Online Appendix to: Raising the Barcode Scanner: Technology and Productivity in the Retail Sector

Emek Basker
University of Missouri

A Measurement Error and Selection Bias

I restrict the sample used in this paper to stores that installed scanners by the end of 1984, and which I was able to match with observations in the LBD, in order to address both omitted-variable and measurement-error biases. In this Appendix, I re-estimate Equations (1)-(3) using the full set of stores classified in SIC 541 to investigate the extent of selection bias and measurement error in the full sample of stores. The results are shown in Table A-1.

The omitted-variable concern is that stores adopted barcodes because of some unobserved factor, which either has itself a direct effect on store productivity, or is correlated with other unobserved factors affecting store productivity. Including store fixed effects in the regression controls for any time-invariant store-level differences in both observed and unobserved factors, but cannot account for time-varying effects. For example, stores that implement scanning may have grown, on average, relative to other grocery stores in this time period, or may have been located in neighborhoods that become relatively more prosperous. If these unobserved characteristics create differential productivity trends in stores that ultimately adopt scanners and stores that do not, a naïve regression would attribute the differential trend to the impact of scanners. In the main text, I argue that this problem is eliminated when I restrict the sample to stores that adopted scanners by December 1984. The specification tests support this argument. While omitted-variable bias could be either positive or negative, in this case it is almost certain to overstate the effect of a scanner.

The measurement-error concern is not present in the restricted sample by construction, since every store included in the sample is known to have installed a scanner. But because I am unable to match more than 6,000 stores listed in the FMI publication to stores in the Census, it is very likely that the full sample of stores includes some for which my scanner variable is set to zero actually had scanners for part of the sample period. This measurement error creates attenuation bias in the estimated coefficients.

In the full sample, the estimate of β from Equation (1), shown in the first column, is nearly 0.15, or three times the magnitude of the estimate in the restricted sample, indicating large omitted-variable bias. Consistent with this conclusion, the estimate of β_{t+1} from Equation (2) is also positive, large, and significant. The statistical significance is not merely an artifact of the large sample. Scanner installation in year $t+1$ is associated with a 10 percent increase in store productivity in year t, an increase that amounts to two thirds of the estimated post-installation productivity increase. The positive and statistically significant estimate of γ from Equation (3) in the full sample, shown in column (3), most likely reflects measurement error.

Table A-1: Scanners' Effect on Labor Productivity: All Grocery Stores

	(1)	(2)	(3)
Scanner	0.1484***	0.1540***	0.1428***
	(0.0068)	(0.0069)	(0.0068)
$\Delta\text{Scanner}_{t+1}$		0.1000***	
		(0.0074)	
CityScanner			0.0330***
			(0.0038)
Year FE	✓	✓	✓
Store FE	✓	✓	✓
Observations	247,242	247,242	247,242

LHS variable is log productivity. Unbalanced panels, 1972–1982.
Robust standard errors in parentheses (clustered by store)
*** significant at 1%; ** significant at 5%; * significant at 10%

B Spline Estimation

As a complement to the difference-in-difference estimates presented in the main part this paper, I also present a piecewise linear regression that exploits the timing of installations. Since most scanners were installed in the second part of my sample, I focus for this section on the years 1981–1983.

This analysis exploits the nonlinearity in the relationship between the timing of installation and the estimated impact on the store's measured productivity in 1982. Specifically, while the impact of scanners installed in 1981 should be fully captured in 1982 productivity, the impact of scanners installed in 1982 should be partially captured in 1982 productivity — more so, the earlier in the year the installation occurred — and the impact of scanners installed in 1983 should have had no effect on 1982 productivity. To test this hypothesis I estimate a spline regression. The left-hand side variable is a store's productivity growth from 1977 to 1982. I limit the sample to the 1,580 stores that installed scanners from January 1981 to December 1983 and for which I have both 1977 and 1982 productivity measures. This limited sample imposes even weaker identifying assumptions than the "installer" scanner used in the main part of the paper, namely that, conditional on a store installing a scanner over this 3-year period, the timing of installation is uncorrelated with other factors related to the store's productivity growth from 1977 to 1982.

For each store, I define **ScanTime** to be the time elapsed, in months, between January 1, 1982 and the store's installation date. For example, if a store installed a scanner in March 1982, this variable equals 2.5; if the store installed a scanner in November 1981, this variable equals −1.5. Since my sample is limited to stores that installed scanners from January 1981 to December 1983, **ScanTime** takes on values from −11.5 to 23.5. To accommodate the nonlinearity discussed above, I create two knots in **ScanTime**, allowing changes in the constant term starting at **ScanTime** = 0, separating stores that installed in 1981 from stores that installed in 1982, and at **ScanTime** = 12, separating stores that installed in 1982 from stores that installed in 1983. I then estimate the piecewise linear relationship between the timing of scanner installation and productivity growth using a spline regression:

$$\Delta \ln(\mathbf{productivity})_{i,1982} = \alpha_0 + \beta_0 \mathbf{ScanTime}_i \cdot \mathbf{1}(\mathbf{ScanTime}_i < 0) + \alpha_1 \mathbf{1}(\mathbf{ScanTime}_i \in [0,12)) + \beta_1 \mathbf{ScanTime}_i \cdot \mathbf{1}(\mathbf{ScanTime}_i \in [0,12)) + \alpha_2 \mathbf{1}(\mathbf{ScanTime}_i \geq 12) + \beta_2 \mathbf{ScanTime}_i \cdot \mathbf{1}(\mathbf{ScanTime}_i \geq 12) + \varepsilon_i, \tag{B-1}$$

where, as before,$\mathbf{1}(\cdot)$ is the indicator function, and $\Delta \ln(\mathbf{productivity})_{i,1982}$ is the difference between establishment i's 1982 productivity and its 1977 productivity.

The results are shown graphically as the dashed (green) line in Figure B-1. I have normalized the coefficients so that the point estimate for the effect of a scanner installed at **ScanTime** = 12 (January 1, 1983) is zero. Dashed vertical lines indicate the knots. The results show a downward-sloping relationship between 1982 productivity gains and the time of installation for the full period from January 1981 to December 1982, followed by a slight upward-sloping relationship. Only the coefficient on the middle segment of the spline, however, statistically significant at the 10 percent level. That coefficient indicates that each additional month *without* a scanner reduces a store's productivity in 1982, relative to its 1977 level, by 0.29 percent, or 3.5 percent annualized.[1]

To increase the power and precision of these estimates, I also estimate an alternative specification in which I force the slopes of the first and last segments to be zero, and omit the level shifters between segments. The estimated coefficients from this regression are shown as the solid (orange) line in Figure B-1. Each month without a scanner reduces the store's 1982 productivity by 0.38 percent. Put differently, scanning for the full year increases the store's productivity growth by 4.6 percent, a figure remarkably close to the difference-in-difference estimate presented in Table 3.

[1] In unreported regressions, I also estimate variants of this regression in which I replace the left-hand side variable with the store's 1982 productivity (instead of productivity growth), both with or without controlling for 1977 productivity. Those results show larger effects, annualized to productivity gains of 7.6 percent and 5.2 percent, respectively. In model variants that allow discontinuities at the knots, the discontinuities are never jointly significant.

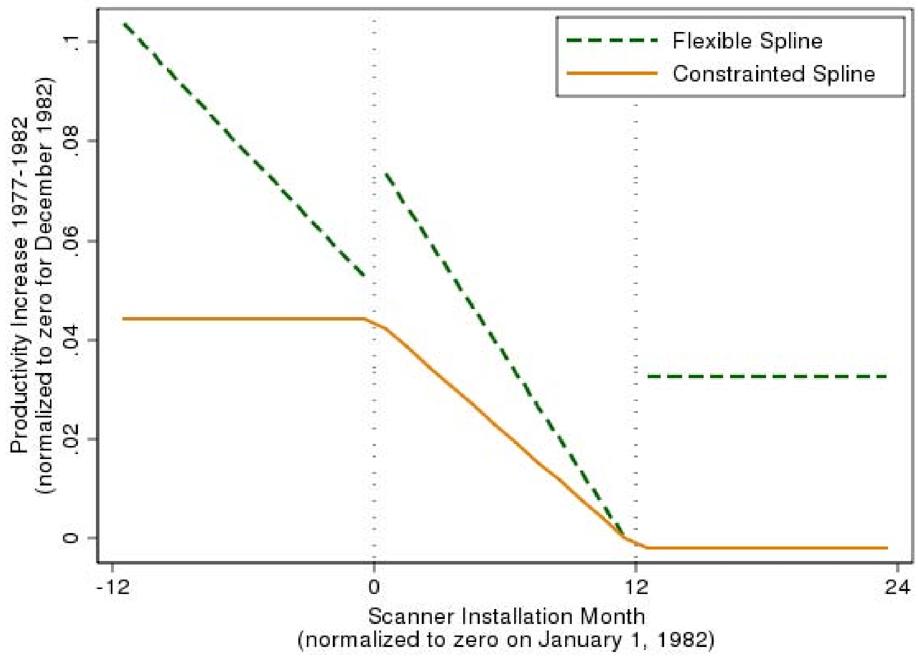

Figure B-1: Spline Estimates of Scanners' Effect on Labor Productivity

C Heterogeneous Effects of Scanners

I extend the basic analysis of Section 4.1 by allowing the effect of a barcode scanner on a store's productivity to vary with the characteristics of the firm to which the store belongs or the city in which it is located. There are many dimensions on which scanner adoption could have differential effects, and a complete catalog of these is beyond the scope of this paper. I focus on one firm-level covariate, the size of the chain, and three city-level covariates: population in 1970, number of food-selling establishments in 1972, and the population growth rate between 1970 and 1980. Summary statistics for these variables are provided in Tables 1 and 2 in the paper.

I estimate

$$\begin{aligned}\ln(\mathbf{productivity})_{it} &= \alpha_i + \delta_t + \gamma_t \times \ln(\mathbf{X}_i) \\ &\quad + \beta\mathbf{Scanner}_{it} + \beta_x\mathbf{Scanner}_{it} \times \ln(\mathbf{X}_i) + \varepsilon_{it} \quad \text{(C-1)}\end{aligned}$$

where $\mathbf{X}$ is, respectively, the firm's size (measured as the number of stores it operates), the city's population in 1970, the number of food-selling establishments (SIC 541) in 1972 in the city, or the city's population growth. Not all cities in the LBD/FMI dataset have city-level variables, so these regressions are estimated with the subset of observations for which I have city-level data.

The results are shown in Table C-1. None of the interaction effects are statistically significant at the 5 percent level; the population interaction is significant at the 10 percent level. The interaction terms are also not statistically significant when I estimate heterogeneous effects on stores' payroll or revenue.

Table C-1: Heterogeneous Effects of Scanners on Labor Productivity

	(1)	(2)	(3)	(4)
Scanner	0.0511***	0.1443**	0.0773*	0.0434***
	(0.0184)	(0.0563)	(0.0441)	(0.0099)
Scanner	−0.0032			
× ln(ChainSize)[a]	(0.0037)			
Scanner		−0.0096*		
× ln(Pop1970)[b]		(0.0054)		
Scanner			−0.0085	
× ln(Estab1972)[c]			(0.0092)	
Scanner				0.0125
× ln(Growth)[d]				(0.0310)
Year FE	✓	✓	✓	✓
Year FE × ln(ChainSize)	✓			
Year FE × ln(Pop1970)		✓		
Year FE × ln(Estab1972)			✓	
Year FE × ln(Growth)				✓
Store FE	✓	✓	✓	✓
Observations	7,865	7,057	3,546	7,057

LHS variable is log productivity. Unbalanced panels, 1972–1982.
Robust standard errors in parentheses (clustered by store)
*** significant at 1%; ** significant at 5%; * significant at 10%
[a] Number of stores in the chain in 1977
[b] 1970 city population from City and County Data Book
[c] 1972 number of food-selling establishments from City and County Data Book
[d] Log 1970 to 1980 population growth