Econometrica Supplementary Material

SUPPLEMENT TO "FROM IMITATION TO INNOVATION: WHERE IS ALL THAT CHINESE R&D GOING?"

(*Econometrica*, Vol. 90, No. 4, July 2022, 1615–1654)

MICHAEL KÖNIG
Department of Spatial Economics, Vrije Universiteit Amsterdam and CEPR

KJETIL STORESLETTEN
Department of Economics, University of Minnesota and CEPR

ZHENG SONG
Department of Economics, Chinese University of Hong Kong

FABRIZIO ZILIBOTTI
Department of Economics, Yale University and NBER

APPENDIX A: THEORY

IN THIS SECTION, we provide the proof of Proposition 1 and an expression for the equilibrium profits (equation (7)) and wage rate.

PROOF OF PROPOSITION 1: From the monotonicity of $Q(a, \tau; \mathcal{A})$ in a, it follows that there exists a threshold function $a^*(\tau; \mathcal{A})$ such that[1]

$$\begin{aligned} Q(a, \tau; \mathcal{A}) &\geq \overline{p} \quad \text{if } a \leq a^*(\tau; \mathcal{A}), \\ Q(a, \tau; \mathcal{A}) &< \overline{p} \quad \text{if } a > a^*(\tau; \mathcal{A}). \end{aligned} \tag{A1}$$

All firms with $a \leq a^*(\tau; \mathcal{A})$ imitate, while some firms with $a > a^*(\tau; \mathcal{A})$ (i.e., those with a sufficiently large p) innovate. To simplify notation, we write $a^*(t) = a^*(\tau; \mathcal{A})$ and $Q(a) = Q(a, \tau; \mathcal{A})$ when this is no source of confusion.

Michael König: m.d.konig@vu.nl
Kjetil Storesletten: kjetil.storesletten@gmail.com
Zheng Song: zheng.michael.song@gmail.com
Fabrizio Zilibotti: fabrizio.zilibotti@yale.edu

[1]Note that this notation involves a slight abuse of notation relative to the function a^* in the text.

 https://doi.org/10.3982/ECTA18586

The difference equation governing the evolution of the log-TFP distribution can then be broken down as follows:

$$
\begin{aligned}
&\mathcal{A}_a(t+1) - \mathcal{A}_a(t) \\
&= \begin{cases}
q\big[\big(1 - F_{a-1}(t)\big)\mathcal{A}_{a-1}(t) - \big(1 - F_a(t)\big)\mathcal{A}_a(t)\big] \\
\quad \text{if } a < a^*(t); \\
\Big(q\big(1 - F_{a-1}(t)\big)\mathcal{A}_{a-1}(t) \\
\quad - G\big(Q(a)\big)\big[q\big(1 - F_a(t)\big)\mathcal{A}_a(t)\big] \\
\quad - \int_{Q(a)}^{\overline{p}} \big[\big(p + (1-p)\delta q\big(1 - F_a(t)\big)\big)\mathcal{A}_a(t)\big]\, dG(p)\Big) \\
\quad \text{if } a = a^*(t) + 1; \\
\Big(G\big(Q(a-1)\big) \times q\big(1 - F_{a-1}(t)\big)\mathcal{A}_{a-1}(t) \\
\quad + \int_{Q(a-1)}^{\overline{p}} \big[\big(p + (1-p)\delta q\big(1 - F_{a-1}(t)\big)\big)\mathcal{A}_{a-1}(t)\big]\, dG(p) \\
\quad - G\big(Q(a)\big) \times q\big(1 - F_a(t)\big)\mathcal{A}_a(t) \\
\quad - \int_{Q(a)}^{\overline{p}} \big[\big(p + (1-p)\delta q\big(1 - F_a(t)\big)\big)\mathcal{A}_a(t)\big]\, dG(p)\Big) \\
\quad \text{if } a > a^*(t) + 1.
\end{cases}
\end{aligned}
\tag{A2}
$$

To understand this law of motion, note that (i) if $a < a^*(t)$, all firms with TFP a and $a - 1$ imitate; (ii) if $a > a^*(t) + 1$, all firms with TFP a facing a realization $p > Q(a)$ and all firms with TFP $a - 1$ facing a realization $p > Q(a - 1)$ innovate, while all other firms with TFP a and $a - 1$ imitate; (iii) if $a = a^*(t) + 1$, all firms with TFP a facing a realization $p > Q(a)$ innovate, and all other firms with TFP a and $a - 1$ imitate. Going from the p.m.f to the corresponding c.d.f. yields

$$
\begin{aligned}
F_a(t+1) - F_a(t) &= \sum_{b=1}^{a} \mathcal{A}_b(t+1) - \mathcal{A}_b(t) \\
&= \begin{cases}
-q\big(1 - F_a(t)\big)\big(F_a(t) - F_{a-1}(t)\big) & \text{if } a \le a^*(t), \\
\Big(-G\big(Q(a)\big)q\big(1 - F_a(t)\big)\big(F_a(t) - F_{a-1}(t)\big) & \\
\quad - \int_{Q(a)}^{\overline{p}} \big[\big(p + (1-p)\delta q\big(1 - F_a(t)\big)\big) & \\
\quad \times \big(F_a(t) - F_{a-1}(t)\big)\big]\, dG(p)\Big) & \text{if } a > a^*(t).
\end{cases}
\end{aligned}
\tag{A3}
$$

Define the complementary cumulative distribution function $H_a(t) = 1 - F_a(t)$. Equation (A3) can be rewritten as

$$
H_a(t+1) - H_a(t) = -\sum_{b=1}^{a} \big(\mathcal{A}_b(t+1) - \mathcal{A}_b(t)\big)
$$

$$= \begin{cases} qH_a(t)\big(H_{a-1}(t) - H_a(t)\big) & \text{if } a \le a^*(t), \\ \Big(G\big(Q(a)\big)qH_a(t)\big(H_{a-1}(t) - H_a(t)\big) & \\ \quad + \int_{Q(a)}^{\overline{p}} \big[\big(p + (1-p)\delta q H_a(t)\big) & \\ \quad \times \big(H_{a-1}(t) - H_a(t)\big)\big]\, dG(p)\Big) & \text{if } a > a^*(t). \end{cases} \tag{A4}$$

Note that $F_a(t+1) \le F_a(t)$ (and conversely, $H_a(t+1) \ge H_a(t)$). Since the probability mass is conserved to 1 (and $\lim_{a\to+\infty} F_a = 1$), the fact that F_a is decreasing over time t for every a implies that the distribution must shift to the right (i.e., toward higher values of a). A distribution that is shifted in this way is called a traveling wave (Bramson (1983)). We now prove that there exists a traveling wave solution of the form $F_a(t) = \tilde{f}(a - \nu t)$ (or, equivalently, $H_a(t) = \tilde{h}(a - \nu t)$) with velocity $\nu > 0$. The formal argument follows Bramson (1983) and König, Lorenz, and Zilibotti (2016). The traveling wave solution above implies that $F_a(t+1) - F_a(t) = \tilde{f}(x-\nu) - \tilde{f}(x)$, where $x \equiv a - \nu t$. For $\nu \approx 0$, we can take the first-order approximation $\tilde{f}(x-\nu) - \tilde{f}(x) \approx -\nu \tilde{f}'(x)$, and thus $F_a(t+1) - F_a(t) \approx -\nu \tilde{f}'(x)$. Therefore, we can rewrite (A3) as

$$-\nu \tilde{f}'(x) = \begin{cases} -q\big(1 - \tilde{f}(x)\big)\big(\tilde{f}(x) - \tilde{f}(x-1)\big) & \text{if } x \le x^*, \\ \Big(-G\big(Q(x)\big)\big[q\big(1-\tilde{f}(x)\big)\big(\tilde{f}(x) - \tilde{f}(x-1)\big)\big] & \\ \quad - \int_{Q(x)}^{\overline{p}} \big[\big(p + (1-p)\delta q\big(1 - \tilde{f}(x)\big)\big) & \\ \quad \times \big(\tilde{f}(x) - \tilde{f}(x-1)\big)\big]\, dG(p)\Big) & \text{if } x > x^*, \end{cases} \tag{A5}$$

or, identically,

$$-\nu \tilde{h}'(x) = \begin{cases} q\tilde{h}(x)\big(\tilde{h}(x-1) - \tilde{h}(x)\big) & \text{if } x \le x^*, \\ \Big(G\big(Q(x)\big)\big[q\tilde{h}(x)\big(\tilde{h}(x-1) - \tilde{h}(x)\big)\big] & \\ \quad + \int_{Q(x)}^{\overline{p}} \big[\big(p + (1-p)\delta q \tilde{h}(x)\big) & \\ \quad \times \big(\tilde{h}(x-1) - \tilde{h}(x)\big)\big]\, dG(p)\Big) & \text{if } x > x^*. \end{cases} \tag{A6}$$

Consider, first, the range $x \le x^*$. Using the upper part of (A5) yields the following Delay Differential Equation (DDE):[2]

$$-\nu \tilde{f}'(x) = -q\big(1 - \tilde{f}(x)\big)\big(\tilde{f}(x) - \tilde{f}(x-1)\big). \tag{A7}$$

[2]See also Asl and Ulsoy (2003), Bellman and Cooke (1963), and Smith (2011).

This equation allows us to characterize the (asymptotic) left tail of the distribution. Taking the limit for $x \to -\infty$, we can take the following first-order (i.e., linear) approximation:

$$\nu \tilde{f}'(x) \simeq q\big(\tilde{f}(x) - \tilde{f}(x-1)\big).$$

Next, we guess that this linear DDE has a solution of the form $\tilde{f}(x) = c_1 e^{\lambda x}$ for $x \to -\infty$. Replacing $\tilde{f}(x)$ by its guess and $\tilde{f}'(x)$ by its derivative, and simplifying terms, allows us to verify the guess as long as the following transcendental equation in λ is satisfied:

$$\lambda \nu \simeq q\big(1 - e^{-\lambda}\big). \tag{A8}$$

The solution to this transcendental equation is given by

$$\lambda = \frac{\nu W\left(-\frac{q e^{-\frac{q}{\nu}}}{\nu}\right) + q}{\nu},$$

where W denotes the Lambert W-function, and we require that $\frac{q e^{-\frac{q}{\nu}}}{\nu} \leq \frac{1}{e}$.

Consider, next, the range of large x where the solution for $x > x^*$ applies in (A6). Then, we can write the following DDE:

$$\begin{aligned} -\nu \tilde{h}'(x) = \bigg(& G\big(Q(x)\big)\big[q\tilde{h}(x)\big(\tilde{h}(x-1) - \tilde{h}(x)\big)\big] \\ & + \int_{Q(x)}^{\overline{p}} \big[\big(p + (1-p)\delta q \tilde{h}(x)\big) \\ & \times \big(\tilde{h}(x-1) - \tilde{h}(x)\big)\big]\, dG(p)\bigg). \end{aligned} \tag{A9}$$

We use this DDE to characterize the right tail of the distribution as $x \to +\infty$. Again, we take a linear approximation:

$$\nu \tilde{h}'(x) \simeq \hat{p}\big(\tilde{h}(x) - \tilde{h}(x-1)\big),$$

where $\hat{p} = \int_0^{\overline{p}} p\, dG(p)$. For the latter, note that $\lim_{x\to\infty} Q(x) = 0$ since as we take x to be arbitrarily large, imitation becomes totally ineffective and firms choose to innovate almost surely. We guess a solution of the DDE of the form $\tilde{h}(x) = c_2 e^{-\rho x}$ for $x \to +\infty$. The guess is verified as long as the following transcendental equation holds:

$$\rho \nu \simeq \hat{p}\big(e^{\rho} - 1\big).$$

The solution to the transcendental equation satisfies

$$\rho = \frac{-\nu W\left(-\frac{\hat{p} e^{-\frac{\hat{p}}{\nu}}}{\nu}\right) - \hat{p}}{\nu}, \tag{A10}$$

where W denotes the Lambert W-function, and we require that $\frac{\hat{p} e^{-\frac{\hat{p}}{\nu}}}{\nu} \leq \frac{1}{e}$. This concludes the proof. *Q.E.D.*

Equilibrium expression for profits and wage rate: We provide the equilibrium expression for profits and for the wage rate, given an aggregate labor supply of $L = 1$, exogenous distributions of wedges and TFP, and the assumption of a small open economy.

The CES production function implies that each firm faces an isoelastic demand $Y_i = P_i^{-\eta} Y$, whence

$$P_i Y_i = Y^{\frac{1}{\eta}} \left(A_i K_i^{\alpha} L_i^{1-\alpha}\right)^{1-\frac{1}{\eta}}.$$

Each firm chooses capital and labor to maximize profits subject to wedges and demand equation:

$$\max_{\{K_i, L_i\}} \pi_i = (1 - \tau_i) P_i Y_i - w L_i - r K_i.$$

Solving the maximization problem by standard methods, substituting in the optimal values of L_i and K_i, using the expression for $P_i Y_i$ above, and rearranging terms yields

$$\pi_i = \frac{1}{\eta} P_i Y_i = \Pi\left(A_i(1 - \tau_i)\right)^{(\eta-1)},$$

where $\Pi \equiv ((1 - \alpha)^{(1-\alpha)} \alpha^{\alpha} (\eta - 1))^{\eta-1} \eta^{-\eta} \frac{Y}{(w^{1-\alpha} r^{\alpha})^{\eta-1}}$. This is equation (7) in the text.

The equilibrium expression for the wage rate is

$$w^{1-\alpha} = \left(1 - \frac{1}{\eta}\right) \frac{\alpha^{\alpha}(1 - \alpha)^{1-\alpha}}{r^{\alpha}} \left(\int_0^1 (1 - \tau_i)^{\eta-1} A_i^{\eta-1}\right)^{\frac{1}{\eta-1}}. \tag{A11}$$

APPENDIX B: DATA AND DESCRIPTIVE STATICS

In this section, we provide some details of the analysis in Section 3.

Alternative Methodology for Estimating TFP, Based on Brandt, Van Biesebroeck, Wang, and Zhang (2017)

We estimate firm-level TFP using the methodology of Hsieh and Klenow (2009). This is consistent with our theoretical model and allows us to directly compare our results with those in the literature on misallocation. However, this approach has been criticized in the empirical industrial organization literature. If firms optimally choose the inputs in the production process to solve a dynamic maximization problem, then the estimation may suffer from an endogeneity problem. The error term of the model can contain determinants of production decisions that are observed by the firm but not by the econometrician, leading to inconsistent estimates of TFP.

In this section, we show that the target moments of our estimation are essentially unchanged if we estimate TFP using the methodology of Ackerberg, Caves, and Frazer (2015) that addresses an endogeneity problem in the estimation of production functions. We follow the implementation of Ackerberg, Caves, and Frazer (2015) proposed by Brandt et al. (2017), which is also related to De Loecker and Warzynski (2012). Because Brandt et al. (2017) postulated a gross production function while we estimate TFP using a value added approach, we perform an adjustment for the two methods to be consistent. The details of the estimation are deferred to the web appendix. The results are shown in Figure A1. The data moments are indistinguishable from those used targets in our estimation. We conclude that our results are robust to using this alternative estimation method for TFP.

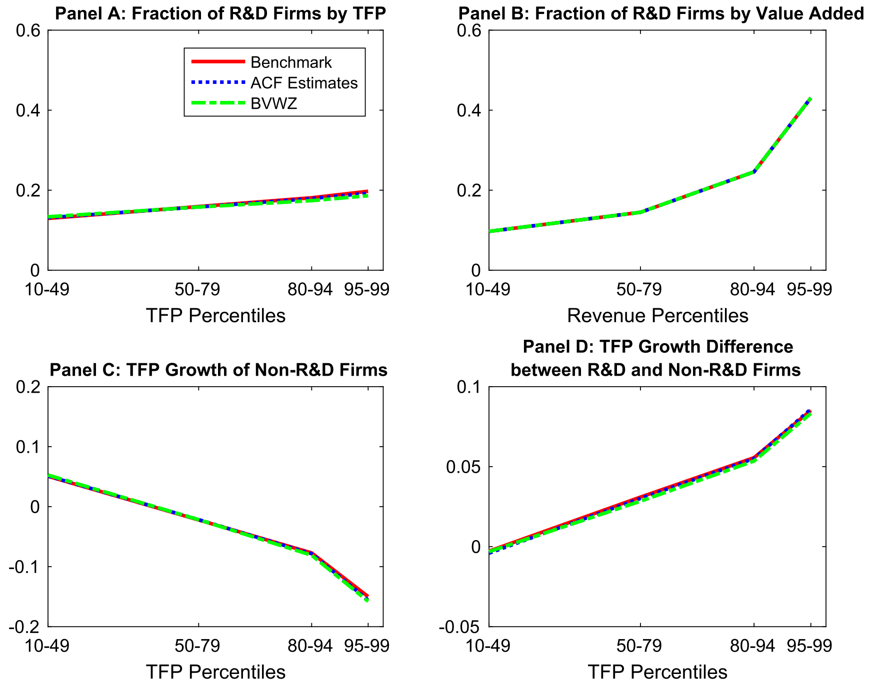

FIGURE A1.—China 2007–2012 sample with alternative TFP measurement. *Note*: The figure shows the equivalent moments to Appendix Figure A4 when TFP is estimated using the methodology of Brandt et al. (2017) based on Ackerberg, Caves, and Frazer (2015).

Regression With Firm Fixed Effects 2001–2007

In this section, we present the result of regressions similar to those in Table II, although for an earlier sample in China, 2001–2007. Since this sample has R&D data for more than one year, this sample allows us to also run regressions with firm fixed effects. Note that the regressions in Table AI are all on annual data, the reason being that we only have R&D data for 2001–2003 and 2005–2007. The regressions in columns (1)–(3) are pooled regressions, while columns (4)–(5) are firm fixed effects (FE) regressions. All pooled regressions include dummies for year, province, industry, and age effects while the FE regressions have year and firm effects. In the FE regressions, the dummies for province, industry, and age, as well as the dummies for export firms and state ownership, are all subsumed in the firm fixed effects.

The results in Table AI show that, first, all the results from our main sample (2007–2012) in Table II hold up for the earlier sample (see the pooled regressions in Table AI). Second, the qualitative results hold up (significantly so) even when controlling for firm fixed effects. When comparing columns (2) and (5), we observe that the coefficients in the FE regressions are about half the size in magnitude but still highly significant. Moreover, the coefficients have always the same sign as in the pooled regressions. We conclude that our main empirical findings on the drivers of firms doing R&D—namely, that R&D is positively associated with TFP and negatively associated with output wedges—hold true both in the cross-section and within firms over time.

TABLE AI

REGRESSIONS WITH FIRM FIXED EFFECTS, 2001–2007. DEPENDENT VARIABLE: R&D DUMMY.

	(1)	(2)	(3)	(4)	(5)
		Pooled regressions		FE regressions	
	R&D$_d$	R&D$_d$	R&D$_d$	R&D$_d$	R&D$_d$
log(TFP)$_t$	0.059	0.377	0.323	0.003	0.186
	(0.0049)	(0.0235)	(0.0191)	(0.0007)	(0.0054)
wedge		−0.432	−0.361		−0.225
		(0.0297)	(0.0240)		(0.0066)
export$_d$			0.045		
			(0.0118)		
SOE$_d$			0.124		
			(0.0148)		
Firm effects	−	−	−	+	+
Year effects	+	+	+	+	+
Industry effects	+	+	+	−	−
Age effects	+	+	+	−	−
Province effects	+	+	+	−	−
R-squared	0.396	0.449	0.456	0.581	0.582

Note: The table shows regressions of an indicator of R&D on annual data for China 2001–2007. The independent variable is R&D$_d$, a dummy variable for R&D that equals 1 if firm R&D expenditure is positive and zero otherwise. log(TFP) is the logarithm of TFP. Wedge refers to the calibrated firm output wedge (see Section 2 for details). export$_d$ is a dummy variable for exports. SOE$_d$ is a dummy variable for state-owned firms. Standard errors are reported in parentheses. The number of observations is 441,039 in columns (1)–(3) and 70,273 in columns (4)–(5). Observations are weighted by employment and standard errors are clustered by industry. Regressions in columns (4)–(5) include firm and year fixed effects. Regressions in columns (1)–(3) include year, industry, age, and province fixed effects. We drop firms with TFP in the bottom 10 percentiles.

Alternative Classification of Innovative Firms

In our main analysis, we classify all firms that report doing some R&D as innovative. However, many firms invest a very small amount of resources in R&D, raising questions of whether innovation is truly a salient strategy for these firms. In this appendix, we propose an alternative classification where firms are deemed innovative only if they invest more than 1.73% of their value added. This threshold is the median R&D intensity among R&D firms in our balanced sample. Conversely, firms investing less than 1.73% are regarded as imitators. Figure A2 shows the data moments corresponding to Figure 2 when applying this more stringent definition of innovators. As one would expect, the percentage of R&D firms is now lower. Moreover, the elasticity of R&D to TFP and size is lower than for the main sample. However, the qualitative patterns are the same in Figures 2 and A2.

Full Sample of Chinese Firms in 2007

Table AII provides a comparison between the descriptive statistics in 2007 for the balanced sample of Chinese firms we use in our analysis (*survivors*) and those that exited the sample before 2012. We ignore which of these firms literally ceased to exist, which ones shrank and fell below the survey threshold in later years, and which ones disappeared

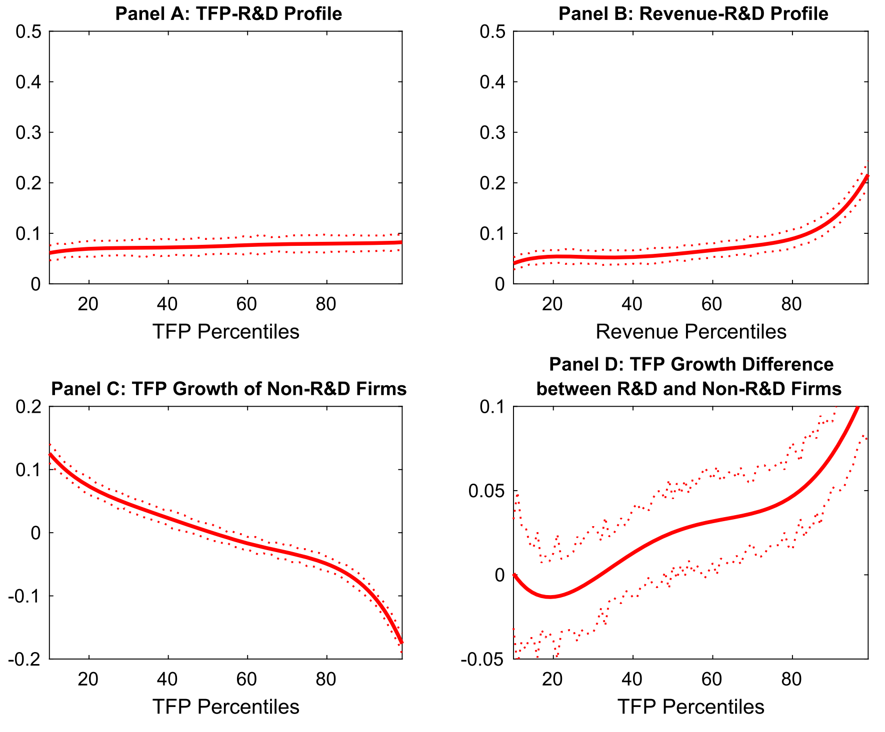

FIGURE A2.—More stringent classification of innovative firms. *Note*: The figure shows the equivalent moments to Figure 2 if only firms with R&D expenditure above 1.73% are classified as R&D firms.

because of mergers and acquisitions.[3] Surviving firms account for ca. 63% of the total manufacturing value added. The median surviving firm is more than twice as large as the median exiting firm. Exiting firms are less likely to perform R&D and, conditional on performing it, they invest less.

Figure A3 displays the analogue of Panels A and B in Figure 2 for the full sample of firms in 2007, including exiters. Both panels show that exiters have a lower propensity to engage in R&D than surviving firms. Note that in both panels A and B, the schedules for the full sample are almost parallel to the sample of survivors which we use. Namely, there is no major difference between survivors and exiters in the selection into R&D by TFP and size. This finding is confirmed by the multiple regressions in Table AIII whose results are very similar to those in Panel A of Table II.

Regression Results for Taiwan

In Table AIV, we report the regression results discussed in the text for Taiwan. By comparing Table AIV with Table II, it is clear that all qualitative results are the same for Taiwan as for our 2007–2012 China data. However, R&D in Taiwan is more highly correlated with TFP and more negatively correlated with wedges, and TFP growth is more

[3]Note that the threshold for being in the NBS data in 2007 is that sales exceed 5 million Yuan, or about 1 million USD. This threshold increased to 20 million Yuan in 2011 and afterwards and some of the exiters are therefore firms with sales below 20 million Yuan in 2012.

TABLE AII

SUMMARY STATISTICS FOR CHINA 2007, SURVIVORS VERSUS EXITERS.

	(1) Survivors	(2) Exiters	(3) Full sample
Number of firms	123,368	172,704	296,072
Median value added (million USD)	1.48	0.62	0.91
Share of R&D firms (in %)	14.7	8.5	11.1
Aggregate value added of R&D firms as share of total v.a. (in %)	42.7	34.4	39.6
Aggregate R&D expenditure as share of total v.a. (in %)	42.7	34.4	39.6
Median R&D Intensity for R&D firms (in %)	1.86	1.26	1.63

Note: Summary statistics for China 2007 for the full sample (including exiters) versus the balanced panel (survivors only). Survivors refers to firms present in both 2007 and 2012. Exiters refers to firms present in 2007 that exit the data before 2012.

strongly related to R&D. In particular, the coefficients on TFP and wedges in Panel A (explaining the R&D decision) of Table AIV are about twice as large in magnitude as in Table II. Morever, the coefficient on R&D in Panel B (explaining TFP growth) of Table AIV are about three times as large in magnitude as in Table II.

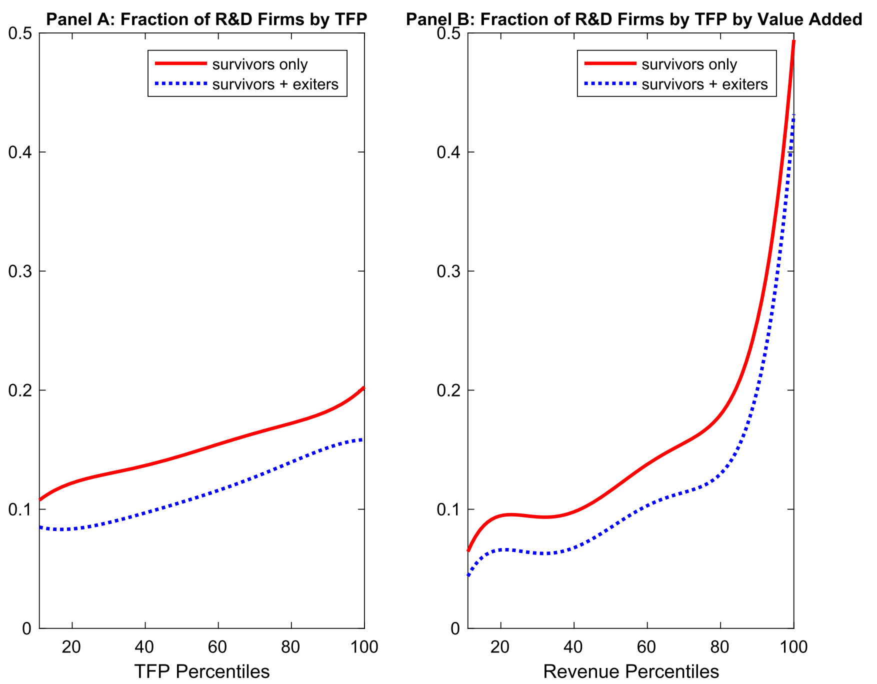

FIGURE A3.—China 2007–2012 non-balanced sample including exiters. *Note*: The dotted lines show the moments equivalent to those of Panels A and B in Figure 2 for the full sample of firms in 2007, that is, including firms that exit the sample before 2012. The solid lines reproduce the moments in Figure 2, calculated for the balanced sample of surviving firms.

TABLE AIII

REGRESSION ANALYSIS FOR ALL CHINESE FIRMS IN 2007. DEPENDENT VARIABLE: R&D DECISION IN 2007.

	(1) R&D$_d$	(2) R&D$_d$	(3) R&D$_d$	(4) R&D$_d$
log(TFP)	0.592	0.353	0.329	0.299
	(0.0063)	(0.0257)	(0.0230)	(0.0204)
wedge		−0.393	−0.364	−0.327
		(0.0319)	(0.0283)	(0.0254)
export$_d$			0.051	0.053
			(0.0134)	(0.0132)
SOE$_d$				0.182
				(0.0202)
R-squared	0.134	0.206	0.209	0.221

Note: The table reports regressions equivalent to those in Panel A of Table II for the full sample of firms in 2007, that is, including firms that exit the sample before 2012. All the explanatory variables are from 2007. Standard errors are reported in parentheses. The number of observations is 263,504. Observations are weighted by employment and standard errors are clustered by industry. All regressions include industry, age, and province fixed effects. We drop firms with TFP in the bottom 10 percentiles.

Target Moments of the Empirical Distribution

Figure A4 displays the empirical moments that we target in the estimation. Each observation represents a quantile of the distribution based on Figure 2.

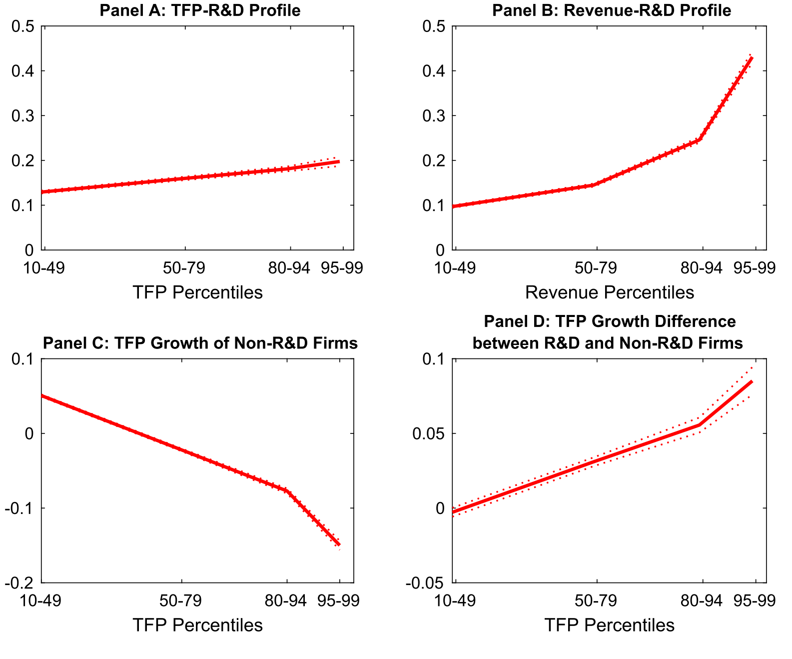

FIGURE A4.—Chinese firms in the balanced panel 2007–2012. *Note*: The figure shows the empirical moments of the balanced panel for China 2007–2012. See also Figure 4. The dotted lines represent standard errors.

TABLE AIV

BALANCED PANEL OF TAIWANESE FIRMS, 1999–2004.

PANEL A: Correlations between firm characteristics and R&D decision

	(1) $R\&D_d$	(2) $R\&D_d$	(3) $R\&D_d$
log(TFP)	0.087	0.614	0.573
	(0.0078)	(0.0184)	(0.0189)
wedge		−0.720	−0.673
		(0.0234)	(0.0236)
$export_d$			0.088
			(0.0227)
Industry effects	+	+	+
Age effects	+	+	+
Year effects	+	+	+
Observations	44,326	44,326	44,326
R-squared	0.219	0.396	0.404

PANEL B: Correlations between firm initial characteristics and TFP growth

	(1) TFP growth	(2) TFP growth	(3) TFP growth
log(TFP)	−0.064	−0.066	−0.066
	(0.0046)	(0.0047)	(0.0055)
$R\&D_d$	0.103	0.093	
	(0.0162)	(0.0168)	
$export_d$		0.039	0.039
		(0.0063)	(0.0063)
R&D $intensity_h$			0.072
			(0.0262)
R&D $intensity_m$			0.106
			(0.0278)
R&D $intensity_l$			0.097
			(0.0150)
Industry effects	+	+	+
Age effects	+	+	+
Observations	9,996	9,996	9,996
R-squared	0.081	0.083	0.084

Note: The table reports the analogues of the regression results in Table II for the Taiwanese firms. There are two differences relative to Table II: (1) Panel A reports pooled regressions with year fixed effects (for Taiwan, data for R&D expenditure are available for multiple years); and (2) there is no province dummy for Taiwan.

APPENDIX C: ESTIMATION

Measurement Error: Mapping

This appendix describes how we incorporate measurement error when calculating the theoretical moments. We provide an analytical mapping from the theoretical distribution of a variable x to the observed distribution of true x plus m.e. This analytical mapping is critical to speed up the structural estimation, avoiding a computational curse that would arise if we had to rely on simulations.

The presentation focuses on TFP. The approach for adding m.e. to the theoretical (true) distribution of value added is equivalent, replacing a and μ with y and μ_y below.

Denote by $\hat{a}$ and a the observed and true log TFP: $\hat{a} = a + \mu$, where μ is m.e. Consider the following discrete state space: $a \in \{\delta, \dots, N\delta\}$, $\hat{a} \in \{\delta, \dots, N\delta\}$, and $\mu \in \{-N^{\mu}\delta, \dots, -\delta, 0, \delta, \dots, N^{\mu}\delta\}$. We set $N^{\mu} = 4$.

Let the theoretical distribution of a be denoted by $\mathcal{A}(a)$. The first task is to convert $\mathcal{A}(a)$ to $\mathcal{A}(\hat{a})$, that is, the distribution of observed TFP with m.e., which can be compared with the data. To this end, we first derive the transition matrix $\mathcal{A}(\hat{a}|a)$. For $j \in \{2, \dots, N-1\}$, we have

$$\mathcal{A}(\hat{a} = a_j | a = a_i) = \mathcal{A}\big(\mu = (j - i)\delta\big). \tag{A12}$$

For $j = 1$ or N, we have $\mathcal{A}(\hat{a} = a_1 | a = a_i) = \sum_{k \geq i-1} \mathcal{A}(\mu = -k\delta)$ and $\mathcal{A}(\hat{a} = a_N | a = a_i) = \sum_{k \geq N-i} \mathcal{A}(\mu = k\delta)$. So, the unconditional probability of $\hat{a}$ is

$$\mathcal{A}(\hat{a} = a_j) = \sum_i \mathcal{A}(\hat{a} = a_j | a = a_i)\mathcal{A}(a = a_i). \tag{A13}$$

Note that when $\mathcal{A}(\hat{a})$ is observable while $\mathcal{A}(a)$ is unknown, one can use $\mathcal{A}(\hat{a} = a_j | a = a_i)$ in (A12) to back out $\mathcal{A}(a)$ by solving the system of equations in (A13).

We now derive the conditional TFP growth. Let us start with observed TFP growth of imitating firms:

$$\begin{aligned}\mathbb{E}^{\mathrm{IM}}[\Delta\hat{a}|\hat{a}] &= \mathbb{E}^{\mathrm{IM}}[\Delta a + \Delta\mu|\hat{a}] \\ &= \mathbb{E}\big[q\big(1 - F(a)\big)|\hat{a}\big] - \mathbb{E}[\mu|\hat{a}] \\ &= \sum_i q\big(1 - F(a)\big)\mathcal{A}(a = a_i|\hat{a} = a_j) - \sum_k k\delta\mathcal{A}(\mu = k\delta|\hat{a} = a_j). \end{aligned} \tag{A14}$$

To go from the theoretical (conditional) distribution of true TFP growth conditional on true a to TFP growth with m.e. conditional on $\hat{a}$, we need conditional probabilities of $\mathcal{A}(a = a_i | \hat{a} = a_j)$ and $\mathcal{A}(\mu = k\delta | \hat{a} = a_j)$.

The posterior distribution of a follows:

$$\mathcal{A}(a = a_i | \hat{a} = a_j) = \frac{\mathcal{A}(\hat{a} = a_j | a = a_i)\mathcal{A}(a = a_i)}{\mathcal{A}(\hat{a} = a_j)}. \tag{A15}$$

To obtain the posterior distribution of μ, first notice that

$$\begin{aligned}\mathcal{A}(\hat{a} = a_j \cap \mu = k\delta) &= \mathcal{A}(\hat{a} = a_j | \mu = k\delta)\mathcal{A}(\mu = k\delta) \\ &= \mathcal{A}(a = a_{j-k})\mathcal{A}(\mu = k\delta),\end{aligned}$$

TABLE AV

MEASUREMENT ERROR MOMENTS.

	Empirical variances				Implied m.e. variance		
	var(ΔY)	var(ΔI)	cov($\Delta Y, \Delta I$)	var(TFP)	$\hat{v}_{\mu y}$	$\hat{v}_{\mu I}$	$\hat{v}_{\mu a}$
2007–2012 China	0.456	0.328	0.156	1.059	0.150	0.086	0.320
2001–2007 China	0.470	0.098	0.043	1.269	0.214	0.027	0.361
1999–2004 Taiwan	1.128	0.124	−0.005	2.363	0.567	0.065	0.950

Note: The first four columns refer to variances of growth in revenue, inputs, and TFP, respectively. Y, I, and TFP represent $\log(P_{it}Y_{it})$, $\log(K_{it}^{\alpha}L_{it}^{(1-\alpha)})$, and $\log(A_{it})$, respectively. The fourth column is cross-sectional dispersion in TFP. We use the full sample (i.e., keeping the firms with initial TFP in the bottom ten percentiles). The results in the trimmed sample are similar. The implied variance of measurement error is derived from equations (12)–(13) and the expression for m.e. in TFP.

for $j \in \{2, \dots, N-1\}$. Note that for $j = 1$ or N, we have the following boundary cases:

$$\mathcal{A}(\hat{a} = a_1 | \mu = -i\delta)\mathcal{A}(\mu = -i\delta) = \sum_{k \le i+1} \mathcal{A}(a = a_k)\mathcal{A}(\mu = -i\delta),$$

$$\mathcal{A}(\hat{a} = a_N | \mu = i\delta)\mathcal{A}(\mu = i\delta) = \sum_{k \ge N-i} \mathcal{A}(a = a_k)\mathcal{A}(\mu = i\delta).$$

Then, the posterior distribution of μ follows:

$$\begin{aligned} \mathcal{A}(\mu = k\delta | \hat{a} = a_j) &= \frac{\mathcal{A}(\hat{a} = a_j | \mu = k\delta)\mathcal{A}(\mu = k\delta)}{\mathcal{A}(\hat{a} = a_j)} \\ &= \frac{\mathcal{A}(\hat{a} = a_j \cap \mu = k\delta)}{\mathcal{A}(\hat{a} = a_j)}. \end{aligned} \tag{A16}$$

We can thus use (A14), together with (A15) and (A16), to generate TFP growth of imitating firms with measurement errors.

Measurement Error: Moments

Table AV reports the empirical moments we use to derive the moments involving measurement error.

Calibration of θ

Figure A5 displays how the ratio of R&D intensity for R&D firms changes with TFP in the data (solid line) and in the benchmark PAM model. The figure shows R&D intensity for each quantile relative to the intensity for firms in the lowest quantile, normalized to unity. The technological parameter θ is calibrated to match the slope between the first and last quantile in Figure A5.

APPENDIX D: RESULTS

In this section, we report robustness estimation results referred to in the text.

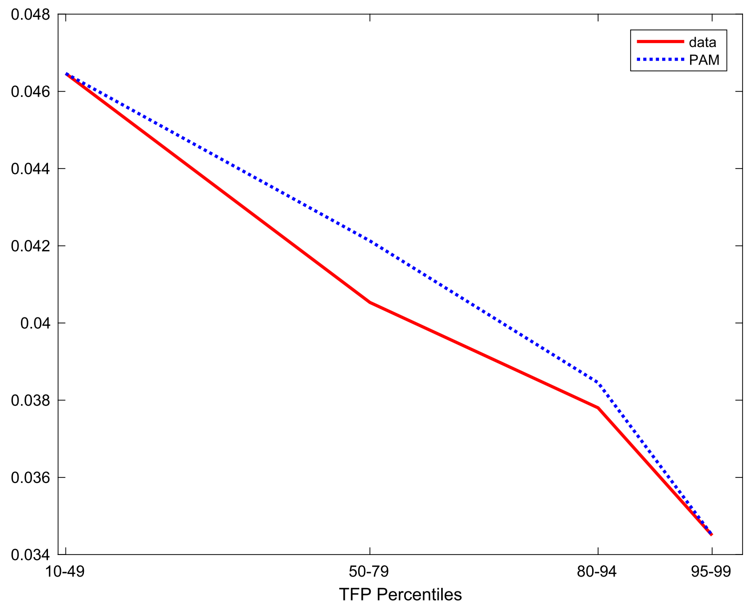

FIGURE A5.—Ratio of R&D to value added, PAM versus data. *Note*: The figure shows the average ratio of R&D to value added for R&D firms in the data (solid line) and the Parsimonious model (dotted line) for four quantiles of the TFP distribution.

D.1. *Estimation of the Fake R&D Model*

Figure A6 shows the fit of the Fake R&D model (FRM). The blue line in each panel represents moments from simulated data based on firms claiming to do (or not to do) R&D. The corresponding moments for a true classification of R&D investments are represented by red lines in the figure.

D.2. *Intensive Margin*

In this section, we lay out and document the fit of the two exercises discussed in Section 5.2 that deal with an intensive margin and heterogeneity in R&D intensities.

High R&D threshold: Figure A7 is the analogue to Figure A4 and shows how the 16 target moments are affected by applying the more stringent classification of innovation, based on Figure A2. Figure A7 also shows the fit of the PAM and IPM when reestimated based on these adjusted moments. The estimated coefficients for this estimation of the PAM and IPM are reported in columns (5)–(6) of Table III.

High- and low-R&D firms: In this section, we introduce a distinction between high- and low-R&D firms. In the data, we assign a firm to the high-R&D group if its R&D expenditure-to-value added ratio is higher than the median 1.73% ratio. Figure A8 displays the data moments. Panels A1 and A2 are the analogues of Panel A in Figure 4 broken down by high- and low-R&D firms. Two features of the data are noteworthy: First, future TFP growth is higher for high-R&D firms. Second, the propensity to engage in R&D conditional on TFP and size is similar for the two groups of firms.

Then, we augment the theory with the assumption that there exist two distinct technologies entailing different costs and success probabilities. More formally, firms are randomly assigned to either of the technologies with parameters $\{\bar{c}_l, \bar{p}_l\}$ and $\{\bar{c}, \bar{p}\}$, respectively.

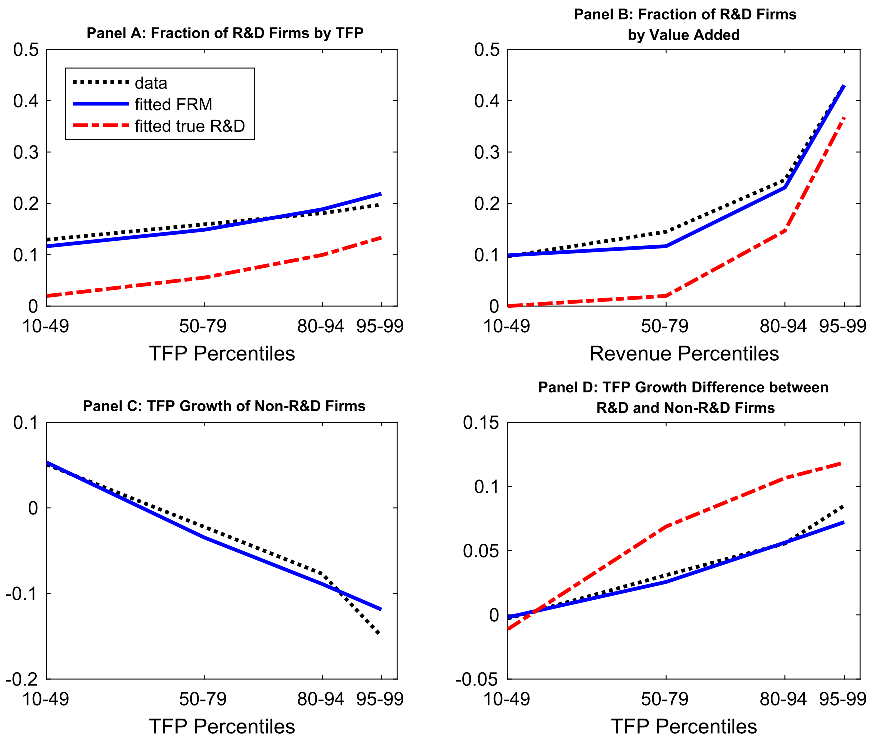

FIGURE A6.—China: Fake R&D Model. *Note*: Each panel of Figure A6 displays three schedules: (i) the dotted line shows the moments in the data, (ii) the dashed line shows the fit of the model (which refers to measured R&D), and (iii) the solid line shows results restricted to the firms which, according to the model predictions, truly perform R&D. See also Figure 4.

Each firm draws a probability of success p from the distribution to which it is assigned. The distribution of wedges is assumed to be independent of the assignment.

The targets of our estimation are now the empirical moments in Figure A8. In addition, we target the ratio of R&D expenditure to value added for high- relative to low-R&D firms, which is a factor of 8.6 in the data. The proportion of high-R&D firms is an additional parameter that we estimate.

The estimates for the PAM and IPM are reported in columns (3)–(4) of Table AVI. In the PAM, firms assigned to the $\{\bar{c}_l, \bar{p}_l\}$ group face both a lower cost and a lower average probability of success if they choose the innovation strategy.[4] In the IPM, the estimated productivities $\bar{p}_l$ and $\bar{p}$ are very similar (in fact, $\bar{p}_l > \bar{p}$.) Still, selection guarantees that, among the firms choosing the innovation strategy, TFP growth is significantly higher for high-R&D than for low-R&D firms, consistent with the data. Intuitively, because of the high investment cost, only the very best firms (i.e., those drawing very high p's) assigned to the $\{\bar{c}, \bar{p}\}$ process choose to innovate. Appendix Figure A8 shows that both the PAM and IPM fit well the target moments.

[4]The proportion of firms drawing from the $\{\bar{c}_l, \bar{p}_l\}$ process is estimated to be 13%. Since the high-R&D firms are by construction 50% of the total R&D firms, this implies that many firms assigned to the $\{\bar{c}, \bar{p}\}$ process choose to imitate because innovation is too costly. In contrast, a large proportion of firms assigned to the $\{\bar{c}_l, \bar{p}_l\}$ process choose the innovation strategy.

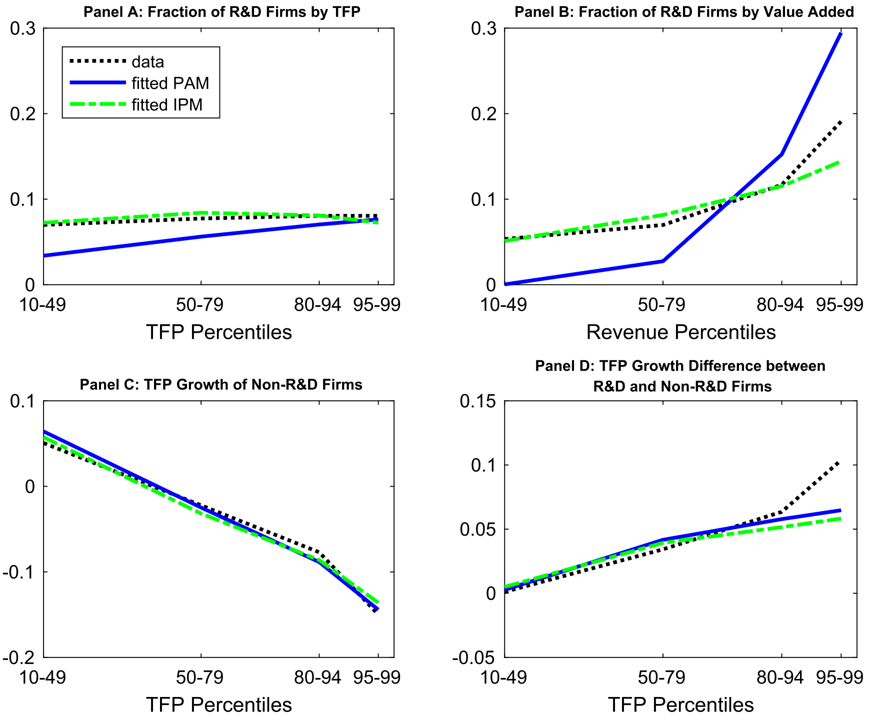

FIGURE A7.—Models with higher R&D cutoff. *Note*: The models are estimated to match empirical moments where only firms with R&D intensity exceeding 1.73% of value added are classified as innovative firms. See Figure 4 for additional information.

Figure A8 shows the fit of the PAM with two R&D technologies (small and large R&D projects). The estimated coefficients for the PAM model with two R&D technologies are reported in columns (3)–(4) of Table AVI. The empirical moments for the high- and low-intensive R&D firms are illustrated by black dotted lines in Figure A8.

APPENDIX E: ESTIMATING THE MODEL ON DIFFERENT SAMPLES

Figures A9 and A10 show the fit of the PAM and IPM for Taiwan and for China in the earlier sample (2001–2007), respectively. Table AVII shows the associated estimated parameters.

APPENDIX F: COUNTERFACTUALS

This section provides robustness analysis for the counterfactuals.

Models With Heterogeneity in Innovation Costs

Figures A11–A12 are the analogues of Figures 8–9 in the manuscript for the IPM that we estimated on the benchmark balanced sample for China, 2007–2012. Table AVIII is the analogue of Table VI in the manuscript for the IPM.

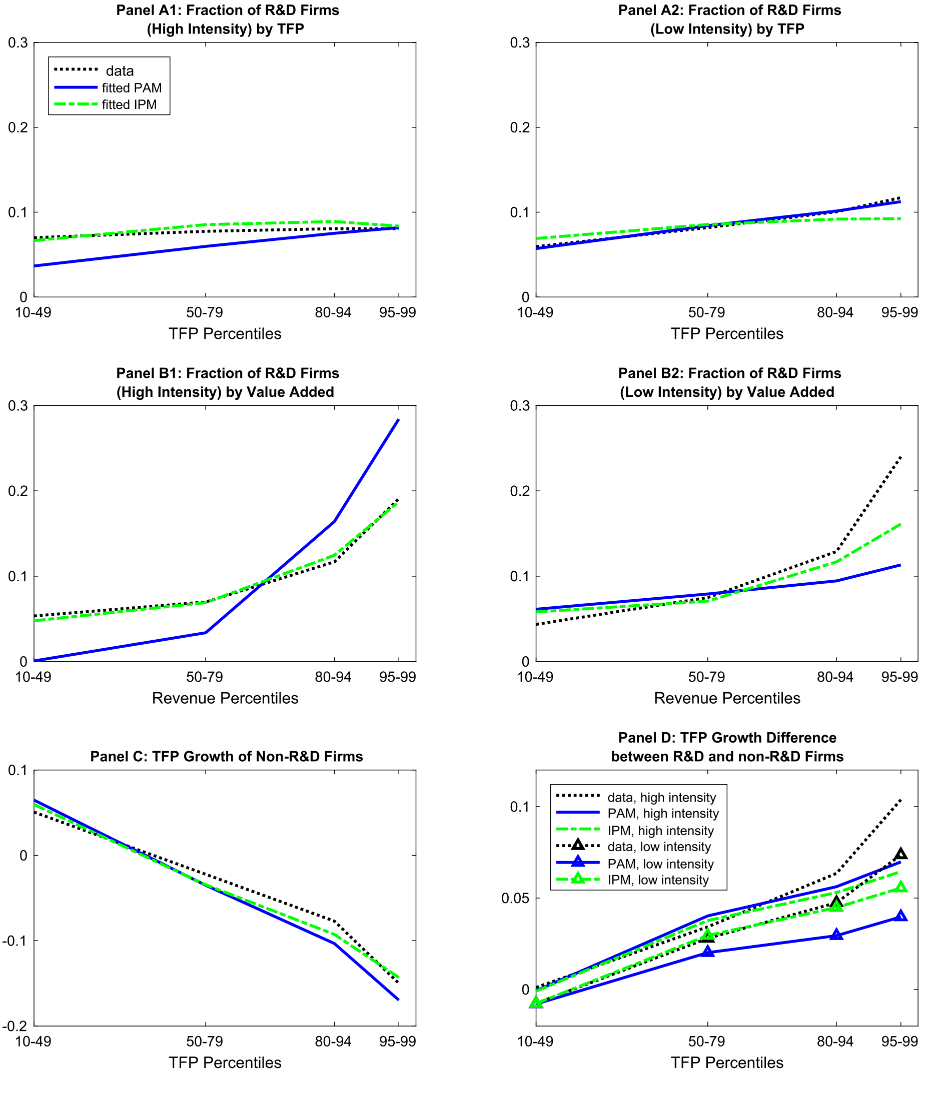

FIGURE A8.—High- and low-R&D firms. *Note*: The figure shows empirical and theoretical moments for the extension to two R&D technologies. Panels A1–A2 and B1–B2 correspond to Panels A and B in Figure 4, reported separately for firms with high- versus low-cost R&D technology. Panels C and D correspond to their counterparts in Figure 4.

TABLE AVI

ESTIMATION OF MODEL WITH TWO R&D TECHNOLOGIES.

	(1)	(2)	(3) High & Low R&D	(4)
	PAM	IPM	PAM	IPM
Imitation prob. q	0.175	0.271	0.223	0.359
	(0.031)	(0.019)	(0.026)	(0.052)
Second chance δ	0.008	0.020	0.058	0.019
	(0.011)	(0.021)	(0.027)	(0.026)
Innov. prod. $\bar{p}$	0.096	0.114	0.103	0.107
	(0.008)	(0.006)	(0.007)	(0.012)
$\bar{p}_l$			0.076	0.115
			(0.009)	(0.010)
Innov. cost $\bar{c}$	1.627	3.374	3.015	3.085
	(0.136)	(0.174)	(0.208)	(0.791)
$\bar{c}_l$			0.094	1.612
			(0.135)	(0.979)
Std.dev. m.e. $\sigma_{\mu a}$	0.549	0.472	0.531	0.433
	(0.014)	(0.008)	(0.015)	(0.024)
Std.dev.innov. subs. σ_c		1.243		1.206
		(0.038)		(0.113)
Policy inter. c_a				2.015
				(0.316)
High $\bar{p}$ share			0.872	0.682
			(0.029)	(0.047)
J-statistic	1.518	0.507	3.310	0.882

Note: Columns (3)–(4) of the table show the estimated parameters for the models with two R&D technologies. For convenience, columns (1)–(2) restate the estimates from Table III for PAM and IPM in the benchmark model. Sample: Chinese Firm Balanced Panel 2007–2012.

TABLE AVII

ESTIMATION FOR CHINA 2001–2007, BALANCED PANEL.

	(1) PAM	(2) FLM	(3) IPM
Imitation prob. q	0.036	0.090	0.093
	(0.032)	(0.046)	(0.039)
Second chance δ	0.033	0.164	0.069
	(0.071)	(0.084)	(0.117)
Innov. prod. $\bar{p}$	0.034	0.046	0.045
	(0.011)	(0.013)	(0.012)
Innov. cost $\bar{c}$	0.530	1.174	0.906
	(0.159)	(0.325)	(0.198)
Std.dev. m.e. $\sigma_{\mu a}$	0.682	0.575	0.580
	(0.038)	(0.025)	(0.026)
Std.dev. innov. subs. σ_c		0.644	0.559
		(0.145)	(0.119)
Policy inter. c_a			0.513
			(0.163)
J-statistic	1.085	0.703	0.371

Note: Estimated parameters of various models using the Chinese 2001–2007 sample. Bootstrapped standard errors in parentheses.

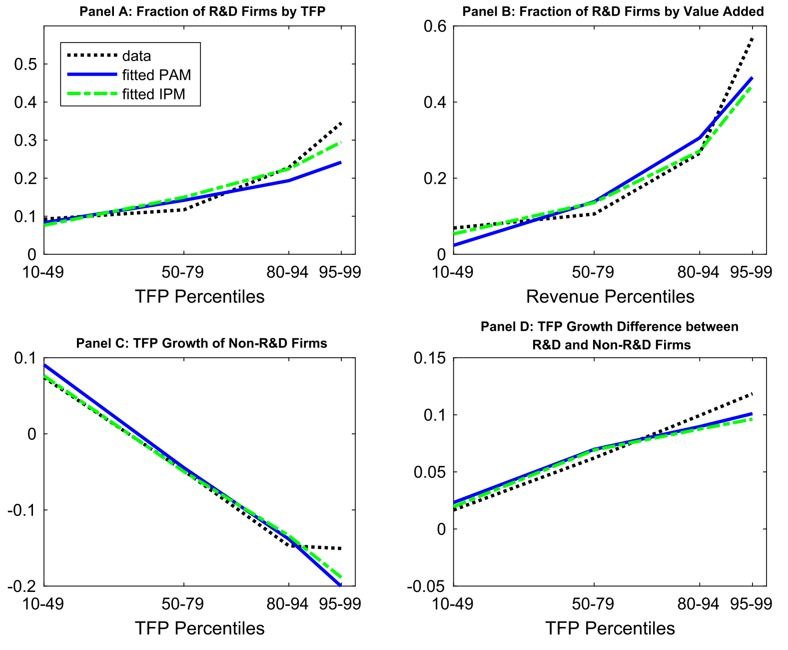

FIGURE A9.—Taiwan 1999–2004: PAM and IPM. *Note*: See Figure 4.

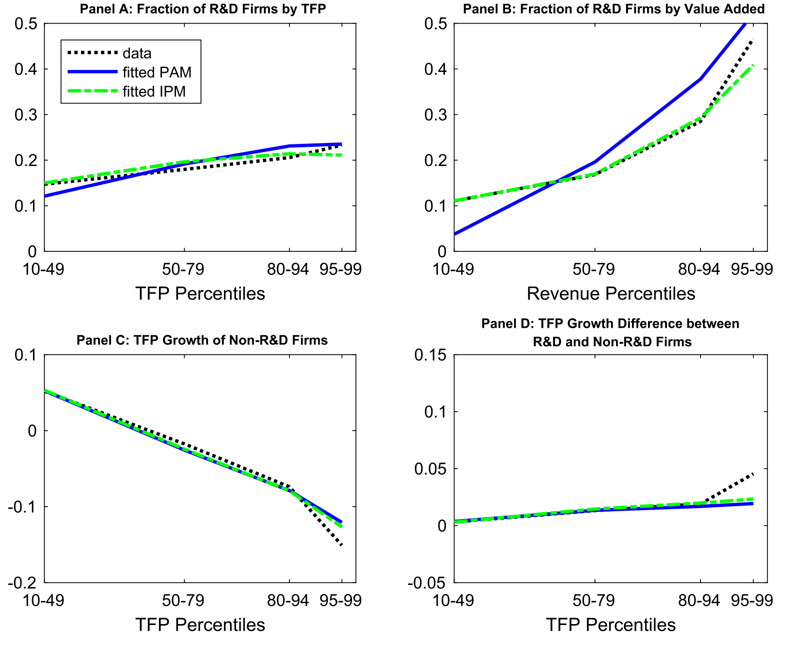

FIGURE A10.—China 2001–2007: PAM and IPM. *Note*: See Figure 4.

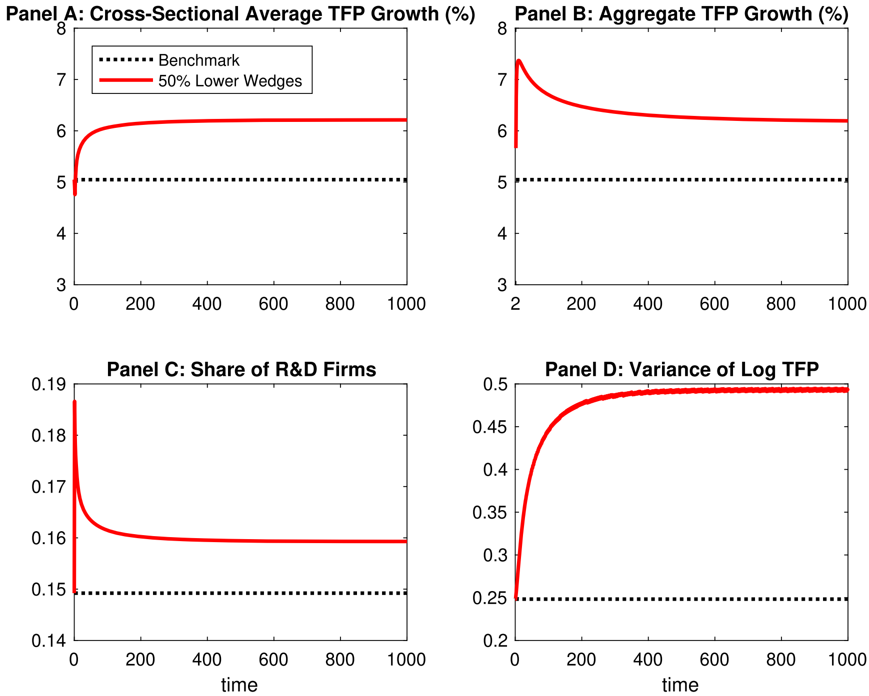

FIGURE A11.—Transition: lower wedges in IPM. *Note*: This figure is the IPM analogue of Figure 8.

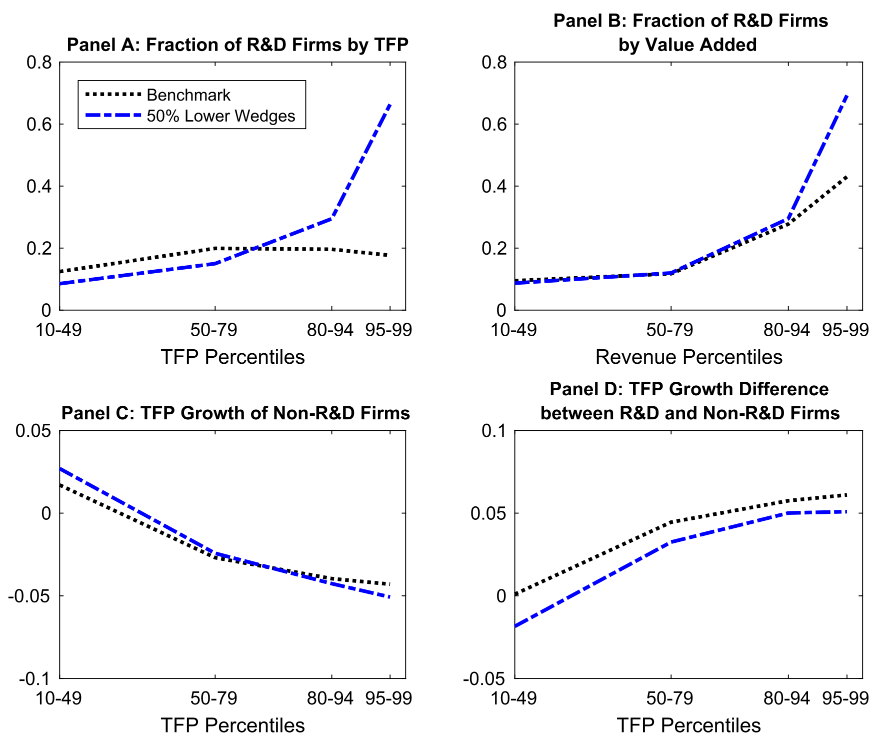

FIGURE A12.—Steady state: lower wedges in IPM. *Note*: This figure is the IPM analogue of Figure 9.

TABLE AVIII

COUNTERFACTUALS, INDUSTRIAL POLICY MODEL.

	(1) IPM estim. model	(2) 50% lower output wedges	(3) Taiwan's q	(4) Taiwan's $\bar{p}$ and $\bar{c}$	(5) Taiwan's $\bar{p}$, $\bar{c}$, and q	(6) Increase $\bar{c}$ so share R&D firms is 5%	(7) Decrease $\bar{c}$ so share R&D firms is 20%	(8) All firms do R&D
Fraction of R&D firms (%)	14.9	16.0	14.8	8.24	8.17	5	20	100
Steady-state TFP growth (%)	5.05	6.20	5.11	5.71	5.78	3.64	5.39	4.41

Note: The table reports statistics for the counterfactual experiments for the IPM discussed in the text. Column (1) reports the predicted moments of the estimated IPM for comparison.

REFERENCES

ACKERBERG, DANIEL A., KEVIN CAVES, AND GARTH FRAZER (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83 (6), 2411–2451. [5,6]

ASL, FARSHID MAGHAMI, AND A. GALIP ULSOY (2003): "Analysis of a System of Linear Delay Differential Equations," *Journal of Dynamic Systems, Measurement, and Control*, 125 (2), 215–223. [3]

BELLMAN, RICHARD, AND KENNETH L. COOKE (1963): *Differential-Difference Equations*. New York: Academic Press. [3]

BRAMSON, MAURY (1983): *Convergence of Solutions of the Kolmogorov Equation to Travelling Waves*, Vol. 285. American Mathematical Soc. [3]

BRANDT, LOREN, JOHANNES VAN BIESEBROECK, LUHANG WANG, AND YIFAN ZHANG (2017): "WTO Accession and Performance of Chinese Manufacturing Firms," *American Economic Review*, 107 (9), 2784–2820. [5, 6]

DE LOECKER, JAN, AND FREDERIC WARZYNSKI (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102 (6), 2437–2471. [5]

HSIEH, CHANG-TAI, AND PETER J. KLENOW (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124 (4), 1403–1448. [5]

KÖNIG, MICHAEL D., JAN LORENZ, AND FABRIZIO ZILIBOTTI (2016): "Innovation vs. Imitation and the Evolution of Productivity Distributions," *Theoretical Economics*, 11 (3), 1053–1102. [3]

SMITH, HAL L. (2011): *An Introduction to Delay Differential Equations With Applications to the Life Sciences*, Vol. 57. New York: Springer. [3]

Co-editor Charles I. Jones handled this manuscript.

Manuscript received 12 June, 2020; final version accepted 23 December, 2021; available online 12 January, 2022.