BETTING ON EXPORTS: TRADE AND ENDOGENOUS HETEROGENEITY*

Alessandra Bonfiglioli, Rosario Crinò and Gino Gancia

We study the determinants of firm-level heterogeneity in a model where innovation choices upon entry affect the variance of productivity draws. In equilibrium, productivity is Pareto distributed with a shape parameter that depends on industry-level characteristics. We show that export opportunities, by increasing the pay-offs in the tail, induce firms to invest in bigger projects with more dispersed outcomes. When more productive firms pay higher wages, trade amplifies wage dispersion by making firms more unequal. These results are consistent with how firm size, innovation and wage heterogeneity vary in a panel of US industries and states.

Current research in international trade puts firm-level heterogeneity at a centre stage. As documented by a growing empirical literature, firms differ in size and productivity even within narrowly defined industries (Syverson, 2004*a*,*b*) and these differences vary systematically with trade participation (Bernard *et al.*, 2012). In particular, exporters are bigger and more productive than non-exporters, and they pay higher wages. Firm heterogeneity also has crucial implications for macroeconomic outcomes, such as aggregate efficiency (Hopenhayn, 2014). Yet, despite the growing attention that firm-level productivity differences have attracted, we still have a limited understanding of the theoretical and empirical underpinnings of this heterogeneity.

Although the distribution of the entire population of existing firms has some common characteristics that have been documented extensively (Axtell, 2001), these aggregate statistics mask significant heterogeneity across sectors and even between countries. For example, Helpman *et al.* (2004) show that cross-sector variation in measures of firm heterogeneity is important for explaining the prevalence of multinational sales relative to exports. Rossi-Hansberg and Wright (2007) find that the standard deviation of establishment size varies with capital shares. Bartelsman *et al.* (2009) and Poschke (2015) document instead differences in the firm-size distribution across countries. Given that more productive firms pay higher wages, firm heterogeneity is likely to map into wage dispersion and wage inequality also varies significantly across countries. Besides these scant observations, systematic evidence and theoretical explanations for differences in firm heterogeneity are still scarce. The goal of this article is to take a step towards filling this gap.

We start our analysis by documenting some little-known facts regarding how a simple measure of firm heterogeneity, the standard deviation of the log of sales across establishments, varies across sectors and time in the US economy. We show that this measure of dispersion can differ by a factor of ten between 6-digit NAICS industries and that it has increased on average by 11.8% between 1997 and 2007. Searching for patterns in the data, we find robust evidence that sales dispersion correlates positively with the export intensity of the industry, average sales per establishment and entry. Similar results hold when using two different strategies for identifying the effect of export intensity and alternative measures of dispersion, computed over total sales, labour productivity and also sales of non-exporters only. Next, we propose a novel explanation based on the idea that the observed heterogeneity stems from technological choices.

To do so, we develop a model in which endogenous investment decisions at the entry stage affect the variance of the possible realisations of productivity. We then explore the implications for the equilibrium distribution of firms and wages in closed and open economies. Leading models of heterogeneous firms take the probability distribution from which firms draw productivity as given and characterise the resulting distribution of firm-level characteristics through the dynamics of entry, exit and possibly growth. Important examples are Melitz (2003), Rossi-Hansberg and Wright (2007), Luttmer (2010) and more recently Jones and Kim (2014) and König *et al.* (2016). This article takes a complementary approach, namely, to recognise that firms can affect the variance of their productivity draws at the entry stage.

Although the success in starting a new enterprise or launching a new product is inherently uncertain, firms can deliberately choose between investing in smaller projects with less variable outcomes and more ambitious projects with higher variance. Such a trade-off is very familiar to anyone pursuing academic research but is also common in the world of business. For instance, designing and assembling a new variety of laptop PCs, which mostly requires the use of established technologies, is safer and less costly than developing an entirely new product, such as tablet computers. In fact, the first tablet-like products date back to the 1980s but did not attain success until the release of the iPad in 2010.[1] Yet, after decades of attempts, Apple was rewarded with the sale of more than 250 million units over a period of only five years.

We formalise these ideas in a multi-industry model *à la* Melitz (2003) in which firms can draw a random productivity level upon paying an innovation cost and there are both fixed and variable export costs. We modify the entry stage by allowing firms to choose the size of their investment in projects of unknown quality, which is shown to affect the variance of the probability distribution from which productivity is drawn. In particular, a complementarity between the size of the investment and the unknown quality of ideas implies that larger innovation projects are associated with more dispersed realisations of productivity. A key insight of the model is that the possibility of exit insures firms from bad realisations and increases the value of drawing

---

[1] In a 1983 speech, Steve Jobs said: 'Apple's strategy is really simple. What we want to do is we want to put an incredibly great computer in a book that you can carry around with you and learn how to use in 20 minutes … and we really want to do it with a radio link in it so you don't have to hook up to anything and you're in communication with all of these larger databases and other computers'. Yet, inventing the iPad required years of investment constellated with failures and unforeseen spin-offs, including the development of the iPhone.

productivity from a more dispersed distribution. However, bigger projects with higher variance require a larger investment, which generates a trade-off.[2]

After solving for the optimal innovation size, the model yields a Pareto distribution for productivity with a shape parameter that depends on industry-level characteristics in a way consistent with the patterns found in the data. In particular, export opportunities induce firms to draw technology from a more dispersed distribution. The reason is that trade reallocates profits in favour of the most productive firms, thereby increasing the pay-offs in the tail. The model also predicts that high fixed costs and low entry barriers increase the value of investing in technologies with more dispersed outcomes by raising the exit cut-off. Hence, it replicates the positive correlation of equilibrium dispersion with export intensity, average sales and entry. Next, we extend the model to show how firm heterogeneity can map into wage inequality, as in Helpman *et al.* (2010). When more productive firms pay higher wages, we obtain another novel result: trade amplifies wage dispersion by inducing firms to invest in technologies with more variable outcomes *ex ante* and hence making them more unequal *ex post*.

In the last Section of the article, we go back to the data. Using individual-level wage data over the period 1997–2007 in the United States, we show that measures of wage dispersion co-vary with industry characteristics in a way that mirrors the pattern found for the dispersion of sales well. Most importantly, export opportunities increase significantly wage inequality at the industry level. We then provide a first attempt at testing a specific mechanism of our model, namely, that export opportunities increase firm heterogeneity by fostering investment in innovation. To do so, we follow Aghion *et al.* (2015) in switching to geographic data and use patent counts to build a measure of innovation intensity for a panel of US states over the period 1989–2007. We also follow Autor *et al.* (2013) in using the industry composition of manufacturing employment of each state to construct a state-level measure of export intensity. With these data, we document two sets of results. First, consistent with the model, innovation intensity increases with export opportunities. Second, the dispersion of firms' sales, computed now at the state level, is correlated with innovation intensity.

Besides the evidence reported in this article, our theory accords well with a number of additional observations. Regarding the main premises of the model, several papers show evidence suggesting that differences in productivity across firms are related to investment in new technologies (Dunne *et al.*, 2004; Faggio *et al.*, 2010; Doraszelski and Jaumandreu, 2013). Our focus on product innovation is also empirically relevant. For instance, Broda and Weinstein (2010) find that almost 50% of the consumer goods sold in US markets in a given year did not exist four years before. And yet, product innovation is subject to high uncertainty. In a survey of existing studies, Castellion and Markham (2013) conclude that around 40% of new products fail.[3] Regarding the main implications of the model, the prediction that low entry barriers stimulate innovation

---

[2] In our model, risk is completely diversified so that investor seek to maximise expected returns. However, expected returns depend on the variance of productivity draws. As we show, this property holds even if firms can choose between productivity distributions that are a mean-preserving spread.

[3] Similarly, the US Census reports that only about 50% of start-up companies are still operating four years after entry. Furthermore, as shown for instance in Cabral and Mata (2003), there is already considerable heterogeneity among new firms.

by fostering competition is consistent with the finding that competition and entry raise firm productivity (Aghion *et al.*, 2004, 2009). The insight that the chance of winning the extra prize of exporting induces firms to bet on bigger projects with a higher variance seems also plausible. Returning to the example of the invention of the iPad, our model suggests that globalisation made Apple's strategy so rewarding. More in general, there is ample anecdotal evidence that firms competing for global markets are aiming at more ambitious innovation projects. To name just one example, in 2010, Google started to invest in the 'Google X' project, a semi-secret laboratory dedicated to making major, high-variance, technological advancements.

This article is related to the vast literature aimed at explaining productivity differences across firms (see Syverson, 2011, for a survey). To the best of our knowledge, the link between innovation choices and the variability of technology has received little attention. Some papers consider the distinction between radical and incremental innovation (Acemoglu and Cao, 2015). But these types of innovations differ more in the degree to which they replace or complement existing technologies, rather than in the variance of the potential outcomes. Some exceptions are Gabler and Poschke (2013), Bartelsman *et al.* (2016) and Caggese (2016), who study how distortions affect the choice between risky technologies. Yet, even these papers do not study the implications for the distribution of firms and wages, which remains an under-explored and promising area of research.

The large literature on trade with heterogeneous firms started by Melitz (2003) does study the implications of export opportunities for the distribution of existing firms (see Melitz and Redding, 2014, for an excellent survey). As is well known, trade can make firms more unequal by reallocating profits and workers from the least to the most productive firms. This effect is however very different from the one we emphasise, in that, it abstracts from the possibility that trade changes the fundamental reason why firms are different, namely, the unconditional productivity distribution.[4] Moreover, the focus of our article is on measures of dispersion of firms' attributes that are scale invariant rather than other characteristics, such as average size or the productivity cut-off for exit, that have been studied more extensively. In this respect, our article is close in spirit to a nascent strand of literature aimed at exploring the effect of trade on higher moments of the distribution of firm characteristics (Mayer *et al.*, 2015).

Some recent papers study the impact of trade on productivity via *ex post* decisions on product scope, quality or innovation. These include Atkeson and Burstein (2010), Lileeva and Trefler (2010), Bernard *et al.* (2011), Bustos (2011), Kugler and Verhoogen (2012) and Dhingra (2013), among others. These papers propose and test different channels through which trade liberalisation can raise firm-level productivity but do not focus on its dispersion. This literature has also shown that trade can help overcome the fixed cost of technology adoption through a scale effect, a result that is very different from our finding that trade induces firms to invest in projects with higher variance.[5]

---

[4] Some papers, including Yeaple (2005) and more recently Grossman and Helpman (2014), trace productivity differences across firms to heterogeneity in ability across workers and managers. We follow the complementary approach that emphasises the role of differences in technology rather than ability.

[5] There is a small literature on trade and risk taking, including interesting work by Vannoorenberghe (2014) and Fillat and Garetto (2015). In this article, we instead study the determinants of technological variability in a model where risk is fully diversified.

Since many technological choices must be made *ex ante*, especially when new products are introduced, combining our model of innovation with *ex post* decisions that can affect an initial realisation of productivity seems a promising step forward to develop a comprehensive theory of how productivity differences emerge and evolve.

Finally, several papers have shown, both theoretically and empirically, that trade impacts wage inequality because exporters pay higher wages (e.g. see all the papers surveyed in section 10 of Melitz and Redding, 2014). In our model, however, the effect of trade works not only through the exporters' wage premium but also by making the entire wage schedule steeper, with different implications. For instance, our mechanism predicts that more export opportunities will increase wage dispersion even among the group of non-exporting firms. This may help explain why the rise in inequality is often found to be 'fractal', that is, to hold across firm size groups (Song *et al.*, 2015).[6]

The remainder of the article is organised as follows. In Section 1, we document some stylised facts regarding how the dispersion of log sales varies across sectors and time in the US economy. Motivated by these empirical observations, in Section 2 we propose a closed-economy model where differences in the dispersion of firm-level outcomes originate from technological choices at the entry stage. Section 3 adds costly trade and shows that more export opportunities induce firms to draw their productivity from more dispersed distributions, thereby generating more heterogeneity in equilibrium. In Section 4, we consider the implications of the model for wage inequality. Section 5 goes back to the data and provides evidence in support of the model's predictions on wage dispersion and innovation. Section 6 concludes.

1. Motivating Evidence: Sales Dispersion and Trade

In this Section, we document how the dispersion of sales of US establishments varies across industries and over time, and how it correlates with a number of industry characteristics. First, we show that the dispersion of sales differs significantly across industries and has increased over time. Second, we present panel regressions suggesting that higher dispersion at the industry level is systematically associated with larger scale in terms of average sales, with higher export intensity and with firm entry. Finally, we provide additional evidence suggestive of a causal effect of export intensity on sales dispersion.[7]

1.1. *Sales Dispersion Across Industries and Over Time*

Our main measure of dispersion is the standard deviation of the logarithm of sales per establishment. We focus on sales because they are an easy-to-observe measure of overall size, and we take logs to make the standard deviation scale invariant. We compute this variable using data from the 'Statistics of US Businesses' of the US Census Bureau for

---

[6] Dunne *et al.* (2004) and Faggio *et al.* (2010) also show that a large fraction of the observed wage dispersion is between firms.

[7] An antecedent of this analysis is Syverson (2004*b*), who studies how various measures of productivity dispersion co-vary with industry characteristics in the US manufacturing sector. Yet, his evidence is limited to the 1977 cross-section.

the years 1997, 2002 and 2007.[8] Data on (receipts of) sales and number of establishments and employees are available for 453 6-digit NAICS industries aggregated into sales-size categories. Since we do not have access to the underlying firm-level data, we follow Helpman *et al.* (2004) in assuming that all establishments falling within the same bin have sales equal to the group mean. Then, we consider each bin in a 6-digit NAICS industry as a single observation and compute the standard deviation of log establishment sales across bins using the number of establishments in each bin as weights.[9] Helpman *et al.* (2004) show that this methodology for computing dispersions approximates well other measures based on the entire population of firms. As an additional check, we have also computed the variance of log sales using firm-level data from Compustat. This database is relatively small, as it only includes listed firms, so we can construct reliable measures of sales dispersion for 21 aggregate sectors, defined at the 3-digit level of the NAICS classification. While this feature makes Compustat not very well suited for our analysis, we find the dispersion of sales computed on Compustat to be highly correlated (0.65) with the one we obtain from Census data.

Table 1 reports some descriptive statistics. For each 3-digit manufacturing sector, the table shows the average, minimum and maximum value of the standard deviation of log establishment sales across the constituent 6-digit industries in 2007. The table also reports the average percentage change in sales dispersion in each 3-digit sector over the previous 10 years.[10] For convenience, sectors are ordered by increasing dispersion. The first column shows that the dispersion of sales varies significantly across sectors, ranging from a minimum of 1.69 (in Paper Manufacturing) to a maximum of 2.71 (in Transportation Equipment Manufacturing). The second and third columns show, however, that the main source of heterogeneity is within 3-digit sectors: among all 6-digit industries, the dispersion of sales varies by a factor of 10, as shown in the last row of the table. The fifth column reports the average number of establishments across the 6-digit industries in 2007. Comparing the first and fifth columns reassures that the dispersion of sales in a sector is not mechanically driven by sample size. Finally, the fourth column shows that the dispersion of sales has increased remarkably between 1997 and 2007, on average by 11.8% (28.5% if we weight industries by sales). Although this rise in dispersion is not a well-known stylised fact, it is consistent with the evidence in Dunne *et al.* (2004), who find that inequality in productivity across US manufacturing plants increased between 1975 and 1992, and in Faggio *et al.* (2010), who find similar results for the United Kingdom between 1984 and 2001.

---

[8] In the 'Statistics of US Businesses', information on firms' sales is released during Census years and is currently available for the years 1997, 2002, 2007 and 2012. The substantial restructuring of the NAICS classification occurred in 2012 makes it impossible to create a mapping between the latest wave of the data and the preceding ones for many industries. We therefore use the first three waves, which also ensures that our results are not contaminated by the Great Recession.

[9] The number of bins and their width is decided every year by the US Census Bureau. In particular, the raw data are disaggregated into 10 bins in 1997, 8 bins in 2002 and 18 bins in 2007. The lowest bin contains firms with revenue below 50,000 US$ while the highest bin contains firms with revenue over 100 million US$. The raw bins are defined in such a way that they can be aggregated into six bins consistently defined throughout the period. We use this consistent definition in most of the analysis. However, in a robustness check, we show that our results are unchanged when using the raw bins.

[10] All reported averages are simple averages. Weighting by sales does not affect the qualitative results.

Table 1

Descriptive Statistics on the Dispersion of Sales in US Manufacturing

| NAICS code | Sector description | SD of log establishment sales | | | | No. of establishments |
		Mean	Min.	Max.	% Ch.	Mean
322	Paper	1.69	0.96	2.27	0.07	252
327	Nonmetallic Mineral Prod.	1.89	0.97	3.56	0.03	728
333	Machinery	1.97	0.48	3.21	−0.03	498
313	Textile Mills	2.00	1.17	2.97	−0.11	258
331	Primary Metals	2.07	1.44	3.10	0.09	203
326	Plastics and Rubber Prod.	2.07	1.36	3.08	0.08	1024
315	Apparel	2.09	1.00	2.94	0.05	546
332	Fabricated Metal Prod.	2.12	1.22	3.33	0.10	1387
324	Petroleum and Coal Prod.	2.17	0.75	4.48	0.47	482
316	Leather and Allied Prod.	2.20	0.89	3.26	0.06	139
339	Miscellaneous Manuf.	2.24	0.95	3.67	0.23	1571
325	Chemicals	2.24	0.55	4.01	0.06	394
337	Furniture and Related Prod.	2.26	0.92	3.18	0.18	1753
334	Computer and Electronic Prod.	2.26	0.88	3.76	−0.01	499
321	Wood Products	2.43	1.46	3.25	0.32	1187
335	Electr. Equip., Appl., and Comp.	2.47	1.79	3.66	0.05	279
312	Beverage and Tobacco Prod.	2.52	1.98	3.12	0.31	452
323	Printing and Rel. Supp. Activ.	2.54	1.33	3.31	0.37	2773
311	Food	2.59	1.07	4.57	0.41	549
314	Textile Prod. Mills	2.59	1.45	3.34	0.32	649
336	Transportation Equip.	2.71	1.72	4.22	0.02	404
	Total	2.23	0.48	4.57	0.12	720

Notes. The standard deviation of log establishment sales is computed separately for 453 6-digit NAICS industries, using data on receipts of sales and number of establishments in six sales-size bins homogeneous over time, and weighting the observations with the number of establishments in each bin. The mean, minimum and maximum value of the standard deviation in each 3-digit sector are computed across the corresponding 6-digit industries and refer to the year 2007. Percentage changes are computed over 1997–2007 for each 6-digit industry, and are then averaged within the corresponding 3-digit sector. The average number of establishments in a 3-digit sector is the mean across the corresponding 6-digit industries in 2007.

1.2. *Exploring the Data: Correlations*

To explore the data further, we now exploit variation across 6-digit industries and over time, and study how the dispersion of sales correlates with a number of industry characteristics. Among the covariates, we consider the logs of average sales per establishment, number of establishments, export intensity, total employment, intensities in physical capital and raw materials, and average educational attainment, as well as the standard deviation of log education across workers. Export intensity is the ratio of exports to total shipments, constructed with export data from Schott (2008) and shipment data from the NBER-CES Manufacturing Industry Productivity Database, and captures engagement in global markets.[11] Total employment is the number of employees at the industry level sourced from the US Census Bureau and is a measure of the size of the sector. Capital and material intensities are computed as in Romalis (2004) with data from the NBER-CES Manufacturing Industry Productivity Database

[11] We have also used import penetration. However, given the collinearity between the two variables, we focus on export intensity, which is found to be statistically much more significant.

and are equal to the ratios of capital compensation and material expenditure, respectively, over the sum of value added and material costs. These variables are meant to capture technological characteristics of industries. Finally, the mean and standard deviation of workers' education are computed with data from the CPS Merged Outgoing Rotation Groups. Including these variables can help account for differences in skill intensity and for the role of sorting between firms and workers based on observables.[12]

We regress the standard deviation of log establishment sales on these industry characteristics. To have a sense of how sales dispersion may vary with the economic cycle, we also control for the growth rate of nominal GDP over the two years prior to each observation.[13] We are interested in exploring both the cross-sectional variation, which may be more informative about the long run, and the time variation in the data. Therefore, we estimate specifications without industry fixed effects, so as to exploit the whole variation in the sample, as well as with industry fixed effects, so as to control for unobserved industry heterogeneity and identify the coefficients through variation over time. Finally, since variables may be trending over time, we also estimate specifications with industry fixed effects and variables in first differences, so as to control for industry-specific time trends.

Table 2 reports the baseline OLS results. Columns (1)–(3) show coefficients from pooled-OLS specifications, columns (4) and (5) control for industry fixed effects, and column (6) reports the estimates from the first-difference specification. Due to missing data on the explanatory variables, our final estimation sample is an unbalanced panel of 364 6-digit NAICS industries observed in 1997, 2002 and 2007. Standard errors are corrected for clustering by 6-digit industry to accommodate autocorrelated shocks at the industry level. In column (1), we start by regressing sales dispersion on export intensity, total employment, capital and material intensity and the mean and dispersion of educational attainment. We find that the dispersion of sales is positively correlated with all these variables. In column (2), we control for two variables that capture important characteristics of the sales distribution: average sales per establishment and the number of establishments. Interestingly, we find that sales dispersion is strongly positively correlated with average establishment sales but not with the number of establishments on which it is computed. Among the other covariates, only export intensity and dispersion of education remain highly significant. The coefficients on the other variables drop in size, suggesting that the effect of these variables might be mediated by average sales. In column (3), we add GDP growth. The coefficient on this variable is positive and precisely estimated, suggesting that the dispersion of sales is higher in periods of economic expansion. All other coefficients are unchanged.

In columns (4) and (5), we add industry fixed effects, including the same control variables as in columns (1) and (3) respectively. The coefficients on establishment sales, export intensity and GDP growth remain positive and statistically significant. These coefficients are larger than before, suggesting that the correlations are stronger within industries. The coefficient on the number of establishments is now positive and significant. This suggests that, while the number of establishments across industries

---

[12] See the Appendix for more details on variables definitions and data sources.
[13] The growth rate of nominal GDP is computed with data from the World Development Indicators.

Table 2

Dispersion of Sales and Industry Characteristics: Baseline Estimates

	Pooled (1)	Pooled (2)	Pooled (3)	Industry FE (4)	Industry FE (5)	Industry FE + First diff. (6)
Log exp. int.	0.090***	0.086***	0.085***	0.132***	0.080*	0.175**
	(0.016)	(0.015)	(0.015)	(0.048)	(0.047)	(0.069)
Log av. est. sales		0.247***	0.257***		0.543***	0.611***
		(0.044)	(0.045)		(0.090)	(0.184)
Log n. of est.		0.074	0.087		0.430***	1.068***
		(0.065)	(0.066)		(0.161)	(0.335)
Log tot. empl.	0.183***	0.086	0.071	0.183***	−0.246**	−0.335
	(0.021)	(0.062)	(0.063)	(0.068)	(0.103)	(0.315)
Log cap. int.	0.535***	0.077	0.052	0.299	−0.332	0.172
	(0.115)	(0.107)	(0.107)	(0.189)	(0.222)	(0.203)
Log mat. int.	0.664***	−0.166	−0.205	0.093	−0.952***	0.326
	(0.136)	(0.151)	(0.152)	(0.290)	(0.354)	(0.476)
Log av. educ.	0.886**	−0.033	−0.015	1.171	−0.054	−0.959
	(0.350)	(0.370)	(0.369)	(0.819)	(0.851)	(1.185)
SD log educ.	1.115***	0.938***	0.940***	1.129**	0.342	−0.343
	(0.379)	(0.353)	(0.353)	(0.516)	(0.508)	(0.651)
GDP gr.			2.211***		3.017***	
			(0.460)		(0.495)	
Observations	1,036	1,036	1,036	1,036	1,036	670
R^2	0.18	0.28	0.29	0.65	0.68	0.45

Notes. The dependent variable is the standard deviation of log establishment sales in a 6-digit NAICS industry. This variable is computed using data on receipts of sales and number of establishments in six sales-size bins homogeneous over time, and weighting the observations with the number of establishments in each bin. All variables except for GDP growth are observed at the 6-digit industry level in the years 1997, 2002 and 2007. All industry-level controls are contemporaneous to the dependent variable. GDP growth is computed over the two years prior to each observation. All specifications are estimated with ordinary least squares. Standard errors are clustered by 6-digit industry and reported in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. See also notes to Table 1.

does not correlate with sales dispersion, entry of new establishments over time is associated with a higher dispersion. Finally, in column (6), we express all variables in first differences and include industry fixed effects to control for industry-specific trends. We still find sales dispersion to be positively correlated with average sales, number of establishments and export intensity.[14]

In Table 3, we perform robustness checks. To exploit both the cross-sectional and the time variation in the data, we use both the specification without fixed effects and the richest specification in first differences. In columns (1) and (2), we re-estimate our specification computing the dependent variable on a sub-sample which excludes the smallest establishments (corresponding to the bottom 25% of the sample). In columns (3) and (4), we instead exclude the largest establishments (top 21% of the sample).[15] Our main evidence is largely unchanged, suggesting that it is not driven by large or

[14] GDP growth is expressed in differences over the previous two years, so we exclude this variable from the first-differenced specifications to avoid double differencing, which would not have a clear interpretation.
[15] The number of observations slightly drops in these specifications as in a handful of industries establishments are concentrated in two or three bins.

Table 3

Dispersion of Sales and Industry Characteristics: Robustness Checks

	No smallest bins		No largest bin		Raw bins	
	Pooled (1)	Ind. FE + FD (2)	Pooled (3)	Ind. FE + FD (4)	Pooled (5)	Ind. FE + FD (6)
Log exp. int.	0.058***	0.095*	0.052***	0.145*	0.083***	0.159**
	(0.011)	(0.050)	(0.014)	(0.078)	(0.015)	(0.067)
Log av. est. sales	0.327***	0.473***	0.022	0.163	0.216***	0.587***
	(0.027)	(0.126)	(0.037)	(0.142)	(0.044)	(0.171)
Log n. of est.	−0.002	0.486**	0.152***	0.876***	0.027	0.918***
	(0.041)	(0.225)	(0.053)	(0.336)	(0.065)	(0.308)
Log tot. empl.	0.077*	−0.259	0.053	−0.120	0.106*	−0.335
	(0.039)	(0.208)	(0.051)	(0.280)	(0.062)	(0.304)
Log cap. int.	0.085	−0.086	0.055	0.489***	0.066	0.090
	(0.079)	(0.149)	(0.091)	(0.167)	(0.105)	(0.180)
Log mat. int.	−0.044	−0.083	0.011	0.889**	−0.183	0.133
	(0.104)	(0.334)	(0.117)	(0.393)	(0.144)	(0.436)
Log av. educ.	0.169	−0.114	−0.145	0.050	−0.047	−1.115
	(0.238)	(0.580)	(0.321)	(1.258)	(0.359)	(1.102)
SD of log educ.	0.160	−0.177	1.018***	−0.160	0.830**	−0.506
	(0.227)	(0.351)	(0.354)	(0.636)	(0.345)	(0.609)
GDP gr.	0.269		2.713***		1.545***	
	(0.307)		(0.446)		(0.449)	
Observations	1,029	660	1,009	646	1,035	670
R^2	0.62	0.53	0.18	0.48	0.28	0.46

Notes. The dependent variable is the standard deviation of log establishment sales in a 6-digit NAICS industry. In columns (1)–(2), this variable is computed excluding observations for the two smallest sales-size bins in each industry. In columns (3)–(4), it is computed excluding observations for the largest sales-size bin in each industry. In columns (5)–(6), it is computed using all sales-size bins available for each industry in a given year, instead of the six homogeneous bins used in the other specifications. All variables except for GDP growth are observed at the 6-digit industry level in the years 1997, 2002 and 2007. All industry-level controls are contemporaneous to the dependent variable. GDP growth is computed over the two years prior to each observation. All specifications are estimated with ordinary least squares. Standard errors are clustered by 6-digit industry and reported in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. See also notes to previous Tables.

small establishments. As a further robustness check, in columns (5) and (6), we re-compute the standard deviation of log sales using all sales-size bins available in each period, so as to exploit the information contained in the data set fully. The coefficients are similar to those in columns (3) and (6) of Table 2, which further suggests that our results are not driven by the number of bins on which sales dispersion is computed.[16]

Our next step is to dig deeper into the empirical results found so far. The positive correlations of sales dispersion with average sales and changes in the number of establishments are interesting properties of the empirical sales distribution. Yet, these characteristics of the distribution are likely to be jointly determined and it is unclear how to identify any causal effect. We can instead investigate further the strong correlation between export intensity and sales dispersion.

---

[16] The fact that our results hold when removing small or large establishments, when controlling for the number of establishments and when changing the number of bins, suggests that they are unlikely to be simply driven by granularity in the data.

1.3. *Identifying the Effect of Trade*

Existing models of trade with heterogeneous firms offer two candidate explanations for the positive correlation between export intensity and sales dispersion. One is reverse causality: industries with higher dispersion of sales may have a higher export intensity because they are more likely to host more productive firms which tend to participate more in export markets (Bernard *et al.*, 2012). A second possibility is trade-driven reallocations of market shares towards larger firms. Since total sales include exports and more productive firms enter more foreign markets, trade makes the sales distribution more fat-tailed, a point explicitly made by di Giovanni *et al.* (2011).[17]

Our finding that export intensity is associated with more dispersion even when removing smaller or bigger establishments may suggest that the correlation is driven neither by exit nor by exporters. To investigate these possibilities further, we now attempt to sort out the direction of causality. Moreover, we show that our findings are robust to using alternative measures of dispersion that are not directly affected by participation in multiple markets. The first variable is the standard deviation of log sales per worker (labour productivity). We compute this variable in the same way as our measure of sales dispersion, using data on sales and employment by sales-size bin from the 'Statistics of US Businesses'. The second variable is the standard deviation of log establishment sales among non-exporting firms. This variable is based on proprietary data from Dun & Bradstreet (Worldbase data set), which uses data from different sources to construct indicators of business activity for firms around the world and provides information on the trade participation of each firm. While we do not have access to the underlying micro data, we obtained some aggregate statistics and in particular the standard deviation of log sales among US non-exporting plants for 4-digit SIC-87 industries in 2005, 2007 and 2009.[18] This sample differs from our baseline in terms of industry classification and time span but is relatively close to it. However, since in some cases sales dispersion is computed over few observations only, when using this variable, we weight the regressions by the number of non-exporters in each industry in 2005. In this way, we make sure that our results are not driven by industries where dispersion may be imprecisely measured.

To identify the effect of trade, we use two alternative identification strategies. First, in the spirit of Autor *et al.* (2013), we follow an instrumental variables approach to identify the effect of exogenous variation in exports on the different measures of dispersion. Second, following Hummels (2007) and Hummels *et al.* (2014), we adopt a difference-in-differences strategy based on changes in transportation costs with heterogeneous effects across industries.

In columns (1)–(6) of Table 4, we estimate the main specifications instrumenting export intensity with the sum of exports from all non-US countries to the destination markets of the United States in each industry and year.[19] Our aim is to identify variation in US exports generated by foreign demand conditions, which would raise both US and

---

[17] However, using French firm-level data, they find that the direct effect of export sales on the distribution of total sales is rather small.

[18] We are grateful to Laura Alfaro and Harald Fadinger for providing this data. See Alfaro *et al.* (2016) for more details on the dataset.

[19] Estimating the specifications in columns (3)–(6) by OLS yields similar results (available upon request).

Table 4

Dispersion of Sales and Industry Characteristics: Identification

| | 2SLS | | | | | | Diff-in-diff | | |
| | Sales | | Lab. prod. | | Sales (non-exp.) | | Sales | Lab. prod. | Sales (non-exp.) |
Dep. var.: SD of log	Pooled (1)	FE + FD (2)	Pooled (3)	FE + FD (4)	Pooled (5)	FE + FD (6)	(7)	(8)	(9)
Log exp. int.	0.134***	0.704***	0.044***	0.310***	0.012	0.404***			
	(0.029)	(0.171)	(0.016)	(0.100)	(0.012)	(0.108)			
Bulk wt × Log oil pr.							−0.098***	−0.033**	−0.044***
							(0.037)	(0.014)	(0.016)
Log av. est. sales	0.288***	0.808***	0.229***	0.288***	0.041***	0.535***	0.571***	0.088	−0.064
	(0.045)	(0.167)	(0.028)	(0.090)	(0.013)	(0.119)	(0.137)	(0.079)	(0.059)
Log n. of est.	0.133**	1.125***	0.300***	0.729***	−0.187***	1.082***	0.566***	0.289**	−0.212
	(0.067)	(0.379)	(0.038)	(0.226)	(0.025)	(0.297)	(0.208)	(0.127)	(0.155)
Log tot. empl.	0.041	−0.483	−0.253***	−0.429***	0.186***	0.161	−0.338**	−0.180*	0.130
	(0.063)	(0.294)	(0.034)	(0.141)	(0.024)	(0.221)	(0.156)	(0.104)	(0.094)
Log cap. int.	0.065	0.035	0.013	0.084	−0.045	0.159	−0.449**	−0.050	0.097
	(0.110)	(0.201)	(0.064)	(0.157)	(0.068)	(0.296)	(0.216)	(0.157)	(0.142)
Log mat. int.	−0.231	−0.318	−0.059	0.033	0.008	−0.347	−1.039***	−0.258	0.071
	(0.159)	(0.459)	(0.092)	(0.255)	(0.089)	(0.608)	(0.382)	(0.289)	(0.188)
Log av. educ.	−0.387	−0.877	−0.349	0.186	−0.003	0.866	−0.765	−0.023	−0.419
	(0.408)	(1.314)	(0.237)	(0.622)	(0.018)	(0.778)	(1.071)	(0.724)	(0.385)
SD of log educ.	0.756**	−0.279	0.224	0.122	0.106	0.381	−0.262	−0.066	−0.174
	(0.375)	(0.713)	(0.230)	(0.325)	(0.097)	(0.430)	(0.610)	(0.400)	(0.178)
GDP gr.	2.077***		0.519**		1.597***				
	(0.461)		(0.212)		(0.144)				

Table 4
(Continued)

| | 2SLS | | | | | | Diff-in-diff | | |
| | Sales | | Lab. prod. | | Sales (non-exp.) | | Sales | Lab. prod. | Sales (non-exp.) |
Dep. var.: SD of log	Pooled (1)	FE + FD (2)	Pooled (3)	FE + FD (4)	Pooled (5)	FE + FD (6)	(7)	(8)	(9)
Observations	1,015	630	1,008	620	1,025	642	776	771	812
R^2	0.29	0.36	0.15	0.46	0.69	0.17	0.66	0.48	0.93
First-stage results									
Log world exports	0.460***	0.519***	0.460***	0.517***	0.750***	0.922***	–	–	–
	(0.067)	(0.061)	(0.068)	(0.061)	(0.095)	(0.227)	–	–	–
Kleibergen–Paap F-stat.	46.8	71.9	45.9	71.7	61.8	16.5	–	–	–

Notes. The dependent variables are indicated in column headings. The standard deviation of log establishment sales (columns (1), (2) and (7)) is computed using data on receipts of sales and number of establishments in six sales-size bins homogeneous over time, and weighting the observations with the number of establishments in each bin. The standard deviation of log labour productivity (sales per worker; columns (3), (4) and (8)) is computed analogously, using data on receipts of sales and number of employees in each bin. The standard deviation of log establishment sales among non-exporting firms (columns (5), (6) and (9)) is based on firm-level data from Worldbase. The bulk weight is expressed in kg per US$ shipped by air and/or vessel, and refers to the year 1995. All variables apart from GDP growth and oil price are defined at the industry level. Except for the bulk weight, the industry controls are contemporaneous to the dependent variable. In columns (1)–(4) and (7)–(8), the industry variables are observed at the 6-digit NAICS level and the sample period includes the years 1997, 2002 and 2007. In columns (5), (6) and (9), the industry variables are defined at the 4-digit level of the SIC-87 classification, the sample period includes the years 2005, 2007 and 2009, and the regressions are weighted by the number of non-exporting firms in each industry in 2005. The specifications in columns (1)–(6) are estimated with two-stage least squares. Export intensity is instrumented with non-US exports to the destination markets of the US in each industry and year. F-statistics are reported for the Kleibergen–Paap test for weak instruments. The specifications in columns (7)–(9) include full sets of industry and year fixed effects. Standard errors are clustered by industry and reported in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. See also notes to previous Tables.

third countries' exports to the same destination market, while cleaning out variation due to US industry-specific technological characteristics, which would only raise US exports and could induce reverse causality. The exclusion restriction is that, conditional on our set of control variables, technological shocks in US industries are uncorrelated with those in other countries' industries. Most of the increase in world exports over the sample period was indeed due to the spectacular growth of low and middle-income countries such as Brazil, Russia, India and China (Autor *et al.*, 2013). It is conceivable that technological shocks occurring in these countries are largely uncorrelated with those hitting the United States.

The first-stage results reported in the bottom of the Table show that the instrument has strong power for predicting US export intensity. The first-stage coefficient is always positive, precisely estimated and large, ranging between 0.46 and 0.92 across specifications, and the Kleibergen-Paap F-statistics for excluded instruments ranges from 16.5 to 71.9. The second-stage results show that the coefficient on export intensity is always positive and is highly statistically significant in all specifications but one, consistent with a causal effect of exports on the dispersion of sales and productivity. The only instance in which the coefficient on export intensity is imprecisely estimated is when using as dependent variable sales dispersion of non-exporters on the pooled data (column 5). Yet, this is not very surprising because industries with higher export intensity have fewer non-exporters.

To have a sense of the size of the effect of trade, note that in our sample, export intensity has a standard deviation of 1.34, while the standard deviation of sales dispersion is equal to 0.58. Then, our coefficients imply that a 1 SD increase in export intensity would raise the dispersion of sales by 0.18–0.31 SD. The observed increase in export intensity over the sample period (21%) explains between 14% and 23% of the increase in sales dispersion over 1997–2007.

Finally, we show that our evidence is qualitatively unchanged when using an alternative strategy, which does not rely on instrumental variables but identifies the effect of exports by exploiting heterogeneous changes in transport costs across industries. In particular, we regress our dispersion measures on the interaction between the log oil price (Brent) and the bulk weight of US shipments in each industry, controlling for industry and year fixed effects. The interaction coefficient is identified by the differential response to a common oil price shock across industries that produce goods of different weight and are thus characterised by a different importance of transport costs. Hence, we refer to this approach as a difference-in-differences strategy. A negative estimate for the interaction coefficient would imply that, when hit by a reduction in oil price, industries shipping heavier goods – which see a larger drop in trade costs (increase in export opportunities) – experience a larger increase in dispersion.

To construct the bulk weights, we use product-level export data. We define the bulk weight of a given industry as the export-weighted average of the bulk weights of its constituent products, which in turn are computed as averages between air and vessel transport in 1995, to ensure that the choice of transport mode does not react to changes in oil price. Columns (7)–(9) of Table 4 report coefficient estimates for three specifications, each using a different dispersion measure as the dependent variable. The interaction coefficient is always negative and precisely estimated, which also points in

the direction of a causal effect of export opportunities on the dispersion of sales and labour productivity.

The stylised facts documented in this Section raise a number of important questions. Why is higher export intensity associated to more heterogeneity across establishments? More generally, what drives changes in the distribution of sales and productivity? In the remainder of the article, we propose a novel explanation based on the idea that firm heterogeneity stems from endogenous technological choices made at the time when new products are introduced. Besides its intuitive appeal and its ability to fit the empirical findings of this Section, our modelling strategy is inspired by several observations. First, Dunne *et al.* (2004) and Faggio *et al.* (2010) show evidence suggesting that changes in productivity dispersion appear to be related to new technologies. Second, by focusing on technological choices made before heterogeneity is realised, our theory will be able to explain changes in dispersion across the whole size distribution. This will allow us to explain the striking finding that trade seems to raise inequality even among non-exporters. Third, in our model, export opportunities will induce all firms to choose technologies with more uncertain outcomes, a prediction that seems consistent with the finding by di Giovanni and Levchenko (2012) that volatility is higher in sectors that are more open to trade.

2. Closed-economy Model

We now build a multi-sector, one factor, model of monopolistic competition between heterogeneous firms along the lines of Melitz and Redding (2014). After investing in innovation at the entry stage, firms draw their productivity from some distribution and exit if they cannot profitably cover a fixed cost of production. Differently from Melitz (2003), we allow the variance of productivity draws to depend on the entry investment. In this Section, we characterise the resulting endogenous distribution of firm-level variables in a closed economy. We defer to the next Section the case in which firms can engage in costly trade. For simplicity, we consider a static model in which entry and production decisions are all simultaneous.

2.1. *Preferences*

Consider an economy populated by a unit measure of identical households of size L with quasi-linear preferences over consumption of a homogenous good x_0 and differentiated goods produced in I industries:

$$U = x_0 + \sum_{i=1}^{I} \frac{\alpha_i X_i^{\zeta_i}}{\zeta_i}, \quad \zeta_i \in (0,1) \quad \alpha_i > 0.$$

Each industry $i \in \{1, \ldots, I\}$ produces differentiated varieties and preferences over these varieties take the constant elasticity of substitution form:

$$X_i = \left[\int_{\omega \in \Omega_i} x_i(\omega)^{\frac{\sigma_i-1}{\sigma_i}} \, \mathrm{d}\omega \right]^{\frac{\sigma_i}{\sigma_i-1}}, \quad \sigma_i > 1.$$

where $x_i(\omega)$ is consumption of variety ω, Ω_i denotes the set of varieties produced in sector i and σ_i is the elasticity of substitution between varieties within an industry. We denote by $p_i(\omega)$ the price of variety ω in industry i and by P_i the ideal price of the consumption basket X_i:

$$P_i = \left[\int_{\omega \in \Omega_i} p_i(\omega)^{1-\sigma_i} \, d\omega \right]^{1/(1-\sigma_i)}.$$

The demand for the differentiated basket X_i is $X_i = (\alpha_i/P_i)^{1/(1-\zeta_i)}$ and the demand for each individual variety is:

$$x_i(\omega) = X_i \left[\frac{P_i}{p_i(\omega)} \right]^{\sigma_i}. \tag{1}$$

The demand for the homogenous good q_0 is residual. We assume that the income of each household is sufficiently high to always guarantee a positive consumption of the homogenous good, which is chosen as the numeraire. In the remainder of the article, we focus on a single sector and derive results that do not depend on general equilibrium effects. For this reason and to save notation, from now on, we remove the index i with the understanding that all parameters can potentially vary across sectors.

2.2. *Problem of the Firm*

Within each sector, every variety ω is produced by monopolistically competitive firms that are heterogeneous in their labour productivity, φ. Since all firms with the same productivity behave symmetrically, we index firms by φ. There are fixed costs of production and of entry, all in units of labour. At the entry stage, a firm can choose how much to invest in innovation, a choice that affects the variance of the possible realisations of productivity. Next, the firm faces standard production and pricing decisions. We solve the problem backwards: first, we describe the strategy of a firm with a given productivity and then solve for investment at the entry stage given rational expectations on the industry equilibrium. We follow the usual convention of identifying firms with varieties. We also assume that labour productivity in the homogenous sector is one so that the wage is one.

A firm with productivity φ chooses its price and whether to exit so as to maximise profit, $\pi(\varphi)$, subject to a downward-sloping demand curve with elasticity σ. The first-order condition for this problem implies that firms set prices equal to a constant markup over the marginal cost,

$$p(\varphi) = \frac{\sigma}{\sigma - 1} \frac{1}{\varphi}, \tag{2}$$

and exit if $\pi(\varphi) < 0$. Using (1) and (2), we can express profit as a function of productivity:

$$\pi(\varphi) = A\varphi^{\sigma-1} - f, \tag{3}$$

where $A = [\sigma/(\sigma - 1)]^{1-\sigma}(XP^\sigma/\sigma)$. Since profits are increasing in φ, the firm will exit whenever its productivity is below the cut-off $\varphi^* = (f/A)^{1/(\sigma-1)}$.

We now consider the entry stage. As in Melitz (2003), firms pay a sunk innovation cost to be able to manufacture a new variety with productivity drawn from some

distribution with c.d.f. $G(\varphi)$. Hence, combining the pricing and exit decision, we can write *ex ante* expected profit as:

$$\mathbb{E}(\pi) = \int_0^\infty \pi(\varphi) \, dG(\varphi) = \int_{\varphi^*}^\infty (A\varphi^{\sigma-1} - f) \, dG(\varphi). \tag{4}$$

We depart from the canonical approach by making the distribution $G(\varphi)$ endogenous. More precisely, we now develop a simple model of investment in innovation projects generating a Pareto distribution for φ with a mean and variance that depend on firms' decisions. The model will formalise the intuitive idea that firms can choose between small projects with relatively low variance and larger projects with more dispersed outcomes.

Before continuing, we pause to discuss briefly why we focus on Pareto distributions. Our choice is based both on empirical and theoretical considerations. First, the Pareto distribution is widely used in the literature and has been shown to approximate well some observed firm-level characteristics.[20] The second reason is analytical tractability. The convenient properties of Pareto distributions allow us to derive closed-form solutions for various measures of firm heterogeneity, which helps in mapping the model to the data. In particular, recall that our motivation is to explain some stylised facts about the standard deviation of the logarithm of firm characteristics, as documented in Section 1. Since the standard deviation of the log of a Pareto-distributed variable is just equal to the inverse of the shape parameter, it is clear why we are interested in endogenising this parameter.

Suppose that, in order to enter, a firm must invest in an innovation project. The outcome of the project is a technology allowing the firm to manufacture a new variety with productivity φ. The realisation of this productivity depends both on the quality of the project, which is uncertain, and the size of the investment, which is a choice variable. More precisely, assume that quality, q, of new projects is random and exponentially distributed:

$$\Pr[q > z] = \exp(-\kappa z),$$

with support $z \in [0, \infty)$ and rate $\kappa > 0$, capturing how 'compressed' the distribution is. Notice that quality is inherently uncertain and exogenous. The firm can instead choose the size of the project, s, with minimum $\underline{s} > 0$.[21] We assume that productivity depends both on the quality and the size of the project as follows:

$$\ln \varphi = sq + \ln \varphi_{\min},$$

with $\varphi_{\min} > 0$. This equation embeds a complementarity between quality and size: resources invested in a bad project ($q = 0$) are wasted, in that they do not increase φ, while even a great idea is useless without some investment to implement it. More

---

importantly, these assumptions imply that φ is Pareto distributed with minimum $\varphi_{\min}$ and shape κ/s, as can be seen from:

$$1 - G(\varphi) = \Pr\left[q > \frac{\ln(\varphi/\varphi_{\min})}{s}\right] = \left(\frac{\varphi}{\varphi_{\min}}\right)^{-\frac{\kappa}{s}}.$$

Hence, by choosing the size of the project, the firm is choosing to draw φ from different Pareto distributions, identified by the new parameter $v \equiv s/\kappa$. Since the standard deviation of the log of φ is equal to v, we can take it as an index of the dispersion of the distribution. As shown below, v will be one of the key determinants of the equilibrium distributions of the log of firm characteristics, such as sales. Moreover, v also affects the expected value of φ, which is equal to $\varphi_{\min}(1 - v)^{-1}$, so that mean and variance are linked. Although we consider this a realistic property, we show in the Appendix that our main results hold in an alternative model in which firms can choose between distributions with the same mean but different variance.

The next step is to study the value of drawing productivity from more or less spread-out distributions. To simplify the notation, from now on, we rewrite the entry problem of the firm as one of choosing directly v, rather than s. Substituting $A(\varphi^*)^{\sigma-1} = f$ into (4), assuming $\varphi^* > \varphi_{\min}$ (so that there is selection), $v < 1/(\sigma - 1)$ (for $\mathbb{E}(\pi)$ to be finite) and using $G(\varphi)$, we can solve for expected profits as a function of v:

$$\mathbb{E}(\pi) = f \int_{\varphi^*}^{\infty} \left[\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} - 1\right] \mathrm{d}G(\varphi) = \frac{f\varsigma}{1/v - \varsigma}\left(\frac{\varphi_{\min}}{\varphi^*}\right)^{1/v}, \qquad (5)$$

where it proves convenient to define $\varsigma \equiv \sigma - 1$. It is easy to see that expected *ex ante* profits are increasing in v with elasticity:

$$\frac{\partial \ln \mathbb{E}(\pi)}{\partial \ln v} = \ln\left(\frac{\varphi^*}{\varphi_{\min}}\right)^{1/v} + \frac{1}{1 - v\varsigma} > 0. \qquad (6)$$

There are three reasons why a higher v, and hence more dispersion in the distribution of productivity draws, implies higher expected profits. First, the possibility of exit insures firms against bad realisations and increases the value of drawing productivity from a more dispersed distribution. Second, more dispersion increases expected profits whenever the profit function is convex in prices and hence in φ. As (3) shows, this is the case when $\sigma > 2$ (i.e. for $\varsigma > 1$). Third, a higher v increases average productivity by raising the mean of φ. To disentangle the second effect from the third, suppose for a moment that $\varphi_{\min} = \bar{\varphi}(1 - v)$ so that the mean of the distribution is constant at $\bar{\varphi}$ and an increase in v corresponds to a mean-preserving spread. Then:

$$\frac{\partial \ln \mathbb{E}(\pi)}{\partial \ln v} = \ln\left(\frac{\varphi^*}{\varphi_{\min}}\right)^{1/v} + \frac{1}{1 - v\varsigma} - \frac{1}{1 - v},$$

which is necessarily positive when $\varsigma > 1$ ($\sigma > 2$), even in the absence of selection effects (i.e. when $\varphi^* \to \varphi_{\min}$).[22]

---

[22] The intuition is that the firms can expand to take advantage of good realisations of productivity and shrink to insure against bad realisations, making them potentially 'risk loving'.

Having characterised the value of drawing productivity from a distribution with higher v, we now turn to the cost. In order to have a well-defined trade-off, we assume that there are diminishing returns to investing in project size so that the total entry cost is a convex function of v. Formally, we denote the entry cost $\lambda F(v)$ with $F : \mathbb{R}_+ \to \mathbb{R}_+$, $F'(v) > 0$, $F''(v) > 0$ and $\lambda > 0$. We interpret the factor λ as a positive shifter parametrising all the costs of financing the entry investment $F(v)$. Furthermore, to make sure that expected profits are finite, we also assume that there exist a $\bar{v} < 1/\varsigma$ such that $\lim_{v \to \bar{v}} F'(v) = \infty$.

We are now in the position to solve the entry stage. The problem is simplified by the fact that all firms entering a given sector are *ex ante* identical and therefore face the same problem of choosing v so as to maximise expected profits minus the entry cost:

$$\max_{v \in [\underline{v}, \bar{v}]} \{\mathbb{E}(\pi) - \lambda F(v)\},$$

where $\underline{v} \equiv \underline{s}/\kappa$. To ensure that the maximand is concave, we assume $\eta_F'(v) > \eta_\pi'(v)$ where $\eta_F(v) \equiv v F'(v)/F(v)$ and $\eta_\pi(v)$ is (6). Then, the first-order condition for an interior v is:

$$\frac{\mathbb{E}(\pi)}{v} \left[\ln\left(\frac{\varphi^*}{\varphi_{\min}}\right)^{1/v} + \frac{1}{1 - v\varsigma} \right] = \lambda F'(v). \tag{7}$$

Concavity and implicit differentiation allow us to sign the comparative statics for v. The equilibrium choice of v is increasing in the elasticity of substitution, ς, average profit, $\mathbb{E}(\pi)$, and the exit cut-off, $\varphi^*/\varphi_{\min}$. However, both $\mathbb{E}(\pi)$ and $\varphi^*/\varphi_{\min}$ are endogenous, and to solve them, we now turn to the industry equilibrium.

2.3. Industry Equilibrium

Free entry implies that *ex ante* expected profits must be equal to the entry cost: $\mathbb{E}(\pi) = \lambda F(v)$. Substituting (5) into this condition, we can solve for the exit cut-off:

$$\left(\frac{\varphi^*}{\varphi_{\min}}\right)^{1/v} = \frac{f}{\lambda F(v)} \frac{\varsigma}{1/v - \varsigma}. \tag{8}$$

We assume that f/λ is sufficiently high to have $\varphi^*/\varphi_{\min} > 1$ in equilibrium. Next, using $\mathbb{E}(\pi) = \lambda F(v)$ and (8), we can rewrite the first-order condition (7) as:

$$\ln\left[\frac{f}{\lambda F(v)} \frac{\varsigma}{1/v - \varsigma}\right] + \frac{1}{1 - v\varsigma} = \frac{v F'(v)}{F(v)}. \tag{9}$$

Given our previous assumptions ($\eta_F'(v) > \eta_\pi'(v)$), (9) has a unique solution over the relevant range $v \in [\underline{v}, \bar{v}]$. As an illustration, we show in the Appendix functional forms yielding simple analytical solutions and continue here with the more general case.

We can now study the equilibrium determinants of v. A higher fixed cost of production, f, or a lower entry cost, λ, increases the exit cut-off and hence raises the benefit of choosing a more dispersed distribution. A higher elasticity of substitution raises the value of v by making profits more convex in productivity and by increasing the exit cut-off. Interestingly, the choice of innovation size and hence v does not depend on the size of the market, captured by the parameter A. The reason is that a

higher demand increases entry so as to keep expected profit per firm constant without affecting the exit cut-off.

The choice of v affects the equilibrium distribution of firm characteristics. Consider the distribution of revenues, which matches closely the variable documented in Section 1. It is easy to show that revenues are a power function of productivity: $r(\varphi) = r(\varphi^*)(\varphi/\varphi^*)^{\varsigma}$. Then, from the properties of the Pareto distribution, $r(\varphi)$ is also Pareto distributed with c.d.f. $G_r(r) = 1 - (r_{\min}/r)^{1/v\varsigma}$, for $r > r_{\min} = \sigma f$.[23] Hence, the log of revenue is exponential with a standard deviation equal to $v\varsigma$. This immediately implies that differences in the choice of v across sectors will translate into differences in the equilibrium distributions of firm characteristics as summarised in the following Proposition.

PROPOSITION 1. *Assume that the solution to (9) is interior. Then, the equilibrium dispersion of firm productivity and revenue, as measured by the standard deviation of the log of φ and $r(\varphi)$, is larger in sectors with a higher fixed cost, f, higher elasticity of substitution between varieties, ς, and a lower entry cost as parametrised by λ.*

These results are broadly consistent with the empirical evidence documented in Section 1. The model reproduces the positive correlation between the standard deviation and the mean of revenues observed in the data, for two main reasons. First, a high v, by raising average productivity, increases directly average revenue. Second, even in the absence of this effect, the average firm size is proportional to the fixed cost, which is a positive determinant of v. Since v is decreasing in the entry-cost shifter, λ, the model also reproduces the positive correlation between entry and dispersion. This prediction seems also consistent with the casual observation that in industries with very low entry barriers, there are many start-ups but only a few giants survive (e.g. Amazon, Google and Facebook, in the universe of 'dot-com' companies). As long as economic expansions are associated with lower entry costs, for instance through cheap access to finance, the effect of λ may explain why faster economic growth correlates to a rise in dispersion.

Although the model implies a positive correlation between firm heterogeneity and the elasticity of substitution, σ, this prediction may be subject to qualifications. A high σ makes the restriction $v < 1/(\sigma - 1)$ more binding.[24] Moreover, product differentiation is likely to be correlated with fixed costs. This may explain why Syverson (2004a,b) finds less productivity dispersion in industries with higher product substitutability.

Furthermore, it is possible to show that revenue-based labour productivity is also an increasing function of φ and hence shares the same properties.[25] Finally, since the model is static, the results in Proposition (1) should apply to the long-run distribution.

----

[23] If φ follows a Pareto(φ^*, z), then $x \equiv \ln(\varphi/\varphi^*)$ is distributed as an exponential with parameter z. Then, any power function of φ of the type $A\varphi^B$, with A and B constant, is distributed as a Pareto$(A(\varphi^*)^B, z/B)$, since $A\varphi^B = A(\varphi^*)^B e^{Bx}$ with $Bx \sim \text{Exp}(z/B)$, by the properties of the exponential distribution.

[24] To make sure that this constraint is always satisfied, the version of the model presented in the Appendix assumes that the quality of ideas is less dispersed in industries producing more homogeneous goods.

[25] In this model, revenue per worker is an increasing function of φ because the fixed cost of production is in units of labour. The model in Section 4 generates variation in revenue per worker across firms through another channel.

Nevertheless, given the high rate of product turnover observed in the data, the effects of changes in parameters may be detectable even in the short-run.

3. Trade and Equilibrium Firm Heterogeneity

We now extend the model by adding the possibility for firms to export their varieties subject to fixed and variable trade costs. This will lead to the familiar results that only the most productive firms export and that trade forces the least productive firms out. This *ex post* reallocation of revenues will have new implications for the *ex ante* entry stage: by increasing the pay-offs in the tail, trade will induce firms to draw their productivity from more dispersed distributions.

Consider a world economy composed, for simplicity, of two symmetric countries. To serve the foreign market, firms must incur a fixed cost f_x in units of labour and an iceberg variable cost such that $\tau > 1$ units must be shipped for one unit to arrive at destination. The presence of a fixed trade cost implies that only the most productive firms choose to serve the foreign market. Formally, notice that, in analogy to (3), profits from exporting are $\pi_x(\varphi) = A(\varphi/\tau)^{\sigma-1} - f_x$. These profits would be negative for firms with productivity $\varphi < \varphi_x^* = \tau(f_x/A)^{1/\varsigma}$. As usual, we restrict attention to the space of parameters such that $\varphi_x^*/\varphi^* = \tau(f_x/f)^{1/\varsigma} > 1$, so that there is a range of firms with $\varphi \in [\varphi^*, \varphi_x^*]$ operating in the domestic market only, while the most productive firms also export.

Under these assumptions, *ex ante* expected profits are:

$$\mathbb{E}(\pi) = f \int_{\varphi^*}^{\infty} \left[\left(\frac{\varphi}{\varphi^*} \right)^{\varsigma} - 1 \right] \mathrm{d}G(\varphi) + f_x \int_{\varphi_x^*}^{\infty} \left[\left(\frac{\varphi}{\varphi_x^*} \right)^{\varsigma} - 1 \right] \mathrm{d}G(\varphi), \qquad (10)$$

where the two terms represent expected profits from the domestic and the foreign market. Solving the integrals yields:

$$\mathbb{E}(\pi) = \frac{\varsigma}{1/v - \varsigma} \left[f \left(\frac{\varphi_{\min}}{\varphi^*} \right)^{1/v} + f_x \left(\frac{\varphi_{\min}}{\varphi_x^*} \right)^{1/v} \right].$$

To study how export opportunities affect the value of drawing productivity from a more dispersed distribution, we compute again the elasticity of expected profits to v:

$$\frac{\partial \ln \mathbb{E}(\pi)}{\partial \ln v} = \frac{1}{1 - v\varsigma} + \ln \left(\frac{\varphi^*}{\varphi_{\min}} \right)^{1/v} + \frac{\ln(\varphi_x^*/\varphi^*)^{1/v}}{(\varphi_x^*/\varphi^*)^{1/v} f/f_x + 1}. \qquad (11)$$

Comparing this derivative to (6), we see that choosing a more spread-out distribution now yields a new advantage: conditional on surviving, it increases the probability of reaching the export cut-off, φ_x^*. Moreover, as it is well known and we show next, $\varphi^*/\varphi_{\min}$ is higher with trade.

As in autarky, we solve for the equilibrium v by imposing the free-entry condition, $\mathbb{E}(\pi) = \lambda F(v)$. This condition allows us to find the exit cut-off:

$$\left(\frac{\varphi^*}{\varphi_{\min}} \right)^{1/v} = \frac{\varsigma}{1/v - \varsigma} \frac{f + f_x[\tau(f_x/f)^{1/\varsigma}]^{-1/v}}{\lambda F(v)}. \qquad (12)$$

As expected, the exit cut-off is higher than in autarky and is decreasing in τ. For convenience, we now define $\rho \equiv \varphi^*/\varphi_x^* = (f/f_x)^{1/\varsigma}/\tau$ and use it as a synthetic measure

of trade openness. This index, which varies between zero and one, only depends on exogenous parameters and determines the fraction of exporting firms, which is equal to $\rho^{1/v}$. Using this notation and substituting (12) into (11), we can show how trade affects the elasticity of expected profits to v, and hence the incentive to draw productivity from a more spread-out distribution:

$$\frac{\partial^2 \ln \mathbb{E}(\pi)}{\partial \ln v \partial \rho} = \frac{f/f_x}{\rho^{1+1/v}} \frac{\ln \rho^{-1/v}}{v(\rho^{-1/v}f/f_x + 1)^2} > 0. \tag{13}$$

In words, more openness raises unambiguously the return from productivity dispersion. This result is intuitive: trade offers new profitable opportunities, but only to the most productive firms and hence reallocates profits to the right tail of the distribution. In turn, a higher v increases the probability mass in that tail. This is one of the main results of the article: the chance of winning the extra prize of exporting induces firms to bet on bigger innovation projects with more variable outcomes.

Following the same steps as in autarky, the equilibrium v is implicitly determined by:

$$\frac{1}{1-v\varsigma} + \ln\left[\frac{\varsigma}{1/v-\varsigma}\frac{f+f_x\rho^{1/v}}{\lambda F(v)}\right] + \frac{\ln \rho^{-1/v}}{\rho^{-1/v}f/f_x + 1} = \frac{vF'(v)}{F(v)}. \tag{14}$$

Since the left-hand side is increasing in openness (this follows from (13)), and assuming again the solution to be interior, more openness leads to a higher equilibrium v and hence more productivity dispersion. This is formalised in the following Proposition.

PROPOSITION 2. *An increase in openness triggered by a fall in the variable cost of trade, τ, induces firms to choose more spread-out productivity draws (higher v) and raises the equilibrium dispersion of firm productivity, as measured by the standard deviation of the log of φ.*

As in the closed-economy case, a higher v affects the equilibrium dispersion of revenue. In particular, revenues from the domestic and the export market are Pareto distributed with c.d.f. $G_r(r) = 1 - (r_{\min}/r)^{1/v\varsigma}$, with $r_{\min} = \sigma f$ for the domestic sales and $r_{\min} = \sigma f_x$ for exports.[26] Hence, the model is consistent with the finding that export opportunities increase sales dispersion also among non-exporting firms.

Of course, the analytical results derived in this Section partly hinge on functional form assumptions and on the convenient properties of Pareto distributions. Yet, we expect the main mechanism to hold more in general. In particular, as long as trade raises the exit cut-off and reallocates profits to the upper tail of the distribution, it will make expected profits more convex in productivity thereby raising the return from increasing technological heterogeneity.

4. From Firm Heterogeneity to Wage Inequality

We now explore the implications of our theory for income and wage inequality. This is a natural step: the distribution of productivity is likely to be a major determinant of the

---

[26] Note also that an increase in foreign demand would have the same qualitative effect as a fall in τ. In particular, denoting foreign demand as A_x, we would have $\rho = (fA_x/f_xA)^{1/\varsigma}/\tau$, which is increasing in A_x. A proportional increase both in A and A_x would instead have no effect on v.

distribution of wages because in the data more productive firms pay higher wages. Moreover, recent studies show that a large fraction of the increase in wage inequality is due to rising inequality between firms (Barth *et al.*, 2014; Song *et al.*, 2015). We therefore extend the model to allow for differences in wages across firms. This will yield two main results: first, it will highlight a new channel through which trade can increase wage inequality and, second, it will identify some additional variables affecting the choice of dispersion at the entry stage.

In principle, our theory can be used to study top-income inequality. An immediate way of doing this is to draw a link between profits and entrepreneurial income. For example, one could assume that there is a class of agents, entrepreneurs, who are the only ones who can enter and start new firms. These agents may be able to finance part of the entry cost externally and will be the residual claimants on a share of profits. Recent models along these lines include Jones and Kim (2014) or Grossman and Helpman (2014). Since trade increases the dispersion of profits, it will also make entrepreneurial income more unequal. Several contributions in corporate finance, such as Gabaix and Landier (2008), have indeed shown that CEO compensations are proportional to firm size and that this can explain why they have increased so much in recent decades. Our theory can then help rationalise some of the changes in the firm size distribution behind this phenomenon.

Another possibility, that we consider more in detail, is to extend the model to study implications for wage dispersion. In the literature, there are several ways of linking firm productivity to wages. With competitive labour markets, wages can vary because of differences in workforce composition across firms (Yeaple, 2005; Monte, 2011; Sampson, 2014). Alternatively, workers could be paid different wages due to labour market frictions (Egger and Kreickemeier, 2009; Helpman *et al.*, 2010; Amiti and Davis, 2012; Felbermayr *et al.*, 2014). For example, in Helpman *et al.* (2010, HIR henceforth), workers matched randomly with heterogeneous firms draw a match-specific ability which is not observed and firms can invest in costly screening. In equilibrium, more productive firms screen workers more intensively to exclude those with lower ability. As a result, they have workforce of higher average ability and pay higher wages. These models yield an exporter wage premium and have been found to have considerable empirical support (Helpman *et al.*, 2016). We therefore now borrow the framework of HIR to study the implications of our theory for wage dispersion. One key advantage of HIR is that it preserves the main equations of the basic Melitz model, thereby allowing us to apply our previous results in a relatively straightforward manner.

We briefly derive the equations of HIR that are relevant for our purpose and refer the reader to the original article for more details. For ease of comparison, we try to follow the original notation whenever possible. Production depends on the productivity of the firm, φ, the measure of hired workers, h, and the average ability of these workers, $\bar{a}$:

$$y = \varphi h^{\gamma} \bar{a},$$

where $\gamma \in (0, 1)$ implies diminishing returns to hired workers. Two important properties of this production function are the complementarity between firm productivity and average worker ability and a trade-off between the quantity and quality of hired workers. Workers' ability is assumed to be independently distributed and drawn form a Pareto distribution with shape parameter $k > 1$ and c.d.f.

$G_a(a) = 1 - (a_{\min}/a)^{-k}$. Search frictions in the labour market imply that a firm has to pay bn units of the numeraire to be matched randomly with a measure n of workers. Ability is unknown. However, once the match is formed, the firm can use a screening technology to identify workers with ability below a_c at the cost of ca_c^δ/δ units of the numeraire, with $c > 0$ and $\delta > k$. Given the distribution of ability, a firm matched with n workers and screening at the cut-off a_c will hire a measure $h = n(a_{\min}/a_c)^k$ of workers with an average ability of $\bar{a} = a_c k/(k - 1)$. Following the notation in HIR, we define $\beta \equiv 1 - 1/\sigma$. Then, total revenue of a firm with productivity φ can be written as:

$$r(\varphi) = (1 + \mathbb{I}\tau^{1-\sigma})^{1-\beta} P X^{1-\beta}(\varphi\bar{a})^\beta h^{\beta\gamma},$$

where $\mathbb{I}$ is an indicator function taking value 1 if the firm decides to export and zero otherwise.

Wages are determined through strategic bargaining between the firm and workers, after the firm has paid all the costs, which are now defined in units of the numeraire. HIR show that the outcome is that the firm retains a fraction of revenues equal to the Shapley value, $1/(1 + \beta\gamma)$, and pays the rest to the workers. Thus, the profit maximisation problem of the firm is:

$$\pi(\varphi) = \max_{n, a_c, \mathbb{I}} \left\{ \frac{r(\varphi)}{1 + \beta\gamma} - bn - \frac{ca_c^\delta}{\delta} - f - \mathbb{I}f_x \right\},$$

and the first-order conditions for n and a_c are:

$$\frac{\beta\gamma}{1 + \beta\gamma} r(\varphi) = bn(\varphi)$$

$$\frac{\beta(1 - \gamma k)}{1 + \beta\gamma} r(\varphi) = ba_c(\varphi)^\delta.$$

Inspection reveals that firms with higher revenue sample more workers (higher n) and screen more intensively (higher a_c). Assuming $\delta > k$ ensures that firms with higher revenue hire more workers.

Substituting the first-order conditions for n and a_c into the profit function yields $\pi(\varphi) = [\Gamma r(\varphi)/(1 + \beta\gamma)] - f - \mathbb{I}f_x$, with $\Gamma \equiv 1 - \beta\gamma - (1 - \gamma k)\beta/\delta$. Since revenues are increasing in productivity, the fixed costs imply that firms with $\varphi < \varphi^*$ exit (where $\pi_{\mathbb{I}=0}(\varphi^*) = 0$) and firms with $\varphi > \varphi_x^*$ export (where $\pi_{\mathbb{I}=0}(\varphi_x^*) = \pi_{\mathbb{I}=1}(\varphi_x^*)$). Moreover, the relative revenue of any two firms only depends on their relative productivity and export status:

$$\frac{r(\varphi)}{r(\varphi^*)} = (1 + \mathbb{I}\tau^{1-\sigma})^{(1-\beta)/\Gamma} \left(\frac{\varphi}{\varphi^*}\right)^{\beta/\Gamma}.$$

Combining these results, we find an expression for *ex ante* expected profits, $\mathbb{E}(\pi)$, which turns out to be identical to the one in the previous Section (10) after the redefinition of the parameter $\varsigma \equiv \beta/\Gamma$ (instead of $\sigma - 1$). Openness is now:

$$\rho = \frac{\varphi^*}{\varphi_x^*} = (f/f_x)^{1/\varsigma} \left[(1 + \tau^{1-\sigma})^{\varsigma(1-\beta)/\beta} - 1 \right]^{1/\varsigma}.$$

The equilibrium v depends on ς, f and ρ as implied by (14) and, in particular, it is still increasing in ς. The difference, however, is that ς corresponds now to a

combination of parameters, $\varsigma = [\beta^{-1} - \gamma - (1 - \gamma k)/\delta]^{-1}$, so that in this extended version of the model, there are more determinants of v. In particular, through their impact on ς, an increase in γ or a fall in k and δ leads firms to draw from more dispersed distributions. These results are intuitive. As already discussed, more heterogeneity is optimal for the firm when profits are more convex in productivity. In the simpler version of the model, convexity only depends on σ. Now, instead, the profit function is more convex also when there are weaker diminishing returns (high γ) and when screening – which is disproportionately beneficial to more productive firms – is more effective, i.e. when worker ability is more dispersed (low k) and the screening cost not too elastic (low δ).

PROPOSITION 3. *The dispersion of firm productivity, as measured by the standard deviation of the log of φ, is larger in sectors with more ability dispersion and weaker decreasing returns to scale.*

As long as ability dispersion correlates with the standard deviation of educational attainment, the model is consistent with the finding in Section 1 that sales tend to be more heterogeneous in sectors where workers have more unequal school attainments.[27] Proposition 3 may also contribute at explaining why the dispersion of firm productivity varies across countries and over time. It suggests that firms in countries with a more heterogeneous labour force will benefit more from high-variance technologies and hence be more unequal in equilibrium.[28] Likewise, the growing evidence on the 'flattening of the firm' may indicate a rise in the span-of-control parameter and this may help explain the generalised increase in productivity dispersion.

What are the implications for wages? Using the expression of wages as a share of revenue per hired worker yields:

$$w(\varphi) \equiv \frac{\beta\gamma}{1 + \beta\gamma} \frac{r(\varphi)}{h(\varphi)} = b \left[\frac{a_c(\varphi)}{a_{\min}} \right]^k.$$

Since $a_c(\varphi)$ is increasing in productivity, more productive firms pay higher wages. Due to the complementarity between average worker ability and productivity, more productive firms have a stronger incentive to be more selective, hire workers with higher ability and pay them higher wages. Moreover, since wages are proportional to revenue, which jumps at the export cut-off $\varphi = \varphi_x^*$, the model implies an exporter wage premium. More precisely, the wage paid by firms with productivity φ can be written as:

$$w(\varphi) = (1 + \mathbb{I}\tau^{1-\sigma})^{\frac{k(1-\beta)}{\delta\Gamma}} \varphi^{\frac{\beta k}{\delta\Gamma}} w(\varphi^*).$$

As shown in HIR, the wages of workers employed by domestic firms and exporters follow Pareto distributions with shape parameter:

$$1 + \delta[(v\varsigma)^{-1} - 1]/k,$$

which is decreasing in v. Thus, heterogeneity in productivity maps into wage dispersion. This allows us to state the following Proposition on the impact of trade on wage inequality.

PROPOSITION 4. *More openness unambiguously raises sectoral wage dispersion among workers employed by domestic firms and among workers employed by exporters. Conditional on not changing export status, more openness increases wage inequality between workers employed by any pair of firms with different productivity.*

Similar results can also be derived for the wage distribution conditional on ability, a. Higher ability workers have higher average wages because they are hired by firms that have on average a higher productivity. They also face higher inequality, since they can be hired by firms in a wider range of productivity. Moreover, following HIR (2008), it can be shown that the wage of workers employed by domestic firms is a truncated Pareto with a shape parameter decreasing in v. However, since in the model ability is unobservable, the predictions on wage inequality conditional on ability are difficult to take to the data.

Before concluding, it is important to highlight the differences between our result and HIR. In HIR and some other existing models, trade affects wage dispersion through the exporter wage premium. The sign of the effect then depends on the fraction of exporters. As long as exporters are a minority, trade increases wage dispersion by raising the share of firms paying high wages. Once exporters are a majority, instead, trade decreases wage dispersion by pushing low-wage domestic firms out and making the surviving firms more equal. Thus, the overall effect of trade on inequality is inverted U-shaped. This effect is present also in our model. But there is now another, potentially more powerful, force: by making all firms more unequal, trade is changing the slope of the entire wage schedule. This second effect, which is absent in HIR, implies that trade now increases wage inequality within exporters, within non-exporters, and also between the two groups of firms.

5. A Further Look at the Evidence

Having studied the theoretical model, we now turn again to the data to test some of its predictions. We start by focusing on the implications for wage inequality. We show that, consistently with the theory, export opportunities make the distribution of wages more dispersed in a panel of US industries. Next, we probe deeper into the mechanism linking trade, innovation and firm heterogeneity.

5.1. *Wage Dispersion and Export Opportunities*

We start our analysis by documenting how wage inequality varies across industries and time. Our first measure of wage dispersion is the standard deviation of log hourly wages. As detailed in the Appendix, we construct this variable using individual-level wage data from the CPS Merged Outgoing Rotation Groups for the years 1997, 2002

and 2007. For each year, we compute the standard deviation of log wages in 74 industries, defined according to the Census industry classification. We next compute a measure of *residual* wage dispersion after controlling for differences in observable worker characteristics. To this end, we follow Helpman *et al.* (2016) and run Mincer wage regressions of log wages on a large set of covariates, separately for each year to allow for changes in the effects of these characteristics over time. Then, we use the residuals from these regressions to construct the standard deviation of log residual wages in each industry and year. The residual component explains the majority (60%) of the variance of log wages in 2007 and essentially all (97%) of its growth between 1997 and 2007.[29] Besides its relevance in explaining overall wage inequality, residual inequality is closely related to our model, in which wages do not depend on observable characteristics of workers.

The upper part of Table 5 reports descriptive statistics on the dispersion of wages and residual wages in the US manufacturing sector. In the first four columns, we show the average, minimum and maximum value of both standard deviations across the 74 industries in 2007. We also report the average percentage change in each standard deviation between 1997 and 2007. The standard deviation of log wages equals 0.54 and that of residual wages 0.43, implying that wage dispersion is smaller than sales dispersion. Similarly to sales dispersion, wage dispersion exhibits large variation across

Table 5

Descriptive Statistics on Wage Dispersion in US Manufacturing

	Standard deviation				Variance decomp.: within-ind. contrib.	
	Mean	Min.	Max.	% Ch.	Level	Change
Log wages	0.54	0.38	1.10	0.12	0.89	1.06
Log res. wages	0.43	0.26	1.10	0.16	–	–

	Standard deviation (% Ch.)			
	Top 20	Top 40	Top 60	Top 80
Log wages	0.07	0.08	0.08	0.08
Log res. wages	0.13	0.08	0.09	0.09

Notes. The standard deviations are computed for 74 Census industries using worker-level data and are weighted by the number of full-time equivalent hours of labour supply. The sample consists of workers aged 18–64. Residual wages are obtained from yearly Mincer regressions of log hourly wages on log age, log age squared and dummies for race, gender, type of job, country of birth, educational level, union membership, full-time/part-time status, 3-digit occupations, Census industries and states. The mean, minimum and maximum values of the standard deviations are computed across the 74 industries in the year 2007; the percentage changes are computed separately for each industry over 1997–2007, and then averaged across the 74 industries. The unreported between component of the variance decomposition is equal to 1 minus the within component. The bottom part of the Table reports percentage changes in the standard deviations computed on the top 20%, 40%, 60% and 80% of the wage distributions.

[29] See Barth *et al.* (2014) for related evidence on residual wage inequality based on matched employer-employee data for the US over 1977–2002.

industries, with the maximum exceeding the minimum by more than three times. Table 5 also confirms the substantial increase in the US wage inequality documented in previous studies. In particular, the standard deviation of log wages and residual wages rose by 12% and 16%, respectively, between 1997 and 2007.

In the next two columns, we perform a variance decomposition exercise analogous to that in Helpman *et al.* (2016). We decompose the total variance of log wages into a within-industry and a between-industry component and show the percentage contribution of the former to the level (in 2007) and growth (between 1997 and 2007) of this variance. This exercise confirms the well-known fact that wage inequality and its growth are not explained by dispersion of wages between industries but rather by wage dispersion within industries: the within-industry component accounts for 89% of the variance of log wages, and within-industry changes in wage inequality explain essentially all of the increase in the variance of wages over the sample period.

In the bottom part of Table 5, we study how wage inequality changed between 1997 and 2007 across different parts of the wage distribution. To this purpose, we compute the average percentage change in the standard deviations of log wages and residual wages after excluding observations in the left tail of each wage distribution. Moving from left to right in the Table, we restrict consideration to the top 20%, 40%, 60% and 80% of the distribution. These exercises show that the increase in wage inequality remains fairly similar when excluding increasingly larger shares of low-wage individuals. This suggests that the recent rise in US wage inequality corresponds to a widening of the entire distribution, a result also found in Song *et al.* (2015).

Next, we study how the dispersion of wages relates to industry characteristics. To this end, we regress the standard deviation of log residual wages on covariates, using the same sample of 6-digit NAICS industries employed in Section 1. We link the Census industries to 6-digit NAICS codes using correspondence tables from the US Census Bureau. Because we attribute the wage dispersion of each Census industry to all 6-digit NAICS codes corresponding to it, we weight the regressions so as to give less weight to smaller industries; as weights, we use the 1997 shares of the 6-digit industries in total manufacturing employment. As in Helpman *et al.* (2016), we restrict attention to residual wage dispersion, which is closer to the model. Nevertheless, using the standard deviation of log hourly wages yields similar results, which are not reported to save space.

Consistently with the theory, our data show that there is a strong positive correlation between the standard deviation of log establishment sales and the standard deviations of log residual wages. In particular, regressing the former variable on the latter yields a coefficient (SE) of 0.821 (0.322). This result echoes recent firm-level evidence, according to which the dispersions of productivity and wages are positively correlated across US firms (Barth *et al.*, 2014). However, given that the dispersion of sales is endogenous, we now focus on its main determinants. We are especially interested in how trade affects inequality, since the model suggests export opportunities to raise wage dispersion.

The results are reported in Table 6. We start with a pooled regression including all the controls used in Table 2. Interestingly, residual wage dispersion covaries with industry characteristics in a way that mirrors well the pattern found for the dispersion of sales. In particular, residual wage inequality tends to correlate positively with average

Table 6

Dispersion of Residual Wages and Industry Characteristics

	OLS				2SLS	
	Pooled (1)	Pooled (2)	Ind. FE (3)	Ind. FE + FD (4)	Pooled (5)	Ind. FE + FD (6)
Log exp. int.	0.005***	0.008*	0.022***	0.027**	0.012***	0.107**
	(0.002)	(0.005)	(0.008)	(0.011)	(0.004)	(0.043)
Log av. est. sales	0.016**	0.016**	0.104***	0.100**	0.018***	0.120***
	(0.007)	(0.007)	(0.022)	(0.041)	(0.007)	(0.041)
Log n. of est.	0.027***	0.027***	0.075**	0.119*	0.033***	0.132*
	(0.010)	(0.010)	(0.038)	(0.068)	(0.010)	(0.069)
Log tot. empl.	−0.029***	−0.029***	−0.073***	−0.005	−0.034***	0.008
	(0.010)	(0.010)	(0.025)	(0.060)	(0.010)	(0.068)
Log cap. int.	−0.012	−0.012	−0.057	−0.045	−0.009	−0.105
	(0.020)	(0.020)	(0.045)	(0.048)	(0.020)	(0.070)
Log mat. int.	0.008	0.008	−0.099	−0.024	0.010	−0.167
	(0.026)	(0.026)	(0.061)	(0.096)	(0.025)	(0.125)
Log av. educ.	0.237***	0.231***	−0.019	0.029	0.182**	0.064
	(0.077)	(0.080)	(0.141)	(0.240)	(0.080)	(0.275)
SD of log educ.	0.232**	0.228**	−0.199	−0.228	0.192**	−0.208
	(0.091)	(0.092)	(0.129)	(0.164)	(0.095)	(0.180)
GDP gr.	0.135	0.133	0.214*		0.131	
	(0.113)	(0.112)	(0.119)		(0.113)	
Log exp. int. sq.		0.000				
		(0.000)				
Observations	1,036	1,036	1,036	676	1,015	636
R^2	0.07	0.07	0.44	0.51	0.06	0.43
First-stage results						
Log world exports	–	–	–	–	0.532***	0.468***
	–	–	–	–	(0.113)	(0.076)
Kleibergen-Paap F-stat.	–	–	–	–	22.3	38.0

Notes. The dependent variable is the standard deviation of log residual hourly wages. This variable is computed for 74 Census industries using worker-level data, and is weighted by the number of full-time equivalent hours of labour supply. The sample consists of workers aged 18–64. Residual wages are obtained from yearly Mincer regressions of log hourly wages on log age, log age squared and dummies for race, gender, type of job, country of birth, educational level, union membership, full-time/part-time status, 3-digit occupations, Census industries and states. The 74 Census industries are mapped into the corresponding 6-digit NAICS industries using correspondence tables from the US Census Bureau. All variables except for GDP growth are observed at the 6-digit NAICS industry level in the years 1997, 2002 and 2007. All industry-level controls are contemporaneous to the dependent variable. GDP growth is computed over the two years prior to each observation. The specifications in columns (1)–(4) are estimated with ordinary least squares, those in columns (5)–(6) with two-stage least squares. Export intensity is instrumented with non-US exports to the destination markets of the US for each industry and year. All regressions are weighted with the shares of the 6-digit NAICS industries in total manufacturing employment in 1997. Standard errors are clustered by 6-digit industry and reported in parentheses. F-statistics are reported for the Kleibergen-Paap test for weak instruments. ***, **, and * denote significance at 1%, 5%, and 10% respectively. See also notes to previous Tables.

establishment sales, number of establishments and the education variables. Most importantly, the coefficient on export intensity is positive and highly significant.

In column (2), we add the square of log export intensity, to allow for possible non linearities in the relation between exports and residual wage inequality. The quadratic term is small and estimated with little precision, suggesting that in our data residual

wage dispersion and export intensity do not display an inverted U-shaped pattern. In column (3), we add industry fixed effects and in column (4), we further take variables in first differences to account for industry-specific trends. The export intensity coefficient remains positive and precisely estimated also in these specifications.

Finally, in columns (5) and (6), we use two-stage least squares. As before, we instrument export intensity with non-US countries' exports to the destination markets of the US.[30] The coefficient on export intensity remains positive and precisely estimated. To conclude, the evidence presented so far paints a consistent picture, suggesting that the distribution of sales and residual wages behave similarly in a panel of US industries and that export opportunities have made both distributions more spread out.

5.2. *Trade, Innovation Intensity and Firm Heterogeneity*

In this subsection, we provide evidence on the mechanism that links export opportunities to the dispersion of sales. In our model, export opportunities raise expected profits in the tail of the productivity distribution and this induces firms to invest in technologies with higher costs and more dispersed outcomes. We now show that specific predictions of this mechanisms are that:

- (*i*) more export opportunities lead to an increase in the share of revenue invested in innovation; and
- (*ii*) a higher investment in innovation is associated with more heterogeneity across firms.

We then provide a first attempt at testing these predictions.

Since aggregate profits must cover exactly the aggregate entry cost, we can express the share of revenue invested in innovation as:

$$i \equiv \frac{\bar{\pi}}{\bar{r}},$$

where $\bar{\pi}$ is average profit and $\bar{r}$ average revenue made by active firms. We use i as our measure of 'innovation intensity'. To solve for it, note that free entry requires expected profits to be equal to the entry cost, which can be written as $\bar{\pi} = (\varphi^*/\varphi_{\min})^{1/v}\lambda F$, where $(\varphi_{\min}/\varphi^*)^{1/v}$ is the probability of successful entry. Next, recall that average profit is a fraction $1/\sigma$ of average revenue minus average fixed costs, which implies $\bar{r} = \sigma[\bar{\pi} + f + f_x(\varphi^*/\varphi_x^*)^{1/v}]$. Finally, combining these expressions for $\bar{\pi}$ and $\bar{r}$ and using (12) yields:

$$i = \frac{\varsigma v}{\varsigma + 1}. \tag{15}$$

Innovation intensity is a positive function of ς and v only. It follows that a fall in trade costs, τ, affects the share of revenue invested in innovation only through its effect on the dispersion of productivity, v. Since a fall in trade costs induces firms to choose

---

[30] Our second identification strategy – based on the interaction of bulk weight and oil price – is less suited in this case because much of the variation in the bulk weight takes place within the 74 Census industries and does not contribute to identification.

technologies with a higher variance, the model predicts export opportunities to have a positive correlation with innovation intensity. But, this correlation should be zero in a model in which v is exogenous. Thus, investigating empirically the determinants of innovation intensity allows us to test a specific prediction of the model with endogenous heterogeneity in productivity.

To make progress, we follow Aghion *et al.* (2015) and switch from industrial to geographic data. In publicly available databases, innovation measures are typically aggregated into a small number of manufacturing sectors, leaving us with insufficient degrees of freedom.[31] Using geographic data, we can instead obtain a measure of innovation for a wide and long panel of US states. In particular, we proxy for innovation intensity, i, using the log number of granted utility patents, by year of application, per 1,000 workers in each US state, using patent data sourced from Aghion *et al.* (2015).[32] Patents correlate strongly with R&D expenditure and are well suited to measure significant innovations. To construct a state-level measure of export intensity, we instead follow Autor *et al.* (2013) and compute the weighted average of export intensities across 6-digit NAICS industries, using as weights the industries' shares in the states' manufacturing employment by year. This measure of export intensity varies across states due to their different patterns of industrial specialisation. A higher value of export intensity is thus associated with US states specialising in more export-oriented industries.

We start our analysis by showing that, consistent with the model, innovation intensity is correlated with export opportunities. The baseline results are reported in Table 7. To maximise sample size, we use data for all years between 1989 and 2007, the period for which we observe both innovation and export intensity. Since annual changes in patent counts are typically noisy, we estimate all specifications in levels. We report results for both pooled specifications, which exploit the whole variation in the data and specifications with state fixed effects, which control for unobserved heterogeneity at the state level and identify the coefficients through within-state variation over time. Standard errors are clustered at the state level.

In column (1), we start with a parsimonious regression of innovation on exports. The correlation is positive and very precisely estimated, with a t-statistic of 8.0. In column (2), we add a set of control variables to account for other factors that may affect innovation intensity. In particular, we include GDP growth to proxy again for cyclical factors, including changes in entry costs. Moreover, to account for heterogeneous trends in other observable characteristics across states, we include two-year changes in the number of bank deposits per capita, the shares of manufacturing and financial sectors in each state's GDP, and the share of working-age population with at least a college degree in the state. In this way, we control for the fact that innovation intensity might increase more in states where access to credit and skilled labour are growing faster, or in states that are experiencing faster structural change. Reassuringly, the inclusion of these controls leaves our main coefficient

---

[31] For instance, the Bureau of Economic Analysis releases data on R&D investment for 13 broad manufacturing sectors.

[32] Utility patents are meant to protect new and useful innovations, and differ from design and plant patents which protect new product designs or new plants respectively. Thus, utility patents are probably the best proxy for innovation. They cover 90% of all patents registered at the US patent office (USPTO). See Aghion *et al.* (2015) for more discussion on this point.

Table 7

Determinants of Innovation: OLS Regressions

	Pooled			State FE		
	(1)	(2)	(3)	(4)	(5)	(6)
Log exp. int. (t)	1.092***	1.104***		0.706***	0.693***	
	(0.137)	(0.138)		(0.096)	(0.098)	
Log exp. int. ($t-1$)			1.092***			0.639***
			(0.143)			(0.092)
GDP gr.		0.768*	1.302***		−0.197	0.815***
		(0.419)	(0.292)		(0.268)	(0.210)
Bank deposits per capita		−0.052	−0.096		0.111	0.097
		(0.207)	(0.208)		(0.075)	(0.071)
Manuf. share of GDP		−0.263	−0.285		0.036	−0.005
		(0.337)	(0.317)		(0.098)	(0.083)
Finance share of GDP		0.065	−0.120		0.159	0.081
		(0.337)	(0.340)		(0.160)	(0.155)
Share of college-educ. pers.		0.224	0.305		−0.007	−0.001
		(0.184)	(0.189)		(0.104)	(0.123)
Observations	969	969	918	969	969	918
R^2	0.24	0.24	0.23	0.90	0.90	0.91

Notes: The dependent variable is the log number of granted utility patents, by year of application, per 1,000 workers in each US state. The sample period is 1989–2007. All variables except for GDP growth are observed at the state level. Export intensity is obtained as the weighted average of export intensities across 6-digit NAICS industries; the weights are the industries' shares in each state's manufacturing employment in each year. The other state variables are computed as changes over the two years prior to each observation. All specifications are estimated with ordinary least squares. Standard errors are clustered by state and reported in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. See also notes to previous Tables.

essentially unchanged. Next, we re-estimate the last specification replacing the current value of export intensity with its first lag as, for some patents, the innovation process could have started in the previous year triggered by past export opportunities. The results, reported in column (3), are very similar to the baseline specification and continue to highlight a strong positive association between innovation and export intensity. In columns (4)–(6), we repeat the previous specifications controlling for state fixed effects. The coefficient on export intensity slightly drops but remains very precisely estimated.

In Table 8, we take a step towards sorting out causality by re-estimating all specifications with two-stage least squares. To construct an instrument that varies across states and over time, we follow the same procedure used in the construction of export intensity. In particular, we compute the weighted average of our industry-level instrument using industries' employment shares as weights. Because innovation may affect the industrial composition of state employment, we use five-year lags rather than current values of the employment shares, so as to mitigate endogeneity concerns. The resulting instrument attributes stronger demand shocks to states that are specialised in industries where non-US exports to the US destination markets grow faster.[33] The

---

[33] We have also computed state-specific bulk weights using a similar approach. This variable varies little across states because industries with high bulk weights are active in most states with similar employment shares. We therefore lack sufficient cross-sectional variation to identify the effect of oil price shocks.

Table 8

Determinants of Innovation: IV Regressions

	Pooled			State FE		
	(1)	(2)	(3)	(4)	(5)	(6)
Log exp. int. (t)	1.462***	1.461***		1.634***	1.769***	
	(0.214)	(0.220)		(0.294)	(0.303)	
Log exp. int. ($t-1$)			1.519***			1.850***
			(0.220)			(0.409)
GDP gr.		−0.754**	0.583		−0.813**	0.812*
		(0.366)	(0.398)		(0.398)	(0.436)
Bank deposits per capita		−0.009	0.049		−0.071	−0.046
		(0.243)	(0.268)		(0.059)	(0.058)
Manuf. share of GDP		−0.069	−0.131		0.310*	0.176
		(0.301)	(0.274)		(0.185)	(0.172)
Finance share of GDP		−0.249	−0.314		0.214	0.043
		(0.378)	(0.369)		(0.179)	(0.166)
Share of college-educ. pers.		0.042	−0.109		0.374**	0.194
		(0.241)	(0.214)		(0.146)	(0.158)
Observations	663	663	612	663	663	612
R^2	0.21	0.22	0.22	0.91	0.91	0.91
First-stage results						
Log world exports	0.328***	0.343***	0.364***	0.097***	0.102***	0.105***
	(0.043)	(0.045)	(0.044)	(0.018)	(0.018)	(0.024)
Kleibergen-Paap F-stat.	58.5	57.0	67.6	29.3	30.8	19.7

Notes: The dependent variable is the log number of granted utility patents, by year of application, per 1,000 workers in each US state. The sample period is 1989–2007. All variables except for GDP growth are observed at the state level. Export intensity is obtained as the weighted average of export intensities across 6-digit NAICS industries; the weights are the industries' shares in each state's manufacturing employment in each year. The other state variables are computed as changes over the two years prior to each observation. All specifications are estimated with two-stage least squares. Export intensity is instrumented with non-US exports to the destination markets of the US. The state-level instrument, available since 1995, is constructed similarly to state-level export intensity, but the weights are given by the industries' shares in each state's manufacturing employment five years prior to each observation. F-statistics are reported for the Kleibergen-Paap test for weak instruments. ***, ** and * denote significance at 1%, 5% and 10% respectively. See also notes to previous Tables.

number of observations drops because the bilateral trade data used to construct the instrument are available since 1995. The first-stage results confirm that the instrument is strongly correlated with export intensity. At the same time, the second-stage estimates confirm that export intensity has a positive and significant effect on innovation. Noting that export and innovation intensity have standard deviations of 0.3 and 0.68, respectively, our coefficients of roughly 1.5 imply that a 1 SD increase in export intensity leads to a 0.66 SD increase in innovation intensity.

Finally, we study the relation between innovation and firm heterogeneity and show that higher innovation intensity is associated with greater sales dispersion across firms. The results are reported in Table 9. The dependent variable is sales dispersion in a state and year, our proxy for v, obtained by averaging at the state level the standard deviations of log establishment sales across industries, using industries' employment shares in each state and year as weights. We start, in column (1), with a parsimonious specification using innovation intensity as the only regressor. The specification is estimated on 153 observations, corresponding to 51 states in 1997, 2002 and 2007. We control for state fixed effects and correct the standard errors for clustering within

Table 9

Innovation, Trade and Sales Dispersion

	(1)	(2)	(3)	(4)
Log n. of patents	0.644***	0.631***		0.462***
	(0.128)	(0.122)		(0.170)
Log exp. int.			0.895***	0.501
			(0.256)	(0.316)
GDP gr.		1.218*	1.224*	1.128
		(0.670)	(0.726)	(0.676)
Bank deposits per capita		−0.092	−0.155	−0.077
		(0.212)	(0.268)	(0.205)
Manuf. share of GDP		0.414*	0.258	0.407*
		(0.232)	(0.225)	(0.209)
Finance share of GDP		−0.711*	−0.731	−0.598
		(0.354)	(0.503)	(0.417)
Share of college-educ. pers.		0.075	−0.093	0.005
		(0.349)	(0.322)	(0.328)
Observations	153	153	153	153
R^2	0.33	0.46	0.42	0.49

Notes. The dependent variable is sales dispersion (the standard deviation of log establishment sales) in each state and year. The sample period consists of the years 1997, 2002 and 2007. Sales dispersion and export intensity in a state and year are obtained as the weighted average of the corresponding variables across 6-digit NAICS industries; the weights are the industries' shares in the state's manufacturing employment in that year. The number of patents is expressed per 1,000 workers. The other state variables are computed as changes over the two years prior to each observation. All specifications are estimated with ordinary least squares and include state fixed effects. Standard errors are clustered by state and reported in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. See also notes to previous Tables.

states. Then, in column (2), we include additional controls. Consistent with the model, higher innovation intensity is associated with a more dispersed distribution of sales. Next, we regress sales dispersion on export intensity, included either alone (column 3) or together with innovation intensity (column 4). While the coefficient on innovation intensity is largely stable, the coefficient on export intensity drops in magnitude and precision when controlling for patent count, suggesting that the effect of export opportunities on the sales distribution may occur through innovation.

6. Conclusions

In this article, we make several contributions to the literature. First, we document some little-known facts regarding how firm heterogeneity varies across US sectors and over time. We find that the standard deviation of log sales across establishments correlates systematically with industry characteristics, especially export opportunities, and has increased significantly over time. Second, we propose one possible explanation, based on the idea that firms can affect the variance of their productivity at the entry stage. The model formalises the idea that firms can choose between larger innovation projects with more dispersed outcomes, and smaller but less variable projects. It shows that export opportunities, by reallocating profits to the most productive firms, increase the return to technological heterogeneity and induce firms to bet on more ambitious project. Third, we explore the implications for wage inequality and find a new channel through which trade liberalisation can affect the entire wage distribution and increase

its dispersion. Fourth, we find evidence that the distribution of wages varies across US industries and time in a way that mirrors well the distribution of firms' sales. Finally, we use patent data for a panel of US states to provide a first attempt at testing a key mechanism of our model, namely, that export opportunities increase firm hetero-geneity by fostering innovation.

Our analysis could be extended in several directions. In a companion paper (Bonfiglioli *et al.*, 2016), we use a similar model to study how financial frictions affect firm-level heterogeneity and trade. By softening competition and lowering the exit cut-off, financial frictions lower the value of investing in bigger projects with more dispersed outcomes and hence compress equilibrium heterogeneity, especially in more financially vulnerable industries. We provide strong support for this prediction using cross-country indicators of credit supply interacted with cross-sector proxies for financial vulnerability and measuring sales dispersion from highly disaggregated US import data. Despite the completely different empirical strategies and data, the evidence in these two papers is strikingly consistent.

To focus on one mechanism shaping the equilibrium distribution of firms and preserve analytical tractability, we abstract from firm dynamics and *ex post* innovation. Yet, our approach could be applied to innovation strategies of existing firms. Making the model dynamic, for instance along the lines of Gabler and Poschke (2013) or Arkolakis (2015), would also allow to study quantitative implications and transitional adjustments. Within our theory, we also restrict the attention to positive implications. Yet, the model suggests interesting normative questions: is equilibrium heterogeneity too high, especially if workers are risk averse and insurance markets are imperfect? Does international trade introduce new externalities in the technology choice at the entry stage? Finally, our look at the data uncovers a number of new findings, but more can be done to deepen our understanding of how and why productivity and wages vary across firms, sectors and time.

Appendix A. Data

Here, we provide details on data sources and variable definitions.

A.1. *Dispersion of Sales and Labour Productivity*

The dispersion of sales and labour productivity are computed with Census data drawn from the 'Statistics of US Businesses'. For Census years 1997, 2002 and 2007, the database contains receipts of sales and number of establishments and employees, disaggregated by industry and sales-size category. The publicly available data aggregate confidential establishment-level information from the Business Registry, which covers the universe of establishments with paid employees in the US. The number of sales-size bins is equal to 10 for 1997, 8 for 2002 and 18 for 2007. The lowest bin contains firms with revenue smaller than 50,000 US$ while the highest bin contains firms with revenue larger than 100 million US$. In our main analysis, we aggregate the data into six bins consistently defined over the sample period. They refer to firms with US$ revenue of: (1) <100,000; (2) 100,000–499,999; (3) 500,000–999,999; (4) 1,000,000–4,999,999; (5) 5,000,000–99,999,999; (6) 100 million or more. As for the industrial breakdown, the 1997 data are reported at the 4-digit level of the SIC-87 classification, while the 2002 and 2007 data are reported at the 6-digit level of the 1997 and 2002 NAICS classifications respectively. We map the original data into 453 6-digit NAICS industries consistently defined over the three years. To this purpose, we use crosswalks provided by the NBER (between SIC-87 and NAICS-97) and the US

Census Bureau (between different versions of the NAICS classification). For each industry, year and bin, we observe total sales and number of establishments and employees, so we can compute average sales per establishment and average sales per worker in each industry-year-bin triplet. Then, we compute the dispersion of sales in each industry and year as the standard deviation of log average sales per establishment across the six bins, weighting the observations with the number of establishments in each bin. The standard deviation of log labour productivity (sales per worker) is computed analogously.

A.2. *Dispersion of Wages and Educational Attainment*

To construct the standard deviations of log hourly wages and log hourly residual wages, we use individual-level data from the CPS Merged Outgoing Rotation Groups (CPS-MORG). We focus on working-age individuals (18–64 years old) and compute hourly wages as weekly earnings divided by the usual number of hours worked per week. Residual wages are the residuals from yearly Mincer regressions of log wages on log age, log age squared and dummies for race, gender, type of job, country of birth, educational level, union membership, full-time/part-time status, 3-digit occupations, Census industries and states. With this data in hand, we compute the yearly standard deviations of log wages and log residual wages for 74 Census industries, consistently defined over the sample period; both standard deviations are weighted with full-time equivalent hours of labour supply as in Autor *et al.* (2003). We map the 74 Census industries into the 6-digit NAICS industries using correspondence tables from the US Census Bureau. We proceed similarly to construct the mean of educational attainment and the standard deviation of log education across workers. Education is a discrete variable ranging from 1 to 16, with higher values corresponding to more advanced degrees.

A.3. *Export Intensity*

Export intensity is the ratio of exports to total shipments in a given industry and year. Shipment data come from the NBER-CES Manufacturing Industry Productivity Database. Export data are sourced from Schott (2008). For the years 1989–2005, the data are provided by 6-digit NAICS industry and destination country. For later years, they are instead disaggregated by destination and product (10-digit HS), so we compute industry-level bilateral exports by summing exports across all products belonging to a given industry and exported to a given destination. After excluding 144 observations with exports greater than total shipments, our final data set contains export intensity for 375 6-digit NAICS industries over 1989–2007.

A.4. *Instrument*

We instrument export intensity using the sum of exports from all non-US countries to the destination markets of the US in each industry and year. To construct the instrument, we use data on bilateral exports from BACI, available since 1995 at the HS 6-digit product level. We convert these data into 6-digit NAICS industries using correspondence tables from the World Integrated Trade Solution and the US Census Bureau.

A.5. *Bulk Weight and Oil Price*

We construct industry-level bulk weights (expressed in kg per US$) as the export-weighted average of individual products' bulk weights across air and vessel shipments in 1995. We source

data from Schott (2008). We combine the bulk weights with data on oil price (Brent) from FRED (Federal Reserve of St. Louis).

A.6. *Patents*

To construct our measure of innovation intensity, we use the number of granted utility patents, by year of application, in each US state and year. The original data are sourced from Aghion *et al.* (2015) and are available for the period 1970–2009. The data contain information on all patents registered at the US Patent Office (USPTO); each patent is located to the state where its inventor works or lives. We normalise patent counts by total employment in each state and year, sourced from the County Business Patterns (CBS).

A.7. *Other Variables*

The number of establishments and workers in each industry and year come from the 'Statistics of US Businesses'. Factor intensities are computed as in Romalis (2004) using data from the NBER-CES Manufacturing Industry Productivity Database. Material intensity is equal to material costs divided by material costs plus value added. Capital intensity is computed as 1 minus material intensity, times the non-labour share of valued added.

We construct sales dispersion, export intensity and the instrument for each US state and year as the weighted average of these variables across 6-digit NAICS industries. The weights are given by the industries' shares in each state's manufacturing employment by year. We obtain yearly information on the industrial composition of state employment from the CBS. For some state-industry-year cells, employment is not disclosed for confidentiality reasons. To estimate employment in these cells, we use information on the number of establishments in nine employment-size bins, available for each industry and year from the CBS. We first estimate employment in each bin as the product between the number of establishments in the bin and the mid-point of the bin itself. Then, we sum the results across bins to obtain an estimate of total employment in the state-industry-year cell.

Data on state population and GDP by sector come from the regional accounts of the BEA. The number of bank deposits in each state and year is instead sourced from the 'USA Counties Database'. Finally, the share of working-age population with at least a college degree in each state and year is constructed with educational attainment data from the CPS-MORG, weighting the individual observations with full-time equivalent hours of labour supply.

Appendix B. A Simple Case with Closed-form Solutions

To guarantee $v < 1/\varsigma$, assume that $\kappa = \alpha\varsigma$ with $\alpha > 1$ and $s \in (0, 1]$. The first assumption implies that quality of potential ideas is more dispersed in an industry producing more differentiated varieties. This is intuitive: there is little scope for technological differences between very homogeneous products. The second assumption normalises the size of the largest project to one. Then, $v = s/(\alpha\varsigma)$. The parameter $\alpha > 1$ pins down the upper bound to v: $\bar{v} = 1/(\alpha\varsigma)$. Next, assume that the entry cost is proportional to $s/(1 - s)$, which is increasing and tends to infinity as s approaches the maximum size. Thus, $F(v) = v\alpha\varsigma/(1 - v\alpha\varsigma)$. Substituting $F(v)$ and $v = s/(\alpha\varsigma)$ into (9) yields:

$$\ln\left(\frac{f}{\lambda}\frac{1-s}{\alpha-s}\right) + \frac{\alpha}{\alpha-s} = \frac{1}{1-s}. \tag{B.1}$$

Note that $f/\lambda > \alpha$ guarantees that (B.1) has a unique interior solution for s. Since $v = s/(\alpha\varsigma)$, the log of revenue has a standard deviation equal to s/α.

Appendix C. Mean-preserving Spreads

We now solve the model under the assumption that $\varphi_{\min} = \bar{\varphi}(1 - v)$ so that the mean of the distribution is constant at $\bar{\varphi}$ and an increase in v corresponds to a mean-preserving spread. Although the model in the main text suggests that the mean and the variance of productivity are likely to be linked, our goal here is to show that the main results in the aarticle do not change if firms can only choose the variance of their initial draw.

Using the expression for $\partial \ln \mathbb{E}(\pi)/\partial \ln v$ derived in the text, the first-order condition for the problem $\max_{v \in [\underline{v}, \bar{v}]} \{\mathbb{E}(\pi) - \lambda F(v)\}$ becomes:

$$\frac{\partial \mathbb{E}(\pi)}{\partial v} = \frac{\mathbb{E}(\pi)}{v} \left[\frac{1}{1 - v\varsigma} - \frac{1}{1 - v} + \ln \left(\frac{\varphi^*}{\varphi_{\min}} \right)^{1/v} \right] = \lambda F'(v).$$

Imposing free-entry, $\mathbb{E}(\pi) = \lambda F(v)$, yields the same expression for the exit cut-off, $\varphi^*/\varphi_{\min}$. Using these results in the new first-order condition yields:

$$\frac{1}{1 - v\varsigma} - \frac{1}{1 - v} + \ln \left(\frac{f}{\lambda F(v)} \frac{\varsigma}{1/v - \varsigma} \right) = \frac{vF'(v)}{F(v)}.$$

This equation, which pins down implicitly the equilibrium v, is identical to (9), except for the new term $-1/(1 - v)$ on the left-hand side. Intuitively, the fact that a higher variance is not associated with higher expected productivity draw lowers the value of v. This immediately implies that firms will choose a lower equilibrium level of v. However, the comparative statics for all the parameters are unchanged. Moreover, revenue, $r(\varphi)$, is still Pareto distributed with c.d.f. $G_r(r) = 1 - (r_{\min}/r)^{1/v\varsigma}$, for $r > r_{\min} = \sigma f$ as in the benchmark case. Hence, all the results in Proposition 1 still hold.

Consider now the model with trade. Deriving expected profit (10) with respect to v when $\varphi_{\min} = \bar{\varphi}(1 - v)$ yields:

$$\frac{\partial \ln \mathbb{E}(\pi)}{\partial \ln v} = \frac{1}{1 - v\varsigma} - \frac{1}{1 - v} + \ln \left(\frac{\varphi^*}{\varphi_{\min}} \right)^{1/v} + \frac{\ln \left(\varphi_x^*/\varphi^* \right)^{1/v}}{\left(\varphi_x^*/\varphi^* \right)^{1/v} f/f_x + 1}.$$

This expression for the return to risk is again, is identical to (11), except for the new term $-1/(1 - v)$ on the left-hand side.

Imposing free-entry, $\mathbb{E}(\pi) = \lambda F(v)$, yields the same expression (12) for the exit cut-off, $\varphi^*/\varphi_{\min}$. Then, the effect of openness, ρ, on the value of risk, as captured by $\partial^2 \ln \mathbb{E}(\pi)/\partial \ln v \partial \rho$ is identical to (13). Following the same steps as in autarky, the equilibrium v is implicitly determined by:

$$\frac{1}{1 - v\varsigma} - \frac{1}{1 - v} + \ln \left(\frac{\varsigma}{1/v - \varsigma} \frac{f + f_x \rho^{1/v}}{\lambda F(v)} \right) + \frac{\ln \rho^{-1/v}}{\rho^{-1/v} f/f_x + 1} = \frac{vF'(v)}{F(v)}.$$

Once again, the lower value of technological variability (the new term $-1/(1 - v)$ on the left-hand side) implies the firms will draw from less dispersed distributions, but the effect of openness and other parameters on the choice of v is qualitatively unchanged.

Universitat Pompeu Fabra, Barcelona GSE and CEPR
Catholic University of Milan, CEPR and CESifo
CREI, Barcelona GSE and CEPR

Additional Supporting Information may be found in the online version of this article:

Data S1.

References

Acemoglu, D. and Cao, D. (2015). 'Innovation by entrants and incumbents', *Journal of Economic Theory*, vol. 157, pp. 255–94.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R. and Hémous, D. (2015). 'Innovation and top income inequality', Working Paper No. 21247, NBER.

Aghion, P., Blundell, R., Griffith, R., Howitt, P. and Prantl, S.(2004). 'Entry and productivity growth: evidence from microlevel panel data', *Journal of the European Economic Association*, vol. 2(2–3), pp. 265–76.

Aghion, P., Blundell, R., Griffith, R., Howitt, P. and Prantl, S.(2009). 'The effects of entry on incumbent innovation and productivity', *Review of Economics and Statistics*, vol. 91(1), pp. 20–32.

Alfaro, L., Conconi, P., Fadinger, H. and Newman, A. (2016). 'Do prices determine vertical integration?', *Review of Economic Studies*, vol. 83(3), pp. 1–35.

Amiti, M. and Davis, D. (2012). 'Firms, trade, and wages: theory and evidence', *Review of Economic Studies*, vol. 79(1), pp. 1–36.

Arkolakis, C. (2015). 'A unified theory of firm selection and growth', Working Paper No. 17553, NBER.

Atkeson, A. and Burstein, A. (2010). 'Innovation, firm dynamics and international trade', *Journal of Political Economy*, vol. 118(3), pp. 433–84.

Autor, D., Dorn, D. and Hanson, G. (2013). 'The China syndrome: local labour market effects of import competition in the United States', *American Economic Review*, vol. 103(6), pp. 2121–68.

Autor, D., Levy, F., and Murnane, R. (2003). 'The skill content of recent technological change: an empirical exploration', *Quarterly Journal of Economics*, vol. 118(4), pp. 1279–333.

Axtell, R. (2001). 'Zipf distribution of US firm sizes', *Science*, vol. 293(5536), pp. 1818–20.

Bartelsman, E., Haltiwanger, J. and Scarpetta, S. (2009). 'Measuring and analyzing cross country differences in firm dynamics', in (T. Dunne, M. Jensen and B. Roberts, eds.) *Producer Dynamics: New Evidence from Micro Data*, pp. 15–82, Chicago: University of Chicago Press.

Bartelsman, E., Gautier, P. and de Wind, J. (2016). 'Employment protection, technology choice, and worker allocation', *International Economic Review*, vol. 57(3), pp. 787–826.

Barth, E., Bryson, A., Davis, J. and Freeman, R. (2014). 'It's where you work: increases in earnings dispersion across establishments and individuals in the US', Working Paper No. 20447, NBER.

Bernard, A., Jensen, B., Redding, S. and Schott, P. (2012). 'The empirics of firm heterogeneity and international trade', *Annual Review of Economics*, vol. 4, pp. 283–313.

Bernard, A., Redding, S. and Schott, P. (2011). 'Multi-product firms and trade liberalisation', *Quarterly Journal of Economics*, vol. 126(3), pp. 1271–318.

Bonfiglioli, A., Crinò, R. and Gancia, G. (2016). 'Trade, finance and endogenous firm heterogeneity', Discussion Paper No. 11175, CEPR.

Bonfiglioli, A. and Gancia, G. (2014). 'Heterogeneity, selection and labour market disparities', Discussion Paper No. 9981, CEPR.

Broda, C. and Weinstein, D. (2010). 'Product creation and destruction: evidence and price implications', *American Economic Review*, vol. 100(3), pp. 691–723.

Bustos, P. (2011). 'Trade liberalisation, exports, and technology upgrading: evidence on the impact of MERCOSUR on Argentinian firms', *American Economic Review*, vol. 101(1), pp. 304–40.

Cabral, L. and Mata, J. (2003). 'On the evolution of the firm size distribution: facts and theory', *American Economic Review*, vol. 93(4), pp. 1075–90.

Caggese, A. (2016). 'Financing constraints, radical versus incremental innovation, and aggregate productivity', Working Paper, Universitat Pompeu Fabra.

Castellion, G. and Markham, S. (2013). 'Perspective: new product failure rates', *Journal of Product Innovation Management*, vol. 30(5), pp. 976–9.

Dhingra, S. (2013). 'Trading away wide brands for cheap brands', *American Economic Review*, vol. 103(6), pp. 2554–84.

di Giovanni, J. and Levchenko, A. (2012). 'Country size, international trade, and aggregate fluctuations in granular economies', *Journal of Political Economy*, vol. 120(6), pp. 1083–132.

di Giovanni, J., Levchenko, A. and Ranciere, R. (2011). 'Power laws in firm size and openness to trade: measurement and implications', *Journal of International Economics*, vol. 85(1), pp. 42–52.

Doraszelski, U. and Jaumandreu, J. (2013). 'R&D and productivity: estimating endogenous productivity', *Review of Economic Studies*, vol. 80(4), pp. 1338–83.

Dunne, T., Foster, L., Haltiwanger, J. and Troske, K. (2004). 'Wage and productivity dispersion in US manufacturing: the role of computer investments', *Journal of Labour Economics*, vol. 22(2), pp. 397–429.

Egger, H. and Kreickemeier, U. (2009). 'Firm heterogeneity and the labour market effects of trade liberalisation', *International Economic Review*, vol. 50(1), pp. 187–216.

Faggio, G., Salvanes, K. and Van Reenen, J. (2010). 'The evolution of inequality in productivity and wages: panel data evidence', *Industrial and Corporate Change*, vol. 19(6), pp. 1919–51.

Felbermayr, G., Impullitti, G. and Prat, J. (2014). 'Firm dynamics and residual inequality in open economies', Working Paper, University of Nottingham.

Fillat, J. and Garetto, S. (2015). 'Risk, returns, and multinational production', *Quarterly Journal of Economics*, vol. 130(4), pp. 2027–73.

Gabaix, X. and Landier, A. (2008). 'Why has CEO pay increased so much?', *Quarterly Journal of Economics*, vol. 123(1), pp. 49–100.

Gabler, A. and Poschke, M. (2013). 'Experimentation by firms, distortions, and aggregate productivity', *Review of Economic Dynamics*, vol. 16(1), pp. 26–38.

Grossman, G. and Helpman, E. (2014). 'Growth, trade, and inequality', Working Paper, Harvard University.

Head, K., Mayer, T. and Thoenig, M. (2014). 'Welfare and trade without Pareto', *American Economic Review*, vol. 104(1), pp. 310–6.

Helpman, E., Itskhoki, O., Muendler, M.-A. and Redding, S. (2016). 'Trade and inequality: from theory to estimation', *Review of Economic Studies*, forthcoming.

Helpman, E., Itskhoki, O. and Redding, S. (2010). 'Inequality and unemployment in a global economy', *Econometrica*, vol. 78(4), pp. 1239–83.

Helpman, E., Melitz, M. and Yeaple, S. (2004). 'Export versus FDI with heterogeneous firms', *American Economic Review*, vol. 94(1), pp. 300–16.

Hopenhayn, H. (2014). 'Firms, misallocation, and aggregate productivity: a review', *Annual Review of Economics*, vol. 6, pp. 735–70.

Hummels, D. (2007). 'Transportation costs and international trade in the second era of globalisation', *Journal of Economic Literature*, vol. 21(1), pp. 131–54.

Hummels, D., Jørgensen, R., Munch, J. and Xiang, C. (2014). 'The wage effects of offshoring: evidence from Danish matched worker-firm data', *American Economic Review*, vol. 104(6), pp. 1597–629.

Jones, C. and Kim, J. (2014). 'A Schumpeterian model of top income inequality', Working Paper, Stanford University.

König, M., Lorenz, J. and Zilibotti, F. (2016). 'Innovation vs. imitation and the evolution of productivity distributions', *Theoretical Economics*, forthcoming.

Kugler, M. and Verhoogen, E. (2012). 'Prices, plant size, and product quality', *Review of Economic Studies*, vol. 79(1), pp. 307–39.

Lileeva, A. and Trefler, D. (2010). 'Improved access to foreign markets raises plant-level productivity ... for some plants', *Quarterly Journal of Economics*, vol. 125(3), pp. 1051–99.

Luttmer, E. (2010). 'Models of growth and firm heterogeneity', *Annual Review of Economics*, vol. 2, pp. 547–76.

Mayer, T., Melitz, M. and Ottaviano, G. (2015). 'Product mix and firm productivity responses to trade competition', Working Paper, Harvard University.

Melitz, M. (2003). 'The impact of trade on intra-industry reallocations and aggregate industry productivity', *Econometrica*, vol. 71(6), pp. 1695–725.

Melitz, M. and Redding, S. (2014). 'Heterogeneous firms and trade', in (G. Gopinath, E. Helpman and K. Rogoff, eds.) *Handbook of International Economics*, vol. 4, pp. 1–54, Oxford: North Holland.

Monte, F. (2011). 'Skill bias, trade, and wage dispersion', *Journal of International Economics*, vol. 83(1), pp. 202–18.

Mrazova, M., Neary, P. and Parenti, M. (2015). 'Technology, demand and the size distribution of firms', Working Paper, University of Oxford.

Poschke, M. (2015). 'The firm size distribution across countries and skill-biased change in entrepreneurial technology', Working Paper, McGill University.

Romalis, J. (2004). 'Factor proportions and the structure of commodity trade', *American Economic Review*, vol. 94(1), pp. 67–97.

Rossi-Hansberg, E. and Wright, M. (2007). 'Establishment size dynamics in the aggregate economy', *American Economic Review*, vol. 97(5), pp. 1639–66.

Sampson, T. (2014). 'Selection into trade and wage inequality', *American Economic Journal: Microeconomics*, vol. 6(1), pp. 157–202.

Schott, P. (2008). 'The relative sophistication of Chinese exports', *Economic Policy*, vol. 23(53), pp. 5–49.

Song, J., Price, D., Guvenen, F., Bloom, N. and von Wachter, T.(2015). 'Firming up inequality', Working Paper No. 21199, NBER.

Syverson, C. (2004*a*). 'Market structure and productivity: a concrete example', *Journal of Political Economy*, vol. 112(6), pp. 1181–222.

Syverson, C. (2004*b*). 'Product substitutability and productivity dispersion', *Review of Economics and Statistics*, vol. 86(2), pp. 534–50.

Syverson, C. (2011). 'What determines productivity', *Journal of Economic Literature*, vol. 49(2), pp. 326–65.

Vannoorenberghe, G. (2014). 'International trade, risk taking and welfare', *Journal of International Economics*, vol. 92(2), pp. 363–74.

Yeaple, S. (2005). 'A simple model of firm heterogeneity, international trade, and wages', *Journal of International Economics*, vol. 65(1), pp. 1–20.