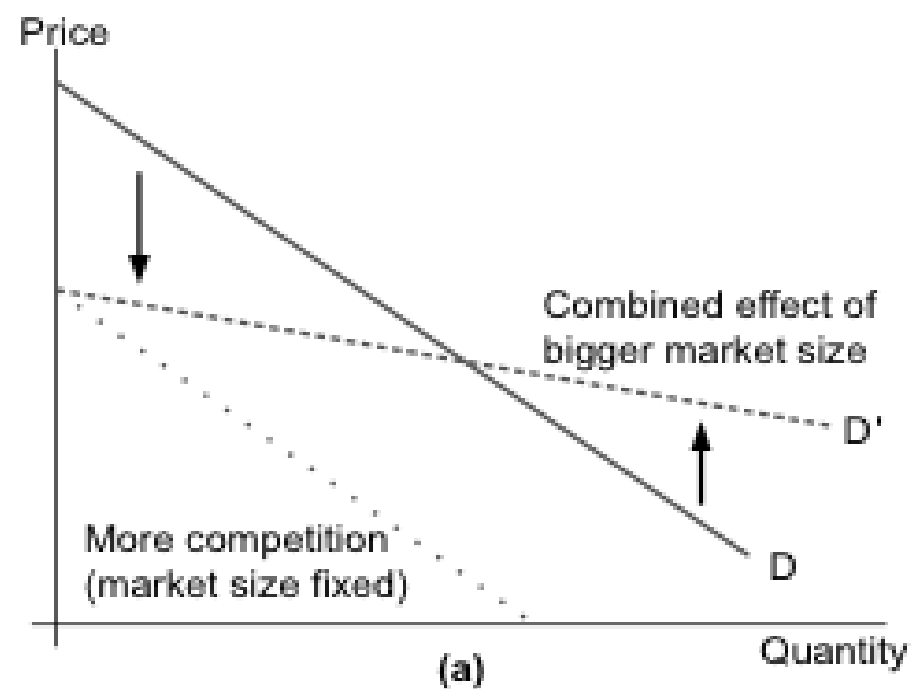
Trefler (2010) find that the fall in Canadian tariffs on inputs that Canadian firms purchased from the United States resulted in a 0.5% rise in Canadian manufacturing productivity. The Canadian impacts are not nearly as large as impacts from developing countries, which suggests that access to a variety of inputs is an essential ingredient in the process of economic development.

More variety means more competition and more competition forces firms to lower their markups and prices. We see evidence of this after the Turkish and Cote d'Ivoire trade liberalizations of 1985 (Levinsohn, 1993; Harrison, 1994) and in Belgium during the 1994-2004 period of increased integration (Abraham et al., 2009; De Loecker 2011). On the other hand, there was no evidence of falling markups in Mexico after the trade liberalization of the early 1980s (Tybout and Westbrook, 1996).

This concludes our discussion of the gains from trade associated with love of variety. We turn next to the gains associ-

ated with re-allocation of resources across heterogeneous firms within an industry.

By the mid-1980s there was a large body of theoretical work demonstrating that freer trade could impact productivity by forcing firms to move up or down their average cost curves. Much of the follow-on empirical work assumed that firms were identical and made a variety of assumptions that allowed inferences to be drawn from industry-level data e.g., Harris (1984). We now know that the heterogeneity of firms even within narrowly defined industries is a central feature of the data that cannot be ignored. See, for example, Bernard et al. (2007).

Our second source of gains from trade is the result of shifting resources away from less productive firms and towards more productive firms. To analyze gains from reallocation of trade between firms, we need a model of trade with heterogeneous firms—that is, in which performance varies across different firms. We can then

capture how firms with different characteristics respond differently to trade. Consider the case from the previous example where opening to trade leads to a transition from production plan a in which each country produces 2 varieties to production plan b in which each country produces 1 variety. In the real world, those varieties are associated with the firms that produce them. Opening up to trade therefore implies that 1 of the 2 firms in each country shuts down, while the remaining firms expand production from 2 units to 8 units. But what factors are to explain

MONOPOLISTIC COMPETITION WITH HETEOGENEOUS COSTS

which firms expand and which ones exit?

Gains from Re-allocation at the Firm Level

dian tariffs on inputs that Canadian firms purchased from the United States resulted in a 0.5% rise in Canadian manufacturing productivity. The Canadian impacts are not nearly as large as impacts from developing countries, which suggests that access to a variety of inputs is an essential ingredient in the process of economic development.

Melitz and Ottaviano (2008) develop a model of trade that allows for differences across firms; we use a simplified version of that model for the discussion here. Consider a monopolistically competitive industry in which many firms compete by offering different products that are relatively close substitutes for one another — at least as compared to products in other industries. For simplicity, we assume that each firm produces a single product, that demand for all products is symmetric, and that firms differ only with respect to productivity. Specifically, firms differ only with respect to their marginal costs of production , where i indexes firms. (A number of authors have developed related models that allow firms to produce multiple products. See Eckel and Neary, 2010, Bernard et al., 2011, and Mayer et al., 2011. Also, demand need not be symmetric: there can be product-quality differences across firms. Such product-quality differences lead to very similar predictions for firm performance as the ones we now discuss for cost differences.) Panel (a) of Figure 2 illustrates the price and quantity choices for two monopolistically competitive firms. Both firms face the same

Figure 3: Winners and Losers from Market Integration (b)

(b)

downward-sloping residual demand curve: residual demand is demand as perceived by the firm, and thus depends on the behavior of other competing firms in the market. On the production side, marginal costs for firm 1 are shown as lower than those for firm 2. In Panel (a), firm 1 has a lower marginal cost ( ) than firm 2 ( ). We also assume that economies of scale exist because of a fixed cost that a firm must incur to develop a product and set up its initial production. In this setting, each firm maximizes profit by choosing an output level q that equalizes marginal cost and marginal revenue. Firm 1 chooses a higher output level than firm 2 ( ), associated with a lower price ( ). Firm 1 also sets a higher markup than firm 2: ; this is a consequence of the marginal revenue curve being steeper than the demand curve. Thus, firm 1 earns a higher operating profit than firm 2: , as represented by the shaded areas in Panel (a) of Figure 2. We assume that all firms face the same set-up cost f so firm 1 also earns higher net profits (subtracting the fixed cost f for all firms). Of course, differences in fixed costs would not affect marginal costs, and thus would not affect firm decisions concerning price and output. We can thus summarize the relevant performance differences that result from marginal cost differences across firms in the following way. Compared to a firm with higher marginal cost, a firm with a lower marginal cost will: 1) Set a lower price, but at a higher markup over marginal cost, 2) produce more output, and 3) earn higher profits.

Panel (b) in Figure 2 shows how firm profit varies with its marginal cost. Both operating and net profit will be decreasing functions of marginal cost, while the difference between the two is the fixed set-up cost f. Going back to panel (a), we see that a firm can earn a positive operating profit so long as its marginal cost is below the intercept of the demand curve on the vertical axis. Let denote this cost cutoff. A firm with a marginal cost above this cutoff is effectively "priced out" of the market and would earn negative operating profits if it were to produce any output (represented by the dotted segment for operating profit in Panel b). Such a firm would choose to shut down and not produce (earning zero operating profit but incurring a net profit loss due to the fixed cost). Why would such a

firm enter in the first place? Clearly, it would not if it knew about its high cost prior both to entry and to paying the fixed cost f.

We assume that entrants face some randomness about their future production cost. This randomness disappears only after the set-up cost f is paid and is sunk. Thus some firms will regret their entry decision, as their net profit is negative (they cannot recover the sunk cost f). This is the case for firm 2 in panel (b); even though its operating profit is positive, it does not cover the sunk cost f. On the other hand, some firms discover that their production cost is very low and earn a high (and positive) net profit.

Firms consider all these possible outcomes, captured by the net profit curve in Panel (b) when they make their entry decision. Firms anticipate that there is a range of lower costs where net profits are positive (shaded area to the left above the horizontal axis), and another range of higher costs where net profits are negative (shaded area to the right below the horizontal axis). In the long run equilibrium, firms enter until their expected net profit across all potential cost levels is driven to zero. If every cost level from 0 to is equally likely, then this equilibrium is reached when the two shaded areas are equal. Panel (b) of Figure 2 summarizes the industry equilibrium for a given market

> "Putting together the downward shift in domestic operating profits and the upward shift in export operating profits, we see that trade liberalization generates both winners and losers – just as in the case of economic integration. Non-exporters lose because they only incur the losses from the lower domestic profits. Exporters, on the other hand, stand to gain as they can make up for the loss of domestic profits with profits earned from exporting."

WHAT CHANGES WHEN ECONOMIES INTEGRATE?

size. It shows which range of firms survive and produce (with cost below), and how their profits will vary with their cost levels .

How will the situation faced by these heterogeneous firms alter when economies integrate into a single larger market? A larger market can support a larger number of firms than a smaller market, which implies more competition in the larger market. Increased competition – absent any increase in market size – leads to an inward shift of each firm's residual demand curve. On the other hand, holding competition fixed, a larger market rotates out the residual demand curves for all firms. Putting these two effects of increased competition and greater market size together gives us the combined effect of international on the residual demand curve perceived by firms. This change is depicted in Panel

(a) of Figure 3, as the shift from demand curve D to. The residual demand curve shifts in from the perspective of the smaller firms with lower output levels that operate on the higher part of the demand curve: here, the effect of tougher competition dominates. However, from the perspective of the larger firms that operate on the lower part of the demand curve, the residual demand curve has shifted out: here, the effect of the larger market size dominates.

Panel (b) of Figure 3 shows the consequences of this demand change for the operating profits of firms with different cost levels. The decrease in demand for the smaller firms translates into a new lower cost cutoff : Firms with the highest cost levels (above) cannot survive the decrease in demand and are forced to exit. On the other hand, the flatter demand curve is advantageous to firms with the lowest cost levels: they can adapt to the increased competition by lowering their markup (and hence their price) and gaining some additional market share. (Recall that the high cost firms are already setting low markups, and cannot lower their prices to induce positive demand, as this would mean pricing below their marginal cost of production.) Thus, the best-performing firms with the lowest cost levels now earn increased operating and net profits. Panel (b) of Figure 3 illustrates how increased market size generates both winners and losers amongst firms in an industry. Low cost firms thrive and increase their profits and market shares, high cost firms contract, and the highest cost firms exit.

In this model, economic integration through market expansion does not directly affect firm productivity. Nevertheless, it generates an overall increase in aggregate productivity as market shares

TRADE COSTS, EXPORT DECISIONS AND TRADE LIBERALIZATION

are reallocated from the low productivity firms with high marginal costs to the high productivity ones with low marginal costs. The discussion to this point has implicitly modeled economic integration as a change in market size from a closed economy with no trade all the way to a single combined market with no trade barriers. In reality, initial trade costs are rarely

so high as to block all trade prior to liberalization; and liberalization reduces trading costs without fully eliminating them. In a number of ways, this kind of partial trade liberalization has a very similar effect to the simpler case of full integration. With partial trade liberalization, the better performing firms expand, the worse performing ones contract, and the worst performing ones

exit. This generates the same type of reallocation effect previously described and leads to a rise in aggregate productivity.

However, adding trade costs also allows us to analyze an additional issue: whether firms choose to export or not. With trade costs, exporting is profitable only for a subset of better-performing firms. Some firms do not export, and

Figure 4: Export Decision and Trade Liberalization
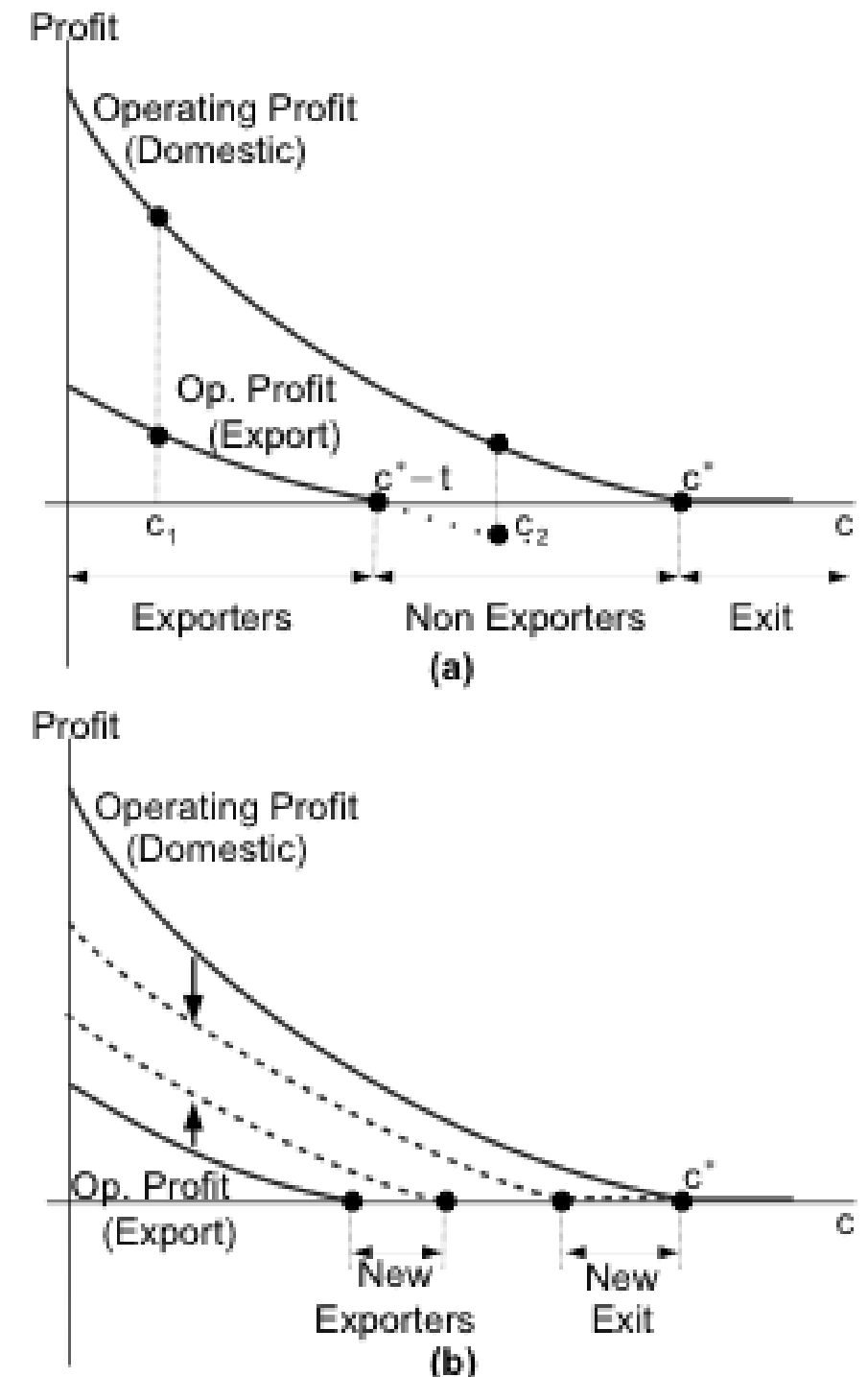
instead only serve domestic consumers. We now extend our theoretical model to incorporate trade costs and firms' export decisions. For this purpose, we can no longer analyze a single market: instead, we need to jointly look at firms' decisions in both the domestic and export markets. For simplicity, we consider a special case where both countries are symmetric, so that demand conditions in both the domestic and export markets will be identical.

Assume that a firm must incur an additional trade cost t for each unit of output that it sells to customers across the border. As a result of this trade cost, each firm will set a different price in its export market relative to its domestic market, which will lead to different quantities sold in each market, and ultimately to different profit levels earned in each market. Because we are assuming that each firm's marginal cost is constant and does not vary with production levels, the decisions regarding pricing and quantity sold in each market can be separated: a decision regarding the domestic market will have no effect on the profitability of different decisions for the export market.

Consider the case of firms located in Home. Their situation regarding their domestic (Home) market is exactly as was illustrated in Figure 2, except that all the outcomes such as price, output, and profit relate to the domestic market only. Now consider the export (Foreign) market. The firms face the same demand curve in Foreign as they do in Home (the two countries are identical). The only difference is that each firm's marginal cost in the export market is shifted up by the trade cost t. What are the effects of the trade cost on the firms' decisions regarding the export market? A higher marginal cost induces a firm to raise its price, which leads to a lower output quantity sold and to lower profits (as highlighted in Figure 2). We also know that if marginal cost is raised above a threshold level , then a firm cannot profitably operate in that market. Thus, when there are trade costs, some firms will find it profitable to operate in the domestic market but not in the export market because the trade cost pushes their marginal cost for that market above the threshold .

Figure 4 helps to visualize the production and export decisions for all firms based on their marginal cost . Panel (a) of Figure 4 separates a firm's operating profit into a portion earned from domestic sales, and a portion earned from export sales. (Both portions are functions of a firm's marginal cost as in Figure 2.) Because the only difference between the domestic and export markets is the additional per-unit trade cost t, the horizontal distance between the two curves is equal to the trade cost t. Firm 1 earns positive operating profits from sales in both the domestic and export markets: it will export and reach consumers in both markets. This will be the case for all firms with cost below . On the other hand, firm 2 only earns positive operating profits from sales in the domestic market—and thus chooses not to export. Any firm with cost above will be in this same situation and therefore will not export: those firms only serve their domestic market. As before, the worst performing firms with cost above cannot profitably operate at all (even in their domestic market) and therefore exit.

Panel (b) of Figure 4 summarizes the effects of trade liberalization – a reduction in the trade cost t – for those firm decisions. The figure shows the same two operating profit curves from Panel (a) both

Figure 5 – Panel (a) Labor Productivity Distribution of All Canadian Manufacturing Plants1988 and 1996 (employment weighted)
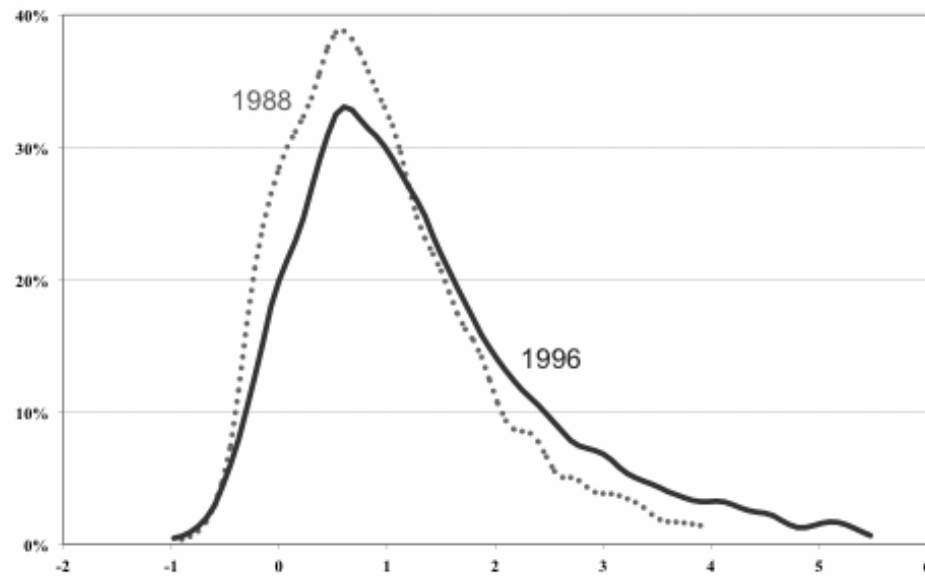
Figure 5 – Panel (b) Labor Productivity Distribution of Entering Canadian Manufacturing Plants 1980-1988 and 1988-1996 (employment weighted)
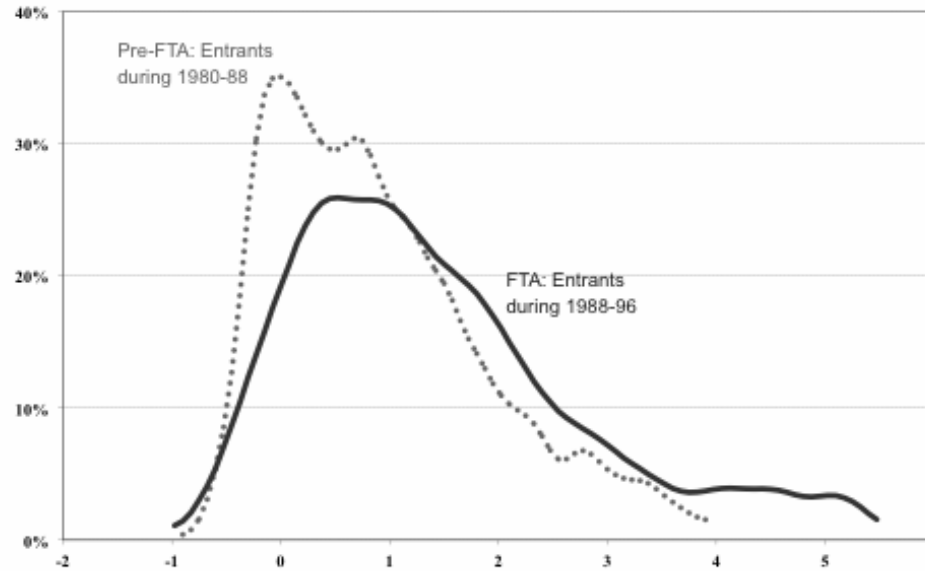
before and after (dashed curves) trade liberalization. The operating profit for the domestic market shifts down due to the increase in competition (which shifts the residual demand curve for the domestic market inward as explained earlier). Some of the higher cost firms that used to produce for domestic consumption no longer earn a positive operating profit after trade liberalization and exit. On the other hand, the operating profit for the export market shifts up due to the lower trade cost. (Increased competition in the export market tends to reduce operating profits there, but this effect is dominated by the direct effect of the trade cost reduction.)

A key empirical prediction is that some firms start exporting. Specifically, among the firms that did not export prior to trade liberalization, only the most productive of these start exporting.

In many ways, the Canada-U.S. Free Trade Agreement is a useful natural experiment for considering the effects of trade integration. The policy experiment is clearly defined: it dealt only with market integration and was not part of a larger package of macroeconomic reforms that often accompany trade liberalization. The enactment of the agreement was largely unanticipated: a Canadian general election was fought on the issue one month before the agreement was to be signed into law and pollsters unanimously predicted that Canada's ruling party – along with the free trade agreement – would be defeated (Brander, 1991; Thompson, 1993). Thus, evidence about the extent of aggregate productivity changes as a result of reallocations among heterogeneous firms can be sought by looking at the distribution of productivity across Canadian manufacturing plants before and after the agreement, at entrants before and after the agreement, and at the productivity distribution of exporters and non-exporters.

The agreement came into effect on January 1, 1989. Panel (a) of Figure 5 shows the distribution of labor productivity as measured by value-added per employee across Canadian manufacturing plants both before the agreement in 1988 and in 1996, when there had been time for firm adjustments to occur. For example, the 1996 curve summarizes the productivity distribution of all 35,000 Canadian manufacturing plants in that year. Clearly, the distribution of firms shifted rightward: Between 1988 and 1996, the share of low-productivity plants in manufacturing declined and the share of high-productivity plants rose.

The horizontal axis is based on a measure of the log of labor productivity. However, to ensure that dispersion is driven by within-industry rather than between-industry differences in labor productivity, we scale each plant's log productivity by subtracting from it the median log productivity of the plant's 4-digit SIC industry. Thus, the median plant in each industry has a score of zero on the horizontal axis. The vertical axis shows the share of plants with the indicated level of productivity. These frequencies are weighted by plant employment; otherwise, tiny plants that account for only a tiny fraction of total employment would dominate the figure. To get a sense of the degree of productivity dispersion, consider the horizontal axis of Figure 5 and suppose that log productivity at plant A is one unit higher than at plant B. This is equivalent to saying that plant A is three times more productive than plant B. If A is four units higher than B then A is 50 times more productive than B.

Obviously, labor productivity as shown in Figure 5 is not an identical concept to the horizontal lines showing levels of marginal cost in the theoretical discussion. When marginal costs are low, we typically expect productivity to be high. Therefore, the inverse of marginal costs (1/c) is often proxied in empirical work by productivity. The productivity heterogeneity shown for

Evidence on Gains from Inter-Firm Reallocation

Canadian manufacturing firms in Figure 5 is a pervasive feature of all economies including, for example, the United States (Bernard, Eaton, Jensen, and Kortum 2003), many European economies (Mayer and Ottaviano, 2007; Bartelsman et al, 2009), as well as China and India (Hsieh and Klenow, 2009). Wagner (2007) surveys related studies from countries all around the world, report-

Figure 5 – Panel (c) Labor Productivity Distribution of Exporters and Non-Exporters, 1996 (employment weighted)
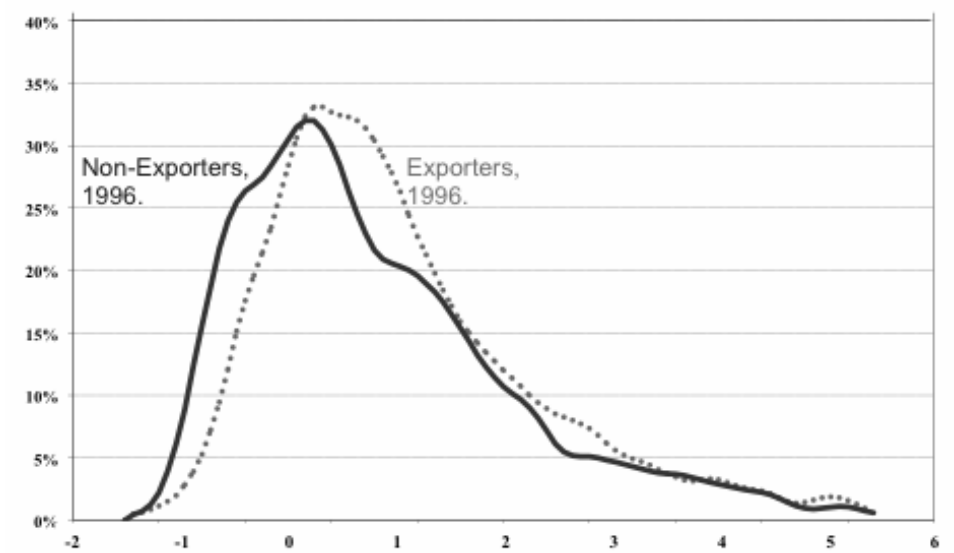
ing similar patterns regarding widespread firm heterogeneity within industries. Thus, the lessons derived from this example are not specific to the Canadian experience.

What caused the change from 1988 to 1996 in the productivity distribution of Canadian manufacturing firms? It was largely due to the reallocation mechanisms across plants described above. The first of these mechanisms that we examine is the fall in the survival threshold of marginal cost i.e., the leftward shift of in panel (b) of Figure 4. The empirical counterpart to a fall in is a rise in the break-even level of productivity. One can examine this mechanism by looking either at exit rates or at entry rates. Since a plant may not exit until it has completed the multi-year process of depreciating its fixed capital, it is best to look at entry rates, which adjust quickly to shocks such as a free trade agreement. Panel (b) of Figure 5 displays the productivity levels of new entrants to Canadian manufacturing both for the pre-agreement period (1980-1988) and for the period immediately after the agreement came into force (1988-1996). There was a striking decline in the entry rates of plants with productivity near or below the median. To use a sports analogy, in the earlier period many low-productivity plants made the cut and joined the team while in the later period a number of such low-productivity plants no longer made the cut.

The pattern here is confirmed by more formal econometric analysis (binary-outcome regressions of exit on plant characteristics as well as industry and time controls). For example, Baggs (2005) and Baldwin and Gu (2006a) show that the free trade agreement tariff cuts raised exit by a large amount. Lileeva (2008) estimates that the free trade agreement tariff cuts raised exit rates by as much as 16 percent, with all of the increase involving the exit

TRADE COSTS AND THE EXPORT DECISION

of non-exporters. Bernard, Jensen and Schott (2006) find similar results for U.S. plants faced with U.S. tariff reductions.

> First, the fall in the U.S. tariffs allowed Canadian plants to export more. This shifted the composition of output towards high-productivity exporters and away from low-productivity non-exporters. Lileeva and Trefler (2010) estimate that the fall in U.S. tariffs causally raised Canadian manufacturing productivity by 4.1 percent via this export-composition channel. Second, the fall in the Canadian tariffs led to a shift in domestic market shares – exporters gained market share at the expense of non-exporters. In the extreme, many non-exporters simply went out of business.

A central prediction of the theory is that in the presence of trade costs, only low-cost, high-productivity firms export. Panel (c) of Figure 5 shows the distribution of Canadian plants separately for exporters and non-exporters. Clearly, the distribution for exporters is to the right of that for non-exporters. On average, Canadian exporters are 40 percent more productive than non-exporters in the same industry (Baldwin and Gu, 2003). Since the seminal work of Bernard and Jensen (1995), a huge body of research covering dozens of countries has found this same pattern of higher productivity for exporters relative to non-exporters.

A much more demanding prediction of the theory deals with who will start exporting in response to falling trade costs. Panel (b) of Figure 4 shows that those who start exporting will be among the most productive of those who never exported before. To test this prediction, Lileeva and Trefler (2010) examined a sample of over 5,000 Canadian manufacturing plants that had never exported prior to the Canada-U.S. free trade agreement. A very large percentage of these plants (40 percent) started exporting after the agreement came into force on January 1, 1989. Lileeva and Trefler examine whether these plants started exporting because of the U.S. tariff cuts and, more importantly, whether those that started exporting because of the tariff cuts were more productive than non-exporters. To this end, Lileeva and Trefler divide up their sample into quartiles of the 1988 distribution of labor productivity (with the quartiles defined separately for each industry, to net out industry characteristics). Only 20 percent of the plants in the bottom quartile of labor productivity started exporting because of the tariff cuts, compared to nearly 60 percent of the plants from the top quartile of labor productivity. (These estimates are from a probit regression in which the dependent variable is 1 if the plant started exporting and 0 if the plant remained a non-exporter. The key independent variable is a plant-specific measure of the change in the U.S. tariff. This measure is described below.) The key conclusion is that, among firms that did not export before trade liberalization, the most productive of these

QUANTIFYING THE GAINS FROM TRADE DUE TO REALLOCATION ACROSS HETEROGENEOUS PLANTS

were three times more likely to start exporting in response to the U.S. tariff cuts. This is as predicted in panel (b) of Figure 4.

> Lower trade costs increase an exporter's sales in the export market, and thus increases the exporter's overall output level q. For some exporters, this increase in output will tip the balance in favor of innovating. For some non-exporters, trade liberalization will tip the balance in favor exporting and innovating.

In the wake of the Canada-U.S. free trade agreement, Canadian manufacturing productivity rose sharply. We have shown that part of this productivity gain was due to the reallocation mechanisms highlighted by the theory. But exactly how important were these for productivity growth and overall welfare?

In the conventional approach to estimating the gains from trade the focus is on measuring welfare, or more specifically, on the income a society would be willing to pay for lower tariffs. These 'compensating variation' gains are typically derived from models that make a large number of parametric assumptions (assuming very

ments in our ability to estimate productivity gains, both because of the creation of high-quality firm-level longitudinal data and because of methodological developments aimed at exploiting these data. Thus, although productivity gains are not the same as welfare gains, we have much greater confidence in our estimates of the productivity gains associated with freer trade.

The productivity gains associated with the reallocation of market shares across firms following the Canada-U.S. free trade agreement are usefully broken into two components. First, the fall in the U.S. tariffs allowed Canadian plants to export more. This shifted the composition of output towards high-productivity exporters and away from low-productivity non-exporters. Lileeva and Trefler (2010) estimate that the fall in U.S. tariffs causally raised Canadian manufacturing productivity by 4.1 percent via this export-composition channel. Second, the fall in the Canadian tariffs led to a shift in domestic market shares – exporters gained market share at the ex-

productivity increase in just a short time – especially when one considers that this productivity gain did not come from productivity improvements at the plant level: it only came from the shifting of market shares from less- to more-productive plants.

Canada is not the only country to have experienced such a substantial productivity boost from reallocations driven by trade liberalization. Bernard and Jensen (2004) find that almost half of all U.S. manufacturing productivity growth during 1983-1992 is explained by the reallocation of resources towards exporters. Pavcnik (2002) studies the response of the Chilean manufacturing sector to a massive trade liberalization episode that took place from 1979-1986. She finds that two-thirds of the ensuing 19 percent increase in productivity (another example of a massive increase in aggregate productivity) was generated by composition changes within industries due to a reallocation of market shares towards more efficient producers. Surveys by Greenaway and Kneller (2007) and Wag-

Gains from Rising Within-Plant Productivity

specific functional forms for preferences that determine the extent of product differentiation, as well as for the utility derived from love-of-variety) and that make use of parameter estimates about which we are highly uncertain. In short, there is a lot of uncertainty surrounding welfare-gain estimates. In the heterogeneous-firms literature the focus has shifted to estimating productivity gains rather than welfare. The last two decades has seen major improve-

pense of non-exporters. In the extreme, many non-exporters simply went out of business. Trefler (2004) calculates that this selection effect increased overall Canadian manufacturing productivity by 4.3 percent.

Putting these numbers together, we see that the reallocation and selection effects induced by the free trade agreement generated a productivity increase of 8.4 percent (= 4.1 + 4.3) for Canadian manufacturing. This represents a massive

ner (2007) summarize the connections between trade liberalization and aggregate productivity—including this reallocation effect across heterogeneous firms—for a wide range of studies and countries.

In the previous section we focused on how trade raises aggregate productivity by allowing productive plants to expand at the expense of less-productive plants. In this section we move from this 'between-plant' effect to a

'within-plant' effect: trade raises productivity of individual plants by raising the returns to innovation. This is our third and last source of gains from trade. At least as far back as Schmookler (1954), economists have known that the larger the market, the more profitable it is for firms to invest in productivity-enhancing activities. Firms in large markets have the large sales volumes needed to justify incurring the high fixed costs of innovation. The U.S. tariff cuts that were part of the U.S.-Canada free trade agreement greatly increased the size of the market faced by Canadian firms. It should therefore have encouraged Canadian firms to increase their exporting and to increase their investments in productivity-enhancing technologies. We start here

A THEORY OF MARKET SIZE AND FIRM INNOVATION

with a short extension to the theoretical model that captures how larger markets generate incentives for some firms to innovate, and then turn to empirical evidence.

Suppose that there is an innovation process that requires an up-front fixed cost and in return generates

weigh this cost saving against the fixed innovation cost , and innovate if   or

In words, only firms with large volumes q (i.e., those with initially lower levels of marginal cost) will find it profitable to innovate. What happens to this firm-level innovation decision when trade is liberalized? Lower trade costs increase an exporter's sales in the export market, and thus increases the exporter's overall

EVIDENCE ON WITHIN-FIRM PRODUCTIVITY GROWTH AND TRADE

output level q. For some exporters, this increase in output will tip the balance in favor of innovating. For some non-exporters, trade liberalization will tip the balance in favor exporting and innovating. For evidence on the link from growth of trade to within-firm productivity, we turn again to Canada's experience with the free trade agreement. Lileeva and Trefler (2010) look at their sample of 5,000 Canadian manufacturing plants that did not export prior to 1988 and divide these plants into those who started exporting after the passage of the free trade agreement and those who did not. In the raw data,

ber does not take into account a serious problem of reverse causality: does exporting lead to increased productivity or does increased productivity lead to exporting?

To answer this question one needs an instrument for exporting. That is, one needs an event that causes a firm to export but that does not directly affect its productivity growth. Lileeva and Trefler (2010) show that 'plant-specific' tariff cuts fit the bill as an instrument. Consider a single Canadian plant (Lumberjack Inc.) and the many products it produces. Empirically, products are defined very narrowly, at the six-digit level of the Harmonized System product classification, so that there are thousands of products in manufacturing. For each product produced by Lumberjack Inc., one can calculate the U.S. tariff cut. Averaging these tariff cuts across all of Lumberjack Inc.'s products yields a plant-specific tariff cut. This plant-specific tariff cut has enormous power in predicting whether a Canadian plant starts exporting and how much it exports. The tariff cut also has no direct impact either theoretically or empirically on a plant's productivity growth. It is thus a novel and valid instrument.

Lileeva and Trefler (2010) actually do something fancier than instrumental

Table 1: Innovation Response to FTA by New Exporters

	Raw Adoption and Innovation Rates			LATE
	New Exporters (1)	Non-exporters (2)	Difference (3)	Difference (4)
Manufacturing Information Systems	16%	6%	10%	7%
Inspection and Communications	18%	10%	8%	8%
Any Product or Process Innovation	30%	20%	10%	8%
Any Product Innovation	26%	14%	12%	7%

a reduction in marginal cost . That is, innovation reduces marginal cost from c to . A firm that produces q units of output and engages in innovation will lower its production costs by . The firm will

the labor productivity of those who started to export rose 29 percent more than for non-exporters: Starting to export was highly correlated with within-plant productivity growth. Of course, this 29 percent num-

variables – they estimate the local average treatment effect (LATE). This is the impact on productivity of starting to export for those plants that started exporting because of the tariff cuts. Thus, unlike all previous

studies of the causal impacts of exporting on productivity, their work only uses information drawn from plants that were likely to be affected by the free trade agreement. Using their plant-specific tariff instrument and the LATE estimator, Lileeva and Trefler (2010) establish that the free trade agreement caused the productivity of new exporters to rise by 15.3 percent. Since this 15.3 percent rise occurred in plants that accounted for 23 percent of Canadian manufacturing output, the 15.3 percent rise in labor productivity raised overall Canadian manufacturing productivity by 3.5 percent (3.5 = 15.3 x 0.23; see Table 2 below).

Lileeva and Trefler (2010) then trace the sources of this productivity gain back to investments in productivity. The authors examine advanced manufacturing technologies and rates of innovation at these plants. Table 1 presents the results. Consider the first row, which deals with management techniques essentially associated with Lean Manufacturing. In the period immediately after implementation of the free trade agreement, 16 percent of new exporters adopted such techniques, 10 percentage points more than non-exporters did. The final column, which reports LATE estimates, shows that 7 of the 10 percentage points was attributable to

report' of Bernard, Jensen, Redding, and Schott (2007) who correctly argue that most careful studies show that exporting does not raise productivity. Over the years however, there have been a few careful studies that find otherwise, as in Canada (Baldwin and Gu, 2003, 2004; Lileeva, 2008), in nine African countries (Van Biesebrock, 2005), and in Slovenia (De Loecker, 2007). See Lopez (2005) for an overall survey.

What has recently buttressed the 'minority view' is a spate of papers isolating the causal mechanisms through which exporting affects productivity. We have already seen the market-size mechanism of Lileeva and Trefler. Bustos (2011a,b) develops a related model of scale-biased technology choice, which she takes to Argentinean firm level data for the 1992-1996 period. Bustos (2011b, table 6) shows that firms that began exporting between 1992 and 1996 also increased their technology spending. Bustos (2011a) shows that technology spending increased most in sectors that experienced improved access to Brazilian product markets (i.e., Mercursor tariff cuts). In a series of studies of Taiwanese electronics exporters, Aw et al. (2007), Aw et al. (2008, 2011) show empirically that there is a complex dynamic interplay between the decisions to export

Table 2: Overall Effect of FTA on Canadian Manufacturing Productivity

Selection/Reallocation (Between Plants)	
Growth of exporters (most-productive plants)	4.1%
Contraction and exit of least-productive plants	4.3%
Within-Plant Growth	
New exporters invest in raising productivity	3.5%
Existing exporters invest in raising productivity	1.4%
Improved access to U.S. intermediate inputs	0.5%
Total	
Total	13.8%

the effects of increased exporting resulting from the U.S. tariff cuts. As shown in Table 1, similar results hold for the adoption of other technologies and for innovation. These results breaks with the 'majority

and conduct R&D, with today's decisions about one affecting tomorrow's decisions about the other – and both affecting future productivity. Extending this dynamic methodology to general equilibrium, Shen

We summarize the causal effects of the free trade agreement on overall Canadian manufacturing productivity in Table 2. As the last row shows, Canadian manufacturing labor productivity rose by 13.8 percent. The idea that a single government policy could raise productivity by such a large amount and in such a short timespan is truly remarkable.

A NEW DIMENSION OF HOMOGENEITY

(2012) also finds strong complementarities between exporting and productivity-enhancing investments among Spanish firms. In our theoretical model above, firms below a certain productivity threshold should not be exporting. Yet in the empirical work reviewed above, we saw that many low-productivity Canadian plants started exporting in response to U.S. tariff cuts. There is a second puzzle that we have not yet noted: Lileeva and Trefler (2010, Table III) report that the plants that gained most from starting to export (both in terms of productivity gains and increased innovation) were primarily plants that initially had low productivity. That is, among plants that started to export, the benefit was greatest for the least-productive plants. To see why, consider a firm that is just indifferent between investing and not investing. From equation (1), indifference means that q = fI /ΔcI, where ΔcI is the reduction in marginal cost or the increase in productivity. Re-arranging (ΔcI = fI /q) and noting that sales q are increasing in initial productivity, we arrive at a simple conclusion. Among the set of firms that are just indifferent between innovating and not innovating, the less-productive, low-q firms must

expect larger productivity gains ΔcI from innovation. Lileeva and Trefler's (2010) results strongly confirm this prediction.

Recent research into the welfare gains from intra-industry trade have focused on three sources of gains: 1) gains from increased variety and economies of scale, 2) productivity gains at the industry level from shifting resources away from low-productivity firms and towards high-productivity firms, and 3) productivity gains at the firm level from innovating for a larger market.. Each of these mechanisms have proven to be highly important empirically in the context of the exhaustively studied Canada-U.S.free trade agreement, and also appear import-

ant in many other less-studied contexts. Indeed, Balistreri et al. (2011) show that adding firm heterogeneity to standard computable equilibrium models of trade raises the gains from trade liberalization by a multiple of four. Empirical confirmation of the gains from trade predicted by models with heterogeneous firms represents one of the truly significant advances in the field of international economics.

We summarize the causal effects of the free trade agreement on overall Canadian manufacturing productivity in Table 2. As the last row shows, Canadian manufacturing labor productivity rose by 13.8 percent. The idea that a single government policy could raise productivity by such a large amount and in such a short timespan is truly remarkable.

In writing this review, we have focused on the net gains from trade. Yet the model we have developed highlights how intra-industry trade will generate both winners and losers. For example, in the context of the Canada-U.S. free-trade agreement, Trefler (2004) shows that 12 percent of the workers in low-productivity firms lost their jobs. Recent research suggests that American workers are similarly struggling in response to the Chinese import surge (Liu and Trefler, 2011; Autor et al., 2011). Clearly, this suggests an important role for

Conclusions

policies that provide an adequate safety net and transitional assistance for those affect-ed workers. There are winners and losers

from trade liberalization not just among firms, but also among their employees.

[1] Financial support for Trefler's research is from the Social Sciences and Humanities Research Council of Canada (SSHRC). His research would not have been possible without the tremendous support of John Baldwin and Alla Lileeva.
[2] Paul Krugman earned the Nobel Prize in 2008 in large part for his work highlighting how economies of scale and product differentiation lead to intra-industry trade and this "new" source of gains from trade. This research is developed in Krugman (1979, 1980) and Helpman and Krugman (1985). The press release by the Nobel committee is available at http://nobelprize.org/nobel_prizes/economics/laureates/2008/press.html. Helpman (2011, Ch. 4) provides an accessible overview of this theory and the associated empirical work.
[3] The data in Figure 1 come from Bruhlhart (2009). That paper documents the many facets of the rise of intra-industry trade

from 1962-2006, and describes all the associated measurement issues. We thank Marius Brulhart for generously sharing his data.
[4] See the OECD Economic Outlook (2002, Ch. 6). All the listed shares of intra-industry trade are based on the aggregated industrial classification code from Figure 1.
[5] Differences in consumer tastes across countries can also be a source of inter-industry trade but most of the world's inter-industry trade is driven by supply-side differences.
[6] The U.S. market was large enough that assembly lines could be dedicated to one particular car model, while Canadian assembly lines had to switch across models, involving costly down-time and reconfiguration costs. It also forced Canadian plants to hold substantially higher inventory levels.
[7] This process is even evident in the Canadian wine industry, an industry that exclusively produced low-end wines that could not possibly compete with Californian giants such as Gallo. In response to the Agreement, Canadian manufactures dramatically reduced the number of varietals produced and focused on the varietals used to produce ice wine. The industry is now healthier than ever. See Beamish and Celly (2003).
[8] The model we introduce in this section is a simplified version of the one developed in Melitz and Ottaviano (2008).
[9] Eckel and Neary (2010), Bernard et al (forthcoming), and Mayer et al (2011) develop related models that further incorporate the firms' production and export decisions across multiple products.
[10] The equation for the demand facing a firm that is used in what follows is , where Q is the quantity of output demanded, S is the total output of the industry, n is the number of firms in the industry, b > 0 is a constant term representing the responsiveness of a firm's sales to its price, p is the price charged by the firm itself, and is the average price charged by its competitors. This demand equation may be given the following intuitive justification: If all firms charge the same price, each will have a market share 1/n. A firm charging more than the average of other firms will have a smaller market share, whereas a firm charging less will have a larger share.
[11] Note that the distance between the AC and MC curves equals the portion of the fixed entry cost f that is paid per unit of output produced.
[12] Product-quality differences between firms would lead to very similar predictions for firm performance as the ones we now derive for cost differences.
[13] This is a consequence of the marginal revenue curve being steeper than the demand curve.

[14] Operating profit can be re-written as the product of the markup times output . Firm 1 sets a higher markup and sells more output, leading to higher operating profits than firm 2.
[15] Firm 2's marginal cost is below the cutoff , so it still earns a positive operating profit and will therefore choose to produce and not exit.
[16] Recall that any firm with marginal cost above the cutoff exits and incurs a net profit loss . This is a lower bound for net profit.
[17] The region above will be important below when we discuss the exporting decision.
[18] Recall that the lower the firm's marginal cost , the higher its markup over marginal cost . High cost firms are already setting low markups, and cannot lower their prices to induce higher demand, as this would mean pricing below their marginal cost of production.
[19] The model developed here is a special case of a more general version analyzed by Melitz and Ottaviano (2008). That paper develops a model with multiple countries of different sizes; and with arbitrary trade costs between any country pair (though the trade costs are proportional to production costs instead of per output unit as in the current version). The paper shows that the effects of multilateral liberalization (all countries proportionally reduce trade costs) are very similar to the case of

full economic integration that leads to a single larger market.
[20] As each firm's marginal cost is constant (does not vary with production levels), those decisions regarding pricing and quantity sold in each market can be separated: a decision regarding the domestic market will have no impact on the profitability of different decisions for the export market.
[21] Because the only difference between the domestic and export markets is the additional per-unit trade cost t, the horizontal distance between the two curves is equal to the trade cost t.
[22] Increased competition in the export market reduces operating profits, but this effect is dominated by the direct effect of the trade cost reduction.
[23] We use plants (a specific production location) as the basic production unit for our empirical analysis; this corresponds to the firms from the theoretical model.
[24] Let φA and φB be productivities of A and B and suppose the log of productivities are 1 point apart i.e., ln(φA) – ln(φB) = 1. From the property of logs, φA / φB = e1 = 2.7 ≈ 3. For a difference of 3 log points, e3 ≈ 20.
[25] To ensure that dispersion is driven by within-industry rather than between-industry differences in labor productivity, we scale each plant's log productivity by subtracting median log productivity for the plant's 4-digit SIC industry. Thus '0' corresponds to the log productivity of the median plant in the industry. Also, the frequencies are weighted by plant employment; otherwise, the figure is dominated by tiny plants that account for only a tiny fraction of total employment. (The median plant had only 12 employees in 1988.) Thus, for example, the average height of the 1996 curve on the interval (-1,0) is about 7%, which means that plants with log productivity between -1 and 0 account for 7% of manufacturing employment.
[26] Similar comments apply below to our discussion of the Figure 8(c) export threshold.
[27] Bernard and Jensen (1999), Trefler (2004), Lileeva (2008) and Lileeva and Trefler (2010) all point out that one must look not just at pre-FTA levels (as in Figure 8), but also at pre-FTA trends. All of the FTA results reported here hold with pre-FTA controls for both levels and trends. For example, variants of some of the panels in Figure 8 with pre-FTA trend controls appear in Lileeva (2008).
[28] Since the seminal work of Bernard and Jensen (1995), a huge body of research covering dozens of countries has found that on average exporters are more productive than non-exporters. This was first confirmed for Canada by Baldwin and Gu (2003) and subsequently by Baggs (2005) and Lileeva (2008).
[29] The more critical reader will wonder why there are so many highly productive non-exporters and whether this contradicts the theory. A simple but prominent example explains why there is no contradiction. Highly productive auto parts plants often cluster around a giant auto assembly plant – Ford, General Motors, and Honda all have major auto assembly plants near Toronto, Canada that are surrounded by parts suppliers. These parts suppliers are highly productive, but do not directly export. This is clearly not a challenge to the theory: these highly productive plants are supplying parts that are built into autos that are ultimately exported to the United States: highly productive parts suppliers are 'indirect' exporters.
[30] In order to net out industry characteristics, the quartiles are defined separately for each industry.
[31] In our model, profits play a key role in all mechanisms. Baggs and Brander (2006) confirm that profits move in the expected directions. In particular, they find that falling Canadian tariffs are associated with declining Canadian profits, especially for import-competing firms, while falling U.S. tariffs are associated with increasing Canadian profits, especially for export-oriented firms.
[32] Specifically, Trefler (2004) regressed labor productivity growth in the FTA period relative to the pre-FTA period on U.S. and Canadian tariff cuts mandated by the FTA. He then showed that the Canadian tariff cuts raised productivity at the industry level, but not at the plant level. This means that the gains in productivity were coming from selection, rather than from improvements at the plant level. Using this approach, he finds that the FTA raised manufacturing labor productivity by 5.8% (t = 3.79) of which 4.3% was due to the exit associated with the Canadian tariff cuts.
[33] Bustos (2011), Aw, Roberts and Whinston (2007), and Aw, Roberts and Xu (2011) have also empirically connected exporting with increased innovation at the firm level. See Burstein and Melitz (forthcoming) for a survey of the literature that in-

corporates exporting and firm-level innovation.
[34] For simplicity, assume that all firms have an initial marginal cost c above so that the marginal cost of production after innovation, , is positive for all firms.
[35] These (5,000) plants accounted for 23% of Canadian manufacturing output and 3.5% (= 15.3% × 23%).
[36] The blow to those workers could also be cushioned by policies that impede the reallocation process across firms. However, such policies (as opposed to policies that provide some form of direct assistance to the affected workers) would also entail

a substantial long run cost. It is precisely this reallocation process that generates some of the long run gains that we have described. In addition, policies that impede the reallocation process (by making firm contractions and expansions costlier) would also reduce the potential gains to firm innovation and hence lead to less innovation and further depress the potential long run gains from trade.

References

Melitz, Marc J., and Trefler, Daniel. 2012. "Gains from Trade when Firms Matter." Journal of Economic Perspectives 26(2): 91-118.
Amiti, Mary, and Jozef Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." The American Economic Review 97: 1611-1638.
Antras, P., and E. Rossi-Hansberg. 2009. "Organizations and Trade." Annual Review of Economics 1(1): 43-64.
Aw, B. Y., M. J. Roberts, and T. Winston. 2007. "Export Market Participation, Investments in R&D and Worker Training, and the Evolution of Firm Productivity." World Economy 30(1): 83-104.
Aw, Bee Yan, Mark J Roberts, and Daniel Yi Xu. 2011. "R&D Investment, Exporting, and Productivity Dynamics." American Economic Review 101(4): 1312-1344.
Baggs, Jen. 2005. "Firm Survival and Exit in Response to Trade Liberalization." Canadian Journal of Economics 38(4): 1364–1383.
Baggs, Jen, and James Brander. 2006. "Trade Liberalization, Profitability, and Financial Leverage." Journal of International Business Studies 37(2): 196–211.
Baldwin, John R., and Wulong Gu. 2003. "Export-Market Participation and Productivity Performance in Canadian Manufacturing." Canadian Journal of Economics 36(3): 634–657.
———. 2006a. "Plant Turnover and Productivity Growth in Canadian Manufacturing." Industrial and Corporate Change 15(3): 417–465.
———. 2006b. The Impact of Trade on Plant Scale, Production-Run Length and Diversification. Statistics Canada,. Working Paper.
———. 2004. "Trade Liberalization: Export-Market Participation, Productivity Growth and Innovation." Oxford Review of Economic Policy 20(3): 372–392.
Baldwin, John R., Desmond Beckstead, and Richard Caves. 2002. Changes in the Diversification of Canadian Manufacturing Firms (1973-1997): A Move to Specialization. Statistics Canada,. Analytical Studies Branch Research Paper Series.
Baldwin, John, Richard E. Caves, and Wulong Gu. 2005. "Responses to Trade Liberalization: Changes in Product Diversification in Foreign and Domestic Controlled Plants." In Governance, Multinationals and Growth, eds. Lorraine Eden and Wendy Dobson. Cheltenham, UK: Edward Elgar Publishing.
Beamish, Paul W., and Nikhil Celly. 2003. Vincor and the New World of Wine. Harvard Business School,. Case Study.
Bernard, A. B., and J. B. Jensen. 1999. "Exceptional Exporter Performance: Cause, Effect, or Both?" Journal of International Economics 47(1): 1-25.
Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum. 2003. "Plants and Productivity in International Trade." American Economic Review 93(4): 1268-1290.
Bernard, A. B., J. B. Jensen, and P. K. Schott. 2006. "Trade Costs, Firms and Productivity." Journal of Monetary Economics 53(5): 917-937.
Bernard, Andrew B., and J. Bradford Jensen. 1995. "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987." Brookings Papers on Economic Activity Microeconomics: 67–112.
———. 2004. "Why Some Firms Export." Review of Economics and Statistics 86(2): 561–569.
Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. forthcoming. "Multi-Product Firms and Trade Liberalization." Quarterly Journal of Economics.
Van Biesebroeck, J. 2005. "Exporting Raises Productivity in Sub-Saharan African Manufacturing Firms." Journal of International Economics 67: 373-391.
Bloom, Nicholas, Mirko Draca, and John Van Reenen. 2011. "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity." NBER Working Paper.
Brander, James A. 1991. "Election Polls, Free Trade, and the Stock Market: Evidence from the 1988 Canadian General Election." Canadian Journal of Economics 24(4): 827–843.
Broda, C., and D. E. Weinstein. 2006. "Globalization and the Gains from Variety." Quarterly Journal of Economics 121(2): 541-585.
Brülhart, Marius. 2009. "An Account of Global Intra-industry Trade, 1962–2006." The World Economy 32(3): 401-459.
Burstein, A., and M. Melitz. forthcoming. "Trade Liberalization and Firm Dynamics." Advances in Economics and Econometrics: Theory and Applications, Econometric Society Monographs.
Bustos, Paula. 2011. "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms." American Economic Review 101(1): 304-340.
Eckel, Carsten, and J. Peter Neary. 2010. "Multi-Product Firms and Flexible Manufacturing in the Global Economy." The Review of Economic Studies 77(1): 188 -217.

Feenstra, Robert C. 2010. Product Variety and the Gains from International Trade. Cambridge MA: MIT Press.
Feinberg, Susan E., and Michael P. Keane. 2006. "Accounting for the Growth of MNC-Based Trade Using a Structural Model of U.S. MNCs." American Economic Review 96(5): 1515–1558.
———. 2009. "Tariff Effects on MNC Decisions to Engage in Intra-Firm and Arms-Length Trade." Canadian Journal of Economics 42(3): 900-929.
Gaston, Noel, and Daniel Trefler. 1997. "The Labour Market Consequences of the Canada-U.S. Free Trade Agreement." Canadian Journal of Economics 30(1): 18–41.
Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." The Quarterly Journal of Economics 125(4): 1727 -1767.
Greenaway, David, and Richard Kneller. 2007. "Firm Heterogeneity, Exporting and Foreign Direct Investment." The Economic Journal 117(517): 134-161.
Grossman, Gene M., and Elhanan Helpman. 1991. Innovation and Growth in the Global Economy. Cambridge, Mass. and London: MIT Press.
Helpman, E., and P. R. Krugman. 1985. Market structure and foreign trade. MIT Press.
Helpman, Elhanan. 2011. Understanding global trade. Cambridge, MA: Harvard University Press.
Holmes, Thomas J., and Jr James A. Schmitz. 2010. "Competition and Productivity: A Review of the Evidence." Annual Review of Economics 2: 619–642.
Kasahara, Hiroyuki, and Joel Rodrigue. 2008. "Does the Use of Imported Intermediates Increase Productivity? Plant-level Evidence." Journal of Development Economics 87(1): 106-118.
Krugman, P. R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." Journal of International Economics 9(4): 469-479.
———. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." American Economic Review 70(5): 950-959.
Lileeva, Alla. 2008. "Trade Evidence and Productivity Dynamics: Evidence from Canada." Canadian Journal of Economics 41(2): 360-390.
Lileeva, Alla, and Daniel Trefler. 2010. "Improved Access to Foreign Markets Raises Plant-level Productivity…For Some Plants." The Quarterly Journal of Economics 125(3): 1051 -1099.
De Loecker, J. 2007. "Do Exports Generate Higher Productivity? Evidence from Slovenia." Journal of International Economics 73: 69-98.
Lopez, R. A. 2005. "Trade and Growth: Reconciling the Macroeconomic and Microeconomic Evidence." Journal of Economic Surveys 19(4): 623-648.
Mayer, T., and G. I. P. Ottaviano. 2008. "The Happy Few: The Internationalisation of European Firms: New Facts Based on Firm-Level Evidence." Intereconomics/Review of European Economic Policy 43(3): 135-148.
Mayer, T., M. J. Melitz, and G. I. P. Ottaviano. 2011. "Market size, Competition, and the Product Mix of Exporters." NBER Working Paper.
Melitz, M. J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." Econometrica 71(6): 1695-1725.
Melitz, M. J., and G. I. P. Ottaviano. 2008. "Market Size, Trade, and Productivity." Review of Economic Studies 75(1): 295-316.
Nunn, N., and D. Trefler. 2008. "The boundaries of the multinational firm: An empirical analysis." In The Organization of Firms in a Global Economy, eds. E. Helpman, T. Verdier, and D. Marin. , p. 55–83.
Pavcnik, N. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean plants." Review of Economic Studies 69(1): 245-276.
Schmookler, Jacob. 1954. "The Level of Inventive Activity." Review of Economics and Statistics 36(2): 183–190.
Thompson, Aileen J. 1993. "The Anticipated Sectoral Adjustment to the Canada - United States Free Trade Agreement: An Event Study Analysis." Canadian Journal of Economics 26(2): 253–271.
Topalova, Petia, and Amit Khandelwal. 2011. "Trade Liberalization and Firm Productivity: The Case of India." Review of Economics and Statistics 93(3): 995-1009.
Trefler, Daniel. 2004. "The Long and Short of the Canada-U. S. Free Trade Agreement." The American Economic Review 94: 870-895.
Wagner, Joachim. 2007. "Exports and Productivity: A Survey of the Evidence from Firm-level Data." World Economy 30(1): 60-82.
Yeaple, S. R. 2005. "Firm Heterogeneity, International Trade, and Wages." Journal of International Economics 65(1): 1-20.

Migration and the World Economy

By Esteban Rossi-Hansberg
Princeton University

The process of globalization, so lauded until recently, is under attack. Many people blame international trade and migration for the stagnation of their incomes and the difficulties that they face in life. Politicians are, naturally, exploiting this discontent. None of the major candidates in the primaries in the U.S. supported expanding the number of free trade agreements, and the nominees of both parties oppose TPP. Migration is, perhaps, suffering even more. Listening to the debate between Reagan and Bush Sr. in the 1980 presidential primary and contrasting it with the current discussion on immigration reforms is a case in point. The idea that migrants are good for the development of a country is simply not present in the current debate in most countries. At best, people advance arguments on why bringing in highly educated people, like Indian programmers, or allowing some students to stay after their post-graduate degrees might be beneficial. It is hard to find commentators or academics, let alone politicians, who favor opening the borders to anyone, independently of their current skill or socio-economic background. Apart from security issues, which of course should be addressed with proper vetting, is there an economic case for opening borders to immigration more broadly? My current research implies that there is.

The topic of immigration is so loaded because it affects so many dimensions of human life. There are the economic implications of migration, but there are also social, cultural, and political implications. I am an economist, so let me exclusively address the broad economic consequences.

It is natural to view migration with some suspicion. After all, we do not normally invite strangers to our home and allow them to live with us. Our home is our property, and we want property rights to be respected. Respecting property rights is also essential for the good performance of an economy. Furthermore, having a stranger at home strains resources, particularly if he does not pay rent, or not fully. The size of our home feels smaller if one extra person lives there partially free. The same could be said for a country. Foreigners come and use the land and other fixed or inflexible factors (like good institutions), and as a result things are more congested. There is less for the original owners, the original citizens, to use. This is the case if migrants get some of these fixed factors for free (like government services or good institutions), or ownership is unevenly distributed across citizens, which is undoubtedly the case everywhere in the world. Of course, one can question who are the original citizens and who can claim property rights of a country. Although perhaps interesting, let's leave that discussion aside and let's just say that the current citizens are the owners. Should they leave the door open and share the fixed endowments of their country? Is this in the interest of current citizens?

If all resources were fixed and in use and technology was given, the answer to these questions would clearly be no. Otherwise, we would break one of the fundamental mechanisms in economics, namely that returns are diminishing in the quantity of a factor when other factors are constant. The rationale is simply based on congestion. The more there are of us, the more we need to share the other factors, which causes our marginal contribution to decline. Owners of private factors could gain, but citizens that do not own factors lose. And if returns to fixed factors and public resources are shared uniformly between all residents, including migrants, everyone loses. This basic economic logic is sound, and I do not attempt to argue in favor of migration by arguing that the large amount of evidence in its favor is somehow flawed. Instead, my point is that the premise of the argument is violated; that is, many factors apart from labor are not fixed, not all factors are fully utilized, and technology evolves endogenously as a result of investments, spillovers, and diffusion.

Clearly, as we increase the number of people in a country, the returns of capital and other complementary factors increase, in turn increasing capital investments. Similarly, more workers translate to greater expenditures, which incentivize a firm's investment in technology since its innovation costs can be shared more widely across consumers. Finally, a substantial amount of factors are not fully utilized. Consider the cost of land in Manhattan relative to the cost of land in the middle of Montana. Montana is a beautiful state, but the price of land anywhere in the state is clearly lower than in virtually every block of Manhattan. Clearly, land is more extensively utilized in Manhattan than in Montana. There would be plenty of space in the U.S. to create many more cities like NY. Furthermore, if everyone lived in cities like NY, most of the

U.S. would be completely empty. In reality, most of the U.S. is essentially empty, so the lack of space, at least in the U.S., cannot be a fundamental factor limiting the productivity of migrants and other workers. There is plenty of land in the world that is not utilized for economic purposes. Can these mechanisms dominate the diminishing returns resulting from congestion and the scarcity of complementary factors? Could it be the case that more immigrants will increase local welfare for current citizens because these forces dominate? The answer requires us to measure and compare them. There is no way we can simply argue our way around this. Measuring these forces is, however, intricate because people move to locations that are extremely heterogeneous. Some places are beautiful, some are extremely productive, and some are simply where your friends, family, or that local vibe that you like so much, are. Accounting for this heterogeneity is important. Perhaps even more important is accounting for the fact that locations in an economy and across the world are linked via trade networks, migration networks, and production networks. The world is incredibly interconnected and these connections are important to truly measure the effect of migrants. The impact of migration is the sum of their effects throughout the network.

One thing is clear; simply measuring the effect that migrants to a city have on the wages of other similarly skilled agents in the same city misses a tremendous number of effects. What if current residents move to other cities and improve those? What if they leave for the suburbs instead?

What if they decide to obtain a degree or become entrepreneurs? The logic of partial equilibrium measurement that treat locations as isolated islands is flawed, and will result in wrong answers. In economics, it is fairly common to focus on certain details and leave the big picture behind. Such an approach can be particularly perilous when analyzing the effects of migration. With these concerns in mind, I have been working on frameworks that account for many of these channels in order to try to measure the welfare effect of liberalizing migration. I want to think about general migration flows, independently of the skill or wealth of the migrant. Ultimately, the skill and wealth of the current generation is just a temporary, and short lived, characteristic. Future generations will decide their own savings and investments in education based on the returns they face.

Together with Klaus Desmet and David Nagy, I have recently measured the implications of relaxing migration flows in such a framework. The exercise requires large amounts of data for thousands of locations in the world. The goal is to assess the evolution of the world economy over time with current migration restrictions as well as with counterfactual migration costs, and compare them. The results are stark. A reform that liberalizes migration so that 10% of the population moves at impact would yield an increase in real world output in present discounted value of 18%. Such a reform would also cause some extra congestion in Europe and the U.S., which implies that average welfare would increase by 9%, a smaller but still impressive figure.

There is no policy that could be readily applied at the world level for which estimated world benefits are as large. Migration is uniquely powerful in generating good effects. So in economic terms, this is a no brainer not because of implausible or abstract theoretical arguments, but because the measurement of the relevant forces tells us so. If one is worried about secular stagnation, namely the slowdown in growth and innovation, liberalizing migration seems like a much more effective policy than any other one we know. It certainly promises to be more effective than monetary or industrial policy.

Clearly, once we start discussing policy, we need to extend the scope of our thinking beyond economics. Can the world incorporate 10% of migrants without suffering important disruptions in institutions and its social fabric? My guess is that it can; in fact, some of these costs are already accounted for in the welfare numbers presented above. Migrants do create congestion and disrupt societies, but our measurement says that the associated increase in real output alleviates, by far, these costs.

Globalization has brought many economic gains. It can, sometimes, also create losers, particularly when people do not react optimally to the new opportunities it offers. Even if we acknowledge these problems, allowing for the flow of goods and people across regions and countries is still one of the best ways we know to create wealth and wellbeing.

See, Desmet, Nagy and Rossi-Hansberg (2016), "The Geography of Development", Princeton University manuscript.

Are the Unfair Trade Laws Fair?

"Antidumping is the fox put in charge of the henhouse – ordinary protection with a good public relations program."

J. Michael Finger (1992)

By Thomas J. Prusa
Rutgers University

In recent decades many long-established firms and industries have struggled with foreign competition. Profits have fallen and in some cases turned negative. Jobs have been lost. Economists argue that such changes are a natural and important part of development and progress. Schumpeter (1942) called this process "creative destruction."

This ivory tower view is often seen by businessmen and politicians as out of touch with reality. Job change and resource reallocation are painful, often slow, and almost always hard to accept. The unhappiness with job losses is intensified when there is a sense the competition might not be playing fair. In fact, it is often unclear, at least in the short run, whether a domestic industry's struggles are the result of imports being lower less costly (or of higher quality) or whether something devious is going on.

Rather than admitting that foreign firms are more efficient, domestic firms often allege that foreign competitors are succeeding because they are trading unfairly.

Partisan disagreement is seemingly a feature of modern politics, yet politicians of every ilk agree on one issue: the need to ensure our foreign partners are trading fairly. While there are a variety of ways firms or governments can trade unfairly, "dumping" is by far the most frequent allegation. Loosely speaking, dumping denotes a situation when a firm charges a lower price in a foreign market than it charges for the same good in its domestic market or when it exports the good at a price below its production costs.

Cognizant of the problems dumped trade "dumping" could have not just only on import competing firms but also on the political support for trade lib-

eralization, provisions to sanction dumping have been a part of every GATT/WTO agreement. Antidumping (AD) laws allow importing countries to offset the unfair trade by imposing a special tariff called an antidumping duty.

AD is undeniably the 800-pound gorilla of trade protection. Bown (2011a, 2011b) shows that AD is the largest source of trade protection for every country studied. Gallaway, Blonigen, and Flynn (1999) and Messerlin (2001) estimate that the welfare costs associated with AD protection are as large as any commercial trade policy, including agricultural protection.

Whether all of this AD activity is sensible or economically desirable depends crucially on exactly what is meant by "unfair". How are the rules implemented? After all, what one person calls unfair pricing could just as easily be considered fair by

another.

As it turns out, the legal definition of unfair trade has virtually nothing to do with the basic economics conception of unfair. Proponents of AD rely on the fact that neither the public nor most politicians have any an idea of what "unfair" pricing means in practice. Instead, AD proponents often refer to scenarios that rarely, or never, happen in practice. For example, Mastel (1988) rationalizes the need for AD using the example of a foreign monopolist operating in a protected "home" market. In this tale the monopolist earns super-normal profits in its protected home market and exports its excess capacity to other countries, often without regard to whether those exports lose money. While such a scenario might indeed warrant AD duties, in my three decades of research involving thousands of AD cases from dozens of countries I have never seen a case consistent with Mastel's example in my three decades of research involving thousands of AD cases from dozens of countries.

The reality is that AD law allows duties to be levied on normal economic transactions. AD requires no proof that the foreign firm's home market is closed to

pricing. And, as I will explain in the next section, AD does not requires neither any evidence that the foreign firm sold at a lower price abroad than it did in its home market nor or that it lost money on the exports. In short, AD allows duties to be imposed on foreign firms when the exact same pricing by domestic rivals would be deemed entirely legal and fair.

The saying, "the devil is in the details" is certainly apropos (Blonigen and Prusa, 2003b). Finger's (1992) statement that AD is "ordinary protection with a good public relations program" is a particularly succinct and apt description of the reality that is AD.

In this paper I will provide two examples of how AD rules work in practice. I will be begin by describing how the AD rules deem "normal" prices as "unfair" and how, according to the legal definition, this means nearly every firm dumps. Then I will review the practice of zeroing, a highly controversial accounting trick that creates margins out of thin air.

WHAT IS UNFAIR?

The question of how firms should set prices in order to maximize profits is addressed in every introductory microeconomics textbook, e.g., Mankiw (2011). In short, a firm will sell a quantity up to the point where marginal revenue equals marginal cost. The price the firm will receive is then given by the demand curve. For some demand and cost configurations a firm will earn a positive economic profit. For other demand and cost configurations the firm will earn a negative economic profit. One of the important lessons in "Eco 100" introductory economics classes is that simply earning a negative profit does not mean a firm should not sell its product. On the contrary, the "not sell" (or shutdown) condition is a special case.

Under AD law, however, any firm earning a negative profit, even if for just a single quarter, can be found to have dumped and therefore can be subject to high AD duties. To the surprise of nearly all students, preeminent firms like Apple, Microsoft, Intel, Maytag, US Steel, etc.

> "The fact that most firms have not been subject to dumping claims just simply means that their foreign competitors have not made the claim against them, not that they trade fairly."

> "For all intents and purposes, according to the legal definition of fair trade, all firms dump. Students are further surprised to discover that compared to businesses in other countries, U.S. firms are among the most frequently targeted in AD investigations."

have all priced, at one time or another, in such a way that they would have legally have been deemed to have dumped. The fact that most firms have not been subject to dumping claims just simply means that their foreign competitors have not

made the claim against them, not that they trade fairly. For all intents and purposes, according to the legal definition of fair trade, all firms dump. Students are further surprised to discover that compared to businesses in other countries, U.S. firms are among the most frequently targeted in AD investigations (Blonigen and. Prusa, 2016).

The bottom line is that the unfair trade that dumping laws sanction is often, perhaps even mostly, trade that would have never been deemed unfair had it been purely based purely on domestic competition. Politicians often say they want to "level the playing field" when it comes to trade. If they truly meant that, they would amend AD so as to make the unit consistent not just with basic microeconomics but also with the many laws we have to protect competition among domestic firms.

competition. AD requires no proof that the foreign firm's below-cost costs sales prices are inconsistent with profit maximizing

ZEROING

While the WTO provides general guidelines and principles for how the dumping margin should be computed, individual countries have substantial latitude in implementing these guidelines. For example, in the process of computing the AD duty, a government must aggregate the results of comparisons between the normal value and export prices. Hundreds or even thousands of individual transactions are aggregated to produce a single AD duty. Zeroing refers to one particular step in the calculation. Zeroing is the practice of replacing the actual amount of dumping that yields negative dumping margins (i.e., export transactions for which the export price exceeds the calculated normal value) with a value of zero prior to the final calculation of a weighted average margin of dumping for the product under investigation with respect to the exporters under investigation. Because the zeroing method drops transactions that have negative margins, it has the effect of increasing the overall dumping margins.

In practice zeroing is much easier to understand than the formal definition suggests. In Table 1, I present an example of a foreign firm's home and export sales in a given month. I assume that the data in Table 1 represent net prices for separate transactions on a series of dates in the month of September. To keep the example as simple as possible I will assume that each transaction is for the same volume, i.e., one unit.

As seen, prices vary from transaction to transaction in both markets. As is often the case in the real world, on some dates the export price is below the home market price on some dates, on others the export price isand above the home market price on other dates. and oOccasionally, the same price is charged in both the markets.

Under WTO rules, a government can calculate the difference in price on a transaction-by-transaction basis and then compute the weighted average of these price differences, i.e., the individual export transactions are compared with the individual domestic transactions made at or at about the same date as the export transactions concerned.

In column (4) of Table 1, I compute the difference for each comparable transaction. Accordingly, for some comparisons the difference is positive (which means dumping) and for other comparisons it is negative. When I sum the weighted price differences, I find that for all comparable transactions, the cumulative difference is zero. Said In other words,-

Table 1 – An Example of Zeroing

(1)	(2)	(3)	(4)	(5)
	Export		Difference: No	Difference:
Sales date	transaction	Home Mkt transaction	Zeroing	Zeroing
2-Sep	75	90	15	15
4-Sep	75	95	20	20
8-Sep	95	95	0	0
10-Sep	100	95	-5	0
12-Sep	105	95	-10	0
16-Sep	105	105	0	0
18-Sep	110	105	-5	0
20-Sep	115	110	-5	0
24-Sep	120	110	-10	0
Wtd Avg. Price	100	100		
Dumping Value			0	35
Dumping Margin			0.0%	3.9%

differently, the dumping amount (35) for the two transactions with positive dumping is exactly equal in absolute value to the amount (-35) for the five transactions with negative dumping. In this example, as long as the dumped and the non-dumped export transactions are allowed to offset each other, the conclusion using the trans-action-to-transaction method will be that there is zero dumping.

As clean and simple as the above calculations are, the U.S. has had a long practice of not computing the margins as described. Instead, in the process of the transaction-to-transaction comparisons, the U.S. employs the practice of zeroing. The use of zeroing leads to dramatically different margins. To see this, in column (5) of Table 1, I have computed the difference for each comparable transaction using zeroing. Each of the five negative margins is set to zero. In our example, the amount of dumping is 35, which implies a dumping margin of 3.9% (35 divided by the total export value of 900 = 0.039).

Several important insights are gleaned from this example. First, zeroing can never lower the margin. Zeroing only drops negative margins. Second, zeroing treats some foreign prices as if they were something different than they actually were. On both September 12th and 16th 16th, the foreign firm charged $105, but a government using zeroing could treat the September 12th price as if it were just $95. Third, zeroing can be the difference between no dumping and a positive dumping margin; i.e., and determine whether an AD duty is applied at all. Bown and Prusa (2011) estimate that zeroing increases the average dumping margin by 17.50 percent-age points (e.g., a change in the margin of dumping from 20.0 to 2.5 percent) and that, in as perhaps as many as half the cases, zeroing is the entire basis for the determination of unfair pricing determination.

Despite more than 20 adverse WTO rulings, the U.S. continues to insist that zeroing is a fair comparison. In fact, the U.S.'s intransigence on this issue is becoming the stuff of legends – Bown and Prusa (2011) found that zeroing is the most litigated issue at the WTO and, as of mid-2016, additional WTO disputes involving the U.S.'s zeroing practice are currently in process.

Despite the U.S.'s position, it is difficult to imagine any economist considering zeroing a fair comparison. By analogy, consider a wage discrimination case where the allegation is that women are paid less than men. However, as part of the empirical test, the researcher "changed" the wage for all women making more than their comparable male worker. Would any objective jury believe the conclusion of such a study (which, given the methodology, would inevitably be that women makes less than men)? Likewise, any dumping margins computed using zeroing reveal little (to nothing) about actual unfair pricing.

CONCLUDING COMMENTS

Readers of this article might think that the two examples of the perversity of modern AD are the exceptions. While they are right to be suspicious that these examples could be representative of the reality, is that the examples cited here are but two of many travesties are embedded in the AD rules. Boltuck and Litan (1991) and Lindsey and Ikenson (2003) document dozens of highly questionable machinations involved in dumping investigations.

Nearly all economists share the view that AD is broken, from Nobel Prize winners (Stiglitz, 1997) to best-selling textbook authors (Mankiw and Swagel, 2005). Readers interested in learning more about the workings of AD law and the economic effects are encouraged to read the one of thea few very fine literature reviews (Bloni-gen and Prusa, 2003a, 2016; Nelson, 2006).

References
Blonigen, Bruce A. and Thomas J. Prusa, (2003a) "Antidumping", in E.Kwan Choi and James Harrigan (Eds.), Handbook of International Trade. Oxford, U.K. and Cambridge, MA: Blackwell Publishers, 251-84.
Blonigen, Bruce and Thomas J. Prusa, (2003b) "The Cost of Antidumping: The Devil is in the Details." Journal of Policy Reform 6(4) December: 233-45.
Blonigen, Bruce A. and Thomas J. Prusa, (2016) "Dumping and Antidumping Duties" forthcoming in Kyle W. Bagwell and Robert W. Staiger (eds.), Handbook of Commercial Policy. 2016.
Boltuck, Richard and Robert E. Litan (Eds.) (1991) Down in the Dumps: Administration of the Unfair Trade Laws, Brookings Institution Press.
Bown, Chad P. (2011a) "Taking Stock of Antidumping, Safeguards and Countervailing Duties, 1990-2009", The World Economy, 34(12), 1955-98.
Bown, Chad P. (2011b) (ed.), The Great Recession and Import Protection: The Role of Temporary Trade Barriers. Washington, DC: The World Bank.
Bown, Chad P, and Thomas J. Prusa. (2011) "U.S. Antidumping: Much Ado About Zeroing", in William J. Martin and Aaditya Mattoo (eds.), Unfinished Business? The WTO's Doha Agenda, London, UK: CEPR and World Bank 355-392.
Gallaway, Michael P., Bruce A. Blonigen, and Joseph E. Flynn. (1999) "Welfare Costs of U.S. Antidumping and Countervailing Duty Laws", Journal of International Economics 49(2): 211-44.
Finger, J. Michael (1992) "Dumping and Antidumping: The Rhetoric and the Reality of Protection in Industrial Countries." The World Bank Research Observer 7(2): 121-43.
Lindsey, Brink and Daniel J. Ikenson (2003) Antidumping Exposed: The Devilish Details of Unfair Trade Law, Cato Institute: Washington, DC.
Mankiw, N. Gregory (2011) Principles of Economics, 5th edition. South-Western Cengage Learning.
Mankiw, N. Gregory, and Phillip Swagel (2005) "Antidumping: The Third Rail of Trade Policy," Foreign Affairs 84(4): 107-19
Mastel, Greg. (1998). Antidumping Laws and the U.S. Economy. Armonk, NY: ME Sharpe.
Messerlin, Patrick (2001) Measuring the costs of protection in Europe: European commercial policy in the 2000s. Washington, DC: Institute for International Economics.
Nelson, Douglas. (2006) "The Political Economy of Antidumping: A Survey", European Journal of Political Economy 22(3): 554-90.
Prusa Thomas J. (2001) "On the Spread and Impact of Antidumping", Canadian Journal of Economics 34(3): 591-611.
Prusa, Thomas J. (2005) "Anti-dumping: A Growing Problem in International Trade", World Economy 28(5): 683–700.
Prusa, Thomas J. and Edwin Vermulst, (2009) "A One-Two Punch on Zeroing: U.S. – Zeroing (EC) and U.S. – Zeroing (Japan)."World Trade Review 8(1) 187-241.
Schumpeter, Joseph A. Capitalism, Socialism, and Democracy. 3rd ed. 1942. New York: Harper and Brothers, 1950.
Stiglitz, Joseph E. (1997) "Dumping on Free Trade: The U.S. Import Trade Laws," Southern Economic Journal, 64(2), 402-24.
Vermulst, Edwin (2006) The WTO Anti-Dumping Agreement: A Commentary, Oxford: Oxford University Press.

" Readers of this article might think that the two examples of the perversity of modern AD are the exceptions. While they are right to be suspicious that these examples could be representative of the reality, is that the examples cited here are but two of many travesties are embedded in the AD rules. "

A Franc Move:

Charting the Ascendency of Unconventional, Post-Crisis Monetary Policy

By Carter Johnson
Brown University

Seven years since the Global Financial Crisis, it has become abundantly clear that global financial markets post-2008 have increasingly been defined not by the crisis itself, but rather by the world's reaction to it. The most fundamental of differences in a world pre- and post-crisis are not to be found in an investment bank's balance sheet or a homeowner's mortgage. Rather, they are to be found in the actions of a lone government institution: the central bank. Institutions of this kind—in brighter days, the sole arbiters of interest rates and money supply—now serve functions as murky and complex as the subprime bonds of days gone by.

ed to the franc in the midst of the Greek sovereign debt crisis. The peg had become too expensive to maintain in the face of rest of the ECB's QE and a weakened European currency. A sudden appreciation occurred, and within one hour the currency rose nearly 30 percent against the Euro (later to fall down to "just" 16 percent) and nearly 20 percent against the dollar.

2000s as the country fought through what has since been termed a "Lost Decade" of falling prices, declining wages, and low growth. The practice, however, marked a radical departure for the Federal Reserve and Bank of England. Between 2008 and 2010, these two central banks collectively purchased nearly $2 trillion worth of mortgage backed bonds and Treasury securities

THE FRANC TAKES A HIT

In December, global financial markets watched as Europe, still experiencing low and stuttering growth, prepared to announce that its own central bank would also introduce Quantitative Easing (QE). The most sudden reaction, however, came not from investors but from another central bank. On December 24th, the Swiss National Bank (SNB) announced the repeal of the euro-Swiss franc cap that had existed since 2011, when investors had flood-

The chart below says it all. The move was enough to put many traders and brokerages out of business. Large U.S. forex brokerage FXCM needed a $300 million bailout to survive. UBS and Citi each tallied $150 million in losses from their franc positions.

The actions of the SNB—and the Swiss franc move—are indicative of the monetary mire in which central bank policies are now stuck, much of it formed with the implementation of Quantitative Easing programs by the U.S. and the United Kingdom in the aftermath of the crisis. QE was first introduced, with little success, by the Bank of Japan in the early

(in the U.S.) and gilts (in the U.K). QE continued in the U.S. until late 2014, at which point the Fed had collectively bought over $3.7 trillion dollars, increasing its asset holdings by more than eight times in the process. Now, Japan, bolstered by Shinzo Abe, has returned to QE as well. These banks present quantitative easing as the only means of stimulating flagging economies with rates already set at or near the zero bound. In introducing these programs, however, central banks have acknowledged that the old rules aren't working. Their answer has been to essentially rewrite them.

GAME OF LOANS

Given all that has happened, perhaps it isn't surprising to see Bill Gross, ex-PIMCO chief, criticize the very system upon which he has based a successful career. In his February, 2015 letter to investors, Gross—now running

bond strategy at money management firm Janus Capital—compared global monetary markets to the board game Monopoly. "Players [in this case, investors or anyone remotely connected to the financial markets]," he wrote, "would be justified in saying that competitive devaluations and the purchase of bonds at near zero interest rates is indeed a significant distortion of the markets and – more importantly – capitalism's rules which have been the foundation of growth for centuries, long before Parker Brothers central bankers came into existence in the early part of the 20th century."

And so recent central bank moves to "rewrite" the rules of the game call to mind several burning questions: Why do these rules need to be rewritten? What aspects of modern, global financial markets and economies are so fundamentally flawed as to be unfixable under decades-long systems and practices? And, perhaps most ominously, what happens when these ag-

> "And so recent central bank moves to "rewrite" the rules of the game call to mind several burning questions: Why do these rules need to be rewritten? What aspects of modern, global financial markets and economies are so fundamentally flawed as to be unfixable under decades-long systems and practices? And, perhaps most ominously, what happens when these aggressive practices (negative interest rates, Quantitative Easing, and currency devaluations) don't work?"

gressive practices (negative interest rates, Quantitative Easing, and currency devaluations) don't work? After all, trillions of dollars have been created and thrown into the global markets in the hopes of stimulating growth. The only problem? Such stimulation has resulted in a significant divergence between the real and financial economies, and has in no way ensured that money created will ever come to bear in useful or growth-producing investments. Growth in Europe and Japan remains dismal, and to quote Gross, "even with the U.S. growing at an acceptable rate for now, its recovery over the past five years has been

anemic compared to historic norms." These questions, and the complexity of their answers, will define the development of financial markets in the coming decades.

WINTER IS COMING

Nearly a year after the Swiss National Bank dropped the euro-franc peg, all eyes are on the Federal Reserve. After delaying in September the first interest rate hike in nearly ten years, December's FOMC meeting is the last possible event of the year in which the Fed can act to raise rates. And, as Fed Chairman Janet Yellen has noted, it promises to be a "live" meeting, particularly after particularly strong job data released on November 6th. The Fed decision, the timing of which has largely shifted from "date dependent" to "data dependent," in the words of Wall Street analysts and former Philadelphia Fed president Charles Plosser, will affect bond, currency and equity markets around the world. If such a hike marks a return to pre-crisis monetary policy, it also marks a divergence from the easing to which so many central banks have resorted since 2008. Such a divergence in monetary policy, just like the unwinding of post-crisis quantitative easing policies, will set the stage for what promises to be an eventful rest of the year for the markets in 2016.

What Is Good for Banks is
not
Good for the Economy

"The fate of the world economy is now totally dependent on the stock market, whose growth is dependent on about 50 stocks, half of which have never reported any earnings."

Former Fed Chairman Paul Volcker.

By Masoud Movahed

Economists and analysts of the 2007-2009 financial meltdown usually take the domestic housing and securities markets as the point of departure in their prognoses of the crisis. While refusing to look beyond the apparent roots of the malaise, they continue to begrudge the decline in housing prices as the bedrock of the financial crunch. And, that, of course, makes perfect sense. With a housing bubble bursting by the end of 2006 that forced the prices of assets down, a deregulated credit market running on an unbridled debt explosion, and numerous banks plunging into failure and insolvency, the ground was rightly laid for the suspicion that the roots of the plague ought to be investigated in the housing and securities markets. But to accurately diagnose the origins of today's economic depression, one should look beyond the U.S. financial and housing markets. The fundamental source of the crisis today—both in finance and beyond—is the declining economic

vitality and dynamism of the advanced industrialized countries, especially the United States. Contrary to the mainstream account, the crisis is deep-rooted in the zero-sum game that the rapid development of some newly emerging economies abroad has entailed for the U.S. economy as well as the advanced world in general. Stemming from intensifying international competition, U.S. manufacturing firms had to struggle ever more with lower rates of profit, which led to system-wide economic distress. In what follows, I explain how the origins of the crunch should be probed in the intensifying competition in the global manufacturing market, which affected the financial sector at home.

It is no longer an esoteric reality that since the 1970s, the American economy has seen a reconstructing so fundamental that its magnitude is hard to overstate. We hear much about financialization of the economy, which has permitted the stratospheric ascent of finance. The "Old

Economy" of complex machinery and laborious manufacturing has given its way to the "New Economy" of finance, software engineering and information technology. Quibbles among economists notwithstanding, the unmistakable broad trend is that the largest share of the aggregate profits in the economy—estimated to be roughly 40 percent of total profits—is generated in the financial sector (see Figure 1). This datum is often taken as the strongest evidence of the salience of finance in the U.S. economy. Financialization is defined in multiple ways, but Gerta Krippner's definition as "a pattern accumulation in which profits accrue through financial channels rather than through trade and commodity production" seems to capture what has changed about investment and capital accumulation in our economy. Provision or transfer of liquid capital in expectation of future interest, dividends and capital gains are only a few of the many other stratagems of finan-

cial activities, in which investment bankers are the most adroit and innovative.

Naturally, such a tectonic economic change has invited a series of explanations aiming to explicate the roots of the phenomenon. There are those, for instance, who would attribute financialization to the natural progression of capitalist development where the productive sector, namely manufacturing, is subject to intensified international competition and witnesses an enormous reduction in profit rates. Thus, the ascent of finance is a natural response to the stagnating tendency of the manufacturing sector. In this vein, economists such as Paul Sweezy and Harry Magdoff argued that with decline of rates in manufacturing due to intensifying global competition, stagnation instead of dynamism and financialization instead of industrialization, become the twin trajectories of the advanced world. It should be noted, however, that financialization is not endemic to the U.S. economy alone. Finance, across the advanced world with the exception of Germany, has become a salient sector of the economy. Indeed, as I will show later in the essay, decline in the rates of profit in U.S. manufacturing has been the primary reason behind the massive explosion of finance worldwide, and the crisis of 2007 to 2009 by no means can be construed independent of the rise of finance.

There are also those who would take not only the decline in manufacturing profits, but also certain macroeconomic policies of the U.S. Federal Reserve as driving forces of financialization. This perspective has been led most notably by Robert Brenner of UCLA whose book *The Economics of Global Turbulence* demonstrated that the titanic fall in the rates of manufacturing profitability of the advanced economies has to do with the over-capacity in global manufacturing. The implication is that since the late 1960s, the manufacturers of consecutively newly emerging economic powers have been able to make use of the latest technology coupled with relatively lower wages of domestic labor markets to manufacture export goods that were already being produced, but can now be manufactured at a lower cost and offered for a lower price. Germany and Japan in the 1960s, the East Asian Tigers (e.g. South Korea and Taiwan) in

the 1970s and 1980s, and the Chinese behemoth in the 1990s and 2000s have all been able to adroitly acquire significant market share in global manufacturing. For U.S. firms, however, to maintain the same

market share that they had in the early 1950s to 1960s and to remain competitive globally, they have had to offer their output to the international market at lower prices, which translates to lower rates of profit on U.S. firms' balance sheet. The up-
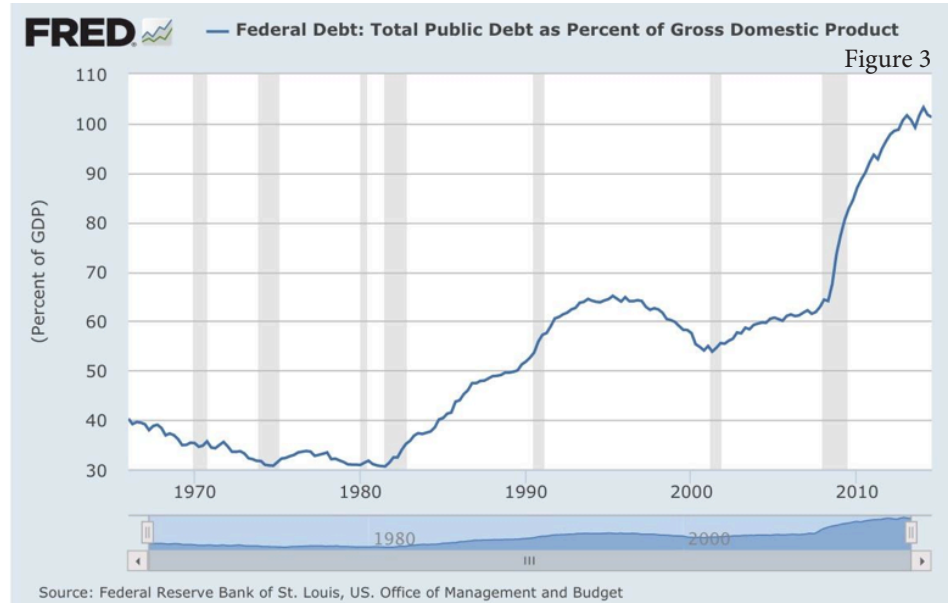
shot was, as Brenner accurately observes, oversupply for low global demand, which depressed not only prices, but also profits.

Hence, the reduction in manufacturing profitability meant that firms had smaller surpluses at their disposal, which itself dampened hiring and labor

demand. This manifested itself conspicuously in the rapid decline of manufacturing employment in the U.S. economy. As a result, by the end of 2010, the sector had lost almost 50 percent of the 22 million

Figure 1

Source: US. Bureau of Economic Analysis

jobs it had at its 1979 postwar peak (see Figure 2). Indeed, slower growth of the sector relative to services (namely finance and information technology), almost daily soaring trade deficits, record manufacturing job losses and factory closures, and

Figure 2

Source: US. Bureau of Labor Statistics
Shaded areas indicate US recessions - 2015 research.stlouisfed.org

massive outsourcing of manufacturing sites to labor-intensive economies (especially China), make it plausible to lament the decline of American manufacturing. Lower rates of profit in the productive sector of the economy, namely manufacturing, ushered in an era of financialization.
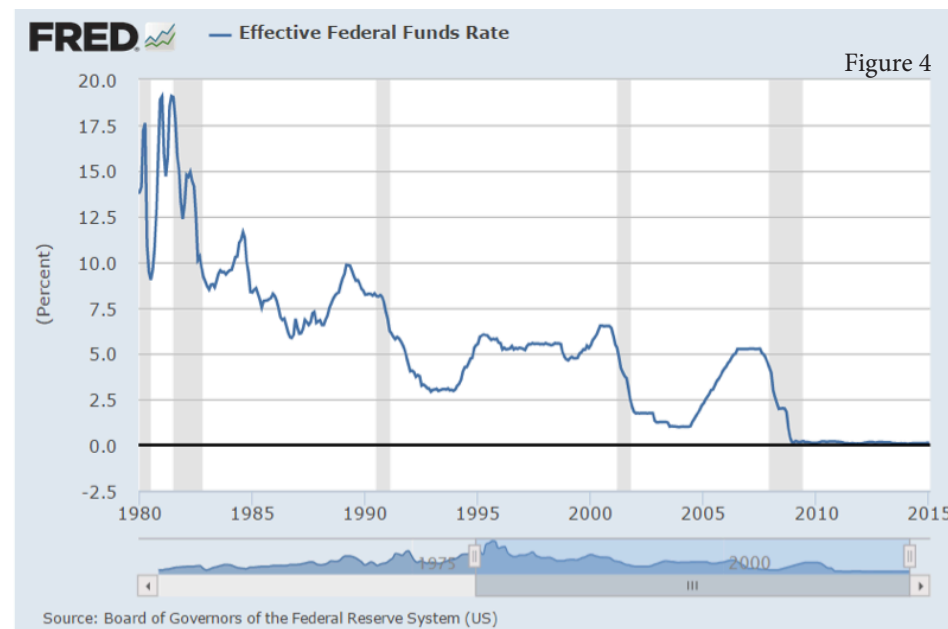
The logic of the rise of finance— as the most thriving and profit-generating sector of the economy—can be explained by the following simple economic rationale. For any given industry to bourgeon, there has to be sufficient demand for the output of that industry. The level of demand—the volume of spending and investment—for a specific industry determines the growth rate of that industry.

because of the growing demand for software and high-speed IT infrastructure, which has invited a mammoth investment in the sector. Little wonder why! Now in such a depressing economic climate, where

Federal Debt: Total Public Debt as Percent of Gross Domestic Product

Figure 3

Source: Federal Reserve Bank of St. Louis, US. Office of Management and Budget

geon, there has to be sufficient demand for the output of that industry. The level of demand—the volume of spending and investment—for a specific industry determines the growth rate of that industry.

the global manufacturing market suffers from over-capacity and as a result reduced profit rates, which by itself disincentives firms from hiring more labor and generating more employment, the economy

Effective Federal Funds Rate

Figure 4

Source: Board of Governors of the Federal Reserve System (US)

That is to say, no sector in an economy—be it in manufacturing or services— can grow if there is no demand for it. For instance, the IT industry has witnessed an unrivaled growth rate in the past two decades simply

continues to show signs of enfeeblement. As Brenner explains, since firms were ever more reluctant to hire labor or raise wages as a result of lower rates of profit, there was no way for the economy to generate

demand—or to encourage spending—other than by way of ever greater borrowing, which meant running the economy on credit. This was essentially dependent upon banks. To boost private spending, the Fed lowered the short-term interest rate in the 1990s, which made highly risky credit available to households, many of which were unqualified (see Figure 3).

As a matter of fact, because of the stagnant growth of wages, many of those households had ever higher debts compared to their incomes. In attempt to pinpoint origins of the housing bubble prior to 2006, two economists at the University of Chicago Atif Mian and Amir Sufi argued that there exists a statistically causal relationship between the massive supply of mortgages and the rapid rise of housing prices which led to the bubble by the end of 2006. Surprisingly enough, they find that the period between 2001 and 2005 is the only one in recent U.S. history where housing prices increased in zip codes that had negative income growth. This is strong evidence that credit was in one way or another supplied in an extraordinarily risky way to ever more unqualified borrowers. In an economy that had already demonstrated sluggish growth rates by the mid 1990s, injections of risky credit by way of lowering short-term interest rates offered a way out of the predicament (see Figure 4). This massive injection of credit became the benchmark economic policy that laid the ground work for the spectacular ascent of finance.

I mentioned earlier that since the 1980s, manufacturing firms have been enfeebled by the decline of profit rates, as were households by the wage stagnation. Corporations along with households were thus enabled to increase their borrowing. Speculative run-ups in asset prices in both the housing and securities markets enabled huge, largely fictitious increases in the wealth of corporations and households. Nurtured by easy credit and deregulation policies of the Federal Reserve Banks, there was a massive run-up in the housing prices between 2000 and 2006. Whenever the run-ups in financial markets led to trouble, the Fed would not hesitate to reduce the short-term interest rates so as to incentivize financial investors to step up their borrowing in

order to correspondingly increase their purchases of housing and financial assets.

CREDIT MARKET EXPLOSION AND BUBBLENOMICS

The key to the complete explosion of the credit market was the Fed's policy of maintaining low short-term interest rates. For banks, that was a license to make money with very little risk, particularly since they can get people to open savings accounts that pay close to nothing. The low short-term rates meant easy access to cheap credit for borrowers who then invested enormously in the stock and securities market. Indeed, what is good for banks is not good for the economy. The incessantly low short-term interest rates in the past two decades thus created an environment conducive to financialization. The flood of easy credit to the stock and housing markets paved the path for the historic run-ups of equity and land prices that ensued during the second half of the decade and provided the increase in paper wealth that was required to enable both corporations and households to step up their borrowing, raise investment and consumption, and keep the economy expanding. The low interest rates of the 1990s and early 2000s created conditions under which firms and households could borrow easily, invest in the housing and stock markets, and push up their prices. So banks took it upon themselves to stimulate growth by enabling corporations and households to increase their borrowing, which precipitated a significant increase in housing and securities prices. With credit made so cheap, and profit-making on lending rendered so easy, banks and non-bank financial institutions could not resist opening the floodgates and advancing funds without limit.

In short, the crisis of 2007 to 2009 was one of the most disastrous financial meltdowns since the Great Depression. While there is a pervasive tendency among economists to look at the financial and securities market to understand the roots of the crunch, a few have departed from them. Viewed against what the mainstream economists put forward, those who

> "The decline of U.S. manufacturing profitability posed serious threats to economic dynamism and vitality as firms were ever more reluctant to hire labor or raise wages. For the Fed to continue generating growth, it had to enable both the public and private sectors—both households and the government—to increase their borrowing."

differ often argue that the origins of the crisis ought to be investigated in the context of the lower rates of profits as a result of intensifying international competition. Decline in the rates of manufacturing profits and certain macroeconomic policies of the Federal Reserve Bank are seen to have provided the impetus towards the financialization of the economy as well as the recent crisis in the financial sector. The decline of U.S. manufacturing profitability posed serious threats to economic dynamism and vitality as firms were ever more reluctant to hire labor or raise wages. For the Fed to continue generating growth, it had to enable both the public and private sectors—both households and the government—to increase their borrowing. This process was entirely dependent on banks. With the policy of low interest rates that provided cheap credit to borrowers, households that struggled with stagnant wages and runaway debt, and housing prices that fell day by day after the bubble burst in 2006, the Fed along with banks continued to stimulate the economy by expanding the credit market. Running the economy on deficit spending in both public and private resulted in the devastating crunch of 2007 to 2009.

Masoud Movahed is a doctoral candidate at the University of Wisconsin-Madison. Previously, he was a Research Associate in Development Economics at New York University. He contributes to, among others, Harvard International Review, Foreign Affairs, Yale Journal of International Affairs, World Economic Forum and

The Value of Economic Statistics

By Diego Perez
New York University

Every month, the Bureau of Labor Statistics and the Bureau of Economic Analysis publish statistics on inflation, employment and several other economic variables. Providing macroeconomic statistics entails an economic cost associated with gathering and processing a large amount of data. At the same time, the public provision of economically relevant information is likely to confer certain benefits on society. After all, the availability of timely and accurate economic data is a common feature of any country with solid institutions. In this column I focus on the benefits associated with the provision of macroeconomic statistics. A proper understanding and measurement of these benefits is key to adequately allocating resources to be invested in the provision of public information.

In a recent paper Price Setting Under Uncertainty About Inflation, joint with Andres Drenik, I quantify the economic effects of providing precise statistics about the aggregate inflation rate. The provision of accurate inflation statistics enhances the availability of information that firms have when setting prices. Using an economic model we show that more precise information is associated with a more efficient allocation of resources in the economy, which in turn increases social welfare. We also find that the increase in welfare is greater in

> When more precise inflation statistics are available, the price of a cup of coffee should be more similar between different stores and brands. The reason is that, with more accurate information about aggregate inflation, firms optimally choose to put more weight in inflation statistics when setting prices; when all firms optimize in this way, their prices are more aligned with each other.

those economies with more economic volatility and less predictable monetary policies.

THE ECONOMIC EFFECTS OF THE PROVISION OF STATISTICS

How does the availability of inflation statistics affect the macro economy? The answer revolves around the information firms must have to efficiently determine the prices they should set for goods and services. In making such decisions, firms must use both idiosyncratic information (information specific to their own revenues and costs) as well as aggregate information (information about economy-wide levels of prices and employment). Precise statistics about inflation allow firms to disentangle the components of demand for their products that are idiosyncratic to them from those components that are aggregate and common to all firms. For example, in a high inflation environment, a high level of sales is more likely to be attributable to aggregate factors such as an expansionary monetary policy; if there are more dollars circulating, families will demand more of all goods in the economy. On the other hand, in a low inflation environment, a high level of sales is more likely to be attributable to a higher real demand for the goods that the firm produces. Therefore, more precise infor-

mation about inflation helps firms disentangle idiosyncratic from aggregate shocks, as well as monetary from real shocks.

With better information regarding the demand for their goods, firms can set prices in a way that better reflects both the production cost and the value assigned by buyers. When prices more accurately reflect the value of demand and the cost of production, inputs of production (labor, physical capital, intermediate inputs) are assigned in a more efficient way. If input factors are more efficiently assigned, the aggregate level of production given the same amount of inputs is higher, which leads to higher social welfare.

A symptom of better access to information about inflation is a lower level of price dispersion for a given type of good. When more precise inflation statistics are available, the price of a cup of coffee should be more similar between different stores and brands. The reason is that, with more accurate information about aggregate inflation, firms optimally choose to put more weight in inflation statistics when setting prices; when all firms optimize in this way, their prices are more aligned with each other.

Once the mechanism through which the availability of public information can affect the macro economy is understood, the next challenge is to quantify it. We can measure this effect if we have: (i) an economic model that fits the macroeconomic behavior of an economy and (ii) an episode of analysis in which the access to information about the inflation rate changed significantly. The second condition is more difficult to find in the real world; is there an episode of significant changes in the availability of information about the inflation rate? There is, in Argentina.

THE ECONOMIC EFFECTS OF THE PROVISION OF STATISTICS

In 2007 the Argentinean government started manipulating official inflation statistics to prevent figures from reflecting accelerating inflation. The manipulation started in January 2007 with the government's intervention in the national statistical agency (INDEC) that included the removal of the authorities in charge of computing and publishing the CPI. Since then, local and international media as well as international institutions and academic circles have discredited the official inflation statistics. In the past decade Argentina has had two regimes controlling access to public information about the level of inflation. Before 2007 the government provided

> Once the mechanism through which the availability of public information can affect the macro economy is understood, the next challenge is to quantify it.

a unique and credible official measure of inflation. From 2007 onwards, official inflation statistics have been discredited and there has been overall uncertainty about the true level of inflation in spite of the presence of alternative noisier measures of inflation.

Using data on disaggregated prices for Argentina and Uruguay, we found that the manipulation of inflation statistics brought about an increase in price dispersion (measured by the coefficient of variation of the prices of similar goods) in Argentina of 13%. With this measured effect in hand, we can estimate an economic model for Argentina tailored to studying our episode of analysis and we can then use it to

quantify the value of providing trustworthy inflation statistics. According to our model estimates, the manipulation of inflation statistics led to a less efficient allocation of inputs of production that in turn led to an equivalent, permanent drop in the aggregate level of output of 1.3%. With less precise information, the firms that generate the most value do not hire more labor or capital.

We then explore the extent to which our estimates can be generalized to the United States. To do so, we re-estimate our model to match the most salient features of the American economy and then re-do our hypothetical exercise that replicates the manipulation of inflation statistics. We find that the negative welfare effects associated with this policy are more than ten times smaller than those for Argentina. The reason is that the availability of inflation statistics provides less value in the US, since aggregate prices do not fluctuate as much as in Argentina. In other words, the US would not incur such large efficiency losses if firms set their prices by taking into account the long run level of inflation rather than by considering the current level of inflation.

In summary, we find that significant welfare gains can be made by providing trustworthy and precise statistical information about the inflation rate, as it helps to allocate resources in a more efficient way. Additionally, these welfare gains are larger in highly volatile economies where prices fluctuate significantly and firms place greater value on information about the aggregate macroeconomic state. The Argentinean episode is ideally suited to measuring these welfare consequences and quantifying their magnitudes.

> We can measure this effect if we have: (i) an economic model that fits the macroeconomic behavior of an economy and (ii) an episode of analysis in which the access to information about the inflation rate changed significantly.

[1] See, for example, "Acusan al Gobierno de manipular datos del INDEC", La Nación, February 10, 2007, and "Lies and Argentine data", The Economist, April 2011.
[2] This analysis ignores any potential feedback that the provision of adequate statistics may have on the volatility of monetary policy, which can be a relevant channel in the case of the US.