Online Appendix for

Production Chain Organization in the Digital Age:

IT Use and Vertical Integration in U.S. Manufacturing

Chris Forman
Charles H. Dyson School of Applied Economics
and Management
Cornell SC Johnson College of Business
Cornell University
Warren Hall, Ithaca, NY 14853

Kristina McElheran
University of Toronto
UTSC Dept of Management
& Rotman School of Management
1265 Military Trail
Toronto, ON M1C 1A4

Appendix A

Table A.1 Correlation Matrix

Variables	1	2	3	4	5	6	7	8	9	10		11	12	13	14
1. External IT	**1**														
2. Internal IT	**0.56*** **	**1**													
3. Production IT	**0.49*** **	**0.61*** **	**1**												
4. Log generic IT	**0.00*** **	**0.04*** **	**0.06*** **	**1**											
5. ITR	-0.05***	0.01	-0.01	-0.03***	1										
6. Log inv/sales	0.007	0.02	0.03***	-0.01	-0.11***	1									
7. Skill mix	0.05***	-0.02	0.00	0.17***	-0.13***	0.06***	-0.13***	1							
8. Log downstream internal demand	0.01	0.06***	0.06***	0.053***	0.21***	-0.07***	0.21***	-0.01***	1						
9. Log downstream external demand	0.033***	0.04***	0.05***	-0.01	0.11***	-0.10***	0.11***	-0.10***	0.50***	1					
10. Has local competitors	0.04***	0.03***	0.04***	0.03***	-0.02***	0.03***	-0.03***	0.06***	-0.03***	-0.003	1				
11. Local industry Herfindahl	-0.02	-0.01	-0.00	0.04***	0.04***	-0.06***	0.04***	-0.13***	0.08***	0.02	-0.52***1				
12. Log # Products	-0.04***	-0.05***	-0.05***	0.12***	-0.09***	-0.01	-0.09***	0.03***	0.05*	-0.01	0.03***	0.04***	1		
13. TFP	0.04***	0.01	0.01	0.08***	-0.04***	-0.08***	-0.04***	0.06***	-0.16***	-0.18***	0.02	-0.05***	-0.11***	1	
14. Log # Firm units	0.05***	0.09***	0.10***	0.01	0.08***	-0.06***	0.08***	-0.03***	0.31***	0.05***	-0.003	0.05***	0.08***	-0.07***	1

Notes: Number of observations is approximately 5,600 per year (rounded per Census disclosure avoidance policy). *significant at 10%; **significant at 5%; ***significant at 1%. Census disclosure clearance date 5/21/2018.

Table A.2 First Stage of Instrumental Variables Estimates of Table 3

	(1)	(2)	(3)	(4)	(5)
	Other Internal Adopters of CAD/CAE	Competitor External IT index	Log of proxy cost	Two instruments	All instruments
Other Firm Adopters of CAD/CAE	0.138*** (0.009)				0.120*** (0.010)
Competitor External IT index		0.505*** (0.060)		0.497*** (0.063)	0.346*** (0.060)
Log of proxy cost			-0.019** (0.006)	-0.016** (0.006)	-0.016** (0.006)
Log inventory-to-sales ratio	-0.049 (0.033)	-0.062* (0.033)	-0.080** (0.034)	-0.072** (0.034)	-0.059** (0.034)
Skill mix	0.014 (0.015)	0.003 (0.016)	0.002 (0.017)	0.002 (0.017)	0.013 (0.016)
Log downstream internal demand	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Log downstream external demand	0.003* (0.002)	0.003* (0.002)	0.002 (0.002)	0.003 (0.002)	0.003 (0.002)
Has local competitors	-0.001 (0.012)	-0.002 (0.012)	-0.001 (0.012)	-0.002 (0.012)	-0.002 (0.012)
Local industry Herfindahl	-0.009 (0.017)	-0.006 (0.017)	-0.009 (0.018)	-0.009 (0.017)	-0.009 (0.017)
Number of products	-0.003 (0.004)	-0.004 (0.004)	-0.002 (0.004)	-0.002 (0.004)	-0.001 (0.004)
Total factor productivity	-0.007 (0.005)	-0.010** (0.005)	-0.014*** (0.005)	-0.012** (0.005)	-0.008* (0.005)
Log number of firm units	-0.003* (0.002)	-0.004** (0.002)	-0.004** (0.002)	-0.004* (0.002)	-0.003 (0.002)
Observations	11,500	11,500	10,500	10,500	10,500
Plants	5,600	5,600	5,200	5,200	5,200
F-statistic	227.6	70.13	9.54	36.39	79.73
Stock-Yogo (2005) critical values	8.96	8.96	8.96	11.59	12.83

Notes: Unit of observation is a plant-year (numbers of observations are approximate per Census disclosure avoidance policy). Sample includes annual data from 1992 and 2002. Regressions include plant fixed effects, differenced out at means, and 1992 year fixed effects. Stock and Yogo (2005) critical values are reported for maximal LIML size > 10%. Robust standard errors, clustered by plant, in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Census disclosure clearance date 5/21/2018.

Table A.3 Second Stage of Instrumental Variables Estimates of Table 3

	(1)	(2)	(3)	(4)	(5)
	Other Firm Adopters of CAD/CAE	Competitor External IT Index	Log of proxy cost	Two Instruments	All instruments
External IT index	-0.301*** (0.099)	-0.237 (0.149)	-1.116* (0.603)	-0.362** (0.159)	-0.334*** (0.097)
Log inventory-to-sales ratio	-0.237*** (0.060)	-0.233*** (0.060)	-0.323*** (0.090)	-0.260*** (0.065)	-0.258*** (0.063)
Skill mix	-0.019 (0.023)	-0.020 (0.023)	-0.028 (0.030)	-0.029 (0.025)	-0.029 (0.025)
Log downstream internal demand	0.003** (0.001)	0.003** (0.001)	0.002 (0.002)	0.003** (0.001)	0.003** (0.001)
Log downstream external demand	-0.004 (0.003)	-0.004 (0.003)	-0.002 (0.004)	-0.004 (0.003)	-0.004 (0.003)
Has local competitors	0.025 (0.018)	0.025 (0.018)	0.018 (0.022)	0.021 (0.019)	0.021 (0.019)
Local industry Herfindahl	0.022 (0.029)	0.022 (0.029)	0.010 (0.036)	0.018 (0.032)	0.018 (0.032)
Log of number of products	0.0004 (0.006)	0.0006 (0.005)	-0.0025 (0.007)	-0.0008 (0.006)	-0.0008 (0.006)
Total factor productivity	-0.024** (0.009)	-0.024** (0.009)	-0.038*** (0.014)	-0.028*** (0.010)	-0.027*** (0.010)
Log number of firm units	0.012*** (0.003)	0.012*** (0.003)	0.007 (0.004)	0.011*** (0.003)	0.011*** (0.003)
Observations	11,500	11,500	10,500	10,500	10,500
Establishments	5,600	5,600	5,200	5,200	5,200
Overidentification test (p-value)	NA	NA	NA	0.114	0.267

Unit of observation is an establishment-year (numbers of observations are approximate per Census disclosure avoidance policy). Sample includes annual data from 1992 and 1999. Regressions include establishment-specific fixed effects, differenced out at means, and 1992 year fixed effects. Robust standard errors, clustered by establishment, in parentheses. Overidentification test uses Hansen J statistic. * significant at 10%; ** significant at 5%; *** significant at 1%. Census disclosure clearance date 5/21/2018.

Table A.4 Additional Robustness

Models	(1) Panel	(2) Panel including Early Adoption of IT	(3) Industry x Time Trends	(4) Industry x Time Trends including Early Adoption	(5) Fractional Probit Model QME GLM	(6) Control Function	(7) One-Sided Tobit	(8) Two-Sided Tobit
Dependent Variables	ITR	ITR	ITR	ITR	ITR	ITR	Log Internal Transfers	ITR
External IT index	-0.029** (0.011)	-0.038** (0.016)	-0.024** (0.011)	-0.032** (0.016)	-0.063*** (0.020)	-0.263*** (0.095)	-1.024*** (0.038)	-0.052*** (0.017)
Early adoption of external IT		-0.014 (0.015)		-0.013 (0.015)				
Other controls from Table 3, column 3	Y	Y	Y	Y	Y	Y	Y	Y
T-Test of Differences between external index and early external index	NA	0.041	NA	0.094				
Observations	59,500	59,500	59,500	59,500	11,500	11,500	11,500	11,500
Plants	5,600	5,600	5,600	5,600	5,600	5,600	5,600	5,600
R-squared	0.008	0.008	0.011	0.011				
R-squared with fixed effects	0.651	0.651	0.651	0.651				

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Columns (1)-(4): Unit of observation is a plant-year (numbers of observations are approximate). Sample includes annual data from 1992 through 2002. Regressions include plant-specific fixed effects and year indicators. Results in columns 3 and 4 include three-digit NAICS industry time trends. All regressions include the controls used in column 3 of Table 3. Robust standard errors, clustered by plant, in parentheses.

Column (5)-(8): Estimates are average partial effects from non-linear models. Average partial effects are computed based on averages of partial effects across all observations in the sample (numbers of observations are approximate). Standard errors are bootstrapped standard errors clustered by establishment. 500 replications are used. Column 6 represents a control-function approach. The exclusion restrictions for this are competitor adoption of external IT, adoption of CAD/CAE by other establishments within the same firm, and the log of proxy cost of delivering telecommunications services to a location.

Census disclosure clearance date 5/21/2018.

Table A.5 Including Changes to Firm Structure as Controls

Models	(1) Change Firm ID	(2) Change Firm ID + IT Controls	(3) Change Firm ID, Pct Change Other Firm ID	(4) Change Firm ID, Pct Change Other Firm ID, + IT Controls	(5) All Firm Structure Controls	(6) All Firm Structure Controls + IT Controls
Dependent Variable	ITR	ITR	ITR	ITR	ITR	ITR
External IT index	-0.060***	-0.061***	-0.059***	-0.060***	-0.059***	-0.059***
	(0.019)	(0.021)	(0.019)	(0.021)	(0.019)	(0.021)
Log inventory-to-sales ratio	-0.218***	-0.218***	-0.217***	-0.217***	-0.217***	-0.217***
	(0.058)	(0.058)	(0.058)	(0.058)	(0.058)	(0.058)
Skill mix	-0.021	-0.020	-0.021	-0.019	-0.021	-0.019
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Log downstream internal demand	0.003**	0.003***	0.003**	0.003**	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log downstream external demand	-0.005*	-0.005*	-0.005*	-0.005*	-0.005*	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Has local competitors	0.024	0.025	0.025	0.025	0.025	0.025
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Local industry Herfindahl	0.024	0.025	0.024	0.025	0.024	0.025
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.018)
Number of products	0.001	0.001	0.001	0.001	0.008	0.007
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Total factor productivity	-0.021**	-0.022**	-0.021**	-0.021**	-0.021**	-0.021**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Log number of firm units	0.012***	0.012***	0.012***	0.012***	0.014*	0.014***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Internal IT index		0.017		0.017		0.017
		(0.012)		(0.012)		(0.012)
Log generic IT investment		-0.002		-0.002		-0.002
		(0.001)		(0.001)		(0.001)
Production IT index		-0.022**		-0.021**		-0.022**
		(0.010)		(0.010)		(0.010)
Changed firm ID	-0.021***	-0.022***	-0.033***	-0.033***	-0.028**	-0.028**
	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)	(0.011)
Pct chg other firm ID			0.023	0.023	0.020	0.020
			(0.016)	(0.016)	(0.016)	(0.016)
Firm growth					-0.001*	-0.001*
					(0.000)	(0.000)
Observations	11,500	11,500	11,500	11,500	11,500	11,500
R-squared	0.020	0.021	0.020	0.021	0.020	0.021
R-squared with fixed effects	0.712	0.712	0.712	0.713	0.712	0.713

Notes: Unit of observation is a plant-year (numbers of observations are approximate per Census disclosure avoidance policy). Sample includes annual data from 1992 and 2002. Regressions include plant-specific fixed effects. Robust standard errors, clustered by plant, in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%. Census disclosure clearance date 6/14/2022.

Appendix Table A.6. Including Firm Structure and Sales as Controls

Models	(1) Includes Total Value Shipped as Control	(2) Adds More IT Variables to (1)	(3) Adds Firm Structure Controls to (1)	(4) Adds More IT Variables and Firm Structure Controls
Dependent Variable	ITR	ITR	ITR	ITR
External IT index	-0.059***	-0.061***	-0.057***	-0.058***
	(0.019)	(0.021)	(0.019)	(0.021)
Log inventory-to-sales ratio	-0.233***	-0.231***	-0.230***	-0.227***
	(0.060)	(0.060)	(0.060)	(0.060)
Skill mix	-0.021	-0.019	-0.022	-0.021
	(0.022)	(0.022)	(0.022)	(0.022)
Log downstream internal demand	0.003**	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Log downstream external demand	-0.005*	-0.005*	-0.005*	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
Has local competitors	0.026	0.027	0.026	0.026
	(0.018)	(0.018)	(0.018)	(0.018)
Local industry Herfindahl	0.027	0.027	0.028	0.028
	(0.029)	(0.029)	(0.029)	(0.029)
Number of products	0.002	0.002	0.002	0.002
	(0.005)	(0.006)	(0.005)	(0.006)
Total factor productivity	-0.019**	-0.020**	-0.019**	-0.019**
	(0.009)	(0.009)	(0.009)	(0.009)
Log number of firm units	0.013***	0.013***	0.014***	0.014***
	(0.003)	(0.003)	(0.003)	(0.003)
Internal IT index		0.017		0.016
		(0.012)		(0.012)
Log generic IT investment		-0.001		-0.001
		(0.001)		(0.001)
Production IT index		-0.021**		-0.022**
		(0.010)		(0.010)
Log of Total Value Shipped	-0.007	-0.006	-0.007	-0.006
	(0.007)	(0.007)	(0.007)	(0.007)
Changed firm ID			-0.028**	-0.028**
			(0.011)	(0.011)
Pct chg other firm ID			0.019	0.019
			(0.016)	(0.016)
Firm growth			-0.001*	-0.001*
			(0.000)	(0.000)
Observations	11,500	11,500	11,500	11,500
R-squared	0.019	0.020	0.021	0.022
R-squared w/ fixed effects	0.712	0.712	0.712	0.713

Notes: Unit of observation is a plant-year (numbers of observations are approximate per Census disclosure avoidance policy). Sample includes annual data from 1992 and 2002. Regressions include plant-specific fixed effects. Robust standard errors, clustered by plant, in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%. Census disclosure clearance date 6/14/2022.

Appendix Table A.7 Expanded Establishment Sample and Firm-Level Analysis

Dependent Variable:	ITR			
Models	(1) Extended Establishment Sample	(2) Extended Establishment Sample, Interaction with Pre-Sample ITR Restriction	(3) Firm Sample	(4) Firm Sample, Interaction with Pre-Sample ITR Restriction
External IT index	-0.013 (0.012)	-0.005 (0.017)	0.001 (0.022)	0.029 (0.026)
Indicator for pre-sample ITR restriction		0.046*** (0.005)		0.051*** (0.012)
External IT index × Indicator for pre-sample ITR restriction		-0.061** (0.024)		-0.120** (0.056)
Total effect of External IT index in ITR-restricted plant/firm		-0.067*** (0.017)		-0.091* (0.047)
Share of firm manufacturing employment			0.006 (0.009)	0.005 (0.009)
Observations	20,000	20,000	2,200	2,200
R-squared	0.040	0.045	0.031	0.038

Notes for columns 1-2: Unit of observation is a plant (numbers of observations are approximate per Census disclosure avoidance policy). Dependent variable is the internal transaction ratio (ITR). Includes indicator variable for whether establishment was in original baseline sample with pre-sample ITR restrictions. Sample includes data from 2002. Controls are the same as those in column 3 of Table 2.

Notes for columns 3-4: Unit of observation is a firm (numbers of observations are approximate per Census disclosure avoidance policy). Firm variables are constructed using weighted averages (based on dollar value of shipments) of plant-level observations in columns 1-2. Dependent variable is the internal transaction ratio (ITR). Includes variable that measures share of sales coming from establishments that were in original baseline sample with pre-sample ITR restrictions. Sample includes data from 2002. Controls are firm-level weighted averages of those in column 3 of Table 3, plus an additional control for the share of manufacturing employment in the firm accounted for by our sample.Robust standard errors, clustered by plant, in parentheses. *significant at 10%; **significant at 5%; *** significant at 1%.

Census disclosure clearance date 6/14/2022.

APPENDIX B: DATA CONSTRUCTION DETAILS AND DISCUSSION OF ROBUSTNESS RESULTS

B.1 Plant-Level Analysis Data Construction

External IT. To construct indices for external and internal uses of the internet, we first identify the number of types of information shared with each of customers, suppliers, or internal units. We add the condition that the plant rely on internet technology.[1] Then, we sum and normalize all of the types of information shared with either external customers or suppliers (a maximum of 7 each) to create an external IT index. We include both upstream and downstream information sharing in our core measure because, empirically, plants across the external IT distribution demonstrate a *mix* of customer- and supplier-focused information exchange. They are highly correlated. Moreover, dimension-reduction efforts related to the CNUS in other work (McElheran and Riggs 2010) emphasizes that upstream and downstream exchange for a given type of information (e.g., demand projections) tend to cluster together, generating a sharper distinction between *types* of information-sharing than between *directions* of information-sharing. This further makes sense in light of prior operations management research indicating the importance of both upstream and downstream coordination in supply chains to reduce inefficiencies and shortages (e.g. Lee et al. 1997). If upstream and downstream coordination are complements, as this clustering and prior research suggest, then examining them separately could bias our findings. That said, we check the robustness of our findings to only focusing on customer-facing coordination.

Internal IT. Similarly, information sharing with internal company units is used to create an internal IT index (up to 7 uses). Internal IT use has been found to be a separate cluster of practices (McElheran and Riggs 2010), though there is a non-trivial correlation between external and internal IT coordination (Appendix Table A.1). This may reflect some greater value of adopting information-sharing with external

[1] We add the internet presence as a key condition not only to identify the IT treatment, but also because older networked technologies for cross-establishment and cross-firm interaction such as EDI may require significant relationship-specific investment in private networks; thus adoption was often driven by a customer with market power sufficient to both demand the investment and dictate the standards governing data transfer. This could dramatically impact the firm organization decision according to the logic of the TCE literature.

partners once an organization has already adopted internal information-sharing (e.g., Barki and Pinsonneault 2005).

Production IT. This index relies on reported use of computer networked business processes for: integrated computer-aided design or computer-aided engineering (CAD/CAE), "design of the production process," production scheduling, production monitoring, product testing and acceptance, order fulfillment, order tracking, and transportation and shipping. We do not restrict on whether the network used is internal or external to the firm for this measure, as many of these are embedded in shop-floor technology. The mean value is 0.49, or roughly four of the eight applications that we consider.[2]

Other Controls:

- **Inventory:** To control for the potential presence of excess buffer stock at the focal plant, which might be sold off without reflecting core changes in governance, we include the value of finished goods inventories within the plant at the beginning of the year. This variable is normalized by the total value of sales, in log terms.[3]
- **Skill mix:** Following prior work using the CNUS (e.g., Atrostic and Nguyen 2005 and Fichman and Melville 2014), we control for the skill mix of workers at the plant with the ratio of production to nonproduction worker wages.
- **Process Complexity:** We take the log of the (reported) total number of products produced by the plant in case the propensity to vertically integrate also varies with complexity (Forbes and Lederman 2009).
- **Internal Demand:** To construct measures of internal downstream demand, we first identify the set of downstream plants within the same firm. For each plant in this set, we weight its total sales (measured in dollars) by the industry-level percentage of inputs required by the focal plant's

[2] The CNUS covers other production- and logistics-related technologies, such as automated warehouse technologies, not all of which we use here for reasons including low adoption rates, high item-non-response, or other quality concerns that would reduce the size of our analysis sample without adding insight. Our results are robust to various ways of constructing this index.
[3] This is winsorized at the 99% level to address some extreme outliers.

industry. This weighting percentage comes from the Detailed Use Tables of the BEA's 2002 Benchmark Input/Output tables. We then sum the value of this variable across all related internal units to estimate the downstream internal demand for the plant's output. Because this variable requires the complete universe of establishments belonging to the firms in our sample, it can only be constructed using Census years – another reason to focus on 1992 and 2002 as endpoints in the analysis.

- **External Demand:** We similarly compute a proxy for the demand for the plant's output in the external market. To do this, we first identify all of the establishments in the U.S. market with a substantial vertical linkage using the methodology described above. As with downstream internal demand, we multiply the value of sales for these local "linked" plants by their industries' percentage of inputs required from the focal plant's primary industry using the Benchmark Input/Output tables. We then sum these values across all vertically-related plants in the U.S. manufacturing sector.
- **Competition:** we include two controls for the presence of competition in the plant's primary industry (e.g., Ray et al. 2009). We do this because local competition may be correlated with the propensity to engage in process innovation (e.g., Kretschmer et al. 2012), which might thus endogenously be related to IT use. First, we include an indicator of whether there are any competitors in the same three-digit NAICS industry code in the same county. Second, using the total value of shipments for each plant, we compute a plant-level Herfindahl index for the three-digit NAICS industry in the plant's county.
- **TFP:** This is essentially the residual of a three-factor log-linear production function controlling for capital, labor, and material inputs, where capital stocks are accounted for and deflated following Cooper, Haltiwanger, and Power (1999). It is now routinely calculated by Census and added to the manufacturing survey data after a lag.
- **Firm Composition:** This is the logged count of all sites belonging to the firm (from the

Longitudinal Business Database (LBD)).

Other IT

- **EDI:** We do not observe variance in adoption of EDI over time, and so examine heterogeneity in the effects of external IT based on cross-sectional variance in whether a plant has EDI in 1999.
- **ERP:** We do not observe variance in adoption of ERP over time, and so examine heterogeneity in the effects of external IT based on cross-sectional variance in whether a plant has ERP in 1999.

B.2. Detailed Discussion of Robustness

Timing Falsification

Because instrumental variables estimations alone are rarely dispositive, in Appendix Table A.4 we further probe the robustness of our baseline results. First, we subject our findings to a timing falsification exercise. While we do not observe the precise year of external IT adoption, the internet's sudden deployment nevertheless provides a useful opportunity to see if the benefits of external IT have the "right" timing. If our assumptions about the nature of unobservables in our model are correct, we should not see any affiliation between external IT adoption and within-firm transfers prior to 1995.[4] To test this hypothesis, we use the ASM yearly data panel to rerun a version of regression equation (1) (in the paper) over the period 1992-2002. The regression equation takes the following form:

$$ITR_{it} = \alpha + \lambda_1 ExternalIT_{it} + \lambda_2 EarlyExternalIT_{it} + \delta Z_{it} + \mu_i + \tau_t + \eta_{it}$$

where $ExternalIT_{it}$ is equal to zero prior to 1995 and 1 thereafter for plants that adopt, and $EarlyExternalIT_{it}$ is set equal to one during the period 1993-1994 among plants that adopt and zero, otherwise. Z_{it} includes the same controls for inventories, skill mix, productivity, and number of products as in Table 2, column 3. We likewise include measures of downstream firm demand, U.S.-wide demand for the plant's outputs, and local competition; however, as these are constructed using the Census of Manufacturers, they change only in Economic Census years of 1992, 1997, 2002.[5]

[4] In this sense, our test is similar to that employed by Forman, et al. (2012) in their examination of the effects of business internet investment on local wage growth.
[5] We experimented with different ways of interpolating these data for missing years and the results are robust.

Column 1 of Table A.4 shows the results of estimating the equation above excluding $EarlyExternalIT_{it}$, to facilitate a comparison with earlier results. The qualitative effects of external IT are similar to those in column 3 of Table 2, if somewhat smaller in absolute value, possibly reflecting the increased measurement error from assuming that external IT was adopted by 1995 among all plants reporting adoption by the 1999 CNUS survey. Column 2 shows the results of estimating the equation above. Here, adoption of external IT is associated with a statistically significant (at the 1% level) decline in the ITR from 1995 on but has no statistically significant effect on internal transfers prior to that period. Further, a t-test of the difference between the effects of external IT before and after 1995 is statistically significant at the 5% level. In columns 3 and 4 we re-estimate both models with 3-digit NAICS linear time trends, producing similar results. Finding that the organizational response did not precede or anticipate the adoption of the IT supports a causal interpretation of the evidence.

Nonlinear Models

To examine the robustness of our results to nonlinear alternatives, we estimate fractional response models (Papke and Wooldridge 1996, 2008), both with and without a control function approach that incorporates the intuition of instrumental variables (Papke and Wooldridge 2008). We also estimate one-sided and two-sided tobit models.

We estimate fractional response models (Papke and Wooldridge 1996, 2008),[6] which can be advantageous for continuous response variables that are bounded and exhibit positive probability mass at one or both corners as our data do. Compared to a tobit model (below), such fractional response models have been shown to provide additional information on the effect sizes at the corner of the distribution for bounded response variables, and have improved fit (Gallani et al. 2015). Column 5 shows the results a fractional probit model using a pooled QMLE approach.[7] We present estimates of the average partial

[6] We adapted the Stata files provided by Leslie Papke at http://www.econ.msu.edu/faculty/papke/.

[7] While fractional response models were originally developed for cross-sectional data (Papke and Wooldridge 1996), they were later extended to control for unobserved heterogeneity in a panel data setting in the fractional probit model of Papke and Wooldridge (2008). They use the approach of Mundlak (1978) and Chamberlain (1980) to model the time-constant unobserved effect conditional on the strictly exogenous variables.

effects (APE) for all nonlinear models, which are estimated by computing the partial effects for every observation and then averaging these across all observations in the sample. We use the bootstrap to obtain standard errors (500 replications). Column 5 shows that the APE is -0.063, so the point estimate of the derivative arrived at in this fashion is similar to that in column 3 of Table 2. Column 6 combines this with the intuition of instrumental variables (IV) estimation in a control function approach (Papke and Wooldridge 2008).[8] It shows an APE of -0.263, significant at the 1% level. This is similar to the IV results in Table 3.

Another alternative to linear models is tobit model estimation.[9] For robustness we report in Columns 7 and 8 of Table A.4 the results of one-sided (for log internal transfers) and two-sided (for ITR) tobit models.[10] The APE in Column 7, in which the dependent variable is log of internal transfers, is significant at the 1% level and comparable to results in column 5 of Table 2. Column 8, in which the dependent variable is ITR, reports an APE of -0.052, similar to the OLS coefficients in Table 2, and also significant at the 1% level. We infer from these tests that the straightforward OLS coefficients provide a reasonable estimate of the causal impact of external IT on the vertical organization of transaction flows in U.S. manufacturing firms.

B.3. Firm-Level Data Construction and Analysis

Our primary establishment-level analysis allows us to isolate the effects of external IT on transaction flows at the establishment level. (Note that in the discussion below we will use the terms "plant" and "establishment" interchangeably. However, our establishment-level data are embedded within a broader environment. By construction, our sample consists of establishments that exist across both years of our

[8] Here, we estimate a linear reduced form model regressing the endogenous variable (external IT index) on our instruments (exclusion restrictions), controls, and averaging the controls within establishment over our time period. The residuals from this reduced form equation are then entered as an additional term in the estimation of the pooled QMLE approach used in column 5. We compute APE and bootstrap standard errors as before. For further details, see Papke and Wooldridge (2008).

[9] While tobit is appropriate when the dependent variable is bounded from either above or below, it produces results that are logically inconsistent when observations "pile up" at only one of the two corners (Gallani et al. 2015). This is an issue in our setting, where there is a large mass of observations at zero.

[10] As in the fractional probit models, we control for time-invariant unobserved heterogeneity across establishments over our time period with a correlated random effects approach (Wooldridge 2013).

data, have received the long survey of the Annual Survey of Manufacturers (ASM) in 1992 and 2002 (which allows us to observe internal plant transfers), and who also received the Computer and Network Use (CNUS) survey in 1999. As a result of this approach, we do not observe activity from all establishments within a firm. Thus, our primary analysis does not seek to estimate the aggregate impact of external IT across the entire firm. This is consistent with other work that has examined plural selling strategies (Atalay et al 2014).

In this section we provide some details on the construction of our firm level data. There are several caveats to the construction of these data. First, as already suggested above, while the US Census data represents the best source available for establishment manufacturing activity, we are unable to observe key variables like interplant transfers and external IT for all establishments in the population. For example, we could observe key variables like interplant transfers in zero, one, or both years of our baseline sample, depending upon what years the establishment is surveyed by the Census. Further, the composition of firms change over time, as establishments enter, exit, and change firm association. While these changes indeed represent changes to the boundaries of the firm and the firm-level organization of economic activity, a two-period panel that incorporated such changes would unable to isolate the effects of external IT on transaction flows from other concerns related to asset ownership.

To summarize, we do not observe all establishments within the firm. Further, for those establishments we do observe, entry and exit makes it difficult to identify the effects of external IT on transaction flows at the firm level using our baseline two-period model. Given these issues, we pursue a different approach, and use cross-sectional variance in our 2002 data to measure the effects of external IT across a broader set of establishments. We proceed in two steps, first creating an establishment-level sample from a broader set of establishments in 2002, and then aggregating these establishments to the firm level.

Construction of expanded establishment-level sample and results

To construct our revised establishment-level sample, we start with the set of establishments from our baseline sample in Table 2 and identify all other establishments from the same firm in 2002 for which we observe interplant transfers (from the long survey of the 2002 ASM) and for which we observe external

IT using the 1999 CNUS data. There are roughly 20,000 such establishments. The dependent variable and RHS variables are all constructed as in our baseline regressions. We then construct cross-sectional regressions of ITR on the same variables as in Table 2. Importantly, these models do not include establishment-level fixed effects that will control for time-invariant establishment-level factors that could be correlated both with external IT and ITR. As a result, they must be viewed with some care. We compare these estimates with our baseline two-period model to assess the potential biases that this approach may create.

Appendix Table A.7 provides results from this expanded establishment-level sample; we include only key parameter estimates in the tables (complete set of results available upon request). Column 1 shows the results of regressing ITR on external IT in a cross section using all of the controls in column 3 of Table 2. While the coefficient estimate is negative, the magnitude is much smaller than that of column 3 of Table 2 but the standard error is similar in size as that prior estimate, and so the estimate is not statistically significant. It is worth comparing these results to those of column 1 of Table 4. In both cases the number of establishments have been expanded to include those that were not involved in plural selling prior to the start of the sample period, though in this case the number of such establishments has been almost doubled (approximately 11,000 establishments in column 1 of Table 4, approximately 20,000 in this table). Once again, we show that when estimating our model over a broader set of establishments – some of whom for which the plural sourcing decision may be less relevant – the marginal effect of external IT on ITR is smaller, and in this case statistically insignificant.

As we did in column 2 of Table 4, we interacted external IT with a binary variable indicating that the plant was in our baseline sample, i.e. those plants that had been engaged in plural selling prior to our sample. This variable is negative and statistically significant, and similar in magnitude to that in column 3 of Table 2. When we compute the linear combination of external IT and the interaction of external IT and this pre-sample restriction, the point estimate is negative and statistically significant at the 1% level, and in magnitude slightly more negative than our baseline estimates (-0.067 compared to -0.060 in column 3 of Table 2).

Construction of firm-level sample and results

To our knowledge, there is no well-accepted theory of how to aggregate establishment-level data to the firm level. Any such aggregation should have several characteristics, however. First, it should allow comparison with other models as closely as possible. Thus, firm-level models linking external IT, firm characteristics, and transaction flows should use variables from establishment-level models. Further, the contribution of large establishments to firm level aggregates should be larger than that of small establishments. This requires some care in our aggregation, because many of our variables are nonlinear transformations such as logs or ratios that do not immediately preserve scale through arithmetic sums or averaging.

We aggregate our establishment-level data to the firm level by starting with the establishment-level cross-sectional data described above. For each firm, we construct the total dollar value of shipments represented by the establishments in this sample as a measure of firm-level manufacturing output. We then compute output shares based upon the share of (dollar value of) shipments represented by this establishment relative to the (constructed) firm total described in the previous sentence. For each variable in the baseline model of column 3 of Table 2, we sum across all of the establishments in the firm, weighting by these shares of output. By construction these output shares will sum to one, allowing us to compute weighted averages for each of our variables.

We note two things about our procedure. First, because establishment-level ITR is itself the total value of internal firm shipments divided by the total value of plant shipments, the weighted average of ITR is simply the total value of interplant transfers in our (constructed) firm divided by the total value of shipments in our (constructed) firm. That is, the weighted average becomes simply the firm-level ratio.

Second, in addition to the variables from our baseline model, in our firm-level regressions we further control for the share of total manufacturing employment within the "true" firm. To do this, we compute total employment from the "true" firm (i.e., using all establishments linked to the firm using the Census concordance in the Longitudinal Business Database (LBD)).[11] We also compute the total employment represented by our constructed firm sample, and compute the fraction of total employment represented by establishments in our constructed firm. We also control for the share of total firm employment (not only manufacturing employment) as a robustness.

Columns 3 and 4 of Appendix Table A.7 provide the results of our firm level models. Column 3 shows that in this sample, the coefficient estimate on firm-level external IT is almost zero (0.001) and not statistically significant. This reflects the implications of aggregating different types of establishments – those that are engaged in plural selling and those which are not – to the firm level.

In column 4 we construct a share-weighted measure of whether the plants in the firm are in our baseline sample, that is, engaged in plural selling. We then interact this variable with external IT index. The interaction of external IT index with this measure of interconnectedness is negative and statistically significant (-0.120, statistically significant at the 5% level). The linear combination of external IT index and its interaction with the firm interconnectedness is equal to -0.091 (statistically significant at the 10% level). Interpreted literally, this means that when the firm is completely composed of plural selling establishments, the coefficient estimate on external IT index is -0.091.

ADDITIONAL REFERENCES

Atalay, E, A Hortacsu, C Syverson. 2014. Vertical integration and input flows. *Am. Econ. Rev*. **104**(4) 1120-48.

Atrostic, BK, Nguyen SV. 2005. IT and productivity in U.S. Manufacturing: Do computer networks matter? *Econom. Inquiry* 43(3): 493-506.

Barki, H, A Pinsonneault. 2005. A model of organizational integration, implementation effort, and performance. *Org. Sci.***16**(2) 165-179.

Chamberlain, G. 1980. Analysis of variance with qualitative data. *Rev. Economic Studies* **47** 225-238.

Cooper, R, J Haltiwanger, L Power. 1999. Machine replacement and the business cycle: Lumps and bumps. *Amer. Econom. Rev.* **89**(4) 921-946.

Fichman, RG and Melville NP. 2014. How posture-profile misalignment in it innovation diminishes

[11] We note that we are able to compute this measure using establishments who are not in our sample because the Census of Manufacturers and LBD data contain data on employment for all manufacturing plants in the population, however these data sources do not contain information for other key variables in our models.)

returns: Conceptual development and empirical demonstration. *J. Mgmt. Info.Systems*. **31**(1): 203-240.

Forbes, S, Lederman M. 2009. Adaptation and vertical integration in the airline industry. *American Economic Review* **99**(5) 1831-49.

Forman, C, A Goldfarb, S Greenstein. 2012. The internet and local wages: A puzzle. *Amer. Econ. Rev.* **102**(1) 556-575.

Gallani, S, R Krishnan, JM. Wooldridge. 2015. Applications of Fractional Response Model to the Study of Bounded Dependent Variables in Accounting Research. Working Paper 16-016, Harvard Business School.

Kretschmer, T, EJ Miravete, JC Pernías. 2012. Competitive pressure and the adoption of complementary innovations. *American Economic Review* **102**(4) 1540-70.

Lee, HL, V Padmanabhan, S Whang. 1997. Information distortion in the supply chain: The bullwhip effect. *Mgmt. Sci.* **43**(4) 546-558.

McElheran K. and L. Riggs. 2010. The role of IT in firm scope choice: Diversification or specialization? International Industrial Organization Conference, May 15. UBC.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica* **46**: 69-85.

Papke, LE, JM Wooldridge. 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *J. Applied Econometrics* **11** 619-632.

Papke, LE, JM Wooldridge. 2008. Panel data methods for fractional response variables with an application to test pass rates. *J. Econometrics* **145** 121-133.

Ray, G, D Wu, and P Konana. 2009. Competitive environment and the relationship between IT and vertical integration. *Info. Systems Research.* **20**(4) 585-603.

Wooldridge, JM. 2013. Correlated Random Effects Panel Data Models. Presentation, IZA Summer School in Labor Economics.