Online Appendices

“Why is Productivity Correlated with Competition?”

Matthew Backus

Appendices

A Empirical Appendix

A.1 Data

There are over five thousand ready–mix concrete establishments observed by the Census of Manufactures (CMF) in each year of my sample. Unfortunately, roughly one third of these establishments are “administrative records” establishments; that is, they are small enough to be exempt from completing long-form census surveys. Census data for these establishments do exist, but they are generated from a combination of administrative records from other agencies and imputation, the latter of which makes them unusable for establishment–level productivity analysis. I therefore exclude all establishments that fall into this category.

A second constraint is that a small handful of establishments in my sample are extensively diversified and operate in multiple SIC codes. This makes it difficult to construct a productivity residual for the ready–mix concrete portion of their business because data on their inputs are pooled at the establishment level, across all lines of production. I deal with this by excluding establishments for which less than fifty percent of their total sales is from ready–mix concrete. For diversified establishments that survive this exclusion, inputs devoted to ready–mix concrete are approximated by scaling the conflated input variable by the fraction of sales revenue from ready–mix concrete.

Finally, I measure the establishment–level price of a cubic yard of concrete by dividing sales

Table A-1: Production Function Parameters

	α_{Lt}	$\alpha_{K^S t}$	$\alpha_{K^E t}$	α_{Mt}	α_{Et}
1982	0.2250	0.0101	0.0193	0.7173	0.0283
1987	0.2217	0.0227	0.0456	0.6889	0.0210
1992	0.2403	0.0176	0.0436	0.6737	0.0248

Notes: This table presents production function parameter estimates, which are used in the construction of ω_{it} for ready–mix concrete plants in CMF years 1982, 1987, and 1992. See Section 3.4 for discussion, and Foster et al. (2008) for further details on estimation.

by quantity. I observe a small number of firms with extremal values, presumably generated by misreporting, and exclude them from the sample.

It is important to note that while these establishments are excluded for regressions that depend on estimates of the productivity residuals, they are not excluded in the calculation of market–level variables; in particular, the competition indexes discussed below in Section 3.3 use data for all establishments.

In order to construct production function residuals as described in Section 3.4, I use data from the CMF at the establishment level for inputs, including energy, materials, equipment capital, structural capital, and labor. Energy and materials inputs are captured by expenditures reported by the CMF divided by two–digit deflators from the NBER productivity database. Equipment and structural capital are the reported book value multiplied by two–digit capital rental rates from the Bureau of Labor Statistics. Finally, labor is taken as the number of production labor hours plus quality-adjusted non–production worker hours. Following Baily et al. (1992), the quality adjustment is the relative wage, so that total labor is simply production worker hours times the ratio of total wages to production wages. Production function input elasticities for the ready–mix concrete industry are presented in Table A-1.

A.2 Additional Tables

In this section, I offer supplementary tables from the analysis of the ready–mix concrete industry. Table A-2 documents the first-stage regressions that are employed throughout the paper. Tables A-4, A-5, and A-6 replicate the main tables of the body of the paper using labor productivity in place of TFPQ. Labor productivity is computed as the ratio of total quantity output to adjusted labor inputs. Finally, Table A-7 weakens Assumption 2 slightly to allow for a second-order Markov process. Because I only have three time periods of data,

Table A-2: First-Stage Regressions

	Dependent Variable:			
	log(No. Estab./mi.^{2})	log(No. Firms/mi.^{2})	log(HHI No. Estab./mi.^{2})	log(HHI No. Firms/mi.^{2})
	(1)	(2)	(3)	(4)
Building Permits/mi.^{2}	0.1456	0.1391	0.2063	0.1909
	(0.0743)	(0.0878)	(0.1062)	(0.1213)
S.F. Building Permits/mi.^{2}	0.0143	-0.0653	-0.0138	-0.1468
	(0.0743)	(0.0892)	(0.1116)	(0.1364)
Road & Hwy $/mi.^{2}	0.4555*	0.4963*	0.4144*	0.4451*
	(0.0301)	(0.0336)	(0.0398)	(0.0572)
Observations (rounded)	7400	7400	7400	7400
Clusters (rounded)	300	300	300	300
R^2	0.7763	0.7191	0.6315	0.4647

Notes: This table presents first-stage OLS regressions for the IV strategy used throughout the paper, using ready–mix concrete plants in CMF years 1982, 1987, and 1992. Year–specific constants are included but not reported and standard errors (in parentheses) are clustered at the CEA level. For reference, * signifies $p \leq 0.05$.

this collapses the dataset to a single period.

A.2.1 First-Stage Regressions

Here, I present the first-stage regressions for the IV strategy used throughout the paper. Estimates are reported in Table A-2. Conditional on the total number or building permits, the negative coefficient on the number of single–family building permits is intuitive, since it implies fewer larger dwellings.

A.2.2 Firm Survival

Here, I consider probits predicting plant–level survival conditional on productivity and other variables, using the same IV strategy as in the rest of the paper. Results are presented in Table A-3.

On net, it appears that increased competition driven by shocks to market size have a positive, rather than a negative effect on survival. This would be consistent with a failure of the supermodularity condition described in Appendix Section B, which would make it possible for the exit threshold to decrease, rather than increase, but that would be a strong conclusion to draw here. More intuitively, we see a positive relationship between productivity and survival, as does the existing capital stock.

Table A-3: Survival Probits

	Dependent Variable: Dummy for Survival			
	(1)	(2)	(3)	(4)
log(No. Estab./mi.^{2})	0.0180			
	(0.0178)			
log(No. Firms/mi.^{2})		0.0179		
		(.0185)		
log(HHI No. Estab./mi.^{2})			0.6685	
			(0.4797)	
log(HHI No. Firms/mi.^{2})				0.4555
				(0.5150)
ω_{it}	0.1962	0.1965*	0.1913*	0.1985*
	(0.0570)	(0.0570)	(0.0574)	(0.0569)
Structural Capital	0.0731*	0.0729*	0.0732*	0.0713*
	(0.0258)	(0.0258)	(0.0258)	(0.0258)
Equipment Capital	0.0549*	0.0551*	0.0543*	0.0562*
	(0.0254)	(0.0254)	(0.0254)	(0.0253)
Observations (rounded)	8500	8500	8500	8500

Notes: This table presents results from an IV probit of survival on a host of firm-specific variables. Survival here is a dummy for whether the firm appears in the LBD five years subsequent. For reference, * signifies $p \leq 0.05$.

A.2.3 Robustness: Labor Productivity

As a robustness check, I revisit the main regressions of the paper using labor productivity instead of gross total factor productivity. Now, the productivity residual is given by the ratio of total physical output to adjusted labor inputs. Results are presented in Tables A-4, A-5, and A-6. The scale of the effects is different because now I am using a value-added production function instead of a gross production function, however, the qualitative results survive: the IV effects are larger than the OLS effects in the conflated, reduced–form approach; the largest effects are found in the highest deciles of the within–CEA productivity distribution, and the structural estimates of the selection effect of competition on productivity are a reasonably precise zero.

A.2.4 Robustness: Second-Order Markov Process

Here I recast the semi–parametric estimation strategy of Section 5.1 where Assumption 2 is weakened to allow for a second-order Markov process.

Table A-4: Competition and Labor Productivity

	Dependent Variable: Labor Productivity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(No. Estab./mi.^{2})	0.1741*				0.2122*			
	(0.0178)				(0.0177)			
log(No. Firms/mi.^{2})		0.1481*				0.2294*		
		(0.0190)				(0.0218)		
log(HHI No. Estab./mi.^{2})			0.1512*				0.2165*	
			(0.0179)				(0.0187)	
log(HHI No. Firms/mi.^{2})				0.1014*				0.2637*
				(0.0199)				(0.0322)
Observations (rounded)	7400	7400	7400	7400	7400	7400	7400	7400
Clusters (rounded)	300	300	300	300	300	300	300	300
R^2	0.0779	0.0625	0.0737	0.0473				
First-Stage F					344.5	196.3	171.4	48.7
p Value					0	0	0	0
Hansen J Statistic					4.639	7.481	3.153	5.702
p Value					0.0983	0.0237	0.2067	0.0578

Notes: Here I present OLS and IV results for the effect of competition on labor productivity residuals, as discussed in Section A.2.3, for ready–mix concrete plants in CMF years 1982, 1987, and 1992. Models (1)-(4) and (5)-(8) use the four distinct competition measures indicated on the left. Year–specific constants are included but not reported and standard errors (in parentheses) are clustered at the CEA level. Instruments for Models (5)-(8) are the number of building permits per square mile, the number of single–family building permits per square mile, and local government road and highway expenditure per square mile. First-stage F tests and Hansen J (over identification) test statistics are reported with associated p values. For reference, * signifies $p \leq 0.05$.

Table A-5: Competition and Labor Productivity by Grouped Quantile IV

	Dependent Variable: Labor Productivity								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(No. Estab./mi.^{2})	0.1731*	0.1774*	0.1811*	0.1987*	0.2019*	0.2011*	0.2105*	0.2326*	0.2720*
	(0.0288)	(0.0256)	(0.0235)	(0.0217)	(0.0216)	(0.0240)	(0.0242)	(0.0262)	(0.0319)
log(No. Firms/mi.^{2})	0.1734*	0.1763*	0.1807*	0.1967*	0.2007*	0.2012*	0.2103*	0.2308*	0.2676*
	(0.0284)	(0.0256)	(0.0239)	(0.0220)	(0.0221)	(0.0246)	(0.0247)	(0.0270)	(0.0334)
log(HHI No. Estab./mi.^{2})	0.1626*	0.1682*	0.1711*	0.1890*	0.1915*	0.1898*	0.1993*	0.2209*	0.2591*
	(0.0274)	(0.0243)	(0.0220)	(0.0203)	(0.0202)	(0.0224)	(0.0226)	(0.0245)	(0.0299)
log(HHI No. Firms/mi.^{2})	0.1725*	0.1761*	0.1805*	0.1972*	0.2010*	0.2012*	0.2101*	0.2305*	0.2682*
	(0.0287)	(0.0258)	(0.0239)	(0.0220)	(0.0222)	(0.0248)	(0.0248)	(0.0271)	(0.0336)
Observations (rounded)	900	900	900	900	900	900	900	900	900
Clusters (rounded)	300	300	300	300	300	300	300	300	300

Notes: This table contains IV regression results for the effect of competition on deciles of the labor productivity residual distribution at the year–CEA level of aggregation, as discussed in Section 5.2 and, for labor productivity, Appendix Section A.2.3, for ready–mix concrete plants in CMF years 1982, 1987, and 1992. Year–specific constants are included but not reported and standard errors (in parentheses) are clustered at the CEA level. Every cell represents an independent IV regression. By column, Model (k) corresponds to IV regressions with the k^{th} decile of the productivity residual distribution as a dependent variable. By row, regressions use the competition measure reported on the left. Instruments for all regressions are the number of building permits per square mile, the number of single–family building permits per square mile, and local government road and highway expenditure per square mile. For reference, * signifies $p \leq 0.05$.

A.2.5 Robustness: Age and Area

Here I revisit Table 2 including the age of the establishment as well as the log of the size of the CEA (in square miles) in which the establishment is situated. The former covariate

Table A-6: Treatment and Selection Effects with Labor Productivity

	Dependent Variable: Labor Productivity			
	(1)	(2)	(3)	(4)
log(No. Estab./mi.^{2}) ($\hat{\beta}_c$)	0.2468*			
	(0.0241)			
log(No. Firms/mi.^{2}) ($\hat{\beta}_c$)		0.2887*		
		(0.0232)		
log(HHI No. Estab./mi.^{2}) ($\hat{\beta}_c$)			0.2876*	
			(0.0493)	
log(HHI No. Firms/mi.^{2}) ($\hat{\beta}_c$)				0.3032*
				(0.0348)
Selection Coeff ($\hat{\alpha}_c$)	-0.0059	-0.0163	-0.0124	0.0042
	(0.0099)	(0.0142)	(0.0527)	(0.0173)
Observations (rounded)	3100	3100	3100	3100
Clusters (rounded)	300	300	300	300
Hansen J Statistic	11.850	20.180	18.010	6.001
p Value	0.0185	0.0005	0.0012	0.1991

Notes: This table presents results for the semi–parametric selection correction procedure detailed in Section 5.1 using four different indices for competition, and labor productivity in place of ω_{it}, for ready–mix concrete plants in CMF years 1982, 1987, and 1992. Year–specific constants are included but not reported and standard errors (in parentheses) are clustered at the CEA level. For reference, * signifies $p \leq 0.05$.

Table A-7: Treatment and Selection Effects with a Second-Order Markov Process

	Dependent Variable: TFPQ (ω_{it})			
	(1)	(2)	(3)	(4)
log(No. Estab./mi.^{2}) ($\hat{\beta}_c$)	0.0290			
	(0.0154)			
log(No. Firms/mi.^{2}) ($\hat{\beta}_c$)		0.0235		
		(0.0168)		
log(HHI No. Estab./mi.^{2}) ($\hat{\beta}_c$)			0.0314*	
			(0.0151)	
log(HHI No. Firms/mi.^{2}) ($\hat{\beta}_c$)				0.0315
				(0.0190)
Selection Coeff ($\hat{\alpha}_c$)	0.0054	0.0094	0.0044	0.0097
	(0.0052)	(0.0066)	(0.0050)	(0.0073)
Observations (rounded)	1000	1000	1000	1000
Clusters (rounded)	250	250	250	250
Hansen J Statistic	10.550	9.328	10.770	8.827
p Value	0.4814	0.5916	0.4620	0.6378

Notes: This table presents results for the semi–parametric selection correction procedure detailed in Section 5.1 using four different indices for competition, extended as described in Appendix Section A.2.4 to allow for a second-order Markov process for the innovation in productivity residuals, for ready–mix concrete plants in CMF years 1982, 1987, and 1992. Year–specific constants are included but not reported and standard errors (in parentheses) are clustered at the CEA level. For reference, * signifies $p \leq 0.05$.

Table A-8: Competition, Productivity, Establishment Age, and CEA Size

	Dependent Variable: TFPQ (ω_{it})							
	OLS				Instrumental Variables			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(No. Estab./mi.^{2})	0.0351*				0.0502*			
	(0.0063)				(0.0075)			
log(No. Firms /mi.^{2})		0.0306*				0.0569*		
		(0.0072)				(0.0087)		
log(HHI No. Estab./mi.^{2})			0.0347*				0.0588*	
			(0.0078)				(0.0085)	
log(HHI No. Firms /mi.^{2})				0.0172				0.0844*
				(0.0092)				(0.0158)
log(Area)	-0.0002	0.0002	0.0068	-0.0022	0.0063	0.0138	0.0211*	0.0466*
	(0.006)	(0.0066)	(0.0067)	(0.0087)	(0.0064)	(0.007)	(0.0075)	(0.0135)
Firm Age	0.0017	0.0019*	0.0018*	0.0018*	0.0017	0.0020*	0.0018*	0.0022*
	(0.0008)	(0.0008)	(0.0007)	(0.0007)	(0.0008)	(0.001)	(0.0009)	(0.0011)
Observations (rounded)	7400	7400	7400	7400	7400	7400	7400	7400
Clusters (rounded)	300	300	300	300	300	300	300	300
R^2	0.1107	0.1078	0.1102	0.1038	0.1098	0.1047	0.1070	0.0852
First-Stage F					359.3	259.9	117.1	40.23
p Value					0.0000	0.0000	0.0000	0.0000
Hansen J Statistic					2.861	1.064	4.430	0.4148
p Value					0.2391	0.5872	0.1091	0.8127

Notes: Here I present OLS and IV results for the effect of competition on productivity residuals, as discussed in Section A.2.5, for ready–mix concrete plants in CMF years 1982, 1987, and 1992. Models (1)-(4) and (5)-(8) use the four distinct competition measures indicated on the left. Year–specific constants are included but not reported and standard errors (in parentheses) are clustered at the CEA level. Instruments for Models (5)-(8) are the number of building permits per square mile, the number of single–family building permits per square mile, and local government road and highway expenditure per square mile. First-stage F tests and Hansen J (over identification) test statistics are reported with associated p values. For reference, * signifies $p \leq 0.05$.

is meant to capture possible issues with capital stock mismeasurement (e.g., due to vintage capital). The latter covariate is included in order to test for the salience of distance to work site as a mechanism for the treatment effect of competition on productivity. Although not a direct measure of distance to worksite, I hypothesize that in smaller CEAs, which may have denser infrastructure and more uniformly distributed establishments, the driving distance to work site will be lower. Results are presented in Table A-8. The inclusion of these covariates exaggerates rather than attenuates the IV estimates *vis-a-vis* those in Table 2. Moreover, the coefficient on log(Area) is positive rather than negative. I infer that neither concerns about capital mismeasurement nor the mechanism of distance to worksite are first–order in the analysis, however I note that this does not completely rule out reduced distance to worksite, a form of efficiencies from differentiation, as some part of the aggregate effect of competition on productivity.

A.3 Mechanisms: Management and Organizational Practices Data

One direction towards understanding the within–firm effects I document here is to look for evidence directly in management practices. This is particularly difficult for lack of data. Here I take advantage of unique data collected in partnership with the Census and documented by **?**.

The MOPS survey is a survey of management and IT practices that accompanied the 2010 Annual Survey of Manufacturers. It included a large set of questions that covered topics on incorporation of management practices, incentive schemes, and information technology.

My design follows Section 6.3: I ask whether exogenous changes in market competitiveness, driven by changes in market size, are correlated with changes in management practices as measured by the monotized scores (between 0 and 1) for each subsection of the survey. Those scores aggregate all of the questions in each area of the survey, see **?** for details.

Results are presented in Table A-9. Disappointingly, for none of the survey areas do I find any statically significant effects—in fact, point estimates are counter-intuitively negative. I conjecture that the questions of the survey are better designed to think about management in large-scale organizations, where questions of delegation, agency, and monitoring are paramount. Most of the ready–mix concrete plants in my sample (though fewer in 2010) are owner-operated, and therefore the management concerns they face are likely to be rather different, from labor practices (especially hiring) to scheduling and coordination. This would be more consistent with the evidence of specialization and managerial inputs documented in Section 6.3.

A.4 Monte Carlo Exercise

Here I construct a simple Monte Carlo exercise to replicate the order statistic bias discussed in Section 5.2.2. In my sample there are 300 markets. In each market there are a handful of firms, distributed $1 + X$, where X is exponential with parameter $\lambda = 10$. Moreover, each market has a geographic area that is distributed exponential with parameter $\lambda = 12,000$. Firms have productivity draws that are iid $N(0, 0.27)$.

With this I can construct (exact) quantiles of the productivity distribution at the market

Table A-9: Competition and Management and Organizational Practices

	Dependent Variable:					
	Management Score		Monitoring Score		Incentives Score	
	(1)	(2)	(3)	(4)	(5)	(6)
log(No. Estab./mi.^{2})	-0.0071		-0.0072		-0.0087	
	(0.0084)		(0.0088)		(0.0105)	
log(No. Firms./mi.^{2})		-0.0068		-0.0066		-0.0088
		(0.0083)		(0.0089)		(0.0105)
Observations (rounded)	1000	1000	1000	1000	1000	1000
Clusters (rounded)	250	250	250	250	250	250
First-Stage F	128.9	82.52	128.9	82.52	128.9	82.52
p Value	0	0	0	0	0	0
Hansen J Statistic	0.3653	0.4202	2.3090	2.407	0.0216	0.0125
p Value	0.5456	0.5168	0.1286	0.1207	0.8831	0.9111

Notes: I consider alternative specifications of the IV regressions with left-hand side variables constructed from the MOPS survey, as detailed in Section A.3. This sample is independent of that used in other tables: it is made up of ready–mix concrete firms sampled in the 2010 ASM that filled out the MOPS survey. Standard errors (in parentheses) are clustered at the CEA level. Instruments include the number of building permits per square mile, the number of single–family building permits per square mile, and local government road and highway expenditure per square mile. For reference, * signifies $p \leq 0.05$.

level, as well as the log number of establishments per square mile, my dependent variable. Since the latter is generated independently of productivity draws, I use OLS estimates in the estimation exercise. I estimate two versions: a straw man version, without any bias correction, and a version with a fifth-degree polynomial series in the number of firms in the market. Each regression is run 10,000 times. Note that the true parameter β_c in this exercise is equal to zero.

Results are presented in Table A-10 for each decile of the productivity distribution. For each parameter I present the mean and standard deviation (the latter in parentheses).

As expected, we see a strong negative bias in the lower deciles and a positive bias in the upper deciles for the uncorrected estimator. Confidence intervals are narrow (between 0.21 and 0.28), and so coverage drops to zero in these regions. It is better, although still shy of 0.95 for deciles 4 through 6. The picture is better for the corrected estimator. Although confidence intervals are between 10 and 20% wider, the coverage is restored to the neighborhood 0.95 for all deciles, and estimates are close to zero. Finally, at the bottom of Table A-10 I report the average p value for the F test corresponding to the joint significance of the coefficients on the polynomial series. We expect this to be close to zero when the bias correction is doing more work, and indeed, this is what we see for the extremal deciles.

Table A-10: Monte Carlo Simulation of Order Statistic Bias Correction

	Dependent Variable: Decile k of the Simulated Productivity Distribution ($\rho_{mt}^{(k)}$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Uncorrected:									
Estimate	-0.0338	-0.0163	-0.0072	-0.0024	0.0060	0.0029	0.0076	0.0178	0.0340
	(0.0073)	(0.0065)	(0.0063)	(0.0063)	(0.0061)	(0.0059)	(0.0063)	(0.0066)	(0.0079)
CI Length	0.0274	0.0245	0.0226	0.0217	0.0217	0.0217	0.0228	0.0240	0.0272
	(0.0017)	(0.0017)	(0.0016)	0.0016)	(0.0016)	(0.0015)	(0.0015)	(0.0017)	(0.0018)
Coverage Rate	0.0000	0.2200	0.7700	0.8900	0.7800	0.9000	0.7200	0.2000	0.0000
Corrected:									
Estimate	-0.0006	-0.0013	-0.0013	-0.0008	-0.0010	0.0004	0.0018	0.0022	0.0017
	(0.0080)	(0.0075)	(0.0077)	(0.0078)	(0.0076)	(0.0068)	(0.0064)	(0.0067)	(0.0077)
CI Length	0.0311	0.0299	0.0285	0.0276	0.0274	0.0276	0.0288	0.0296	0.0310
	(0.0025)	(0.0025)	(0.0025)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Coverage Rate	0.9500	0.9500	0.9500	0.9200	0.9200	0.9700	0.9700	0.9900	0.9500
F Test p Value	0.0000	0.0031	0.2394	0.4370	0.3002	0.4969	0.2038	0.0187	0.0000
	(0.0000)	(0.0159)	(0.2843)	(0.2993)	(0.2864)	(0.3063)	(0.2608)	(0.0606)	(0.0000)

Notes: This table presents results for the Monte Carlo simulation of the correction for order statistic bias. Means and standard deviations (the latter in parentheses) are reported for the parameter estimate, the length of the confidence interval, and the coverage rate, for both the uncorrected and the corrected estimates. Finally, at the bottom of the table, I report the p value for the F test corresponding to the joint significance of the coefficients on the polynomial series.

B Theoretical Motivation

Consider the following variation on an entry game. In the first stage of the game there is an infinite mass of potential entrants. They are ex-ante identical; potential entrants do not know their type, but they do have rational expectations. If they enter, they pay a cost of entry c_e. Let λ denote the endogenous mass of entrants in state 1. In the second stage entrants learn their idiosyncratic types, denoted ϕ, which are distributed i.i.d. according to a continuous distribution G on $[0,1]$. At this point they make a second choice. They may exit or stay active. If they exit they receive a payoff normalized to zero. Let μ denote the measure of active, non–exiting firms on the type space $[0,1]$. Those firms obtain profits given by $\pi(\phi, C(\mu), D)$, where D is a demand shifter and C is a continuous function mapping measures μ into $\mathbb{R}$. C represents an ideal competition index. I assume that C is increasing in the following partial order: if $\mu \geq \mu'$ on $[0,1]$ and $\mu' > \mu$ for some open set in $[0,1]$ then $C(\mu') > C(\mu)$. I also assume that π is continuously differentiable in all of its arguments, strictly increasing in ϕ and D, and strictly decreasing in the same partial order on C.

This model is in the spirit of Hopenhayn (1992) and Asplund and Nocke (2006), however I have abstracted away from dynamics for simplicity. Consistent with those models, monotonicity of π implies that the exit decision in the second stage follows a threshold rule; let us call it $\bar{\phi}$. An equilibrium is a pair $< \lambda, \bar{\phi} >$ such that:

$$\int_{\bar{\phi}}^{1} \pi(\phi, C(\mu), D) dG(\phi) = c_e \tag{E1}$$

$$\pi(\bar{\phi}, C(\mu), D) = 0. \tag{E2}$$

Equilibrium condition (E1) reflects optimal choice by potential entrants ex ante, while condition (E2) reflects the ex post exit choice of entrant firms. First, lets note that a nontrivial equilibrium exists and is unique under mild conditions.

Proposition 1. *Let* $\psi \equiv \{\phi : \pi(\phi, C(0), D) = 0\}$. *If*

1. $\int_{\psi}^{1} \pi(\phi, C(0), D) dG(\phi) > c_e$, *and*
2. *there exists a finite measure* $\tilde{\mu}$ *on* $[0, 1]$ *such that* $\pi(1, C(\tilde{\mu}), D) < 0$,

then there exists a unique equilibrium with a nonempty market (i.e., such that $\mu \neq 0$*).*

Proof. I omit the proof of existence, which follows directly from Conditions 1 and 2 and application of the Schauder Fixed Point theorem.

To see uniqueness, suppose by way of contradiction that there were two equilibria at $< \lambda, \bar{\phi} >$ and $< \lambda', \bar{\phi}' >$. From (E2), we have $\bar{\phi}' > \phi \Leftrightarrow \lambda' > \lambda$. Suppose without loss of generality that $\bar{\phi}' > \bar{\phi}$. However, now $\int_{\bar{\phi}'}^{1} \pi(\phi, C(\mu'), D) < \int_{\bar{\phi}}^{1} \pi(\phi, C(\mu), D) dG(\phi) = c_e$, which contradicts the claim that $< \lambda', \bar{\phi}' >$ is an equilibrium. □

The key result is a comparative static in D, which stands in for demand shifters. I am interested in showing how the optimal threshold $\bar{\phi}$ moves as the market size grows. In order to prove this I need two more assumptions. The first assumption says that when the market grows, the profits of high type firms grow no less than proportionately.

Assumption 1. *For* $\phi' > \phi$, $D' > D$, *and any* C,

$$\frac{\pi(\phi', C, D')}{\pi(\phi, C, D')} \geq \frac{\pi(\phi', C, D)}{\pi(\phi, C, D)}.$$

The second assumption embodies the reallocation hypothesis:

Assumption 2. *For $\phi' > \phi$, $C' > C$, and any D,*

$$\frac{\pi(\phi', C', D')}{\pi(\phi, C', D)} > \frac{\pi(\phi', C, D)}{\pi(\phi, C, D)}.$$

On the theory side, Asplund and Nocke (2006) argue that it is consistent with many standard models. Going a step further, Boone (2008) argues that this is in fact constitutive of our very idea of competition. Finally this mechanism has become central in trade and productivity analysis literatures as well. In Proposition 2 I show that it implies the selection effect hypothesis.

Proposition 2. *If $D' \geq D$, then $\bar{\phi}' \geq \bar{\phi}$.*

Proof. First, note that $C(\mu') \geq C(\mu)$. Suppose, by way of contradiction, otherwise. Then (E2) implies that $\bar{\phi}' < \bar{\phi}$. Now, $\int_{\bar{\phi}'}^{1} \pi(\phi, C(\mu'), D') > \int_{\phi^*}^{1} \pi(\phi, C(\mu), D) dG(\phi) = c_e$, which generates a contradiction.

Next observe that, for (E1) to hold, there must exist $\tilde{\phi}$ such that $\pi(\tilde{\phi}, C(\mu'), D') = \pi(\tilde{\phi}, C(\mu), D) > 0$. Moreover, for $\phi' < \tilde{\phi}$,

$$\begin{aligned} \frac{\pi(\phi', C(\mu'), D')}{\pi(\tilde{\phi}, C(\mu'), D')} &< \frac{\pi(\phi', C(\mu), D)}{\pi(\tilde{\phi}, C(\mu), D)} \\ &\Rightarrow \pi(\phi', C(\mu'), D') < \pi(\phi', C(\mu), D). \end{aligned}$$

The first line follows from log increasing differences and the complementarity of D and $C(\mu)$. The second line follows from $\pi(\tilde{\phi}, C(\mu'), D') = \pi(\tilde{\phi}, C(\mu), D) > 0$, and implies that $\bar{\phi}' < \bar{\phi}$. $\square$