NBER WORKING PAPER SERIES

LOOKING BACKWARD, INNOVATING FORWARD:
A THEORY OF COMPETITIVE CASCADES

Kevin Lim
Daniel Trefler
Miaojie Yu

Working Paper 30455
http://www.nber.org/papers/w30455

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2022

Trefler thanks CIFAR's Program in Institutions, Organizations and Growth, the Social Sciences and Humanities Research Council of Canada (SSHRC) and the Bank of Canada for generous financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Looking Backward, Innovating Forward: A Theory of Competitive Cascades
Kevin Lim, Daniel Trefler, and Miaojie Yu
NBER Working Paper No. 30455
September 2022
JEL No. F01,F12,F14,O3

ABSTRACT

Innovation depends on exporting and, in particular, on scale and competition in export markets. We develop a theory featuring (1) quality-segmented markets, (2) step-by-step innovation that moves firms forward along the quality ladder, and (3) escape-the-competition motives for innovation. We derive four predictions about the impact on innovation of scale and competition: a firm with a large and less-competitive quality segment ahead or forward of it will have strong incentives to innovate into this profitable segment, while a firm with a small and more-competitive quality segment behind it will also have strong incentives to innovate for fear of facing firms in this segment in the future. We take these predictions to Chinese firm-level data during a period of explosive export growth (2000-2006). Using information about scale and competition by quality segment in China's export markets, we confirm all four hypotheses. By implication, and unlike in standard CES models, the impact of trade on innovation depends critically on how it drives scale and competition in high- versus low-quality segments.

Kevin Lim
University of Toronto
150 St George Street
Suite 323
Toronto, ON M5S3G7
Canada
kvn.lim@utoronto.ca

Daniel Trefler
Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON M5S 3E6
and NBER
dtrefler@rotman.utoronto.ca

Miaojie Yu
Office of the President
Liaoning University
Shenyang, Liaoning Province, China
China
mjyu@nsd.pku.edu.cn

A data appendix is available at http://www.nber.org/data-appendix/w30455

1. Introduction

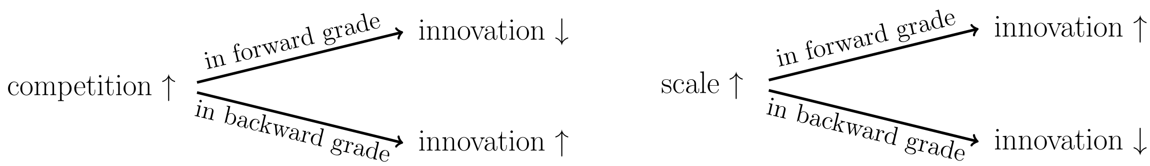
Improved access to foreign markets alters the innovation decisions of firms in two ways. First, it provides firms with a larger market (or *scale*) that increases the returns to innovation. Second, greater scale encourages entry, meaning greater *competition*, which makes innovation less attractive for some firms (business-stealing competition) but more attractive for other firms seeking to escape the competition through innovation. Examples of this research include Lileeva and Trefler (2010), Bustos (2011), and Aghion et al. (2022) on scale, Atkeson and Burstein (2010), Impulliti and Licandro (2018), and Aghion et al. (2018) on business-stealing competition, and Aghion et al. (2001) and Aghion et al. (2005) on escaping the competition. This paper uses firm-level Chinese data for 2000–2006 to investigate econometrically how improved access to foreign markets affects innovation via these scale and competition channels. It does so using a new model that combines step-by-step innovation, escape the competition, and a type of quality segmentation observed in many markets that we now describe.

Consider the mobile phone market. The world's three hundred mobile phone manufacturers can be divided heuristically into three quality segments: a high grade (e.g., Apple and Huawei), a middle grade (e.g., China's Xiaomi), and a low grade (e.g., China's Tecno Mobile focused on Africa). Grades are segmented both because consumers have heterogeneous demands for quality and because today's grade determines the set of grades that can be reached through successful innovation tomorrow ('step-by-step' innovation). After China's entry into the the WTO in 2001, China's mobile phone manufacturers quickly dominated the global low grade before becoming major players in the middle grade, and are now entering the high grade.

Consider Xiaomi's innovation decision. Looking *forward* in quality space, if Xiaomi innovates into the high grade it will open up new export markets (scale); however, this will put it in direct competition with Apple. Whether Xiaomi should innovate depends on the scale and competition forces in its middle grade relative to the high grade. Looking *backward* in quality space, Xiaomi is also nervously tracking low-grade firms such as Tecno Mobile who may soon innovate into the middle grade where they will compete neck-and-neck with Xiaomi. This puts pressure on Xiaomi to innovate in order to escape tomorrow's competition. We refer to these effects of competition on innovation as *competitive cascades* because competition in the high grade discourages innovation in the middle grade, and more firms remaining in the middle grade creates competition that discourages innovation in the low grade.

The need to simultaneously look backward and forward creates a complicated problem for Xiaomi. Yet the solution is characterized by a remarkably clear and testable set of

Figure 1: Four Predictions: Impact of Competition and Scale on Innovation



predictions summarized in figure 1. To describe these succinctly, denote the low, middle, and high grades by $g-1$, g, and $g+1$, respectively. The first two predictions are about the impact of competition on innovation. Firms want to innovate out of more competitive grades and into less competitive grades. This implies that Xiaomi's incentive to innovate is high when grade $g+1$ is *less* competitive. It is also high when grade $g-1$ is *more* competitive, since if Xiaomi fails to innovate, it may be competing tomorrow with firms in $g-1$. In other words, whether competition increases or reduces innovation depends on whether the competition comes from forward or backward grades. A similar logic holds for scale, but in the opposite direction: Xiaomi's incentive to innovate is high when grade $g+1$ has *larger* scale and when grade $g-1$ has *smaller* scale. Hence, even though Xiaomi's problem is complex, there is a sharp characterization of how its innovation decision depends on scale and competition in backward and forward grades.

Despite the fact that many markets feature step-by-step innovation, quality grades, and competitive cascades, we are the first to examine these intuitive and important predictions about exporting and innovation. We test the predictions using China's rapid expansion into foreign markets during 2000–2006. In this period, China's exports grew by a stunning 22% a year and this growth was shared equally by continuing exporters (incumbents) and by first-time exporters (entrants). These firms benefited from the increased scale associated with improved access to foreign markets that came with China's WTO accession in 2001. Chinese firms became the dominant driver of increased competition in many foreign markets, often with Chinese firms competing head-on with each other. We will show that during 2000–2006, these changes in export-market scale and competition contributed in important ways to China's rapid expansion of R&D expenditures, patenting, and new-product sales. (They grew by 45%, 31%, and 16% per year, respectively.) Understanding this contribution is crucial given China's continued central role in global value chains and the impacts of Chinese technological change on global welfare. (See di Giovanni et al., 2014 for a discussion.)

This paper proceeds in three steps. First, we develop a theoretical framework in which products are segmented into quality-based grades and firms must invest in innovation to raise the grade of their product. We use the model to derive an estimating equation that relates firm innovation to export-market size and competition in grades that are forward and backward in quality space relative to a firm's existing grade. Second, we develop a method for estimating each Chinese firm's grade, leveraging a structural demand equation implied by our model. This is an entirely novel approach to a novel problem so we spend time developing and validating it. Third, we use detailed data on trade, production, and innovation outcomes for Chinese firms to estimate the effects of export-market size and competition on innovation, focusing on R&D spending, patents, the value of new-product sales, and the principal component of all three measures. We confirm all four of the predictions in figure 1. Further, for the competition predictions we find that increased competition associated with entry is much more important for innovation than is increased competition from incumbents. Finally, we find these results to be robust to a wide range of alternative specifications, including ones that address endogeneity concerns.

Literature

Our paper contributes to several literatures. First, we add to the body of research documenting how firm-level innovation is affected by exporting (*scale*). There is growing empirical evidence that improvements in export-market access have positive effects on firm-level innovative activity. For example, tariff cuts in export markets have been shown to lead to greater product and process innovation (Lileeva and Trefler, 2010, for Canada), higher spending on technology transfers and high-tech equipment (Bustos, 2011, for Argentina), and patenting (Coelli et al., 2022, for multiple countries). Maican et al. (2021) and Peters et al. (2020) also estimate that export-market profits are a key component of expected returns to R&D for Swedish and German firms respectively.[1] See Shu and Steinwender (2019), Melitz and Redding (2021) and Akcigit and Melitz (2022) for surveys. We contribute to this evidence by showing that the effect of greater export-market size on firm innovation – both in our model and the data – depends on whether the increase occurs backward or forward in quality space relative to the firm's current quality. Our findings also complement branches of this literature that emphasize heterogeneity in export-market size effects with respect to other firm characteristics, for example firm size (Akcigit and Kerr, 2018) and productivity (Aghion et al., 2022).

Second, we contribute to the empirical literature examining how firm-level innovation

[1]There is evidence of such effects outside of the international trade context as well. For example, Acemoglu and Linn (2004) provide evidence that increases in potential market size lead to greater entry of new pharmaceutical products, while Beerli et al. (2018) show that an increase in domestic market size raises productivity for Chinese manufacturing firms.

is affected by competition. Here the evidence is much more mixed, with some studies finding negative impacts of import competition on innovation (e.g., Autor et al., 2020, Liu et al., 2021) and others finding positive effects (e.g., Bloom et al., 2016). Several papers have highlighted that the effects of competition on innovation may depend on firm characteristics such as size (Zhang, 2018) and productivity (Bombardini et al., 2018, Cusolito et al., 2021). We add to this discussion by showing that the effects of competition on innovation depend on where in quality space the competition occurs.

Third, we offer a new theory of what innovation actually buys a firm. Existing frameworks mostly model innovation by incumbent firms as a means of improving production efficiency ("process innovation") and differ in how firm innovation decisions interact.[2] Some frameworks also model innovation as a means by which firms can obtain new products ("product innovation"), by creating these products from scratch,[3] stealing existing products from incumbents,[4] or possibly both.[5] Our model captures elements of both product and process innovation, since a firm that successfully innovates not only creates a new product of higher quality but also changes the distribution of its future productivity shocks. However, the key difference in our model relative to the literature is that successful innovation also changes the *market* in which a firm operates, which then exposes the firm to different export-market conditions. This concept of innovation as a means of ascending through a sequence of ordered markets is key for rationalizing the opposite effects on innovation of competition in forward versus backward grades that we find empirically.

In particular, if all firms operated in a single grade, stronger competition would always discourage innovation due to standard business-stealing effects, whereas when firms can innovate to move up the grade ladder, investing in innovation to *escape* from competition in backward grades becomes possible. In this sense, we also contribute to the more specific discussion about competition associated with escape-the-competition motives for innovation. In seminal work, Aghion et al. (2001) and Aghion et al. (2005) develop models of competition between two firms in which innovation allows one firm to increase

[2]These include partial equilibrium models that abstract from such cross-firm interactions (e.g., Aw et al., 2011, Aghion et al., 2022, Maican et al., 2021, Peters et al., 2020), models with atomistic firms in which firm innovation decisions interact only through general equilibrium price indices (e.g., Atkeson and Burstein, 2010, Costantini and Melitz, 2007, Bustos, 2011, Chen and Xu, 2022, König et al., 2021, Lentz and Mortensen, 2008), duopolistic models that study interactions between two innovating firms (e.g., Acemoglu and Akcigit, 2012, Aghion et al., 2005, Aghion et al., 2001, Akcigit et al., 2021), and oligopolistic models with symmetric firms (e.g., Impulliti and Licandro, 2018).

[3]As in Arkolakis et al. (2018), Atkeson and Burstein (2010), and Bloom et al. (2021).

[4]As in Grossman and Helpman (1991), Klette and Kortum (2004), Acemoglu et al. (2018), and Acemoglu and Linn (2004).

[5]See Atkeson and Burstein (2019).

market share by improving its production capabilities relative to its competitor. However, both firms always remain in the same market, and an increase in competition is modeled as an increase in the substitutability between firms' products instead of growth in the measure of firms operating in each market, as in our case. In other work, Fieler and Harrison (2022) develop a model where firms can choose to compete in nests with many competitors or invest to create new nests with fewer competitors, while in Bloom et al. (2021), firms facing greater import competition can respond by reallocating trapped factors toward innovative activities that create new products. This is similar to our mechanism, in the sense that innovation changes the market in which a firm operates, except that there is no notion of orderedness in these markets, hence the concept of "forward" and "backward" market shocks does not apply.

The outline of the paper is as follows. Section 2 presents the theoretical framework. Section 3 reviews the data. Section 4 describes a critical step in our analysis, namely the estimation of grades and product quality. Section 5 presents our main findings about the impacts on innovation of changes in export-market scale and competition in forward and backward grades. Section 6 addresses endogeneity concerns. Section 7 provides robustness checks. Section 8 concludes with implications for policy, including WTO subsidies reform.

2. Model

We consider an economy at two points in time, $t-1$ and t. In each period, there is a set of firms that each produce a unique product. These products are heterogeneous in a fundamental characteristic that we refer to as the *grade* of the product, indexed by $g \in \{1, \cdots, G\}$ with $G < \infty$.[6] For our purposes, there are three important features of grades. First, products that are of higher grades have higher quality, where quality is modeled as a consumer taste shifter. Second, the domestic and export profits that a firm earns are dependent on the grade in which the firm produces. Third, the grade of a firm at time t depends on the firm's grade at time $t-1$ and the firm's investment in innovation.

Since each firm produces a unique product at a given point in time, we can refer to the grade of firm i at time t without ambiguity and denote this by $g(i,t)$. Furthermore, as we describe below, the innovation decision of firm i at time t will depend on its *lagged* grade at time $t-1$, which for brevity we simply denote by $g(i) \equiv g(i, t-1)$.

[6]We treat G as exogenous. While microfounding this (e.g., in the spirit of Perla and Tonetti, 2014) would provide modelling elegance, it would not add any additional insights for our empirical analysis below. Atkin et al. (2021) adopt a similar assumption.

2.1. Firm heterogeneity

At the end of period $t-1$, firms are heterogeneous in three dimensions: grade $g(i)$, total factor productivity (TFP) which we denote by $\Omega_{i,t-1}$, and export status which we denote by $\delta_{i,t-1}$, where $\delta_{i,t-1}=1$ if the firm exports and $\delta_{i,t-1}=0$ if not. We assume that TFP follows a first-order Markov process with the cumulative density of Ω_{it} conditional on $\Omega_{i,t-1}$ denoted by $F(\cdot|\Omega_{i,t-1})$. We assume that export status also follows a first-order Markov process. Specifically, the probability that a firm which exported in grade g at time $t-1$ exports in period t is denoted by:

$$p_t^{XX,g} \equiv \Pr\left[\delta_{it}=1 \mid \delta_{i,t-1}=1, g(i)=g\right] \tag{1}$$

while the corresponding probability for a non-exporter in period $t-1$ in grade g is:

$$p_t^{NX,g} \equiv \Pr\left[\delta_{it}=1 \mid \delta_{i,t-1}=0, g(i)=g\right] . \tag{2}$$

We make four observations. First, unlike Melitz (2003), exporting is not a choice variable. This simplification will allow us to have incredibly rich heterogeneity across grades (e.g., demand and the cost of innovation will vary by grade) and across firms. We will exploit this empirically. Second, exporting depends on grade, which depends on innovation. Since we will find higher exporting propensities in higher grades, the choice to innovate leads directly to a higher probability of exporting, a feature that the firm internalizes in making its innovation decision. Third, given the innovation process described below, the distribution of TFP depends on grade. Hence, TFP and exporting are correlated because the distribution of both depends on grade. Fourth, we can allow the probability of exporting to depend on directly on TFP, but this does not enrich the empirics below and so we forego the extra notation required.

2.2. Innovation

At the start of period t, before draws of $\{\Omega_{it}, \delta_{it}\}$ are realized, each firm makes a decision about how much to invest in a costly innovation good. This good may be interpreted as R&D, but can also reflect other inputs into the innovation process. A firm i that ended period $t-1$ in grade $g(i)$ chooses an investment level a_{it}, successfully innovates with probability $M^{g(i)}(a_{it})$ and fails to innovate with probability $1-M^{g(i)}(a_{it})$. We refer to M^g as the *innovation success function* in grade g and denote its first derivative by m^g. We assume that M^g is strictly increasing, strictly concave, and satisfies $\lim_{a\to 0} m^g(a)=+\infty$ and $\lim_{a\to+\infty} m^g(a)=0$. We also allow the per-unit nominal cost of the innovation good, denoted by b_t^g, to vary by grade e.g., innovation in higher grades may be more costly.

Conditional on successfully innovating, the firm transitions forward to a grade $g' > g(i)$ with exogenous probability $p_F^{g(i),g'}$, where $\sum_{g'>g(i)} p_F^{g(i),g'} = 1$. Similarly, conditional on unsuccessful innovation, the firm remains in or regresses to a grade $g' \leq g(i)$ with some exogenous probability $p_B^{g(i),g'}$, where $\sum_{g'\leq g(i)} p_B^{g(i),g'} = 1$. Hence, greater investment in innovation makes it more likely that a firm will transition to a higher grade.[7]

The transition probabilities $p_F^{g(i),g'}$ and $p_B^{g(i),g'}$ determine industry dynamics, but we know from multiple data sets that industry dynamics are not entirely controlled by the innovation choices of firms. For example, Foster et al. (2008) find that exit is often preceded by negative demand shocks, which Griliches and Regev (1995) refer to as the "shadow of death." More generally, Foster et al. (2008) show that negative demand shocks play a key role for industry dynamics. Since quality will be defined here as a demand shifter, the Foster et al. demand dynamics are closely related to our grade dynamics. That is, grade regression appears to be an important feature of US firm-level data. It is also likely important in China where we estimate that 12% of firms in a given year experience grade regression. This number is even higher when it includes exit by firms in grade $g = 1$. To accommodate demand/grade dynamics that are often happening independently of firm innovation decisions, we make the following assumption: With exogenous probability η a firm receives an obsolescence shock that drops its grade by one step.[8]

Reviewing the timing, at the start of period t the innovation investment is made, innovation success or failure is realized, and the obsolescence shock is realized. Together, these determine the firm's period-t grade. The firm then draws $\{\Omega_{it}, \delta_{it}\}$, makes production decisions, and receives profits. Let $\pi_t^g(\Omega_{it}, \delta_{it})$ denote the profit that a firm with TFP Ω_{it} and export status δ_{it} captures when it produces in grade g at time t. Since firms choose innovation investments before observing $\{\Omega_{it}, \delta_{it}\}$, the optimal innovation investment for a firm i depends on *expected* profits:

$$\bar{\pi}_{it}^g \equiv \mathbb{E}\left[\pi_t^g(\Omega_{it}, \delta_{it}) \,|\Omega_{i,t-1}, \delta_{i,t-1}, g(i)\right] . \tag{3}$$

[7]For most of this paper we follow the literature in assuming that failed innovation leaves a firm in its current grade. That is, $p_B^{g(i),g(i)} = 1$. We nevertheless allow failed innovation to move a firm backward to capture in the simplest way possible a large management literature documenting and explaining innovation strategies that worsen a product's attractiveness. Christensen (1997) famously described how leaders in the hard drive industry missed the shift in demand towards smaller, more mobile drives. Misguided innovation then led to a loss of market share. Likewise, IBM was a leader in artificial intelligence, but its focus on expert systems left it far behind competitors working on machine learning. Finally, a brief review of the Chinese mobile phone sector reveals many cases where a phone's new function was so beset with problems that it reduced the manufacturer's quality reputation and market share.

[8]One can also allow for a firm to exit with some positive probability conditional on failing to innovate. This is inconsequential for our purposes.

For simplicity, we assume that firms making innovation investment decisions at the start of period t fully discount outcomes beyond that period. Firm i therefore chooses $a_{it} \geq 0$ to maximize the following objective function:

$$-b_t^{g(i)} a_{it} + M^{g(i)}(a_{it}) \sum_{g' > g(i)} \bar{p}_F^{g(i),g'} \bar{\pi}_{it}^{g'} + \left[1 - M^{g(i)}(a_{it})\right] \sum_{g' < g(i)} \bar{p}_B^{g(i),g'} \bar{\pi}_{it}^{g'} \\ + \left[M^{g(i)}(a_{it})\, p_F^{g(i),g(i)+1} \eta + \left[1 - M^{g(i)}(a_{it})\right] p_B^{g(i)g(i)} (1 - \eta)\right] \bar{\pi}_{it}^{g(i)} \tag{4}$$

where $\bar{p}_F^{gg'} \equiv p_F^{gg'}(1-\eta) + p_F^{g,g'+1}\eta$ for $g' > g$ and $\bar{p}_B^{gg'} \equiv p_B^{gg'}(1-\eta) + p_B^{g,g'+1}\eta$ for $g' < g$ are the net probabilities of advancing and regressing given the outcomes of both innovation and obsolescence.[9] The corresponding first-order condition for the firm's problem is then:

$$b_t^{g(i)} = m^{g(i)}(a_{it}) \left[\sum_{g' > g(i)} \bar{p}_F^{g(i),g'} \bar{\pi}_{it}^{g'} - \sum_{g' < g(i)} \bar{p}_B^{g(i),g'} \bar{\pi}_{it}^{g'} + \bar{p}_O^{g(i)} \bar{\pi}_{it}^{g(i)}\right] \tag{5}$$

where $\bar{p}_O^{g} \equiv \eta p_F^{g,g+1} - (1-\eta)\, p_B^{gg}$ (o subscript for *o*wn grade) is the net probability of a firm remaining in its own grade g conditional on the outcome of innovation.

In appendix A.2, we show formally that a model with forward-looking firms leads to a similar first-order condition for innovation investments (see equation A.27), except that the weights $\left\{\bar{p}_F^{gg'}, \bar{p}_B^{gg'}, \bar{p}_O^{g}\right\}$ have a different interpretation: instead of only reflecting the probabilities of transitioning from g to g' in a single period, these weights also reflect the discounted probability of transitioning to g' over multiple periods. Intuitively, firms in g care about profit opportunities in g' if the latter can be reached either in one period through jumps of potentially multiple steps (as in our model) or through jumps over multiple periods (if firms are forward-looking). Hence, in this sense, our model captures the key implications of forward-looking behavior even without allowing for it explicitly. We revisit this discussion in section 7.

Equation (5) is central and anticipates one of our main results. It shows that optimal innovation for firm i depends on outcomes in three types of grades: the firm's own grade

[9] To understand the expression for $\bar{p}_F^{gg'}$ note that a firm moves forward from g to g' in one of two ways: It successfully innovates to g' and does not face obsolescence (with probability $p_F^{g,g'}(1-\eta)$) or it successfully innovates to $g'+1$ and faces obsolescence (with probability $p_F^{g,g'+1}\eta$). The sum of these probabilities is $\bar{p}_F^{gg'}$. Likewise, a firm moves backward from g to g' in two ways whose probabilities sum to $\bar{p}_B^{gg'}$. Furthermore, the term in front of $\bar{\pi}_{it}^{g(i)}$ is the sum of two terms: The first part is the probability of successfully innovating to $g(i)+1$ followed by obsolescence and the second term is the probability of failed innovation that leaves the firm in the same grade $g(i)$ followed by no obsolescence.

$g(i)$, *forward grades* $g' > g(i)$, and *backward grades* $g' < g(i)$. Given the properties of M^g described above, if g' can be reached from $g(i)$ with strictly positive probability then *optimal innovation investments are strictly increasing in* $\bar{\pi}_{it}^{g'}$ *for forward grades and strictly decreasing in* $\bar{\pi}_{it}^{g'}$ *for backward grades.* On the other hand, since the sign of $\bar{p}_O^g$ is ambiguous, so too is the effect of own-grade expected profits $\bar{\pi}_{it}^{g(i)}$ on innovation.

2.3. Profits, scale and competition

Equation (5) is a structural relationship between optimal innovation investments and expected profits in each grade. To link innovation investments to export-market size and competition, we must therefore take a stand on how these variables affect firm profits. We can generally express the profits earned by a firm with TFP Ω_{it} and export status δ_{it} in grade g as the sum of profits from domestic (D) and export (X) markets:

$$\pi_t^g(\Omega_{it}, \delta_{it}) = \pi_t^{D,g}(\Omega_{it}, \delta_{it}) + \pi_t^{X,g}(\Omega_{it}, \delta_{it}) \ . \tag{6}$$

Anticipating that we do not observe export destinations for a large number of Chinese firms in our data, we treat the export-market as a single market. In addition, we assume that preferences for products within a grade g are identical across markets and take a constant elasticity of substitution form with product substitution elasticity σ^g. This implies that domestic and export profits can be expressed respectively as:

$$\pi_t^{D,g}(\Omega_{it}, \delta_{it}) = \left(\frac{1}{\sigma^g}\right) \bar{R}_t^{D,g} \left(\frac{1}{N_t^{D,g}}\right) \left(\frac{\Omega_{it}}{\bar{\Omega}_t^{D,g}}\right)^{\sigma^g - 1} \tag{7}$$

$$\pi_t^{X,g}(\Omega_{it}, \delta_{it}) = \delta_{it} \left(\frac{1}{\sigma^g}\right) \bar{R}_t^{X,g} \left(\frac{1}{N_t^{X,g}}\right) \left(\frac{\Omega_{it}}{\bar{\Omega}_t^{X,g}}\right)^{\sigma^g - 1} \tag{8}$$

where $\bar{R}_t^{D,g}$ ($\bar{R}_t^{X,g}$) is total domestic sales (export revenues) for all firms (exporters) in grade g, $N_t^{D,g}$ ($N_t^{X,g}$) is the number of firms (exporters) in grade g, and $\left\{\bar{\Omega}_t^{D,g}, \bar{\Omega}_t^{X,g}\right\}$ are measures of average TFP among firms and exporters in grade g:

$$\bar{\Omega}_t^{D,g} = \left[\frac{1}{N_t^{D,g}} \sum_{i \in \mathcal{N}_t^{D,g}} \Omega_{it}^{\sigma^g - 1}\right]^{\frac{1}{\sigma^g - 1}}, \qquad \bar{\Omega}_t^{X,g} = \left[\frac{1}{N_t^{X,g}} \sum_{i \in \mathcal{N}_t^{X,g}} \Omega_{it}^{\sigma^g - 1}\right]^{\frac{1}{\sigma^g - 1}} \tag{9}$$

with $\left\{\mathcal{N}_t^D, \mathcal{N}_t^X\right\}$ denoting the set of firms and exporters in grade g. Note that a firm's export status δ_{it} does not matter for its domestic profits, whereas the firm captures positive profits from exporting if and only if $\delta_{it} = 1$.

The intuition for the profit equations (7) and (8) is straightforward. With CES preferences within each grade, profits for a firm in grade g are a constant fraction $\frac{1}{\sigma^g}$ of the firm's sales in that market. Firm sales can then be decomposed into three terms.

First, $\bar{R}_t^{D,g}$ and $\bar{R}_t^{X,g}$ capture the *scale* of the grade g domestic and export markets. These terms account for: (i) total consumer expenditure in the domestic and export markets across all grades; (ii) how consumers differentiate and allocate expenditure across grades in each market; (iii) the extent of foreign competition faced by domestic producers in the domestic market (import competition) and in the export market; and (iv) any variable trade costs associated with exporting. In our empirical application below, we will use data on $\bar{R}_t^{X,g}$ to control for export-market scale. This approach allows us to proceed with the empirical analysis without having to first identify the underlying drivers of scale described above.

Second, $N_t^{D,g}$ and $N_t^{X,g}$ capture the *extensive margin of competition* in the grade g domestic and export markets. Absent firm heterogeneity, each firm (exporter) would capture a constant fraction $\frac{1}{N_t^{D,g}}$ $(\frac{1}{N_t^{X,g}})$ of the domestic (export) market in grade g. Hence, *ceteris paribus*, more competitors reduce the profits that an individual firm is able to capture.

Third, $\bar{\Omega}_t^{D,g}$ and $\bar{\Omega}_t^{X,g}$ capture the *intensive margin of competition* in the grade g domestic and export markets. Firms that are more productive than the average firm in the market capture more than an equal share of the market, whereas firms that are less productive than the average firm capture less than an equal share. Hence, as the productivity of the average competitor in a market increases, sales and profits for each firm fall conditional on the firm's own productivity.

Taking expectations of the profit expressions (7) and (8) conditional on firm i's lagged TFP and export status, we then obtain:

$$\bar{\pi}_{it}^{D,g} = \left(\frac{1}{\sigma^g}\right) \bar{R}_t^{D,g} \left(\frac{1}{N_t^{D,g}}\right) \left(\frac{1}{\bar{\Omega}_t^{D,g}}\right)^{\sigma^g - 1} H^g\left(\Omega_{i,t-1}\right) \tag{10}$$

$$\bar{\pi}_{it}^{X,g} = p_{it} \left(\frac{1}{\sigma^g}\right) \bar{R}_t^{X,g} \left(\frac{1}{N_t^{X,g}}\right) \left(\frac{1}{\bar{\Omega}_t^{X,g}}\right)^{\sigma^g - 1} H^g\left(\Omega_{i,t-1}\right) \tag{11}$$

$$\bar{\pi}_{it}^{g} = \bar{\pi}_{it}^{D,g} + \bar{\pi}_{it}^{X,g} \tag{12}$$

where $H^g\left(\Omega_{i,t-1}\right) \equiv \int \Omega_{it}^{\sigma^g-1} dF\left(\Omega_{it}|\Omega_{i,t-1}\right)$ is the expectation of $\Omega_{it}^{\sigma^g-1}$ conditional on $\Omega_{i,t-1}$ and p_{it} is firm i's expected probability of exporting at t conditional on its export status and grade at time $t-1$:

$$p_{it} \equiv \delta_{i,t-1} p_t^{XX,g(i)} + \left(1 - \delta_{i,t-1}\right) p_t^{NX,g(i)} . \tag{13}$$

2.4. Innovation and export-market scale and competition

We are interested in explaining firm i's expenditure on innovation, denoted by $Y_{it} \equiv b_t^{g(i)} a_{it}$. We need two preliminaries. First, denote the set of *innovation-relevant factors* as:

$$\left\{ \bar{R}_t^{D,g}, \bar{R}_t^{X,g}, N_t^{D,g}, N_t^{X,g}, \bar{\Omega}_t^{D,g}, \bar{\Omega}_t^{X,g} \right\}_{g=1}^{G} . \tag{14}$$

Second, in the following propositions the forward and backward grades of firm i are restricted to those grades that can be reached with positive probability. Plugging equations (10)–(12) into the innovation first-order condition (5) and totally differentiating delivers our two main comparative static results about innovation spending Y_{it}.

PROPOSITION **1.** *Let g' be a forward grade for firm i and suppose $p_{it} > 0$. Holding constant all other innovation-relevant factors, firm i's optimal innovation expenditure Y_{it} is (i) strictly increasing in export-market scale $\bar{R}_t^{X,g'}$, (ii) strictly decreasing in the extensive margin of export-market competition $N_t^{X,g'}$, and (iii) strictly decreasing in the intensive margin of export-market competition $\bar{\Omega}_t^{X,g'}$.*

PROPOSITION **2.** *Let g' be a backward grade for firm i and suppose $p_{it} > 0$. Holding constant all other innovation-relevant factors, firm i's optimal innovation expenditure Y_{it} is (i) strictly decreasing in export-market scale $\bar{R}_t^{X,g'}$, (ii) strictly increasing in the extensive margin of export-market competition $N_t^{X,g'}$, and (iii) strictly increasing in the intensive margin of export-market competition $\bar{\Omega}_t^{X,g'}$.*

In short, less scale and more competition in forward grades make innovation less attractive, whereas less scale and more competition in backward grades have the opposite effect. As highlighted above, we refer to the effects of competition on innovation as *competitive cascades*. Returning to our mobile phone example, more competition from firms such as Apple in a high-quality grade discourages innovation by Xiaomi and makes it more likely that Xiaomi continues competing in a middle-quality grade. This in turn discourages innovation by firms in low-quality grades. In this sense, competition in forward grades cascades downward along the quality ladder. Similarly, looking backward, more competition from firms such as Tecno Mobile in low-quality grades encourages Xiaomi to innovate to escape the competition. This increases the likelihood that Xiaomi becomes a high-quality producer, which in turn increases competition in high-quality grades and encourages firms like Apple to innovate to escape. In this sense, competition in backward grades cascades upward along the quality ladder.

Interestingly, the effects of scale and competition on a firm's own grade are ambiguous because a firm remains in the same grade either through successful innovation plus obsolescence or failed innovation without obsolescence. Mathematically, $\bar{p}_O^g \equiv p_F^{g,g+1}\eta - p_B^{gg}(1-\eta)$ has an ambiguous sign. This is very clear in the special case of *one-step innovation* where successful innovation moves a firm up one grade and unsuccessful innovation leaves the firm in the same grade. Then equation (5) reduces to:

$$\begin{aligned} b_t^{g(i)} =& m^{g(i)}(a_{it})\left[(1-\eta)\bar{\pi}_{it}^{g(i)+1} + \eta\bar{\pi}_{it}^{g(i)} - (1-\eta)\bar{\pi}_{it}^{g(i)} - \eta\bar{\pi}_{it}^{g(i)-1}\right] \\ =& m^{g(i)}(a_{it})\left[(1-\eta)\bar{\pi}_{it}^{g(i)+1} - \eta\bar{\pi}_{it}^{g(i)-1} + (2\eta-1)\bar{\pi}_{it}^{g(i)}\right] . \end{aligned} \tag{15}$$

Combining this with equations (10)–(12) yields the following proposition.

PROPOSITION **3.** *Assume that there is one-step innovation. Let $g = g(i)$ be firm i's grade and suppose $p_{it} > 0$. Holding constant all other innovation-relevant factors, larger scale $\bar{R}_t^{X,g}$, lower extensive-margin competition $N_t^{X,g}$, and lower intensive-margin competition $\bar{\Omega}_t^{X,g}$ (i) raise firm i's innovation if $\eta > \frac{1}{2}$, (ii) lower firm i's innovation if $\eta < \frac{1}{2}$, and (iii) have no effect on firm i's innovation if $\eta = \frac{1}{2}$.*

It is thus useful to have an empirical sense of η's value. In the one-step innovation case, the probabilities of transitioning from $g(i)$ to either $g(i)-1$, $g(i)$ or $g(i)+1$ sum to one and are pinned down by η and the average probability of successful innovation $\bar{M}$. Calibrating η and $\bar{M}$ to match observed transitions in our data generates $\eta = 0.594$, which is close to 0.5.[10] This motivates our baseline specification in which we include expected export-market scale and competition regressors for grades $g-1$ and $g+1$, but not for grade g. In section 7 we then show that when regressors for grade g are included they are statistically and economically insignificant.

3. Data

Production data: Production and sales data (including firm-level exports) are from the 2000–6 Chinese Manufacturing Enterprises (CME) database. We link firms across time following Brandt et al. (2012, 2014, 2017), who have generously published their sophisticated programs. Not suprisingly, our firm counts are almost identical to theirs. See online appendix E.1.1. Appendix B describes how we clean the data and choose which

[10]Consider firms in grades $g = 2, \ldots, G-1$. In the data the probability of moving up for these firms is 0.301 and occurs in the one-step model when there is successful innovation without obsolescence: $\bar{M}(1-\eta) = 0.301$. The probability of moving down for these firms is 0.154 and occurs in the model when innovation fails with obsolescence: $(1-\bar{M})\eta = 0.154$. Solving for $\bar{M}$ and η yields $\eta = 0.594$.

observations to include in the sample. Notably, we always omit processing firms and in robustness checks also exclude state-owned enterprises and foreign-invested firms.

Innovation data: Firms engage in innovation in a variety of formal and informal ways that we measure using data on patents, R&D, and the value of new-product sales. Patent data are from the China National Intellectual Property Administration (CNIPA) and are matched to our CME database using firm names and addresses. R&D and new-product sales are reported directly in the CME. R&D is available for 2001–3 and 2005–6. The three innovation measures are winsorized. Appendix B contains details. Each measure is zero for many firms. To deal with this sparsity we combine the three measures by estimating their common principal component. This principal component of innovation is our baseline measure of innovation, though we always report results separately for R&D, new-product sales, and patents. The principal component is estimated by 2-digit industry. Our sample has 28 industries. Details of estimation, including a table of factor loadings, appear in appendix B. The loadings show that for each industry, the principal component is positively correlated with all three innovation measures. The loadings also show that patents carry the least weight of the three.[11]

TFPR and Markups: We estimate revenue total factor productivity for each firm as in Orr et al. (2019). They show that translog gross-output production functions estimated by 2-digit industry perform well on our data. We use the proxy variable approach in Ackerberg et al. (2015), but with three modifications: the law of motion for firm-level productivity depends on export status as in the learning-by-exporting approach of De Loecker and Warzynski (2012) and De Loecker (2013); the Olley and Pakes (1996) selection correction method is used to correct for attrition bias; and we add lagged capital and its square as additional (over-identified) instruments in estimation of the production functions.[12] With TFPR estimates in hand, we estimate markups using De Loecker and Warzynski (2012) with material inputs. The distribution of our markup estimates are shown in appendix figure A.2. The log of markups has a sensible median of 0.17, with 5th and 95th percentiles of 0.01 and 0.36, respectively.

Quantity and price data: We use data on constant and current dollar output for 2000–3

[11]The patent and R&D data are known to be distorted by government incentive schemes. In this paper we always use these data within narrow bins defined not just by industry and year, but also by quality grade. This purges some if not most of these biases. Chen et al. (2021a) also document the presence of R&D tax 'notches', but we cannot see how these would affect our within-industry-year-grade results.

[12]Details appear in appendix B. Appendix figure A.2 shows that our log TFPR estimates are tightly distributed with an interquartile range of 0.13. The figure also shows that whether or not we make the three proxy-variable modifications does not affect these distributions. Online appendix figure B.1 reports the distributions of our estimated output elasticities for labour, capital and materials, as well as our estimates of returns to scale. These are all sensible.

from the CME to recover quantities. We then recover prices as sales per unit of quantity.[13] In addition, we extrapolate quantities and prices to 2004–6 using the procedure described in online appendix E.1.4. The appendix cross-validation test establishes that our extrapolation procedure is very accurate.

4. The Assignment of Firms to Grades

4.1. Estimation Algorithm

Crucial for our paper is the notion of quality grades. This section defines grades by estimating a *grade assignment* function $g(i)$ that maps firm i in year $t-1$ into grade $g \in \{1, \ldots, G\}$. We assume that consumer preferences in any market take the following CES form across grades (for brevity in this section we suppress industry and time subscripts):

$$U = \left[\sum_{g=1}^{G} \left(\theta^g Q^g \right)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}} \tag{16}$$

where θ^g denotes the quality of grade g products, Q^g is a CES aggregate of products within grade g with elasticity of substitution σ^g (which generates the profit equations 7 and 8), and ρ is the elasticity of substitution across grades. Importantly, we assume that θ^g is strictly increasing in g, which is key for identification of a firm's grade and captures the fact that firms must invest in innovation to improve product quality.

From equation (16), total demand faced by a firm i operating in grade g is:

$$q_i = A^g \left(\theta^g \right)^{\sigma^g - 1} \left(p_i \right)^{-\sigma^g} \tag{17}$$

where A^g is a grade-specific general equilibrium demand shifter reflecting demand from both domestic and export markets and p_i is the firm's output price.[14] In equation (17), q_i and p_i are data while θ^g, σ^g and A^g are unknown parameters that we need in order to estimate $g(i)$.

If we knew each firm's grade assignment, then we could identify the sample of firms in grade g and use the sample to estimate θ^g, σ^g and A^g using standard techniques e.g., Berry (1994). Unfortunately, we do not know the grade assignment and have therefore developed a novel iterative approach. Let n index iterations and let $g_n(i)$ be the estimated

[13] Brandt et al. (2017) use these data to build up the price indexes they (and we) use to construct TFPR.

[14] The form of A^g is straightforward to derive given consumer CES consumer preferences, but this is inconsequential for our analysis.

grade assignment function at iteration n. For the purposes of the algorithm we treat quality as a firm-specific variable. Let $\theta_{i,n}$ be a firm's quality at iteration n.

We start with an initializing choice $\theta_{i,0}$. In iteration $n \geq 1$, we k-means cluster the $\theta_{i,n-1}$ into G clusters. The clustering is the grade assignment $g_n(i)$. Graphically, we are dividing the $\theta_{i,n-1}$ into G intervals that cover the real line. Firms whose $\theta_{i,n-1}$ fall into the g^{th} interval are assigned to grade g.

Before moving to iteration $n+1$ we need $\theta_{i,n}$. Inverting equation (17):

$$\ln \theta_{i,n} = \frac{1}{\sigma^{g_n(i)} - 1} \left(\ln q_i + \sigma_n^{g_n(i)} \ln p_i \right) - \frac{1}{\sigma^{g_n(i)} - 1} \ln A_n^{g_n(i)} \tag{18}$$

where $\sigma_n^{g_n(i)}$ and $A_n^{g_n(i)}$ are the iteration-n estimates of the product substitution elasticity and export-market demand shifter in grade $g_n(i)$, respectively. We estimate σ_n^g using our markup estimates μ_i as follows. Letting S_n^g be the set of firms assigned to grade g in iteration n, the average grade-g markup is $\mu_n^g \equiv \sum_{i \in S_n^g} \mu_i / |S_n^g|$. Appealing to properties of CES, we set $\sigma_n^g \equiv \mu_n^g / (\mu_n^g - 1)$.

To estimate the demand shifters A_n^g, recall that the largest quality in grade g is the smallest quality in grade $g+1$:

$$\max_{i \in S_n^g} \ln \theta_{i,n} = \min_{i \in S_n^{g+1}} \ln \theta_{i,n} \ . \tag{19}$$

Combined with equation (18) this implies the following relation between the demand shifters A_n^{g+1} and A_n^g:

$$\ln A_n^{g+1} = \underline{B}_n^{g+1} - \left(\frac{\sigma_n^{g+1} - 1}{\sigma_n^g - 1} \right) \overline{B}_n^g + \frac{\sigma_n^{g+1} - 1}{\sigma_n^g - 1} \ln A_n^g \tag{20}$$

where $\overline{B}_n^g \equiv \max_{i \in S_n^g} \{ \ln q_i + \sigma_n^g \ln p_i \}$ and $\underline{B}_n^{g+1} \equiv \min_{i \in S_n^{g+1}} \{ \ln q_i + \sigma_n^{g+1} \ln p_i \}$. At this point in iteration n, we know q_i and p_i for all i and σ_n^g and S_n^g for all g. That is, we know everything except the demand shifters. It follows that equation (20) is a first-order difference equation in the A_n^g and these are easily solved as a linear function of A_n^1. Since quality is only meaningful up to a constant, without loss of generality we normalize log quality in grade 1 to zero. Equation (17) then pins down the demand shifter in grade 1:

$$\ln A_n^1 = \ln \bar{q}_n^1 + \sigma_n^1 \bar{p}_n^1 \tag{21}$$

where $\ln \bar{q}_n^1$ and $\ln \bar{p}_n^1$ denote the average log quantity and price in grade 1, respectively.

Now that we know all of the σ_n^g and A_n^g, equation (18) gives us $\theta_{i,n}$. We move to

Table 1: External Validation of Grade Assignment

	k-means with quality				k-means with sales	
Grade	Share of firms	Berry-Khandelwal (trade data)	Schott (price)	Kugler-Verhoogen (Input Defl.)	Share of firms	ln(sales)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.15	0.00	0.00	0.000	0.8029	0.00
2	0.18	0.07	0.16	0.017	0.1398	1.65
3	0.16	0.13	0.21	0.024	0.0374	2.56
4	0.15	0.22	0.23	0.025	0.0122	3.27
5	0.13	0.30	0.26	0.034	0.0049	3.93
6	0.11	0.38	0.27	0.036	0.0018	4.57
7	0.09	0.47	0.32	0.052	0.0007	5.42
8	0.04	0.43	0.40	0.078	0.0003	6.11

Notes: To facilitate comparison across columns, in columns 3, 4, 5, and 7 grade 1 is normalized to zero i.e., the grade 1 value is subtracted from all numbers in the column. Column 4 uses 2000–3 price data.

iteration $n+1$, which starts with k-means clustering on $\theta_{i,n}$ to get $g_{n+1}(i)$. The algorithm continues until $g_{n+1}(i) = g_n(i)$ for all i i.e., until we have a stable estimate of the grade assignment function.[15]

We empirically implement the algorithm as follows. We set the number of grades at $G = 8$ and show in section 7 that our findings are not sensitive to the choice of G. We then implement the algorithm separately for each 2-digit industry, pooling across years in order to track how firms transition across grades over time. See appendix C for details.

4.2. External Validation of Grade Assignment

Features of the grade assignment appear in table 1. Column 2, which sums to one, shows that higher grades have fewer firms, which is consistent with our key assumption that it is costly for firms to advance to higher grades.

To externally validate our assignment of firms to quality grades, we estimate alternative measures of quality by firm and year and then calculate the mean of these measures

[15]Here are some technical details. (*i*) k-means clustering is invariant to adding a constant to the $\theta_{i,n}$ and hence is invariant to the normalization of the export demand shifter A_n^1. (*ii*) Price and quantity data are doing very limited work in our algorithm. As opposed to traditional methods of estimating demand where price and quantity data are required to estimate the main object of interest σ^g, in our setting the σ^g parameters are retrieved from markups. Hence, the role of price and quantity is confined here to identification of the demand shifters A^g.

by grade. The results appear in columns 3–5 of table 1. In the international trade literature, Khandelwal (2010) provides the best-known procedure for estimating quality. He uses US trade data at the Harmonized System level. We adapt his procedure by using firm-level export data matched with our firm-level production data. This allows us to construct firm-level instruments for firm-level prices. See online appendix E.1.5 for details. Column 3 of table 1 reports our Khandelwal-inspired measure of quality averaged across firms in each grade. As is apparent, there is a high degree of correlation between this measure and our assignment of firms to quality grades.

In column 4, we use Schott's (2004) influential quality measure, namely unit value import prices. In column 5, we follow Kugler and Verhoogen (2012) in arguing that a firm's output quality is correlated with its input quality as measured by input prices. These we measure with the Brandt et al. (2017) input deflators. Considering columns 3–5, these measures of quality are all highly correlated with our assignment of firms to grades.

Our quality and grade assignments are, at bottom, firm-specific demand shifters and so can be compared to those in Foster et al. (2008). In online appendix E.1.7 we show that our grades display almost identical dynamics to the demand shifters examined so carefully in Foster et al. (2008).

Finally, one might conjecture that our estimation procedure simply groups together firms that have similar sales, so that what we interpret as heterogeneity in quality is largely driven by heterogeneity in firm size. To examine this, we implement our estimation procedure targeting sales rather than quality in the k-means clustering. Comparing columns 2 and 6, clustering on sales leads to a substantially different distribution of firms across grades. In particular, the distribution based on sales is highly skewed with 80% of firms in grade 1 and only 6% of firms in grades 3–8. Skewness is further evident in column 7 where the average firm in grade 4 is 26 times larger than in grade 1 ($= e^{3.27}$) and the average firm in grade 8 is 450 times larger. In other words, except for a small handful of very large firms, clustering on sales relegates most firms to a small number of grades, whereas clustering on quality produces a more discriminating distribution of firms across grades.

4.3. Assessing Two Key Premises of our Theory

Our theory has two key premises. The first premise is that innovation is an activity that allows firms to move up the quality grade ladder. In appendix D we show this by regressing a firm's grade change $g(i,t) - g(i,t-1)$ on measures of its innovation. The results are strikingly strong: firms that innovate have a much higher probability of moving up the grade ladder. Our second premise is that export-market shocks vary across grades so that firms in different grades are exposed to different shocks. In appendix D we also show that during 2000-6, the highest grades experienced the fastest growth in exports and number of exporters. Hence, we conclude that our iterative procedure delivers a grade assignment that is consistent with the key premises of our theory.

5. Baseline Empirical Results

As explained in the discussion following Proposition 3, the starting point for our baseline empirical analysis is the innovation first-order condition (equation 15) with $\eta = 1/2$ together with the decomposition of export-market profits into scale and competition (equation 11). Since we focus on export markets and work in logs, we drop X superscripts and define:

$$\bar{r}^g_{jt} = \ln(1 + \bar{R}^{X,g}_{jt}) \,, \quad n^g_{jt} = \ln(1 + N^{X,g}_{jt}) \,, \quad \bar{\omega}^g_{jt} = \ln(1 + (\bar{\Omega}^g_{jt})^{\sigma-1}) \text{ and } \omega_{it} = \ln(\Omega_{it}) \tag{22}$$

where we add 2-digit industry subscripts (j) to industry-level variables. We add 1 to logs to deal with the 0.2% of the sample where there are no exporters within a grade-industry-year bin.

Plugging the profit functions of (10)–(12) into the first-order condition (15) and log-linearizing yields our estimation equation (see appendix A.1 for the derivation):

$$\begin{aligned} y_{ijt} = {} & \beta^+_r p_{ijt} \bar{r}^{g(i)+1}_{jt} + \beta^+_n p_{ijt} n^{g(i)+1}_{jt} + \beta^+_\omega p_{ijt} \bar{\omega}^{g(i)+1}_{jt} \\ & + \beta^-_r p_{ijt} \bar{r}^{g(i)-1}_{jt} + \beta^-_n p_{ijt} n^{g(i)-1}_{jt} + \beta^-_\omega p_{ijt} \bar{\omega}^{g(i)-1}_{jt} \\ & + \gamma_p p_{ijt} + \gamma_\omega \omega_{i,t-1} + \alpha_i + \alpha^{g(i)}_{jt} + \varepsilon_{ijt} \,. \end{aligned} \tag{23}$$

where α_i is a firm fixed effect, $\alpha^{g(i)}_{jt}$ is a grade-industry-year fixed effect, and ε_{ijt} accounts for approximation error. A firm's probability of exporting p_{ijt} multiplies each of the export-market variables $\left\{ \bar{r}^{g'}_{jt}, n^{g'}_{jt}, \bar{\omega}^{g'}_{jt} \right\}$ because, intuitively, these variables are only relevant for a firm's innovation investments if the firm expects to be an exporter. We therefore often refer to these products as *expected* export-market scale and competition.[16]

[16] This footnote is intended to show as briefly as possible that we have exact theoretical expressions for each coefficient in equation (23). The β coefficients are defined in appendix equation (A.16). γ_p is defined in (A.17). γ_ω is also defined in (A.17) and appears here because $\omega_{i,t-1}$ is a log-linear approximation of $H^{g(i)}(\Omega_{i,t-1})$ i.e., of expected productivity conditional on productivity in period $t-1$. We experimented with approximations of $H^{g(i)}(\Omega_{i,t-1})$ involving higher-order polynomials in $\omega_{i,t-1}$, but this made no difference to our empirical results. α_i is defined in (A.18). $\alpha^{g(i)}_{jt}$ is defined in (A.19) and, significantly, is a function of the marginal cost of innovation $b^{g(i)}_{jt}$ as well as the the part of domestic profits that is common to

Proposition 1 implies that for forward grades:

$$\beta_r^+ > 0, \quad \beta_n^+ < 0, \quad \beta_\omega^+ < 0\,. \tag{24}$$

Greater expected export-market size and less expected export-market competition in forward grades make them more profitable and hence encourage firms in grade $g(i)$ to innovate. Proposition 2 implies that backward grade signs are opposite of forward grade signs:

$$\beta_r^- < 0, \quad \beta_n^- > 0, \quad \beta_\omega^- > 0\,. \tag{25}$$

Larger expected export-market size and less expected export-market competition in backward grades discourage firms in grade $g(i)$ from innovating.

In section 7 we consider an unrestricted specification where we allow for (1) grade transitions of more than one step, (2) $\eta \neq 1/2$ and the corresponding inclusion of grade-g scale and competition regressors, and (3) regression coefficients that vary by grade e.g., β_r^+ replaced by $\beta_r^{+,g}$. In each case, our baseline conclusions are re-confirmed.

5.1. Baseline Results

We begin by estimating equation (23) using the principal component of innovation as the dependent variable, where each of the variables on the right-hand side are constructed exactly as described in the theory section.[17] Table 2 presents the results. There are 621,879 firm-year observations covering 170,302 firms. All specifications include firm fixed effects. Our baseline specification appears in column 7 and includes the $\alpha_{jt}^{g(i)}$ fixed effects which we henceforth refer to as grade-industry-year fixed effects or gjt fixed effects for short. While the theory is clear that estimation must occur within these detailed gjt bins, the within-gjt sample variation is narrow so in columns 1–6 we build up from simpler sets of fixed effects.

Column 1 includes industry-year fixed effects to control for industry trends. Five conclusions emerge from this column. First, the coefficient signs for expected export-market

all firms, $\bar{R}_{jt}^{D,g(i)}/(\sigma^{g(i)} N_{jt}^{D,g(i)} (\bar{\Omega}_{jt}^{D,g(i)})^{\sigma^{g(i)}-1})$. The reader can easily verify that all terms in the first-order condition (15), and in equations (10)–(12) that are substituted into (15), appear in equation (23).

[17] $\bar{r}_{jt}^{g(i)-1}$, $\bar{r}_{jt}^{g(i)+1}$, $n_{jt}^{g(i)-1}$, $n_{jt}^{g(i)+1}$, $\bar{\omega}_{jt}^{g(i)-1}$ and $\bar{\omega}_{jt}^{g(i)+1}$ are constructed using the discussion surrounding equations (8), (9) and (22). p_{ijt} is constructed using equation (13) together with the sample moments corresponding to equations (1)–(2). We explored many refinements for constructing p_{ijt} e.g., conditioning on a firm's quantile within the productivity distribution ($\omega_{i,t-1}$) or conditioning on a firm's province. Our results are not sensitive to these alternatives.

Table 2: Baseline Specification: Principal Component of Innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exports (r) Backward	-0.84*	-0.90*	-0.78*	-0.84*	-0.90*	-0.97*	-0.62*	-0.61*
$p_{ijt}\bar{r}_{jt}^{g(i)-1}$ $(-)$	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)
	[0.12]*	[0.13]*	[0.12]*	[0.12]*	[0.12]*	[0.12]*	[0.11]*	[0.10]*
Exports (r) Forward	0.41*	0.39*	0.37*	0.36*	0.42*	0.44*	0.22*	0.18*
$p_{ijt}\bar{r}_{jt}^{g(i)+1}$ $(+)$	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
	[0.07]*	[0.07]*	[0.07]*	[0.07]*	[0.08]*	[0.08]*	[0.06]*	[0.07]*
Competition (n) Backward	0.76*	0.78*	0.68*	0.70*	0.78*	0.79*	0.59*	0.57*
$p_{ijt}n_{jt}^{g(i)-1}$ $(+)$	(0.08)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)
	[0.11]*	[0.12]*	[0.11]*	[0.11]*	[0.11]*	[0.11]*	[0.11]*	[0.09]*
Competition (n) Forward	-0.48*	-0.44*	-0.39*	-0.34*	-0.36*	-0.32*	-0.29*	-0.25*
$p_{ijt}n_{jt}^{g(i)+1}$ $(-)$	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
	[0.07]*	[0.07]*	[0.06]*	[0.06]*	[0.07]*	[0.07]*	[0.07]*	[0.05]*
Competition (ω) Backward	0.21*	0.23*	0.22*	0.24*	0.24*	0.27*	0.13*	0.12*
$p_{ijt}\bar{\omega}_{jt}^{g(i)-1}$ $(+)$	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
	[0.05]*	[0.05]*	[0.05]*	[0.05]*	[0.05]*	[0.05]*	[0.04]*	[0.04]*
Competition (ω) Forward	-0.06	-0.09	-0.05	-0.09	-0.07	-0.12*	0.06	0.05
$p_{ijt}\bar{\omega}_{jt}^{g(i)+1}$ $(-)$	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
	[0.04]	[0.05]	[0.04]	[0.04]	[0.05]	[0.05]	[0.04]	[0.04]
Lagged productivity	-0.18*	-0.16*	-0.18*	-0.16*	-0.14*	-0.11*	-0.11*	-0.08*
$\omega_{i,t-1}$	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	[0.03]*	[0.03]*	[0.03]*	[0.03]*	[0.03]*	[0.03]*	[0.02]*	[0.02]*
Prob. of exporting	0.16*	0.16*	0.07	0.08	0.03	0.00	0.03	0.04
p_{ijt}	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	[0.05]*	[0.05]*	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.03]
Fixed effects #2	*jt*	*jpt*	*jt*	*jpt*	*jt*	*jpt*	*gjt*	*gjpt*
Fixed effects #3			*g*	*g*	*gj*	*gjp*		
F tests								
$p\,\bar{r}^{g-1} = p\,\bar{r}^{g+1}$	107.05*	112.75*	89.15*	95.74*	102.85*	114.54*	44.52*	32.66*
$p\,n^{g-1} = p\,n^{g+1}$	130.81*	131.27*	99.54*	100.31*	102.34*	97.35*	65.75*	54.34*
$p\bar{\omega}^{g-1} = p\,\bar{\omega}^{g+1}$	16.34*	23.97*	15.20*	22.84*	16.53*	25.28*	1.38	1.23
All Six = 0	33.84*	32.81*	29.21*	27.32*	33.44*	31.66*	23.04*	19.23*
Observations	621,879	620,018	621,879	620,018	621,879	619,707	621,852	613,799
R^2	0.725	0.736	0.726	0.736	0.726	0.742	0.729	0.755

Notes: This table reports estimates of equation (23). Predicted signs appear next to the row labels. The dependent variable is the principal component of innovation. All specifications include firm fixed effects (170,302 firms). Additional fixed effects are listed in the 'Fixed effects' rows where g indexes firm i's grade in period $t-1$ (8 grades), j indexes 28 2-digit industries, t indexes years (2000–6), and p indexes 31 provinces. Standard errors clustered two-way by firm and gj are in parentheses with a * next to the coefficient indicating significance at the 1% level. Bootstrapped standard errors with two-way clustering (firm and gj) are in square brackets, with a * next to the bracket indicating significance at the 1% level. F tests are for the null that the backward and forward coefficients are equal and for the null that all six coefficients are zero. Fs are based on the standard errors in parentheses and a * indicates rejection of the null at 1%.

size, expected extensive-margin competition, and expected intensive-margin competition are all as predicted. See equations (24) and (25) or the signs next to each row label. Second, all but one of the coefficients is statistically significant. To limit p-hacking, throughout this paper we only report statistical significance at the 1% level (indicated by a single *) and ignore statistical significance at the 5% and 10% levels. Standard errors two-way clustered by firm and grade-industry (gj) are reported in parentheses and a * next to a coefficient is based on these standard errors. Bootstrapped standard errors two-way clustered by firm and gj appear in square brackets and a * there indicates significance based on bootstrapped standard errors. In online appendix E.1.6 we provide a detailed justification for our clustering and show that it produces the largest standard errors among the 12 most reasonable clustering candidates.

Third, the magnitudes of most of our key regression coefficients are large. In this table each dependent and independent variable is pre-scaled by its interquartile range (iqr). Thus, a one iqr increase in backward expected export-market scale leads to a 0.84 iqr decrease in the principal component of innovation. Fourth, since our regressors are correlated, it is possible for them to be individually significant but not jointly significant. The F-statistic for the null that all six regressors are zero appears in the table and is 33.84, well above the cut-off of about 3. Fifth, our key hypothesis implies not only that forward and backward coefficients have different signs, but also that these coefficients are statistically different. The F-test rows report results of tests that the backward and forward coefficients are equal, from which we see that the null of equality is soundly rejected (critical F around 7). Sixth, the extensive-margin competition variables $n_{jt}^{g(i)-1}$ and $n_{jt}^{g(i)+1}$ are much more important than the intensive-margin competition variables $\bar{\omega}_{jt}^{g(i)-1}$ and $\bar{\omega}_{jt}^{g(i)+1}$. Indeed, the latter will sometimes be insignificant in the results reported below. In simpler language, it is the number of competitors that matters more so than the productivity of these competitors.

Moving across the columns in table 2 we increase the number of fixed effects with a view to controlling for omitted demand- and supply-side shocks to innovation. In column 1 we allow for industry trends that might be correlated with our six key variables. For example, our variables might be correlated with (*a*) demand shocks that differ by industry and hence drive differential industry growth or (*b*) supply shocks that differ by industry such as rising spillovers from increased multinational presence in some of China's sectors, which might lower the cost of innovation. (*a*) and (*b*) are controlled for by the industry-year fixed effects in column 1. In column 2 we allow industry trends to vary by province e.g., (*a*) rising incomes in Guangdong drive rising demand for consumer electronics there, (*b*) a new provincial technical institute lowers the local cost of hiring R&D workers or (*c*) a provincial policy encourages spillovers from foreign multinationals. In columns 3–4 we add grade fixed effects to allow levels of innovation to vary across grades e.g., innovation may be more important or more costly in higher grades.[18] In column 5 we allow for grade fixed effects to vary by industry e.g., the innovation-grade gradient is steeper for higher-tech industries. In column 6 the grade-industry-province (gjp) fixed effects allow innovation-grade gradients to vary flexibly across industry-province pairs.

When we move to column 7 we introduce the full set of fixed effects implied by the theory. This is our baseline specification. We see that the coefficients shrink (though all but one remain statistically significant at the 1% level), which tells us that the gt component is important or, put simply, that high and low grades are trending differently and this matters for understanding the relationship between innovation and our six key variables.[19] Finally, in column 8 we allow for grade-industry-province-year ($gjpt$) fixed effects e.g., as Guangdong incomes rise faster than in other regions, consumer demand in that province tilts towards higher-quality electronics. This adds 27,525 fixed effects and leaves us with just 16 observations per fixed effect.[20] Remarkably, the addition of these fixed effects has little impact on our coefficients or their standard errors. Five of our six coefficients have the signs predicted by the theory and are economically and statistically significant. We believe that our saturation of the regression with so many fixed effects goes considerably further than the standard research on China and points to a high level of confidence in our results.

Finally, all specifications in table 2 include two additional time-varying, firm-level variables, $\omega_{i,t-1}$ and p_{ijt}. The theory does not predict their signs.[21] p_{ijt} plays an additional econometric role. Pick any one of our key variables, say $p_{ijt}\bar{r}_{jt}^{g(i)-1}$. We want to know if its importance is due to the interaction of p_{ijt} with $\bar{r}_{jt}^{g(i)-1}$, as implied by the theory, or due to each term separately, which is inconsistent with the spirit of our theory. To answer this, we can include all three of $p_{ijt}\bar{r}_{jt}^{g(i)-1}$, p_{ijt}, and $\bar{r}_{jt}^{g(i)-1}$, which is in fact what we do in our baseline specification since $\bar{r}_{jt}^{g(i)-1}$ is implicitly controlled for by our baseline gjt fixed effects. Further, we have included p_{ijt} in all specifications and its coefficient is tiny both economically and statistically. So it is the interaction $p_{ijt}\bar{r}_{jt}^{g(i)-1}$ that matters, as implied

[18]Our estimated coefficients on the grade fixed effects are higher for higher grades (not shown). This illustrates that in our baseline specification of column 7 with grade-industry-year fixed effects we are *not* exploiting cross-grade differences in innovation or cross-grade differential trends in innovation.

[19]Bøler et al. (2015) provide evidence that Norwegian firms increased R&D spending in response to lower R&D costs (lower b_{jt}^{g}). The average of such an effect is captured here by the gjt fixed effects.

[20](613,799 observations − 170,302 firm fixed effects) / 27,525 $\approx$ 16.

[21]See footnote 16 above or appendix equation (A.17). While the theory does not predict the sign of the coefficient on $\omega_{i,t-1}$, its negative sign is likely due to the within-grade result that higher quality requires more expensive inputs and so comes at the cost of lower productivity (Jaumandreu and Yin, 2018).

by our theory. Finally, excluding p_{ijt} and/or $\omega_{i,t-1}$ makes absolutely no difference to the remaining coefficients. See online appendix table B.5.

Summarizing, we have provided robust evidence that both scale and competition matter for innovation, and in particular that innovation cascades are economically and statistically significant.

5.2. R&D, New-Product Sales, and Patents Separately

In table 3 we move from using the principal component of innovation as our dependent variable to separately using the log of one plus R&D expenditures, the log of one plus new-product sales, and the log of one plus patents. Consider R&D. Column 1 is our baseline specification (same as column 7 of table 2). The pattern of coefficient signs and statistical significance are exactly the same as for the principal component of innovation. That is, the theory correctly predicts five of our six key variables. Turning to magnitudes, the skewness of R&D, new-product sales and patents makes the interquartile range uninformative and so we do not scale these dependent variables. Looking at column 1, the -0.21 coefficient means that a one iqr increase in backward expected export-market size leads to a 0.21 log decrease in R&D expenditures. This is about a third of the mean of log R&D (0.597) and hence is a sizeable effect. Means of the dependent variables appear in the final row of the table. The remaining coefficients on our key variables are smaller in magnitude, but those that are statistically significant imply non-trivial changes of between 0.06-0.18 log points. Column 2 reports results using grade-industry-year-*province* fixed effects. This does not alter our conclusions.[22]

Results for new-product sales are again very similar to those based on the principal component of innovation. For example, in our baseline specification (column 3), a one iqr increase in backward expected export-market size leads to a 0.22 log decrease in new-product sales, which is about a quarter of the mean of the dependent variable (0.949).

Results for patents are weaker. Our baseline specification appears in column 6. All six of the coefficients have the signs predicted by theory and are economically large, but none are statistically significant. In column 5 we retreat from our baseline fixed effects (grade-industry-year) to fewer fixed effects (grade-industry and industry-year). Now all six coefficients are statistically significant.[23]

[22]The full table 2 for each of R&D, new-product sales, and patents appears in online appendix tables B.6–B.8.

[23]The weaker patent results are consistent with Autor et al. (2020) who estimate small impacts on patents of US imports from China. Our results are also consistent with Chen et al. (2021b) who estimate that the majority of benefits from R&D investment by high-tech Chinese manufacturing firms arise from non-patent activities.

Table 3: Baseline Specification: Individual Measures of Innovation

	R&D		New-Product Sales		Patents	
	(1)	(2)	(3)	(4)	(5)	(6)
Exports (r) Backward	-0.21*	-0.21*	-0.22*	-0.19*	-0.026*	-0.011
$p_{ijt}\bar{r}_{jt}^{g(i)-1}$ $(-)$	(0.03)	(0.03)	(0.03)	(0.03)	(0.006)	(0.005)
	[0.04]*	[0.04]*	[0.04]*	[0.03]*	[0.007]*	[0.006]
Exports (r) Forward	0.05*	0.04	0.15*	0.12*	0.018*	0.007
$p_{ijt}\bar{r}_{jt}^{g(i)+1}$ $(+)$	(0.02)	(0.03)	(0.03)	(0.03)	(0.005)	(0.004)
	[0.02]	[0.03]	[0.04]*	[0.03]*	[0.006]*	[0.005]
Competition (n) Backward	0.18*	0.18*	0.23*	0.18*	0.019*	0.009
$p_{ijt}n_{jt}^{g(i)-1}$ $(+)$	(0.03)	(0.03)	(0.04)	(0.03)	(0.005)	(0.004)
	[0.03]*	[0.03]*	[0.05]*	[0.03]*	[0.006]*	[0.005]
Competition (n) Forward	-0.07*	-0.07*	-0.18*	-0.14*	-0.011*	-0.006
$p_{ijt}n_{jt}^{g(i)+1}$ $(-)$	(0.02)	(0.02)	(0.03)	(0.03)	(0.004)	(0.003)
	[0.02]*	[0.02]*	[0.04]*	[0.03]*	[0.005]	[0.004]
Competition (ω) Backward	0.06*	0.04*	0.04*	0.04*	0.009*	0.004
$p_{ijt}\bar{\omega}_{jt}^{g(i)-1}$ $(+)$	(0.01)	(0.01)	(0.01)	(0.01)	(0.002)	(0.002)
	[0.02]*	[0.02]*	[0.01]*	[0.01]*	[0.003]*	[0.002]
Competition (ω) Forward	0.02	0.03	-0.00	-0.01	-0.008*	-0.002
$p_{ijt}\bar{\omega}_{jt}^{g(i)+1}$ $(-)$	(0.01)	(0.01)	(0.01)	(0.01)	(0.002)	(0.002)
	[0.02]	[0.01]	[0.02]	[0.01]	[0.002]*	[0.002]
Lagged productivity	-0.04*	-0.03*	-0.05*	-0.04*	-0.004*	-0.003*
$\omega_{i,t-1}$	(0.01)	(0.01)	(0.01)	(0.01)	(0.001)	(0.001)
	[0.01]*	[0.01]*	[0.01]*	[0.01]*	[0.001]*	[0.001]
Prob. of exporting	-0.02	-0.01	0.07*	0.05*	-0.001	-0.000
p_{ijt}	(0.01)	(0.01)	(0.01)	(0.01)	(0.002)	(0.002)
	[0.02]	[0.01]	[0.02]*	[0.02]*	[0.003]	[0.002]
Fixed effects #2	*gjt*	*gjpt*	*gjt*	*gjpt*	*gj* and *jt*	*gjt*
F tests						
$p\,\bar{r}^{g-1} = p\,\bar{r}^{g+1}$	29.72*	21.23*	46.75*	41.48*	20.93*	5.57
$p\,n^{g-1} = p\,n^{g+1}$	38.23*	38.24*	52.36*	39.40*	12.69*	4.88
$p\bar{\omega}^{g-1} = p\bar{\omega}^{g+1}$	1.49	0.18	3.44	6.05	17.86*	4.13
All Six = 0	16.04*	16.12*	15.84*	12.28*	5.88*	3.06*
Observations	517,156	510,600	611,361	602,937	622,169	622,142
R^2	0.687	0.714	0.740	0.772	0.566	0.571
Mean of dep. var.	0.597	0.597	0.949	0.949	0.042	0.042

Notes: This table is the same as table 2 except the dependent variable is now one of the following: the log of one plus R&D expenditures (columns 1–2), the log of one plus new-product sales (columns 3–4), or the log of one plus the number of patents (columns 5–6). The table reports estimates of equation (23). Predicted signs appear next to the row labels. All specifications include firm fixed effects. Additional fixed effects are listed in the 'Fixed effects #2' row (g for grade, j for industry, t for year, and p for province). See the notes to table 2 for a discussion of standard errors (clustering and bootstrapping) and F-statistics. In particular, a * indicates significance at the 1% level.

5.3. A Simple Placebo Test

We conclude this section with a placebo test to assess whether our findings are somehow mechanically driven by the partitioning of firms into grades. To investigate, we randomly assign firm-year observations to grades, recalculate all of the variables used in our regressions, and re-estimate our baseline specification with the principal component of innovation as the dependent variable (the specification in column 7 of table 2). We repeat this process 1,000 times, each time for a different random draw of grades. We obtain the correct sign pattern in only 10 of these draws, that is, only 0.1% of the time. Thus, grade assignment is playing a key role in our results.[24]

6. Threats to Identification

Our six key variables $p_{ijt}\bar{r}_{jt}^{g(i)-1}$, $p_{ijt}\bar{r}_{jt}^{g(i)+1}$, $p_{ijt}n_{jt}^{g(i)-1}$, $p_{ijt}n_{jt}^{g(i)+1}$, $p_{ijt}\bar{\omega}_{jt}^{g(i)-1}$, and $p_{ijt}\bar{\omega}_{jt}^{g(i)+1}$ are potentially correlated with the error term. In this section we examine potential sources of such correlations.

Correlation of Six Key Variables with Observable Firm Characteristics: Good firms are characterized by a cluster of correlated attributes, e.g., they tend to be large, productive, high-quality exporters. It is possible that our six key variables are correlated with attributes in this cluster and hence that our results are due to omitted variable bias. Table 4 suggests otherwise. The dependent variable is the principal component of innovation. Column 1 repeats our baseline specification from column 7 of table 2. This specification already includes many controls for firm characteristics: firm fixed effects and time-varying controls for lagged productivity ($\omega_{i,t-1}$), lagged export status (p_{ijt} depends on lagged export status $\delta_{i,t-1}$), and lagged quality (grade $g(i)$). In column 2 we add two other correlates of 'good' firms, namely, the lagged log of both domestic sales and employment. As elsewhere, we scale these by their interquartile ranges. Both variables are economically and statistically significant. More importantly, their inclusion makes absolutely no difference to the coefficients on our six key variables. The same conclusion holds when the dependent variable is R&D, new-product sales, or patents. See online appendix table B.10. Thus, our results survive the inclusion and, as we have seen in previous tables, the exclusion of firm fixed effects and five time-varying lagged firm characteristics: productivity, export status, grade, sales, and employment. A simplistic appeal to omitted variable bias

[24] Very similar results obtain when we impose on the randomization that it replicate the actual shares of firm-year observations by grade.

does not explain our results.[25]

Correlation of p_{ijt} with Observable Firm Characteristics and Potential Alternative Mechanisms: Our main hypothesis is that shocks to export-market scale and competition in forward and backward grades affect innovation for firms that are likely to export. Since the probability of exporting p_{ijt} is correlated with many firm attributes, it is possible that the real mechanism driving our results has less to do with exporting and more to do with a correlate of exporting such as productivity or size. To examine this, we replace p_{ijt} with a correlate of exporting. Consider column 3 of table 4. We include all the variables in our column 1 baseline specification as well as six variables of the form $\omega_{i,t-1}\bar{r}_{jt}^{g(i)-1}$.[26] The results are stark. Comparing our baseline column 1 with column 3 shows that the coefficients on our six key variables are unchanged by the inclusion of these alternative interactions. Further, the alternative interactions are not statistically significant individually or jointly. The joint hypothesis that all six are zero has a tiny F-statistic of 1.31. In columns 4 and 5, we replace interactions based on productivity with interactions based on the lagged log of domestic sales and employment, respectively. Again, our baseline interactions are unchanged and the additional interactions are jointly and individually statistically insignificant. Online appendix table B.10 repeats table 4, but for the dependent variables R&D, new-product sales, and patents. The conclusions are the same for each of these. In summary, table 4 lends support to the conclusion that our results are driven by an exporting mechanism and not by a mechanism involving an observable correlate of exporting.

Correlation of p_{ijt} with Unobservable Firm Characteristics: It remains possible that p_{ijt} is correlated with *unobserved* firm characteristics. To purge p_{ijt} of any such correlation, recall from equation (13) that p_{ijt} only depends on firm i's characteristics via i's lagged export status $\delta_{i,t-1}$. We therefore purge p_{ijt} of its i-specific information by replacing $\delta_{i,t-1}$ with a prediction of it based on data from other similar firms. Specifically, we replace $\delta_{i,t-1}$ with its average across the set of firms that share i's grade, industry, year, and productivity quartile. One objection to this is that similar firms have similar unobservables so that our alternative to $\delta_{i,t-1}$ is not purged of the correlation with unobservables. To address this, suppose that firms sort regionally based in part on unobservable characteristics e.g.,

[25] Given the similarity of bootstrapped and regular standard errors, we no longer report the former.

[26] We include $\{\, \omega_{i,t-1}\bar{r}_{jt}^{g(i)-1}, \omega_{i,t-1}\bar{r}_{jt}^{g(i)+1}, \omega_{i,t-1}n_{jt}^{g(i)-1}, \omega_{i,t-1}n_{jt}^{g(i)+1}, \omega_{i,t-1}\bar{\omega}_{jt}^{g(i)-1}, \omega_{i,t-1}\bar{\omega}_{jt}^{g(i)+1} \,\}$. Since our baseline interactions use p_{ijt}, which must lie between 0 and 1, we facilitate comparison by requiring the $\omega_{i,t-1}$ used in the interactions to lie between 0 and 1 and do so by using their percentiles. Likewise in columns 4–5 where we replace lagged productivity with lagged log domestic sales and employment.

Table 4: Endogeneity of p_{ijt} and Alternative Mechanisms

		Add Sales,	Add Interactions:		
	Baseline	Employment	Productivity	Domestic Sales	Employment
	(1)	(2)	(3)	(4)	(5)
Interact with prob. of exporting					
Exports (r) Backward	-0.62*	-0.61*	-0.63*	-0.63*	-0.62*
$p_{ijt}\bar{r}_{jt}^{g(i)-1}$ $(-)$	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Exports (r) Forward	0.22*	0.22*	0.25*	0.25*	0.25*
$p_{ijt}\bar{r}_{jt}^{g(i)+1}$ $(+)$	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Competition (n) Backward	0.59*	0.58*	0.60*	0.61*	0.60*
$p_{ijt}n_{jt}^{g(i)-1}$ $(+)$	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)
Competition (n) Forward	-0.29*	-0.28*	-0.32*	-0.31*	-0.31*
$p_{ijt}n_{jt}^{g(i)+1}$ $(-)$	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Competition (ω) Backward	0.13*	0.13*	0.13*	0.14*	0.13*
$p_{ijt}\bar{\omega}_{jt}^{g(i)-1}$ $(+)$	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Competition (ω) Forward	0.06	0.06	0.06	0.06	0.06
$p_{ijt}\bar{\omega}_{jt}^{g(i)+1}$ $(-)$	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Additional Firm characteristics					
Lagged ln(Domestic Sales$_{i,t\text{-}1}$)		0.19*		0.19*	
		(0.02)		(0.02)	
Lagged ln(Employment$_{i,t\text{-}1}$)		0.39*			0.43*
		(0.03)			(0.03)
Interact with lagged productivity, domestic sales, or employment					
Exports (r) Backward			0.10	-0.08	0.02
$z_{i,t-1}\bar{r}_{jt}^{g(i)-1}$ $(+)$			(0.52)	(0.53)	(0.53)
Exports (r) Forward			-1.03	-1.14	-1.05
$z_{i,t-1}\bar{r}_{jt}^{g(i)+1}$ $(+)$			(0.64)	(0.64)	(0.64)
Competition (n) Backward			0.18	0.41	0.28
$z_{i,t-1}n_{jt}^{g(i)-1}$ $(+)$			(0.53)	(0.54)	(0.54)
Competition (n) Forward			2.21	2.36	2.25
$z_{i,t-1}n_{jt}^{g(i)+1}$ $(-)$			(0.93)	(0.92)	(0.92)
Competition (ω) Backward			-0.42	-0.38	-0.41
$z_{i,t-1}\bar{\omega}_{jt}^{g(i)-1}$ $(+)$			(0.45)	(0.45)	(0.46)
Competition (ω) Forward			-0.68	-0.67	-0.71
$z_{i,t-1}\bar{\omega}_{jt}^{g(i)+1}$ $(-)$			(0.54)	(0.54)	(0.53)
F-test for 6 interactions with lagged prod, sales or employ			1.31	1.48	1.38
Observations	621,852	619,315	621,852	621,852	619,315
R^2	0.729	0.730	0.729	0.729	0.730

Notes: The dependent variable is the principal component of innovation. Column 1 repeats column 7 of table 2. Included but not reported in all specifications are lagged productivity ($\omega_{i,t-1}$), the probability of exporting (p_{ijt}), firm fixed effects, and gjt fixed effects. Predicted signs appear next to row labels. Standard errors are two-way clustered by firm and gj. A * indicates significance at the 1% level.

electronics firms of a particular grade and productivity quartile co-locate in the province of Guangdong. Then we can further purge unobservables by replacing $\delta_{i,t-1}$ with its mean among the set of firms in i's grade-industry-year-productivity bin *excluding* firms that are in the same province as i.[27] Let $\delta_{-p(i),t-1}$ be this leave-one-province-out predictor of $\delta_{i,t-1}$. Thus, for example, we are replacing the lagged export status of a high-quality, high-productivity electronics, firm in Guangdong with its mean counterpart outside of Guangdong i.e., outside of China's main electronics cluster. Firms that choose to locate outside the main cluster are different in terms of observables (not reported) and hence are likely different in terms of unobservables. Restated, $\delta_{-p(i),t-1}$ is unlikely to be correlated with i's unobservables. Table 5 reports the results when the dependent variable is the principal component of the innovation. Column 1 uses the same specification as our baseline (column 7 of table 2). Column 2 recalculates p_{ijt} using $\delta_{-p(i),t-1}$ in place of $\delta_{i,t-1}$. This barely affects the signs and statistical significance of the coefficients on our six key variables. We conclude from this that endogeneity bias arising from the correlation of p_{ijt} with unobservable firm characteristics is unlikely to be a problem.

IV provides an alternative approach to dealing with unobservables. We treat each of our six key variables as endogenous variables and each of their six counterparts based on $\delta_{-p(i),t-1}$ as instruments. This approach is fraught because the small-sample properties of IV deteriorate rapidly as the number of instruments grow. Keeping this caveat in mind, column 3 of table 5 reports the IV results. They are similar to our baseline results. Minor exceptions are the coefficients on backward export-market size and extensive-margin competition which are somewhat larger, though not statistically so. Similar results hold for R&D, new-product sales, and patents, as shown in appendix table B.11. The conclusion from IV is again that our results are not likely driven by a correlation of p_{ijt} with unobservable firm characteristics.[28]

Correlation of Observed and Unobserved Firm Characteristics with $\bar{r}_{jt}^{g-1}$, $\bar{r}_{jt}^{g+1}$, n_{jt}^{g-1}, n_{jt}^{g+1}, $\bar{\omega}_{jt}^{g-1}$ and $\bar{\omega}_{jt}^{g+1}$: Consider one of our key variables, say $p_{ijt}\bar{r}_{jt}^{g-1}$. We have argued above that endogeneity of p_{ijt} is unlikely to be a problem. We now consider whether $p_{ijt}\bar{r}_{jt}^{g-1}$ is endogenous because $\bar{r}_{jt}^{g-1}$ is endogenous. In what follows we make the argument using $\bar{r}_{jt}^{g-1}$ and $\bar{r}_{jt}^{g+1}$, but the same holds for n_{jt}^{g-1}, n_{jt}^{g+1}, $\bar{\omega}_{jt}^{g-1}$, and $\bar{\omega}_{jt}^{g+1}$ as well. We

[27] The quartiles of the distribution of productivity among firms in firm i's grade-industry-year bin.

[28] Here are some details of IV. (*i*) The Kleibergen-Paap weak instruments F-statistic is well above the Stock-Yogo heuristic of 20. See the last column of table 5. (*ii*) Each of the six first stages is sensible in that it loads heavily on the 'own' instrument e.g., in the first stage for $p_{ijt}\bar{r}_{jt}^{g(i)-1}$ the largest coefficient is $p'_{ijt}\bar{r}_{jt}^{g(i)-1}$ where p'_{ijt} is p_{ijt} with $\delta_{i,t-1}$ replaced by $\delta_{-p(i),t-1}$. (*iii*).Throughout table 5 we do not include p_{ijt} as a separate regressor. This is because it then becomes a seventh endogenous variable with a seventh instrument. We drop it to reduce the number of instruments. Including it makes no difference to our results.

Table 5: Endogeneity Revisited: Leave-One-Province-Out, SOEs, and Agglomeration

		Leave one province out when constructing:					
		Lagged export status $\delta_{i,t\text{-}1}$		Each component of 6 key variables		Drop SOEs and FIEs	Add city to gjt FEs
	Baseline	OLS	IV	OLS	IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exports (r) Backward	-0.60*	-0.73*	-0.91*	-0.47*	-0.52*	-0.82*	-0.57*
$p_{ijt}\bar{r}_{jt}^{g(i)-1}$ $(-)$	(0.08)	(0.14)	(0.22)	(0.08)	(0.09)	(0.10)	(0.09)
Exports (r) Forward	0.23*	0.23*	0.21	0.17*	0.20*	0.32*	0.18*
$p_{ijt}\bar{r}_{jt}^{g(i)+1}$ $(+)$	(0.06)	(0.09)	(0.10)	(0.06)	(0.07)	(0.08)	(0.07)
Competition (n) Backward	0.59*	0.60*	0.83*	0.41*	0.47*	0.77*	0.55*
$p_{ijt}n_{jt}^{g(i)-1}$ $(+)$	(0.08)	(0.15)	(0.24)	(0.07)	(0.08)	(0.10)	(0.08)
Competition (n) Forward	-0.28*	-0.25*	-0.20	-0.20*	-0.23*	-0.41*	-0.24*
$p_{ijt}n_{jt}^{g(i)+1}$ $(-)$	(0.06)	(0.10)	(0.13)	(0.06)	(0.06)	(0.07)	(0.06)
Competition (ω) Backward	0.13*	0.24*	0.24*	0.14*	0.14*	0.18*	0.08
$p_{ijt}\bar{\omega}_{jt}^{g(i)-1}$ $(+)$	(0.04)	(0.07)	(0.09)	(0.04)	(0.04)	(0.04)	(0.03)
Competition (ω) Forward	0.06	0.00	0.01	0.03	0.04	0.04	0.06
$p_{ijt}\bar{\omega}_{jt}^{g(i)+1}$ $(-)$	(0.03)	(0.07)	(0.09)	(0.03)	(0.04)	(0.04)	(0.03)
Lagged productivity	-0.83*	-0.85*	-0.78*	-0.85*	-0.84*	-0.12*	-0.08*
$\omega_{i,t-1}$	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.02)	(0.02)
Fixed effects #2	*gjt*	*gjt*	*gjt*	*gjt*	*gjt*	*gjt*	*gjtc*
Observations	621,729	621,729	621,729	621,729	621,729	438,504	604,095
R^2	0.729	0.729		0.729		0.729	0.775
Weak instrumnets F (KP)			459.30		842.20		

Notes: The dependent variable is the principal component of innovation. Column 1 is our baseline specification. (It is ever so slightly different from column 7 of table 2 because 123 observations are lost to missing information about province and because the p_{ijt} regressor is dropped.) In column 2 we recompute p_{ijt} leaving out data from firm i's province and use the resulting six leave-one-province-out regressors as exogenous variables. In column 3 we use these six new regressors as instruments for our original six key variables. In column 4 we recompute not only p_{ijt}, but also $\bar{r}_{jt}^{g(i)-1}$, $\bar{r}_{jt}^{g(i)+1}$, $n_{jt}^{g(i)-1}$, $n_{jt}^{g(i)+1}$, $\bar{\omega}_{jt}^{g(i)-1}$ and $\bar{\omega}_{jt}^{g(i)+1}$ leaving out data from firm i's province and use the resulting six leave-one-province-out regressors as exogenous variables. In column 5 we use these six new regressors as instruments for our original six key variables. In column 6 we do exactly as in our baseline specification (column 7 of table 2), but omit SOEs and FIEs. In column 7 we again do exactly as in our baseline specification, but use grade-industry-year-*city* fixed effects. See footnote 31 for details.. All specifications include firm and gjt fixed effects. Standard errors clustered two-way by firm and gj are in parentheses. A * indicates significance at the 1% level.

begin by outlining four criteria that must be satisfied for endogeneity to occur.

Criteria: First, $\bar{r}^{g}_{jt}$ is gjt-level exports, not firm-level exports. Hence, for $\bar{r}^{g}_{jt}$ to be correlated with a grade-g firm's residual means the firm is either large or part of a large set of firms with correlated residuals. We refer to this as the 'granularity' criterion. Second, the set cannot be too large because then the correlated component of the set's residuals would be absorbed by the gjt fixed effect. So the set must be sub-gjt. Third, we care about the endogeneity of $\bar{r}^{g-1}_{jt}$ and $\bar{r}^{g+1}_{jt}$, not $\bar{r}^{g}_{jt}$, so our grade-g firm must have a residual that is correlated with the residuals from granular sets of grade-$(g-1)$ and grade-$(g+1)$ firms. Fourth, our grade-g firm must have a residual that is *negatively* correlated with a granular set of grade-$(g-1)$ firms and *positively* correlated with a granular set of grade-$(g+1)$ firms; otherwise, the endogeneity cannot explain the distinctive sign pattern of our six key variables. In what follows we provide two examples of how such correlations could arise and show that they are empirically irrelevant.

Example 1: Assume that there are only three grades, let g be the middle grade, and let i be a state-owned enterprise (SOE) in grade g. Suppose the Chinese government subsidizes the exports and R&D expenditures of SOEs while taxing the exports and R&D expenditures of private firms. The subsidies raise i's R&D$_{it}$ and probability of exporting p_{ijt}. Now make the additional assumption that SOEs dominate in the high grade so that the export subsidy raises the grade's export sales $\bar{r}^{g+1}_{jt}$. Finally, assume that private firms dominate in the low grade so that the export tax lowers the grade's export sales $\bar{r}^{g-1}_{jt}$. That is, the policies raise R&D$_{it}$, raise $p_{ijt}\bar{r}^{g+1}_{jt}$, and possibly lower $p_{ijt}\bar{r}^{g-1}_{jt}$, thereby providing a very different explanation of our results. One can construct a similar argument to the above, but with subsidized SOEs replaced by foreign-invested enterprises (FIEs) that disproportionately engage in exporting and R&D.[29] If this example is driving our results, then the results will deteriorate when we drop SOEs and/or FIEs from our sample. To investigate, we re-estimate our baseline specification without SOEs and FIEs. The results appear in column 6 of table 5 and, rather than deteriorate, they are even stronger.[30] Thus, our results cannot be explained away by the types of correlations with unobservables described here.

Example 2: Suppose again that there are only three grades and that government export and R&D policies benefited regions that export high-grade goods and hurt those regions that export low-grade goods. Think of the former as coastal cities. This might explain why a firm in a middle grade coastal city has a residual that is positively correlated with $\bar{r}^{g+1}_{jt}$

[29]As in the Bilir and Morales (2020) discussion of R&D choices of foreign affiliates of US multinationals.

[30]Results for R&D, new-product sales, and patents appear in appendix table B.12. These results do not deteriorate when SOEs and/or FIEs are omitted.

and negatively correlated with $\bar{r}_{jt}^{g-1}$. Regional agglomeration effects might also explain the correlations. Since we showed in tables 2–3 that our results survive grade-industry-year-*province* fixed effects, the regions would have to be sub-provincial i.e., cities. To address endogeneity concerns, we adopt two approaches. First, we re-estimate our baseline model using grade-industry-year-*city* fixed effects.[31] The results appear in column 7 of table 5. Comparing columns 1 and 7 of table 5 , it is clear that the inclusion of more geographically disaggregated fixed effects has only a modest effect on our main findings. The same holds for R&D, new-product sales, and patents as shown in online appendix table B.13. Thus, we again find that all but our intensive-margin competition results are robust.

Example 3: Here we put aside the issue of how the correlation of firm i's residual with $\bar{r}_{jt}^{g-1}$ and $\bar{r}_{jt}^{g+1}$ could have the right sign pattern and simply focus on the possibility that the correlations are non-zero. At this abstract level, the previous two examples assumed that firm i shared unobservables with firms in grades $g(i)-1$ and $g(i)+1$. Supposing again that firms sort regionally based on unobservables, we can purge the unobservables by computing $\bar{r}_{jt}^{g(i)-1}$, $\bar{r}_{jt}^{g(i)+1}$, and p_{ijt} without using any data from firms in i's province. This is the leave-one-province-out approach we used before, but now extended to include not just lagged export status $\delta_{i,t-1}$ but all variables on the right-hand side.[32] Note that if firms in a region learn from one another, as in Fernandes and Tang (2014) where firms learn about exporting opportunities from others in the same region, this will show up in our data as a correlation of unobservables across firms in the same region. Restated, the endogeneity concern here can be couched in terms of peer effects. This is of interest because leave-one-out estimators are a core statistical technique in that literature (Angrist, 2014). Table 5 reports results for the principal component of innovation. Column 4 replaces our original regressors with the leave-one-province-out regressors. Column 5 instruments our original regressors with the leave-one-province-out regressors. In both cases, we expect coefficients to be smaller in absolute value because correlations with unobservables have been purged. Comparing the baseline column 1 with columns 4–5, we do indeed see slightly smaller coefficients, but changes are modest and in no way explain

[31] City codes changed during our sample. We build a crosswalk between the two codes. Specifically, neighbouring cities that grew into each other during 2000–6 are amalgamated into a single code, leaving us with 341 city codes. There are a 147,461 grade-industry-year-city fixed effects or a paltry 3 observations per fixed effect. To reduce the number of fixed effects, for each city we define a fixed effect that is the actual city if the city has at least 7,500 firm-year observations and that is the province if the city has less than 7,500 firm-year observation. This leaves us with 50 'cities' and just under 10 observations per fixed effect. The approach gives China's largest industrial cities their own city fixed effects, as our example 2 requires.

[32] These are $\bar{r}_{jt}^{g-1}$, $\bar{r}_{jt}^{g+1}$, n_{jt}^{g-1}, n_{jt}^{g+1}, $\bar{\omega}_{jt}^{g-1}$, and $\bar{\omega}_{jt}^{g+1}$ as well as $p_{jt}^{XX,g}$ and $p_{jt}^{NX,g}$ that go into the computation of p_{ijt} (see equations 1–2 and 13).

away our results.[33]

Summarizing, we have explored many potential sources of endogeneity and showed that our results are robust to each.

7. Robustness

In this section we show that our results are robust to a large number of alternative specifications.

Timing of a firm's innovation decision. Our model assumes that firms make innovation decisions at the start of period t based on expectations about period-t productivity ω_{it} and exporting δ_{it}. Alternatively, we could have assumed that innovation decisions are made at the end of period $t-1$, in which case each dependent variable y_{ijt} is replaced by its one-period lag $y_{ij,t-1}$. Online appendix tables B.14-B.15 repeat the specifications in tables 2 and 3 with this alternative timing assumption. The results are generally better than the y_{ijt}-based results reported above.

The choice of G. Our main results assume that there are $G = 8$ grades in total. Recall that G is an exogenously set parameter for the k-means clustering used to assign firms to grades. To assess the implications of the choice of $G = 8$, we re-estimate the grade of each firm under three alternative choice of $G \in \{6, 10, 12\}$. We then re-estimate our baseline specification for each of our measures of innovation (principal component of innovation, R&D, new-product sales, and patents). Online appendix table B.16 shows that our estimated coefficients are largely insensitive to the choice of G.

Heterogeneous effects. Our baseline specification constrains the coefficients on our six key variables to be the same across grades. The theory states that a coefficient can vary across grades (see appendix equation A.16), but cannot switch signs across grades (propositions 1–2). To investigate the stability of coefficient signs across grades we interact each of our six key variables with grade dummies so that each coefficient now varies by grade. The large number of coefficients are plotted in online appendix figure B.3, which shows that coefficient signs are stable across grades.

Grade transitions of more than one step. In our baseline specification we assumed one-step innovation so that failed innovation leaves firm i in grade $g(i)$ and successful

[33] Similar conclusions hold for R&D, new-product sales, and patents. See appendix table B.11.

innovation takes the firm to grade $g(i)+1$. However, Propositions 1–2 hold for multi-step innovation: Successful innovation takes the firm forward k grades with probability $p_F^{g(i),g(i)+k}$ while failed innovation takes the firm backward k grades with probability $p_B^{g(i),g(i)-k}$. Proposition 1 then states that the coefficients on variables such as $p_{ijt}\bar{r}_{jt}^{g(i)+k}$ should be positive for all $k \geq 1$, while Proposition 2 states that the coefficients on variables such as $p_{ijt}\bar{r}_{jt}^{g(i)-k}$ should be negative for all $k \geq 1$. Further, we expect that if a grade has a low probability of being reached then shocks in that grade should have little impact on innovation. As discussed mathematically at the end of appendix A.1, if the forward and backward probabilities $p_F^{g(i),g(i)+k}$ and $p_B^{g(i),g(i)-k}$ are declining sufficiently rapidly in k, then the effects of $p_{ijt}\bar{r}_{jt}^{g(i)+k}$ and $p_{ijt}\bar{r}_{jt}^{g(i)-k}$ on firm i's innovation must decline with k. Turning to an empirical examination of this, we note that neighbouring shocks such as $p_{ijt}\bar{r}_{jt}^{g(i)+1}$ and $p_{ijt}\bar{r}_{jt}^{g(i)+2}$ are too highly correlated to estimate their coefficients separately.[34] Instead we sum the $p_{ijt}\bar{r}_{jt}^{g(i)+k'}$ from $k'=1$ to k and use the sum as a regressor. We expect that the larger is k, the smaller in absolute value is the coefficient on the sum. This logic holds for all six key variables. Online appendix table B.17 implements this. It starts with our baseline specification and replaces each of the six key variables with their corresponding sums. It then shows the stark result that all six sums have the predicted signs and, more importantly, as k rises, the six coefficients all shrink monotonically in absolute value. This holds for the principal component of innovation, R&D, new-product sales, and patents. This is an interesting prediction about multi-step innovation that has never been examined before. As we shall now see, the logic behind it explains why our results hold in infinite-horizon models with forward-looking firms.

Forward-Looking Firms. In appendix A.2 we develop an infinite-horizon model with forward looking firms and discounting. To isolate the role of forward-looking behaviour, we make four simplifying assumptions. (*i*) Grade attributes such as the innovation success function $M^g(a)$ are identical across grades. (*ii*) $M^g(a) = \frac{1}{\gamma}a^\gamma$ for $\gamma \in (0,1)$. (*iii*) The economy is in steady-state. (*iv*) To focus as simply as possible on the role of grades $g-1$ and $g+1$ found in our empirics, we assume that successful (unsuccessful) innovation increases (decreases) a firm's grade by one step and there is no obsolescence ($\eta = 0$). Under these assumptions we show that innovation is strictly increasing in

$$\sum_{g'=1}^{G} w^{gg'}\bar{\pi}^{g'} \qquad (26)$$

[34]This is expected. Firms in the same grade are similar so firms in neighbouring grades are fairly similar.

where the $w^{gg'}$ are constants that depend only on the discount factor and γ. Therefore, our key insights (Propositions 1 and 2) hold if $w^{gg'} > 0$ for $g' > g$ and $w^{gg'} < 0$ for $g' < g$. While proving this for general values of the discount factor and γ is difficult, we can easily compute the coefficients $w^{gg'}$ numerically. Figure A.1 shows these coefficients for $\beta = 0.95$ and different values of $\gamma \in (0, 1)$. In all cases, we find that $w^{gg'} > 0$ for $g' > g$ and $w^{gg'} < 0$ for $g' < g$. That is, Propositions 1 and 2 hold in a model with forward-looking firms.

The intuition for why the first-order condition for optimal innovation depends on profits in all grades in this forward-looking model (i.e., equation 26) and in our model (i.e., equation 5) is simple. Firms in grade g care about profit opportunities in other grades if either (1) these grades can be reached in one period through grade jumps of more than one step (as in our model) or (2) these grades can be reached through grade jumps of one step over multiple periods (as in the forward-looking model).

The intuition for why Propositions 1 and 2 hold in the forward-looking model is also simple. First, readers familiar with the innovation literature will immediately recognize that the first-order condition for optimal innovation with forward-looking firms states that innovation is increasing in $V^{g+1} - V^{g-1}$, where V^g is the (discounted) continuation value of being in grade g. Then, for any grade $g' \geq g + 1$, a firm is more likely to reach g' faster from $g+1$ than from $g-1$. As a result, V^{g+1} is more sensitive than V^{g-1} to profits in g'. It follows that innovation is increasing in the profits of any grade $g' \geq g+1$. This is our Proposition 1. Applying a symmetric logic to any $g' \leq g - 1$, it follows that innovation is decreasing in the profits of any grade $g' \leq g - 1$. This is our Proposition 2.

To conclude, we can re-interpret our multi-step model of innovation as a one-step model with forward-looking firm behaviour. See appendix A.2 for details.

Own-Grade Effects: Online appendix table B.18 adds the own-grade variables $p_{ijt}\bar{r}_{jt}^{g(i)}$, $p_{ijt}n_{jt}^{g(i)}$, and $p_{ijt}\bar{\omega}_{jt}^{g(i)}$ to our baseline specification. This is done for the principal component of innovation, R&D, new-product sales, and patents. We find that the coefficients on $p_{ijt}\bar{r}_{jt}^{g(i)}$ and $p_{ijt}n_{jt}^{g(i)}$ are tiny. For example, when the principal component of innovation is the dependent variable, the coefficients (standard errors) are 0.13 (0.17) and -0.10 (.12), respectively. That is, they are economically and statistically small and simply do not belong in the regression. Proposition 3 together with our empirical finding that the obsolescence rate η is close to 1/2 provide a clear explanation of this result. In contrast, the coefficient on $p_{ijt}\bar{\omega}_{jt}^{g(i)}$ is large and statistically significant (0.26 with a standard error of 0.07). Further, when it is included, the coefficient on forward intensive-margin competition becomes statistically insignificant. We have already observed that the forward and backward intensive-margin coefficients are fragile and this table provides further evidence of this. Interestingly, more intensive-margin competition in a firm's own grade induces innovation, meaning that a firm responds to high levels of productivity among its own-grade competitors by innovating to escape the competition as in Aghion et al. (2001, 2005). This is an interesting result, but not part of our focus on innovation cascades.

8. Conclusion

We have examined how innovation depends on exporting and, in particular, scale and competition in export markets. Our theory features (1) quality-segmented markets, (2) step-by-step innovation that moves firms forward and backward through laddered grades, and (3) escape-the-competition motives for innovation. Using data for Chinese firms in a period of explosive export growth (2000–2006), we have verified that innovation is promoted by (1) larger scale and less competition in forward export-market grades and (2) smaller scale and more competition in backward export-market grades. In addition, the component of competition that matters for innovation is entry (the extensive margin) rather than increased productivity of existing competitors (the intensive margin). Our findings highlight that a proper understanding of the impact of exporting on innovation must take into account how the former affects scale and competition at different points in the quality distribution.

Our theory provides a natural framework for examining the impact of China's rise on innovation in the rest of the world. Our research suggests that the effects of Chinese import competition on innovation in other countries will depend on where in the quality distribution Chinese exporters are entering. We will investigate this in future work by embedding our theory within a multi-country model of trade and making use of innovation and production data for firms outside of China.

Our research also has important policy implications for the most contentious source of international trade disputes, namely subsidies and especially subsidies to innovative industries. In a number of important WTO disputes, low- and high-quality goods have been treated as separate markets and the Appellate Body has ruled that WTO panels should ignore future damages to high-quality firms from current industrial subsidies to low-quality firms. See *Lead and Bismuth II* and discussions in Grossman and Mavroidis (2003) and Bown and Hillman (2019). Our work shows how these separate markets are closely linked by innovation choices and provides a framework for revisiting WTO precedent on this extremely important issue.

To illustrate, consider the market for commercial aircraft, which is segmented by characteristics such as range (regional versus long-haul). Subsidies provided by the Chinese government to COMAC (the main producer of regional jets in China) have little immediate impact on foreign firms that do not currently compete in the same quality segment as COMAC's flagship regional jet, the ARJ21. However, it is clear that industry leaders Boeing and Airbus view COMAC as a serious *future* competitive threat, so much so that they have ended their decades-old dispute at the WTO in order to deal with growing Chinese competition. In our language, Boeing and Airbus are more concerned with competition from behind than competition from within their own grade. Furthermore, the threat of future damages has already become relevant at the US International Trade Commission, where Boeing successfully appealed for countervailing duties against Canada's Bombardier (a regional jet producer) based on the argument that Bombardier is only several innovative steps away from becoming a direct competitor. As the WTO currently renegotiates its Agreement on Subsidies and Countervailing Measures (ASCM), our research provides a timely and valuable framework for moving the negotiations forward.

References

Acemoglu, Daron and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics*, August 2004, *119* (3), 1049–1090.

_ **and Ufuk Akcigit**, "Intellectual Property Rights Policy, Competition, and Innovation," *Journal of the European Economic Association*, February 2012, *10* (1), 1–42.

_ **,** _ **, Harun Alp, Nicholas Bloom, and William Kerr**, "Innovation, Reallocation, and Growth," *American Economic Review*, November 2018, *108* (11), 3450–3491.

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer, "Identification properties of recent production function estimators," *Econometrica*, November 2015, *83* (6), 2411–2451.

Ackerberg, Daniel, Benkard Lanier, Steven Berry, and Ariel Pakes, "Econometric Tools for Analyzing Market Outcomes," in J. J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Vol. 6, Elsevier, 2007.

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, and Marc J. Melitz, "The Impact of Exports on Innovation: Theory and Evidence," Working papers 678, Banque de France 2018.

_ **,** _ **,** _ **, and Marc Melitz**, "The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports," *Review of Economics and Statistics*, 2022.

_ **, Christopher Harris, Peter Howitt, and John Vickers**, "Competition, Imitation and Growth with Step-by-Step Innovation," *Review of Economic Studies*, July 2001, *68* (3), 467–492.

_ **, Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt**, "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics*, May 2005, *120* (2), 701–728.

Akcigit, Ufuk and Marc Melitz, "International Trade and Innovation," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics: International Trade*, Vol. 5, New York: Elsevier, 2022, chapter 6, pp. 377–404.

_ **and William R Kerr**, "Growth through Heterogeneous Innovations," *Journal of Political Economy*, August 2018, *126* (4), 1374–1443.

__ , **Sina Ates, and Giammario Impullitti**, "Innovation and Trade Policy in a Globalized World," Working paper. 2021.

Angrist, Joshua D, "The Perils of Peer Effects," *Labor Economics*, October 2014, *30*, 98–108.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple, "Innovation and Production in the Global Economy," *American Economic Review*, August 2018, *108* (8), 2128–2173.

Atkeson, Andrew and Ariel Burstein, "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, June 2010, *118* (3), 433–484.

__ **and** __ , "Aggregate Implications of Innovation Policy," *Journal of Political Economy*, December 2019, *127* (6), 2625–2683.

Atkin, David, Arnaud Costinot, and Masao Fukui, "Globalization and the Ladder of Development: Pushed to the Top or Held at the Bottom?," Working Paper No. 29500, National Bureau of Economic Research, 2021.

Autor, David, David Dorn, Gordon H Hanson, Gary Pisano, and Pian Shu, "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents," *American Economic Review: Insights*, September 2020, *2* (3), 357–374.

Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu, "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, June 2011, *101* (4), 1312–1344.

Beerli, Andreas, Franziska J Weiss, Fabrizio Zilibotti, and Josef Zweimüller, "Demand Forces of Technical Change: Evidence from the Chinese Manufacturing Industry," Working paper. 2018.

Berry, Steven T., "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, Summer 1994, *25* (2), 242–262.

Bilir, L Kamran and Eduardo Morales, "Innovation in the Global Firm," *Journal of Political Economy*, April 2020, *128* (4), 1566–1625.

Bloom, Nicholas, Mirko Draca, and John van Reenen, "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity," *Review of Economic Studies*, January 2016, *83* (1), 87–117.

__ , **Paul Romer, Stephen J Terry, and John Van Reenen**, "Trapped Factors and China's Impact on Global Growth," *Economic Journal*, January 2021, *131*, 156–191.

Bombardini, Matilde, Bingjing Li, and Ruoying Wang, "Import Competition and Innovation: Evidence from China," Working paper. 2018.

Bown, Chad P and Jennifer A Hillman, "WTO'ing a Resolution to the China Subsidy Problem," *Journal of International Economic Law*, December 2019, *22* (4), 557–578.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing," *Journal of Development Economics*, March 2012, *97* (2), 339–351.

_ , _ , and _ , "Challenges of working with the Chinese NBS firm-level data," *China Economic Review*, 2014, *30* (C), 339–352.

_ , Johannes Van Biesebroeck, Luhang Wang, and Yifan Zhang, "WTO Accession and Performance of Chinese Manufacturing Firms," *American Economic Review*, September 2017, *107* (9), 2784–2820.

Bustos, Paula, "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American Economic Review*, February 2011, *101* (1), 304–340.

Bøler, Esther Ann, Andreas Moxnes, and Karen Helene Ulltveit-Moe, "R&D, International Sourcing, and the Joint Impact on Firm Performance," *American Economic Review*, December 2015, *105* (12), 3704–3739.

Chen, Yanyou and Daniel Xu, "A Structural Empirical Model of R&D Investment, Firm Heterogeneity, and Industry Evolution," Working paper. 2022.

Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu, "Notching R&D Investment with Corporate Income Tax Cuts in China," *American Economic Review*, July 2021, *111* (7), 2065–2100.

Chen, Zhiyuan, Jie Zhang, and Yuan Zi, "A Cost-benefit Analysis of R&D and Patents: Firm-level Evidence from China," *European Economic Review*, April 2021, *133*, 1–28.

Christensen, Clayton, *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*,, Cambridge, MA: Harvard Business Review Press, 1997.

Coelli, Federica, Andreas Moxnes, and Karen Helene Ulltveit-Moe, "Better, Faster, Stronger: Global Innovation and Trade Liberalization," *Review of Economic Statistics*, March 2022, *104* (2), 205–216.

Costantini, James and Marc Melitz, "The Dynamics of Firm-Level Adjustment to Trade Liberalization," in Elhanan Helpman, Dalia Marin, and Thierry Verdier, eds., *The Organization of Firms in a Global Economy*, Cambridge: Harvard University Press, 2007, chapter 4, pp. 107–141.

Cusolito, Ana Paula, Alvaro Garcia-Marin, and William F Maloney, "Proximity to the Frontier, Markups, and the Response of Innovation to Foreign Competition: Evidence from Matched Production-Innovation Surveys in Chile," Working paper. 2021.

De Loecker, Jan, "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 09 2011, *79* (5), 1407–1451.

_ , "Detecting Learning by Exporting," *American Economic Journal: Microeconomics*, August 2013, *5* (3), 1–21.

_ **and Frederic Warzynski**, "Markups and Firm-Level Export Status," *American Economic Review*, October 2012, *102* (6), 2437–71.

_ **, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik**, "Prices, Markups, and Trade Reform," *Econometrica*, March 2016, *84*, 445–510.

di Giovanni, Julian, Andrei A Levchenko, and Jing Zhang, "The Global Welfare Impact of China: Trade Integration and Technological Change," *American Economic Journal: Macroeconomics*, October 2014, *6* (3), 153–183.

Fernandes, Ana P and Heiwai Tang, "Learning to Export from Neighbors," *Journal of International Economics*, September 2014, *94*, 67–84.

Fieler, Ana Ceccilia and Ann E Harrison, "Escaping Import Competition in China," Working paper. 2022.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, March 2008, *98* (1), 394–425.

Griliches, Zvi and Haim Regev, "Firm Productivity in Israeli Industry: 1979-1988," *Journal of Econometrics*, January 1995, *65* (1), 175–203.

Grossman, Gene M and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, January 1991, *58* (1), 43–61.

— **and Petros C Mavroidis**, "US – Lead and Bismuth II: United States – Imposition of Countervailing Duties on Certain Hot-Rolled Lead and Bismuth Carbon Steel Products Originating in the United Kingdom: Here Today, Gone Tomorrow? Privatization and the Injury Caused by Non-Recurring Subsidies," *World Trade Review*, 2003, *2*, 170–200.

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse,Toolkit, and Cookbook," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Trade*, Amsterdam: Elsevier, 2014, pp. 131–195.

Hsieh, Chang-Tai and Zheng Michael Song, "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China," *Brookings Papers on Economic Activity*, 2015, pp. 295–346.

Impulliti, Giammario and Omar Licandro, "Trade, Firm Selection and Innovation: The Competition Channel," *Economic Journal*, February 2018, *128* (608), 189–229.

Jaumandreu, Jordi and Heng Yin, "Cost and Product Advantages: Evidence from Chinese Manufacturing Firms," Working paper. 2018.

Khandelwal, Amit, "The Long and Short (of) Quality Ladders," *Review of Economic Studies*, October 2010, *77* (4), 1450–1476.

Klette, Tor Jakob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, October 2004, *112* (5), 986–1018.

König, Michael, Kjetil Storesletten, Zheng Michael Song, and Fabrizio Zilibotti, "From Imitation to Innovation: Where Is all that Chinese R&D Going?," Working paper. 2021.

Kugler, Maurice and Eric Verhoogen, "Prices, Plant Size, and Product Quality," *Review of Economic Studies*, January 2012, *79* (1), 307–339.

Lentz, Rasmus and Dale T Mortensen, "An Empirical Model of Growth through Product Innovation," *Econometrica*, November 2008, *76* (6), 1317–1373.

Lileeva, Alla and Daniel J. Trefler, "Improved Access to Foreign Markets Raises Plant-Level Productivity... for Some Plants," *Quarterly Journal of Economics*, August 2010, *125* (3), 1051–1099.

Liu, Qing and Hong Ma, "Trade policy uncertainty and innovation: Firm level evidence from China's WTO accession," *Journal of International Economics*, November 2020, *127*, 1–20.

— **and Larry D Qiu**, "Intermediate Input Imports and Innovations: Evidence from Chinese Firms' Patent Filings," *Journal of International Economics*, November 2016, *103*, 166–183.

— **, Ruosi Lu, Yi Lu, and Tuan Anh Luong**, "Import Competition and Firm Innovation: Evidence from China," *Journal of Development Economics*, June 2021, *151*, 1–17.

Maican, Florin G, Matilda Orth, Mark J Roberts, and Van Anh Vuong, "The Dynamic Impact of Exporting on Firm R&D Investment," Working paper 2021.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, November 2003, *71* (6), 1695–1725.

Melitz, Marc J and Stephen J. Redding, "Trade and Innovation," Working Paper No. 28945, National Bureau of Economic Research, June 2021.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, November 1996, *64* (6), 1263–1297.

Orr, Scott, Daniel Trefler, and Miaojie Yu, "Estimating Productivity Using Chinese Data: Methods, Challenges and Results," in Lili Yan Ing and Miaojie Yu, eds., *World Trade Evolution: Growth, Productivity and Employment*, Routledge, 2019, pp. 229–260.

Perla, Jesse and Christopher Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, February 2014, *122* (1), 52–76.

Peters, Bettina, Mark J Roberts, and Van Anh Vuong, "Firm R&D Investment and Export Market Exposure," Working paper. 2020.

Schott, Peter, "Across-Product versus Within-Product Specialization in International Trade," *Quarterly Journal of Economics*, May 2004, *119* (2), 647–678.

Shu, Pian and Claudia Steinwender, "The Impact of Trade Liberalization on Firm Productivity and Innovation," in Josh Lerner and Scott Stern, eds., *Innovation Policy and the Economy*, Vol. 19, Cambridge:MA: National Bureau of Economic Research, 2019, chapter 2, pp. 39–68.

Yu, Miaojie, "Processing Trade, Tariff Reductions and Firm Productivity: Evidence from Chinese Firms," *Economic Journal*, June 2015, *125* (585), 943–988.

Zhang, Linyi, "Escaping Chinese Import Competition? Evidence from U.S. Firm Innovation," Working paper. 2018.

A. Theory Appendix

A.1. Derivation of Empirical Specification from the Model

For brevity, let us suppress the $g(i)$ notation and denote firm i's lagged grade simply by g. To economize on notation, we will also assume here that $\bar{p}_O^g = 0$ as in our baseline specification, although none of the following derivations require this. Our starting point is the innovation first-order condition (5). Using equations (10)-(11) to substitute for expected profits, we can rewrite the first-order condition as:

$$\log b_t^g = \log\left[m^g\left(a_{it}\right)\right] + \log \Psi_{it}^g \tag{A.1}$$

where:

$$\Psi_{it}^g \equiv \sum_{g'>g} \bar{p}_F^{gg'} \left(\psi_t^{D,g'} + p_{it}\psi_t^{X,g'}\right) H^{g'}\left(\Omega_{i,t-1}\right) - \sum_{g'<g} \bar{p}_B^{gg'} \left(\psi_t^{D,g'} + p_{it}\psi_t^{X,g'}\right) H^{g'}\left(\Omega_{i,t-1}\right) \tag{A.2}$$

with $\psi_t^{D,g} \equiv \bar{R}_t^{D,g}/[\sigma^g N_t^{D,g}(\bar{\Omega}_t^{D,g})^{\sigma^g-1}]$ and $\psi_t^{X,g} \equiv \bar{R}_t^{X,g}/[\sigma^g N_t^{X,g}(\bar{\Omega}_t^{X,g})^{\sigma^g-1}]$.

Now let us first log-linearize the term $\log\left[m^g\left(a_{it}\right)\right]$ around a value for a_{it} that only depends on g, denoted by a^g (e.g. the value of a_{it} for the average firm in grade g averaged across time):

$$\log\left[m^g\left(a_{it}\right)\right] \approx c^g - \xi^g \log a_{it} \tag{A.3}$$

where $\xi^g \equiv -\frac{m^{g\prime}(a^g)a^g}{m^g(a^g)} > 0$ and $c^g \equiv \log\left[m^g\left(a^g\right)\right] - \xi^g \log a^g$.

Next, we log-linearize Ψ_{it}^g around time-invariant values $\left\{\psi^{D,g}, \psi^{X,g}\right\}$ for the domestic and export-market potentials for each grade (e.g. the time averages of these values):

$$\log \Psi_{it}^g \approx \log \bar{\Psi}_{it}^g + \sum_{g'\neq g} \tilde{\beta}_{it}^{D,gg'}(\log \psi_t^{D,g'} - \log \psi^{D,g'}) + \sum_{g'\neq g} \tilde{\beta}_{it}^{X,gg'} p_{it}(\log \psi_t^{X,g'} - \log \psi^{X,g'}) \tag{A.4}$$

where:

$$\bar{\Psi}_{it}^g \equiv \sum_{g'>g} p_F^{gg'}(\psi^{D,g'} + p_{it}\psi^{X,g'})H^{g'}\left(\Omega_{i,t-1}\right) \quad - \sum_{g'<g} p_B^{gg'}(\psi^{D,g'} + p_{it}\psi^{X,g'})H^{g'}\left(\Omega_{i,t-1}\right) \tag{A.5}$$

$$\tilde{\beta}_{it}^{D,gg'} \equiv \begin{cases} \frac{\bar{p}_F^{gg'}\psi^{D,g'}H^{g'}(\Omega_{i,t-1})}{\bar{\Psi}_{it}^g}, & g' > g \\ -\frac{\bar{p}_B^{gg'}\psi^{D,g'}H^{g'}(\Omega_{i,t-1})}{\bar{\Psi}_{it}^g}, & g' < g \end{cases} \tag{A.6}$$

$$\tilde{\beta}_{it}^{X,gg'} \equiv \begin{cases} \frac{\bar{p}_F^{gg'}\psi^{X,g'}H^{g'}(\Omega_{i,t-1})}{\bar{\Psi}_{it}^g}, & g' > g \\ -\frac{\bar{p}_B^{gg'}\psi^{X,g'}H^{g'}(\Omega_{i,t-1})}{\bar{\Psi}_{it}^g}, & g' < g\,. \end{cases} \tag{A.7}$$

Note that the $\left\{\tilde{\beta}_{it}^{D,gg'}, \tilde{\beta}_{it}^{X,gg'}\right\}$ coefficients in equation (A.4) are firm-time specific only through $\{p_{it}, \Omega_{i,t-1}\}$. Hence, let us first write these terms as:

$$\tilde{\beta}_{it}^{D,gg'} = \bar{\beta}^{D,gg'} + \left(\tilde{\beta}_{it}^{D,gg'} - \bar{\beta}^{D,gg'}\right) \tag{A.8}$$

$$\tilde{\beta}_{it}^{X,gg'} = \bar{\beta}^{X,gg'} + \left(\tilde{\beta}_{it}^{X,gg'} - \bar{\beta}^{X,gg'}\right) \tag{A.9}$$

where $\left\{\bar{\beta}^{D,gg'}, \bar{\beta}^{X,gg'}\right\}$ are values of $\left\{\tilde{\beta}_{it}^{D,gg'}, \tilde{\beta}_{it}^{X,gg'}\right\}$ replacing $\{p_{it}, \Omega_{i,t-1}\}$ with values that depend only on grade, $\{p^g, \Omega^g\}$ (e.g. the respective values for the average firm in grade g averaged across time). Substituting (A.8) and (A.9) into equation (A.4) and dropping second-order terms (i.e. those involving products of deviations such as $\left(\tilde{\beta}_{it}^{D,gg'} - \bar{\beta}^{D,gg'}\right) \times \left(\log \psi_t^{D,g'} - \log \psi^{D,g'}\right)$), we then obtain:

$$\log \Psi_{it}^g \approx \log \bar{\Psi}_{it}^g + \sum_{g' \neq g} \bar{\beta}^{D,gg'} (\log \psi_t^{D,g'} - \log \psi^{D,g'}) + \sum_{g' \neq g} \bar{\beta}^{X,gg'} p_{it} (\log \psi_t^{X,g'} - \log \psi^{X,g'})\,.$$

Similarly, we can linearize $\log \bar{\Psi}_{it}^g$ around time-invariant values $\{p_i, \log \Omega_i\}$ for each firm (e.g. the time-averages of $\{p_{it}, \log \Omega_{it}\}$ within each firm), obtaining:

$$\log \bar{\Psi}_{it}^g \approx \log \bar{\Psi}_i^g + \tilde{\alpha}_{p,i}^g (p_{it} - p_i) + \tilde{\alpha}_{\Omega,i}^g (\log \Omega_{i,t-1} - \log \Omega_i) \tag{A.10}$$

where $\bar{\Psi}_i^g$ is the value of $\bar{\Psi}_{it}^g$ replacing $\{p_{it}, \Omega_{i,t-1}\}$ with $\{p_i, \Omega_i\}$ and $\left\{\tilde{\alpha}_{p,i}^g, \tilde{\alpha}_{\Omega,i}^g\right\}$ are given by:

$$\tilde{\alpha}_{p,i}^g \equiv \frac{1}{\Psi_i^g} \left[\sum_{g' > g} \bar{p}_F^{gg'} \psi^{X,g'} H^{g'}(\Omega_i) - \sum_{g' < g} \bar{p}_B^{gg'} \bar{\psi}^{X,g'} H^{g'}(\Omega)_i\right] \tag{A.11}$$

$$\tilde{\alpha}_{\Omega,i}^g \equiv \frac{\Omega_i}{\bar{\Psi}_i^g} \left[\sum_{g' > g} \bar{p}_F^{gg'} \left[\psi^{D,g'} + p_i \psi^{X,g}\right] h^{g'}(\Omega_i) - \sum_{g' < g} \bar{p}_B^{gg'} \left[\psi^{D,g'} + p_i \psi^{X,g}\right] h^{g'}(\Omega_i)\right] \tag{A.12}$$

with $h^{g'}$ denoting the first derivative of $H^{g'}$. As above, we can write the coefficients $\left\{\tilde{\alpha}_{p,i}^g, \tilde{\alpha}_{\Omega,i}^g\right\}$ as:

$$\tilde{\alpha}_{p,i}^g = \tilde{\alpha}_p^g + \left(\tilde{\alpha}_{p,i}^g - \tilde{\alpha}_p^g\right) \tag{A.13}$$

$$\tilde{\alpha}_{\Omega,i}^g = \tilde{\alpha}_\Omega^g + \left(\tilde{\alpha}_{\Omega,i}^g - \tilde{\alpha}_\Omega^g\right) \tag{A.14}$$

where $\left\{\tilde{\alpha}_p^g, \tilde{\alpha}_\Omega^g\right\}$ are the values of $\left\{\tilde{\alpha}_{p,i}^g, \tilde{\alpha}_{\Omega,i}^g\right\}$ replacing $\{p_i, \Omega_i\}$ with $\{p^g, \Omega^g\}$. Substituting into equation (A.10) and ignoring second-order terms gives:

$$\log \bar{\Psi}_{it}^{g} \approx \log \bar{\Psi}_{i}^{g} + \tilde{\alpha}_{p}^{g} \left(p_{it} - p_{i}\right) + \tilde{\alpha}_{\Omega}^{g} \left(\log \Omega_{i,t-1} - \log \Omega_{i}\right) . \tag{A.15}$$

Finally, combining equations (A.1), (A.3), (A.10), and (A.15), we obtain the empirical specification (23) after imposing one-step innovation and $\eta = \frac{1}{2}$. The coefficients on export-market variables are:

$$\begin{aligned} \beta_{r}^{g+} &\equiv \frac{\bar{\beta}^{X,g,g+1}}{\xi^{g}}, & \beta_{n}^{g+} &\equiv -\frac{\bar{\beta}^{X,g,g+1}}{\xi^{g}}, & \beta_{\omega}^{g+} &\equiv -\frac{\bar{\beta}^{X,g,g+1}}{\xi^{g}} \\ \beta_{r}^{g-} &\equiv \frac{\bar{\beta}^{X,g,g-1}}{\xi^{g}}, & \beta_{n}^{g-} &\equiv -\frac{\bar{\beta}^{X,g,g-1}}{\xi^{g}}, & \beta_{\omega}^{g-} &\equiv -\frac{\bar{\beta}^{X,g,g-1}}{\xi^{g}} \end{aligned} \tag{A.16}$$

while the coefficients on expected export status and lagged TFP are:

$$\gamma_{p}^{g} \equiv \frac{1}{\xi^{g}} \left[\tilde{\alpha}_{p}^{g} - \sum_{g'=1}^{G} \bar{\beta}^{X,gg'} \log\left(\sigma^{g'} \bar{\psi}^{X,g'}\right)\right] \qquad \text{and} \qquad \gamma_{\omega}^{g} \equiv \frac{\tilde{\alpha}_{\Omega}^{g}}{\xi^{g}} . \tag{A.17}$$

The firm fixed effect is given by:

$$\alpha_{i} \equiv \frac{1}{\xi^{g}} \left(\log \bar{\Psi}_{i}^{g} - \tilde{\alpha}_{p}^{g} p_{i} - \tilde{\alpha}_{\Omega}^{g} \log \Omega_{i}\right) \tag{A.18}$$

while the grade-time fixed effect is given by:

$$\alpha_{t}^{g} \equiv \frac{1}{\xi^{g}} \left[c^{g} + \left(\xi^{g} - 1\right) \log b_{t}^{g} + \sum_{g' \neq g} \bar{\beta}^{D,gg'} \left(\log \psi_{t}^{D,g'} - \log \bar{\psi}^{D,g'}\right)\right] . \tag{A.19}$$

Note in particular that α_{t}^{g} absorbs the cost of innovation b_{t}^{g} as well as domestic market innovation-relevant factors. Furthermore, the sign restrictions in (24) and (25) follow from the signs of the coefficients in equation (A.7).

To see how the magnitudes of the export-market coefficients $\left\{\beta_{r}^{gg'}, \beta_{n}^{gg'}, \beta_{\omega}^{gg'}\right\}$ vary with g', first note from equation (A.7) that this variation depends only on three terms: market potentials $\xi^{X,g'}$, TFP conditional expectation functions $H^{g'}$, and transition probabilities $\left\{p_{F}^{gg'}, p_{B}^{gg'}\right\}$. While our theory imposes no restrictions on the first two variables, it is natural to assume that the transition probability between two grades g and g' is declining in the distance $|g' - g|$. It is then clear from equation (A.7) that imposing this assumption would tend to make the export-market coefficients declining in $|g' - g|$.

A.2. Model with Forward-looking Firm Behaviour

We illustrate here the implications of forward-looking firm behaviour in our model and show that our key insights are robust to allowing for this. To isolate the role of forward-looking behaviour, we make several simplifying assumptions. First, all firm-specific shocks (TFP and export status) are *iid*. Second, all grades are identical except for the fact that reaching higher grades requires successful innovation. Third, the economy is in steady-state. Fourth, the innovation success function is of the form $M(a) = \frac{1}{\gamma}a^{\gamma}$ with a constant elasticity $\gamma \in (0, 1)$. Finally, $p_F^{g,g+1} = 1$, $p_B^{g,g-1} = 1$, and $\eta = 0$, so that successful innovation increases a firm's grade by one step, failed innovation decreases a firm's grade by one step, and there is no obsolescence.

Now let V^g denote the expected value of a firm that is able to produce in grade g. This value satisfies the following Bellman equation:

$$V^g = \bar{\pi}^g + \beta \max_{a^g} \left\{ -ba^g + M(a^g) V^{g+1} + [1 - M(a^g)] V^{g-1} \right\} \tag{A.20}$$

where β is the temporal discount factor. The first-order condition for the firm's innovation investment at an interior solution is:

$$b = m(a^g) \left(V^{g+1} - V^{g-1} \right) . \tag{A.21}$$

Comparing this with the first-order condition (5) from our model, it is clear that forward-looking behaviour simply implies that optimal innovation depends on the difference in firm values $V^{g+1} - V^{g-1}$ rather than the difference in static profits $\bar{\pi}^{g+1} - \bar{\pi}^{g-1}$. If $\beta = 0$, then the two are equivalent.

Given the assumed properties of the innovation success function M, we can generally solve equation (A.21) for a^g to obtain:

$$a^g = F\left(\frac{V^{g+1} - V^{g-1}}{b} \right) \tag{A.22}$$

where $F(x) \equiv m^{-1}\left(\frac{1}{x}\right)$ is a strictly increasing function. Using equation (A.22) to substitute for a^g in the Bellman equation (A.20) then makes it clear that in general, the value in grade g, V^g, depends on expected profits in *all* grades, since firms in grade g internalize the fact there is a non-zero probability of reaching any other grade at some point in the future. Consequently, optimal innovation for firms in grade g also depends on profits in all grades. Note from the first-order condition (5) that this result is also allowed for by our model without forward-looking behaviour through the forward and backward transition probabilities, $p_F^{gg'}$ and $p_B^{gg'}$. Intuitively, firms in g care about profit opportunities in other grades if these grades can be reached in one period through grade jumps of more than one step (as in our model) or through grade jumps of one step over multiple periods (as with forward-looking behaviour).

However, to establish that the insights in Propositions 1 and 2 are robust to allowing for forward-looking firm behaviour, we must show that profits in forward grades affect innovation positively, while profits in backward grades affect innovation negatively. To make progress here, let us linearize the function F around some constant value c. We can then turn the Bellman equation (A.20) into a linear system in firm values and profits:

$$V^g = \bar{\pi}^g - f + \beta\nu V^{g+1} + \beta\left(1-\nu\right)V^{g-1} \tag{A.23}$$

where $\nu \equiv \frac{1}{\gamma}c^{\frac{\gamma}{1-\gamma}}$ and $f \equiv \beta b c^{\frac{1}{1-\gamma}}$. Note in particular that since we are assuming interior solutions, then $\frac{1}{\gamma}\left(a^g\right)^\gamma < 1$, which from (A.22) implies $c < \gamma^{\frac{1-\gamma}{\gamma}}$ and hence $\nu \in (0,1)$.

To solve for firm values as a function of expected profits, we must specify the boundary conditions for this system. Hence, let us normalize the value of being in grade 0 (e.g.., exit) to zero and suppose that successful innovation in grade G results in staying in grade G rather than advancing a grade. Hence, equation (A.23) for $g = 1$ becomes:

$$V^1 = \bar{\pi}^1 - f + \beta\nu V^2 \tag{A.24}$$

and for $g = G$ becomes:

$$V^G = \bar{\pi}^G - f + \beta\nu V^G + \beta\left(1-\nu\right)V^{G-1}. \tag{A.25}$$

Equations (A.23), (A.24), and (A.25) now define a system of G linear equations in the firm values $\{V^g\}_{g=1}^G$ given expected profits $\{\bar{\pi}^g\}_{g=1}^G$ in each grade. Hence, we can generally write the solution for the value in grade g as:

$$V^g = \sum_{g'=1}^{G} \tilde{w}^{gg'}\bar{\pi}^{g'} + \tilde{f}^g \tag{A.26}$$

where $\tilde{w}^{gg'}$ and $\tilde{f}^g$ are constants that are independent of profits. Substituting this result into the first-order condition (A.21) then gives:

$$b = m\left(a^g\right)\left(\sum_{g'=1}^{G} w^{gg'}\bar{\pi}^{g'} + \tilde{f}^{g+1} - \tilde{f}^{g-1}\right) \tag{A.27}$$

where $w^{gg'} \equiv \tilde{w}^{g+1,g'} - \tilde{w}^{g-1,g'}$. Therefore, the key insights of Propositions 1 and 2 hold if $w^{gg'} > 0$ for $g' > g$ and $w^{gg'} < 0$ for $g' \leq g$.

While proving this for general values of β and ν is difficult, we can easily compute the coefficients $w^{gg'}$ numerically. Figure A.1 shows these coefficients for $\beta = 0.95$ and three different values of $\nu \in \{0.25, 0.5, 0.75\}$, where we omit the endpoint grades $g = 1$ and $g = 8$. Note that one can think of different values of ν as arising from holding c constant and choosing three different values for the success function elasticity γ. In the figure, g' appears on the horizontal axis and, for each g' there are six bars corresponding to $g = 2, \ldots, 7$. In all cases, we find that $w^{gg'} > 0$ for $g' > g$ and $w^{gg'} < 0$ for $g' < g$.

The intuition for this result is the following. Optimal innovation in grade g is increasing in the difference between the values of grades $g + 1$ and $g - 1$. All grades $g' > g$ are "closer" to $g + 1$ than $g - 1$ and hence profits in these grades matter more for V^{g+1} than V^{g-1}, so that the effect of profits in these grades on the difference $V^{g+1} - V^{g-1}$ is positive. Similarly, all grades $g' < g$ are "closer" to $g - 1$ than $g + 1$ and hence profits in these grades matter more for V^{g-1} than V^{g+1}, so that the effect of profits in these grades on the difference $V^{g+1} - V^{g-1}$ is negative. Note that since grade g is neither closer to $g + 1$ nor closer to $g - 1$, the effect of $\bar{\pi}^g$ on innovation in grade g is ambiguous, as can be seen in figure A.1. In sum, the core predictions of our model – that profits in forward grades affect innovation positively while profits in backward grades affect innovation negatively – are robust to allowing for forward-looking firm behaviour (as long as one excludes g as a backward grade, as we do in our baseline specification).

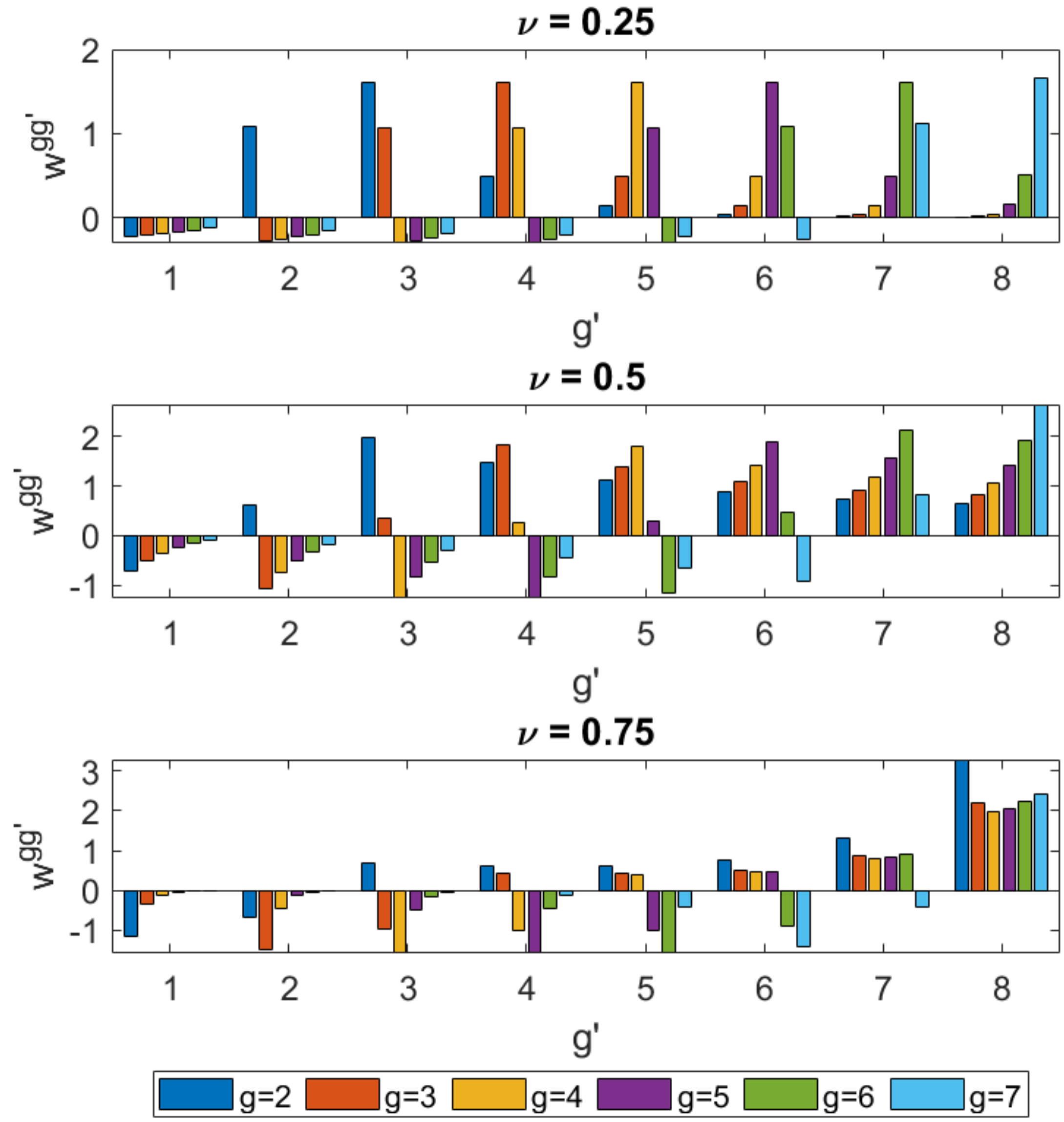
B. Data Appendix

In the raw Chinese Manufacturing Enterprises (CME) database we have virtually identical numbers of firms per year as in Brandt et al. (2014). See online appendix E.1.1. We clean the raw data using the data-cleaning algorithms and supplementary data described in Brandt et al. (2012) and Brandt et al. (2017), graciously made available by the authors. These include an algorithm to link firms across time, a crosswalk between old and new industry classifications, and data on capital stocks and input deflators.

Sample selection. After dropping the 2-digit industry "Waste resources and recycling", we have 1,434,565 firm-years and 434,344 firms. Dropping processing firms leaves us with 1,239,996 firm-years and 385,455 firms.[35] We next drop firms based on four criteria

[35] We define a processing firm as a firm with an exports-to-output ratio in excess of 1 in any year. This identifies 52,000 firms, comparable to what has been found in the (accurate) customs data. We have experi-

Figure A.1: Coefficients $w^{gg'}$ in the innovation first-order condition (A.27)



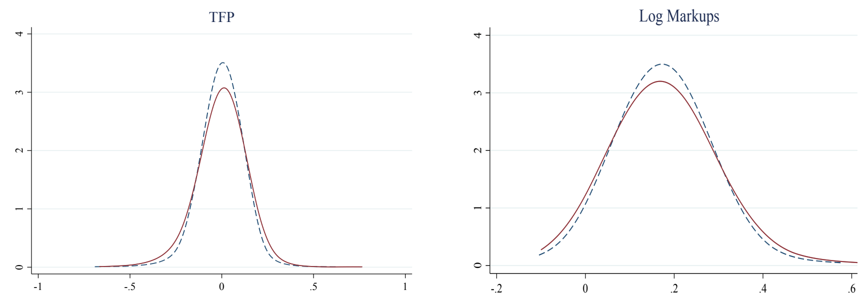
that are relevant for productivity analysis. First, the firms must have complete data on sales, employment, material costs, and capital. Second, we delete firms that report less than 8 employees in every year. Third, we drop the 2-digit industry "Tobacco". Fourth, if a firm has missing data in one year but available data in the surrounding years, we drop all data before this 'hole'. This is needed to take lags when estimating the law of motion for productivity and affects much less than 0.5% of our sample. This leaves us with 1,076,018 firm-years and 342,061 firms in 28 industries. We use this sample to compute grade aggregates such as $\bar{r}^{g}_{ijt}$. As in Brandt et al. (2017), we estimate productivity at the 2-digit industry level and drop firms that switch 2-digit industries. This leaves us with 1,047,177 firm-years and 264,168 firms in 28 industries. When running regressions, we delete firms with only one year of data because these observations are collinear with their firm fixed effects. This leaves us with 733,957 firm-years and 241,880 firms for the regressions. Finally, when running regressions we lose a firm's first observation to lags (where possible, we use 1999 data to prevent this loss) and we lose observations when the principal components measure of innovation is missing, leaving us with regressions estimated on 621,879 firm-year observations and 170,301 firms.

TFP and markup estimation details. We estimate productivity using both value-added and gross-output production functions, for both Cobb-Douglas and translog functional forms. The Cobb-Douglas elasticities (coefficients on capital, labour and materials) are very similar to those reported in Brandt et al. (2014) and Brandt et al. (2017). See Orr et al. (2019). As discussed there, the translog gross-output production function estimates are most sensible as judged by input elasticities, returns to scale, and stability across specifications. We therefore use these. In Orr et al. we considered five different variants of the proxy-variable approach. Here we consider the simplest and most complex of these. The simplest is exactly as in Ackerberg et al. (2015). The most complex adds to this (1) the Olley and Pakes (1996) selection correction terms to correct for attrition bias, (2) a law of motion for firm-level productivity that depends on lagged export status so as to control for learning-by-exporting as in De Loecker and Warzynski (2012) and De Loecker (2013), and (3) lagged capital and its square as extra (over-identified) instruments. We calculate markups using the De Loecker and Warzynski (2012) method with material cost shares.

Figure A.2 reports distributions of revenue TFP (left panel) and log markups (right panel) for the 'simple' case (dashed line) and 'complex' case (solid line). As is apparent, the two cases produce very similar results and in the main text we only report results for the 'complex' case.

mented with different export-to-output ratio cutoffs and this makes no difference to our findings.

Figure A.2: Revenue TFP and Markups



Notes: This figure displays the distributions of revenue TFP (left panel) and log markups (right panel) for our estimated translog, gross-output production functions. Each panel displays two distributions. The dashed-line distribution is the 'simple' case and the solid-line distribution is the 'complex' case that are described in the text. Revenue TFP is demeaned at the 2-digit industry level. Log markups are not demeaned and have a median of 0.17, indicating that the median markup is 17%. Since revenue TFP and log markups vary across firm-year observations, the distributions in the figure are based on 1,047,177 firm-year observations.

Online appendix figure B.1 reports the distributions of returns to scale as well as the output elasticities for labour, capital, and materials. The labour and capital output elasticities tend to be close to zero, which is standard for these data, e.g., De Loecker et al. (2016) and Brandt et al. (2017). The returns to scale tend to be strongly concentrated around 1 (the mean and median are close to 1.03), which is reassuring.

Patent data. Patent data from CNIPA are matched to our CME database using firm names and addresses. This is the same criteria used by Liu and Qiu (2016) and our match rates are comparable to those reported in Liu and Ma (2020). CME firms that have no match in the patent data are assumed to have no patents.

Winsorization of innovation data. We winsorize patents, R&D, and new-product sales to deal with extreme values. For patents, we top code at 52, which is the 99th percentile of the distribution of non-zero patents. Only 391 firm-year observations are top coded. For R&D, we top code firm-years with R&D-sales ratios greater than 0.20. Only 413 firm-year observations are top coded. For new-product sales, we top code firm-years with (new product sales)/(total sales) greater than one. Only 7,369 firm-year observations are top coded. Our results are not sensitive to omitting these firm-year observations or doubling the top coding levels.

Computing the principal component of innovation. To compute this we use three factors: the log of one plus the number of patents, the log of one plus R&D expenditures, and the log of one plus new-product sales. Since R&D data are missing in 2000 and 2004, for the purpose of computing the principal component (and only for this purpose) we linearly impute 2000 and 2004 R&D data. The principal component is computed separately by 2-digit industry. The factor loadings appear in online appendix table B.2.

Definition of SOEs and FIEs: We define an SOE as a firm which (*i*) has a state capital share in excess of 50%, as in Hsieh and Song (2015) or (*ii*) has one of the following ownership codes: 110 (domestic SOE), 141 (state-owned joint venture), or 143 (state-owned and collective joint venture enterprises), as in Yu (2015). We define foreign-invested firms (FIEs) as firms with ownership codes 310, 320, 330, 340, 210, 220, 230, or 240, as in Yu (2015).

C. Convergence of Grade Assignment Algorithm

The algorithm is initialized with $\theta_{i,0} = (\ln q_i + \sigma \ln p_i)/(\sigma - 1)$ (see equation 18). We choose $\sigma = 5.03$, which is the median estimate of the elasticity of substitution across a large sample of international trade papers (Head and Mayer, 2014). The algorithm stops when less than 0.1% of firm-year observations change grades between iterations n and $n+1$.[36] At the point of stopping, the Neyman and Spearman correlations between $\theta_{i,n+1}$ and $\theta_{i,n}$ are both above 0.999.

D. Assessing Two Key Premises of our Theory

The model is based on two key assumptions. First, innovation raises the probability of moving up the grade ladder. If this is not true then we are misinterpreting all of our results. Second, export-market shocks differ across grades. If this were not true, we would be unable to separately identify forward and backward effects. There would be only a single effect. We provide additional evidence about these points here.

Innovation raises the probability of moving up the grade ladder: To investigate how innovation affects quality, we regress changes in a firm's grade $g(i,t) - g(i,t-1)$ and quality $\theta^{g(i,t)} - \theta^{g(i,t-1)}$ on the firm's innovation in period t. Table A.1 reports our results, where columns 1–4 (5–8) report changes in a firm's grade (quality) as the dependent variable and each column examines one of four measures of innovation: the principal component

[36] For some industries, a very small number of observations cycle between adjacent grades. This is most pronounced for Rubber Products (CIC 29), but even in this industry there are only 22 firm-years that cycle.

Table A.1: Innovation Leads to Changes in Grades and Quality

	Grade Change = $g(i,t)$ - $g(i,t\text{-}1)$				Quality Change = $\theta^{g(i,t)}$ - $\theta^{g(i,t\text{-}1)}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Principal Component, t	0.037*				0.031*			
	(0.002)				(0.002)			
R&D, t		0.026*				0.026*		
		(0.002)				(0.002)		
New-Product Sales, t			0.023*				0.023*	
			(0.001)				(0.002)	
Patents, t				0.079*				0.069*
				(0.008)				(0.010)
Observations	785,302	621,689	748,350	796,945	551,009	497,868	534,674	551,009
R^2	0.458	0.569	0.510	0.458	0.512	0.561	0.543	0.511

Notes: Each column is a specification involving a regression of either grade change or quality change on a measure of innovation. The independent variable R&D enters as $\ln(1 + RD_{it})$. Likewise for new-product sales and patents. All specifications include firm and gjt fixed effects. Standard errors clustered two-way by firm and gj are in parentheses. A * indicates significance at the 1% level.

innovation, R&D, new-product sales, and patents. All specifications include firm and grade-industry-year fixed effects. "Grade" refers here to lagged grade. In all cases, we find that innovation is strongly related to changes in grade and quality. Online appendix table B.9 shows that we obtain very similar results when regressing grade and quality changes between $t-1$ and t on a firm's *lagged* innovation in $t-1$ and when regressing the level of a firm's grade or quality on measures of its innovation (either contemporaneous or lagged).

Export-market shocks vary across grades: Table A.2 provides evidence that the trade shocks impacting China during our sample period were in fact heterogeneous across grades. The table shows the average annual growth rates between 2000 and 2006 of our export-market size and competition variables in each grade: $\bar{R}_t^{X,g}$, $\bar{N}_t^{X,g}$ and $(\bar{\Omega}_t^{X,g})^{\sigma^g-1}$. Evidently, total exports and the number of firms competing for export-market profits expanded most rapidly in higher quality grades. For instance, total exports and the number of exporters grew at average rates of 38% and 25% per year respectively in the highest grade but both contracted slightly in the lowest grade. On the other hand, the average TFP of exporters as measured by $(\bar{\Omega}_t^{X,g})^{\sigma^g-1}$ grew more slowly in higher grades, suggesting that higher quality comes at the cost of lower productivity, as documented by Jaumandreu and Yin (2018).

Table A.2: Grade Characteristics: Annual Growth Rates, 2000–2006

Grade	Exports	Exporters (N^X)	TFP index ($\bar{\Omega}$)
(1)	(2)	(3)	(4)
1	-0.03	-0.05	0.13
2	0.07	0.09	0.15
3	0.12	0.14	0.14
4	0.14	0.15	0.14
5	0.17	0.17	0.12
6	0.18	0.17	0.12
7	0.20	0.20	0.09
8	0.38	0.25	0.07
All	0.23	0.15	0.12

Notes: For the 621,879 observations used in our regressions, we compute total exports in 2000 and 2006, then take the log change divided by 6. (For R&D we use 2001 in place of 2000.)