# Internet Appendix for

# "Innovation Activities and Integration through Vertical Acquisitions"

(*Not for publication*)

Laurent Frésard, Gerard Hoberg, and Gordon Phillips

December 20, 2018

This appendix contains material briefly discussed in the paper, but reported here to conserve space. Section  lists words from the BEA commodity vocabulary that we exclude because they are used in a large number of commodities. Section  lists the phrase exclusions from firm 10-Ks that we apply to construct vertical links between firms. Section  provides validation tests for our text-based measure of firm-pair vertical relatedness and firm-level vertical integration. Finally, Section  reports additional tests that assess the robustness of our main results (reported in Table VII of the paper).

# I Excluded BEA words

Because they are used in a large number of commodities and are not specific, we exclude the following words from the BEA commodity vocabulary we use to compute vertical relatedness:

Accessories, accessory, air, airs, attachment, attachments, commercial, commercials, component, components, consumer, consumers, development, developments, equipment, exempt, expense, expenses, ga, gas, industrial, industrials, net, part, parts, processing, product, products, purchased, purchase, receipt, receipts, research, researches, sale, sales, service, services, system, systems, unit, units, work, works, tax, taxes, oil, repair, repairs, aids, aid, air, apparatuses, apparatus, applications, application, assemblies, assembly, attachments, attachment, automatic, auxiliary, bars, bar, bases, base, blocks, block, bodies, body, bulk, business, businesses, byproducts, byproduct, cares, care, centers, center, collections, collection, combinations, combination, commercials, commercial, completes, complete, components, component, consumers, consumer, consumption, contracts, contract, controls, control, covers, cover, customs, custom, customers, customer, cuts, cut, developments, development, directly, distributions, distribution, domestic, dries, dry, equipments, equipment, establishments, establishment, exempt, expenses, expense, facilities, facility, fees, fee, fields, field, finished, finish, finishings, finishing, gas, generals, general, greater, hands, hand, handling, high, hot, individuals, individual, industrials, industrial, industries, industry, installations, installation, lights, light, lines, line, maintenances, maintenance, managements, management, manmade, manufactured, manufacture, materials, material, naturals, natural, nets, net, offices, office, only, open, operated, operate, organizations, organization, others, other, pads, pad, paid, pay, parts, part, permanent, portable, powers, power, processing, products, product, productions, production, public, purchased, purchase, purposes, purpose, receipts, receipt, reclassified, reclassify, repairs, repair, researches, research, sales, sale, self, services, service, sets, set, shares, share, shipped, similar, singles, single, sizes, size, small, soft, specials, special, stocks, stock, storages, storage, supplies, supply, supports, support, surfaces, surface, systems, system, taxes, tax, taxable, technical, this, trades, trade, transfers, transfer, types, type, units, unit, used, without, work, works.

# II 10-K Phrase Exclusions

Because we use 10-K text to identify a firm's own-product market location (vertically related vocabulary is identified using BEA data), we exclude any part of a sentence that follows any of the following 81 phrases:

Buy, buys, sells its, are sold, buying, products for, for sale, for their, used in, used by, used as, used for, used with, used primarily, used mainly, used commonly, primarily used, mainly used, commonly used, for use, uses, utilized, serve, serving, serves, sold to, sold primarily, sold mainly, sold commonly, designed for, supply of, supply for, supplier to, supplied to, service to, purchase, purchaser, purchasers, customer, customers, user, users, for application, equipment for, equipment to, equipment by, product for, product to, product by, solution for, solution to, solution by, component for, component to, component by, application for, application to, application by, system for, system to, system by, equipments for, equipment for, equipment to, equipments to, equipments by, products for, products to, products by, solutions for, solutions to, solutions by, components for, components to, components by, applications for, applications to, applications by, systems for, systems to, systems by.

# III Further Validation Tests

We present four analyses to provide additional external validation for the text-based measures of vertical relatedness between firms and vertical integration within firms introduced and used in the paper. First, we present a rigid test based on firms' sensitivity to trade credit shocks discussed in the paper (Section III.E). Second, we compare our measure of vertical integration to (industry) measures of related-party trade (RPT). Third, we present the correlation between our measure of vertical integration and firms' mentions of vertical integration in their 10-Ks. Fourth, we investigate whether and how our text-based measure of firm-level vertical integration and firm-pair vertical relatedness changes around vertical and non-vertical acquisitions.

# A Correlation of Trade Credit shocks

First, as presented in the paper, we construct a test of the extent to which any proposed vertical relatedness network is vertical based on the extent to which accounts receivable ($AR$) and accounts payable ($AP$) respond to shocks in a way that is consistent with adjacency along a supply chain (as opposed to being consistent with horizontal links). Intuitively, our test is based on how firms that are related vertically versus horizontally should respond to trade-credit shocks. Firm pairs that are vertically related will experience *negatively* correlated shocks in accounts payable versus accounts receivable due to their supply chain adjacency. For example, a shock to an upstream industry's *receivables* should be associated with a similar shock to the downstream industry's *payables*. In contrast, firms that are horizontally related should experience trade-credit shocks in either accounts payable or accounts receivable that are positively correlated. We define trade credit as accounts payable minus accounts receivable for each firm. We then regress changes in trade credit of upstream firms on the changes in the trade credit of downstream firms.

To operationalize these predictions in our setting, we consider trade-credit shocks among firm pairs. When $AR$ increases for a supplier, one should expect an adjacent increase in the $AP$ of its customers. We first compute for each firm-year $\triangle AR$ as $\frac{AR_t - AR_{t-1}}{AR_t + AR_{t-1}}$ and $\triangle AP$ as $\frac{AP_t - AP_{t-1}}{AP_t + AP_{t-1}}$.[1] Critical to our examination, we then compute the difference $(\triangle AR - \triangle AP)$. To measure firm pairwise trade credit correlations for a given network, we estimate the following regression, where one observation is one firm-pair that is a member of a given network:

$$(\Delta AR - \Delta AP)_{i,t} = \alpha + \gamma \cdot (\Delta AR - \Delta AP)_{j,t} + \eta_t + \epsilon_{i,t}. \quad (1)$$

The subscript $i$ corresponds to an upstream firm and $j$ to a downstream firm indicated by the given network being tested. We account for time variation in aggregate trade credit shocks (e.g. macroeconomic conditions) by including year fixed effects ($\eta$). In more refined tests, we then focus on sub-samples of firm-pair observations where (1) $| \triangle AR_{i,t}| > | \triangle AP_{i,t}|$, or (2) $| \triangle AR_{j,t}| < | \triangle AP_{j,t}|$). The former condition focuses on positive shocks to the $AR$ of upstream firms, while the latter focuses on positive shocks to

[1] By construction, $\triangle AR$ and $\triangle AP$ can take values between +1 and -1 and are thus not influenced by outliers.

the $AP$ of downstream firms. The prediction is that the coefficient $\gamma$ should be positive for horizontal networks, and negative for vertical networks.

The results in Table IA.1 show that $\gamma$ is systematically negative for the vertical networks we construct. However, the estimates of $\gamma$ are far more negative, and are also statistically different from zero, only for our text-based networks. Not surprisingly, results are strongest of all for the text-1% network (the $t$-statistic ranges from 2.40 to 4.57), where the likelihood of contamination due to breadth is minimized. None of the estimates of $\gamma$ are significant for the NAICS-based vertical network, and the coefficient estimates are an order of magnitude smaller. In the last column we see that the estimates of $\gamma$ for the TNIC-3 horizontal network are significantly positive, as is predicted for horizontal relationships. The results of these tests show that horizontally related firms experience positively correlated responses in accounts payable and accounts receivable, whereas vertically-related firm pairs experience negatively correlated responses. These results provide a strong validation test of our new measure of vertical linkages.

## B Related-Party Trade

As an alternative way of to provide external validation, we relate our text-based measure of vertical integration to industry measures of related-party trade (RPT) provided by the U.S. Census Bureau.[2] The data measure the intensity of trade (both imports or exports) that occurs between related parties, where "related party trade" is defined as trade with an entity located outside the United States in which the importer (exporter) holds at least a 6% (10%) equity interest (as defined by the Census). The data thus captures the intensity of international transactions that occur within firm boundaries. Arguably, related party trade could capture both horizontal and vertical flows of goods. Yet, to the extent that our text-based measure of vertical integration builds on vertical relations between products described in firm 10Ks, any correlation between our measure and RPT should be related to international transactions that are vertical in nature (see Antras (2013), or Antras and Chor (2013) for instance).

The RPT data is available over the 2000-2013 sample period at the NAICS 6-digit level. We aggregate the data to the NAICS 4-digit and 5-digit levels to map it to our

[2] http://sasweb.ssd.census.gov/relatedparty/

Compustat sample. For each industry, we compute the share of related-party imports to total imports to capture the propensity of firms to integrate foreign supplier activities (RPT(import)). Similarly, we compute the share of related-party exports in total exports to capture the propensity of firms to integrate foreign customers (RPT(export)). We also consider the average share between the import and export shares (RPT). We then aggregate $VI$ and $VI_{segment}$ at the industry-level (NAICS 4-digit and 5-digit levels) using equally-weighted averages.

Table IA.2 presents the results of OLS regressions of industry-level $VI$ (or $VI_{segment}$) on the three measures of related-party trade. Across all specifications, we observe a positive correlation between our text-based measure vertical integration and measures of RPT. Focusing on the average level of RPT in the first column, the correlation with $VI$ is 0.490 at the NAICS 4-digit industry level, and 0.791 at the NAICS 5-digit industry level. Both coefficients are statistically significant at the 5% confidence level. At both aggregation levels, our measure of vertical integration is also more strongly related to related-party import transactions compared to related-party export transactions (columns (2) and (3)). The coefficients on related-party import are 0.508 and 0.626, and they are 0.116 and 0.547 for related-party export. Moreover, columns (4) to (6) indicate that related-party trade is negatively and only weakly related to vertical integration when measured using Compustat segments and the NAICS-10% vertical network as an alternative.

## C Mentions of Vertical Integration

We also consider a direct validation test for our measure of vertical integration. We do so by searching for the terms 'vertical integration' and 'vertically integrated' in each firm's 10-K (excluding cases where the firm indicates it is not integrated or lacks integration). We define the dummy variable $VI_{10k}$ to be one when a firm indicates that it is vertically integrated in a given year, and zero otherwise. Because this measure is direct and does not rely on the BEA input-output matrix, it is a strong validation test.

Table IA.3 presents results from probit regressions estimating the probability that a firm explicitly indicates that it is vertically integrated ($VI_{10k} = 1$) as a function of our text-based measure of vertical integration ($VI$) and the existing NAICS-based measure ($VI_{segment}$). We standardize both independent variables prior to running this regression so that their coefficients can be compared directly. The first column indicates that our text-based measure's coefficient is more than three times larger than the Compustat NAICS-based measure's coefficient (0.217 versus 0.066). Statistical significance levels show the same pattern. The superior performance of $VI$ continues to hold when we include $VI$ and $VI_{segment}$ separately (columns (2) and (3)). Columns (4) to (6) reveal that these results are robust to including year and industry fixed effects.

## D Changes in Vertical Position following Acquisitions

Finally, we investigate whether and how our text-based measure of firm-level vertical integration and firm-pair vertical relatedness change following vertical and non-vertical acquisitions. We perform two tests. First, we regress (including firm and year fixed effects) our text-based ($VI$) and the Compustat-based measure of firm-level vertical integration ($VI_{segment}$), measured in year $t$, on binary variables indicating whether the firm made a vertical ($D(vertical) = 1$) or non-vertical ($D(nonvertical) = 1$) acquisition in year $t$, $t-1$ or $t-2$, identified using our text-based vertical network (as in the paper).

Panel A of Table IA.4 presents the results. For both measures, the intensity of vertical integration increases following vertical acquisitions, and decreases following non-vertical acquisitions. Yet, comparing coefficients in columns (1) and (2), we observe that the magnitude of the coefficients indicate that our text-based measure of integration is about two times more responsive to actual acquisitions compared to the COMPUSTAT-based measure. Panel B of Table IA.4 reports a similar analysis using the number of vertical peers for each firm-year as independent variable, computed as the number of pairs for a given firm in a vertical network. When using our text-based vertical networks (1% or 10% granularity), we estimate that the number of vertical peers is sensitive to acquisition events. The number of vertical peers measured using the NAICS-based networks, in contrast, is largely insensitive to acquisition events.

# IV Additional Results

This section contains additional analyses that are mentioned and described in the paper but were not reported there to preserve space. Specifically, this appendix includes:

- **Table IA.5**: Probit regressions whose dependent variable is the probability of being a target in a vertical acquisition. We identify vertical transactions using the NAICS-10% vertical network instead of our Text-10% vertical network (as in the paper).

- **Table IA.6**: We assess the potential role of multicollinearity. We estimate Probit regressions whose dependent variable is the probability of being a target in a vertical acquisition, identified using our Text-10% vertical network (as in the paper). We include firms' R&D and patenting intensity individually, and estimate the baseline specifications on sub-samples where the correlation between firms' R&D and patenting intensity is (artificially) low.

- **Table IA.7**: We assess the potential role of scaling firms' R&D and patents differently to compute intensities. We estimate Probit regressions whose dependent variable is the probability of being a target in a vertical acquisition, identified using our Text-10% vertical network (as in the paper). We measure firms' R&D and patenting intensity by scaling them using either sales or assets.

- **Table IA.8**: We assess the potential role of missing values of R&D expenses reported in Compustat. We estimate Probit regressions whose dependent variable is the probability of being a target in a vertical acquisition, identified using our Text-10% vertical network (as in the paper). We use different specifications to assess the role of missing reported R&D expenses.

- **Table IA.9**: We assess the potential role of the measurement of patenting intensity by considering patents' stock instead of flows. We estimate Probit regressions whose dependent variable is the probability of being a target in a vertical acquisition, identified using our Text-10% vertical network (as in the paper). We measure firms' patenting intensity using patent stocks (using different definitions) as opposed to patent flows as in the paper.

- **Table IA.10**: We present summary statistics of firms' R&D and patenting intensity across quartiles of firm-level vertical integration.

- **Table IA.11**: List of the 30 most vertically integrated firms in 2008 based on our firm-level measure of vertical integration.

- **Figure IA.1**: We present the distribution of estimated coefficients on firms R&D and patenting intensity and their respective $t$-statistics based on 1,000 estimations made on random samples of 3,000 firms using probit models in which the dependent variable is the probability of being a target in a vertical acquisition, identified using our Text-10% vertical network (as in the paper).

## Table IA.1: Correlation of Trade Credit Shocks

| Network: | Text-10% | Text-1% | NAICS-10% | NAICS-1% | TNIC |
|---|---|---|---|---|---|
| $\gamma$ (unconditional) | -0.0006$^c$ | -0.0024$^c$ | -0.0001 | -0.0001 | 0.0071$^c$ |
| ($t$-statistic) | (-3.37) | (4.57) | (-0.93) | (-0.03) | (15.91) |
| $\gamma$ (if $\mid \triangle AR_{i,t} \mid > \mid \triangle AP_{i,t} \mid$) | -0.0006$^b$ | -0.0030$^c$ | -0.0002 | -0.0007 | 0.0071$^c$ |
| ($t$-statistic) | (-2.40) | (-3.71) | (-0.90) | (-0.92) | (12.37) |
| $\gamma$ (if $\mid \triangle AR_{j,t} \mid < \mid \triangle AP_{j,t} \mid$) | -0.0006$^b$ | -0.0027$^c$ | -0.0001 | -0.0005 | 0.054$^c$ |
| ($t$-statistic) | (-2.47) | (-3.83) | (-0.051) | (-0.06) | (8.37) |

*Note*: This table displays characteristics of our new Text-based vertical network and the existing NAICS-based vertical network. Our sample is based on annual firm observations from 1996 to 2013. We consider five networks: Text-10% and Text-1% networks correspond to vertical networks based on textual analysis set at a 10% and respectively 1% granularity level, NAICS-1% and NAICS-5% correspond to vertical networks based in the 2002 BEA Input-Output Table with relatedness cutoffs of 1% and 5% respectively, and TNIC corresponds to the horizontal Text-based Network Industry Classification developed by Hoberg and Phillips (2016). The coefficient $\gamma$ is obtained from OLS regressions of trade credit shocks of upstream firms on trade credit shocks of downstream firms. We report $t$-statistic below the coefficients. Symbols $^a$, $^b$, and $^c$ indicate statistical significance at the 10%, 5%, and 1% confidence levels

## Table IA.2: VI and Related-Party Trade

| Dep. Variable: | $VI$ | | | $VI_{segment}$ | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A: NAICS 4-digit industries** | | | | | | |
| RPT | 0.490$^{b}$ | | | -0.037 | | |
| | (0.244) | | | (0.244) | | |
| RPT(import) | | 0.508$^{a}$ | | | 0.366$^{b}$ | |
| | | (0.185) | | | (0.185) | |
| RPT(export) | | | 0.116 | | | -0.816$^{c}$ |
| | | | (0.262) | | | (0.261) |
| #.Obs. | 820 | 820 | 820 | 820 | 820 | 820 |
| Pseudo $R^2$ | 0.005 | 0.013 | 0.001 | 0.001 | 0.016 | 0.004 |
| **Panel B: NAICS 5-digit industries** | | | | | | |
| RPT | 0.791$^{c}$ | | | -0.455$^{c}$ | | |
| | (0.171) | | | (0.172) | | |
| RPT(import) | | 0.626$^{c}$ | | | -0.076 | |
| | | (0.130) | | | (0.131) | |
| RPT(export) | | | 0.547$^{c}$ | | | -0.847$^{c}$ |
| | | | (0.179) | | | (0.178) |
| #.Obs. | 1,422 | 1,422 | 1,422 | 1,422 | 1,422 | 1,422 |
| Pseudo $R^2$ | 0.014 | 0.015 | 0.006 | 0.004 | 0.001 | 0.014 |

*Note*: Columns (1) to (3) report OLS estimations where the dependent variable is our new text-based measure of vertical integration $VI$. Columns (4) to (6) report OLS estimations where the dependent variable is a measure of vertical integration based on Compustat segments $VI_{segment}$. In Panel A, all variables are aggregated at the NAICS 4-digit industry level (averages). In Panel B, all variables are aggregated at the NAICS 5-digit industry level (averages). The independent variables are standardized for convenience. Standard errors are clustered by industry and year and are reported in parentheses. Symbols $^{a}$, $^{b}$, and $^{c}$ indicate statistical significance at the 10%, 5%, and 1% confidence levels.

## Table IA.3: Validation: Vertical Integration Detection

| Dep. Variable: | Prob($VI_{10k} = 1$) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $VI$ | 0.217$^c$ | 0.229$^c$ | | 0.125$^c$ | 0.133$^c$ | |
| | (0.007) | (0.007) | | (0.010) | (0.010) | |
| $VI_{segment}$ | 0.066$^c$ | | 0.101$^c$ | 0.053$^c$ | | 0.064$^c$ |
| | (0.007) | | (0.006) | (0.008) | | (0.008) |
| Year FE | No | No | No | Yes | Yes | Yes |
| Industry FE | No | No | No | Yes | Yes | Yes |
| #.Obs. | 51,012 | 51,012 | 51,012 | 51,012 | 51,012 | 51,012 |
| Pseudo $R^2$ | 0.038 | 0.035 | 0.008 | 0.131 | 0.130 | 0.126 |

*Note*: This table reports Probit estimations where the dependent variable is $VI_{10k}$, a dummy that equals one if a firm mentions being vertically integrated in its annual 10-K report, and zero otherwise. The independent variables are standardized for convenience. Standard errors are clustered by industry and year and are reported in parentheses. Symbols $^a$, $^b$, and $^c$ indicate statistical significance at the 10%, 5%, and 1% confidence levels.

## Table IA.4: Changes in Vertical Measures Following Acquisitions

| Dep. Variable: | VI | $VI_{segment}$ | # Vertical Peers | | | |
|---|---|---|---|---|---|---|
| Vertical Network: | | | Text-10% | Text-1% | NAICS-10% | NAICS-1% |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $D(vertical)_t$ | $0.120^c$ | $0.094^b$ | $0.090^c$ | $0.169^c$ | 0.010 | -0.007 |
| | (0.02) | (0.04) | (0.01) | (0.01) | (0.01) | (0.01) |
| $D(vertical)_{t-1}$ | $0.085^c$ | $0.096^c$ | $0.087^c$ | $0.121^c$ | 0.021 | -0.004 |
| | (0.02) | (0.03) | (0.01) | (0.01) | (0.01) | (0.01) |
| $D(vertical)_{t-2}$ | $0.066^c$ | $0.054^a$ | $0.064^c$ | $0.090^c$ | 0.012 | -0.010 |
| | (0.02) | (0.03) | (0.01) | (0.01) | (0.01) | (0.01) |
| $D(nonvertical)_t$ | $-0.087^c$ | -0.011 | $-0.457^c$ | $-0.105^c$ | 0.014 | -0.002 |
| | (0.01) | (0.03) | (0.01) | (0.01) | (0.01) | (0.01) |
| $D(nonvertical)_{t-1}$ | $-0.048^c$ | -0.027 | $-0.043^c$ | $-0.062^c$ | $0.018^a$ | -0.006 |
| | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) |
| $D(nonvertical)_{t-2}$ | $-0.043^c$ | $-0.034^a$ | $-0.042^c$ | $-0.050^a$ | 0.015 | -0.003 |
| | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| #Obs. | 44,461 | 44,461 | 44,461 | 44,461 | 44,461 | 44,461 |
| Adj. $R^2$ | 0.826 | 0.730 | 0.779 | 0.815 | 0.855 | 0.867 |

*Note*: This table presents results from OLS models in which the dependent variables are firm-level measures of vertical integration (columns (1) and (2)) and a firm's number of vertical peers (columns (3) to (6)). We consider our text-based measure of vertical integration ($VI$) in column (1) and the Compustat-based measure ($VI_{segment}$) in column (2). We compute the number of vertical peers for a given firm by counting its the number of vertical pairs in a given vertical network. The independent variables are binary variables indicating whether the firm made a vertical ($D(vertical) = 1$) or non-vertical ($D(nonvertical) = 1$) acquisition in year $t$, $t-1$ or $t-2$, identified using our text-based vertical network (as in the paper). All specifications include firm and year fixed effects. Standard errors are clustered by industry and year and are reported in parentheses. Symbols $^a$, $^b$, and $^c$ indicate statistical significance at the 10%, 5%, and 1% confidence levels.

Table IA.5: The Determinants of Vertical Acquisitions based on NAICS network

| Dependent Variable: | Prob(Vertical Target) | | | | | | |
|---|---|---|---|---|---|---|---|
| Specification: | Main | Main | Ind×Yr | lags | OLS | Text | Ind. |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| R&D/sales | -0.057[b] | -0.045 | -0.006 | -0.018 | -0.001[a] | 0.054[b] | -0.092[c] |
| | (0.02) | (0.03) | (0.03) | (0.03) | (0.00) | (0.02) | (0.03) |
| Patents/assets | 0.027[a] | 0.042[b] | 0.023 | 0.039[a] | 0.001[b] | -0.024 | 0.084[c] |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.00) | (0.03) | (0.02) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Ind×Year FE | No | No | Yes | No | No | No | No |
| #Obs. | 61,463 | 61,463 | 58,257 | 52,257 | 61,463 | 55,833 | 61,463 |
| Pseudo. $R^2$ | 0.015 | 0.129 | 0.150 | 0.127 | 0.033 | 0.123 | 0.133 |

*Note*: This table presents results from probit models in which the dependent variable is a dummy indicating whether the given firm is a target in a vertical transaction in a given year. Vertical transactions are identified using the Vertical NAICS-10% network. The first two columns are the baseline models without and with control variables. Column (3) includes industry × year fixed effect, where industries are defined using FIC-100 industries. Column (4) considers (one-year) lagged independent variables. Column (5) reports estimates from a linear probability model (OLS) instead of probit. Column (6) considers firms R&D and patenting intensities directly from 10-K mentions. In column (7), all independent variable are computed as industry (equally-weighted) averages, based on TNIC-3 industries. All independent variables are defined in the Appendix. The independent variables are standardized for convenience. All estimations include year fixed effects. Standard errors are clustered by FIC-300 industry and year and are reported in parentheses. Symbols [a], [b], and [c] indicate statistical significance at the 10%, 5%, and 1% confidence levels.

Table IA.6: The Determinants of Vertical Acquisitions - Multi-Collinearity?

| Dependent Variable: | Prob(Vertical Target) | | | | |
|---|---|---|---|---|---|
| Specification: | Main | Main | Mcol1 | Mcol2 | Mcol3 |
| | (1) | (2) | (3) | (4) | (5) |
| R&D/sales | -0.047$^b$ | | -0.136$^a$ | -0.129$^b$ | -0.148$^b$ |
| | (0.02) | | (0.07) | (0.06) | (0.06) |
| Patents/assets | | 0.101$^c$ | 0.090$^c$ | 0.111$^c$ | 0.109$^c$ |
| | | (0.01) | (0.02) | (0.02) | (0.02) |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| #Obs. | 61,463 | 61,463 | 33,087 | 39,233 | 42,595 |
| Pseudo. $R^2$ | 0.124 | 0.128 | 0.157 | 0.145 | 0.136 |
| corr(R&D,patent) | 0.32 | 0.32 | -0.08 | -0.05 | -0.02 |

*Note*: This table presents results from probit models in which the dependent variable is a dummy indicating whether the given firm is a target in a vertical transaction in a given year. Vertical transactions are identified using the Vertical Text-10% network. In column (1) and (2), we include firms' R&D and patenting intensity individually. Columns (3) to (5) consider subsamples created so that the correlation between industry R&D and patenting intensity is small. Every year, we independently assign observations into three, four, or five groups based on tercile, quartile, or quintile splits for firms R&D and patenting intensity. We then keep observations that are not assigned in similar groups (e.g. low tercile for R&D and high tercile for patenting). The las row report the correlation between firms R&D and patenting intensity for each subsample. All independent variables are defined in the Appendix. The independent variables are standardized for convenience. All estimations include year fixed effects. Standard errors are clustered by FIC-300 industry and year and are reported in parentheses. Symbols $^a$, $^b$, and $^c$ indicate statistical significance at the 10%, 5%, and 1% confidence levels.

Table IA.7: The Determinants of Vertical Acquisitions - Different Scaling of R&D and patents

| Dependent Variable: | **Prob(Vertical Target)** | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| R&D/sales | -0.092$^c$ | -0.049$^b$ | | |
| | (0.03) | (0.02) | | |
| Patents/assets | 0.112$^c$ | | 0.088$^c$ | |
| | (0.01) | | (0.01) | |
| Patents/sales | | 0.010$^c$ | | 0.011$^b$ |
| | | (0.00) | | (0.01) |
| R&D/assets | | | -0.052$^b$ | -0.096$^c$ |
| | | | (0.02) | (0.02) |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| #Obs. | 61,463 | 61,463 | 61,463 | 61,463 |
| Pseudo. $R^2$ | 0.129 | 0.124 | 0.128 | 0.125 |

*Note*: This table presents results from probit models in which the dependent variable is a dummy indicating whether the given firm is a target in a vertical transaction in a given year. Vertical transactions are identified using the Vertical Text-10% network. We measure R&D and patenting intensities using different scaling, by sales or assets. All estimations include control variables similar to our baseline model, defined in the Appendix of the paper. The independent variables are standardized for convenience. All estimations include year fixed effects. Standard errors are clustered by FIC-300 industry and year and are reported in parentheses. Symbols $^a$, $^b$, and $^c$ indicate statistical significance at the 10%, 5%, and 1% confidence levels.

Table IA.8: The Determinants of Vertical Acquisitions - Missing R&D

| Dependent Variable: | Prob(Vertical Target) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| R&D/sales | -0.067[a] | | -0.048[a] | -0.069[b] |
| | (0.03) | | (0.03) | (0.03) |
| Patents/assets | 0.096[c] | 0.137[c] | 0.109[c] | 0.099[c] |
| | (0.01) | (0.05) | (0.01) | (0.01) |
| Missing(R&D) | | | | -0.130[c] |
| | | | | (0.03) |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| #Obs. | 38,053 | 23,410 | 61,463 | 61,463 |
| Pseudo. $R^2$ | 0.159 | 0.089 | 0.128 | 0.132 |

*Note*: This table presents results from probit models in which the dependent variable is a dummy indicating whether the given firm is a target in a vertical transaction in a given year. Vertical transactions are identified using the Vertical Text-10% network. In column (1) we restrict the sample to include only firm-year observations for which R&D expenses are non-missing in Compustat. In column (2) we restrict the sample to include only firm-year observations for which R&D expenses are missing in Compustat. In column (3) we replace missing R&D by zeros instead of industry medians. In column (4) we replace missing R&D by zeros instead of industry medians and include a binary variable that equals one if R&D is missing. All estimations include control variables similar to our baseline model, defined in the Appendix of the paper. The independent variables are standardized for convenience. All estimations include year fixed effects. Standard errors are clustered by FIC-300 industry and year and are reported in parentheses. Symbols [a], [b], and [c] indicate statistical significance at the 10%, 5%, and 1% confidence levels.

## Table IA.9: The Determinants of Vertical Acquisitions - Patent Stock

| Dependent Variable: | Prob(Vertical Target) | | | |
|---|---|---|---|---|
| Annual depreciation rate: | 0% | 10% | 15% | 20% |
| | (1) | (2) | (3) | (4) |
| R&D/sales | -0.081$^c$ | -0.062$^b$ | -0.065$^c$ | -0.068$^c$ |
| | (0.03) | (0.03) | (0.03) | (0.03) |
| Patent Stock/assets | 0.108$^c$ | 0.078$^c$ | 0.085$^c$ | 0.089$^c$ |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| #Obs. | 61,463 | 61,463 | 61,463 | 61,463 |
| Pseudo. $R^2$ | 0.131 | 0.126 | 0.126 | 0.127 |

*Note*: This table presents results from probit models in which the dependent variable is a dummy indicating whether the given firm is a target in a vertical transaction in a given year. Vertical transactions are identified using the Vertical Text-10% network. Instead of our baseline of baseline measure of industry-level patenting intensity (based on annual flow of new patents) we use various measures of patent stock to measure patenting intensity. Patent stock for each firm-year observation is based on cumulative patent stocks, computed based on depreciation rates ranging from 0% to 20% per year (see Hall, Mairesse, and Mohnen (2010)), and then averaged by industry. All estimations include control variables similar to our baseline model, defined in the Appendix of the paper. The independent variables are standardized for convenience. All estimations include year fixed effects. Standard errors are clustered by FIC-300 industry and year and are reported in parentheses. Symbols $^a$, $^b$, and $^c$ indicate statistical significance at the 10%, 5%, and 1% confidence levels.

## Table IA.10: Averages by Quartiles of VI

| Variable | Quartile 1 (Low VI) | Quartile 2 | Quartile 3 | Quartile 4 (High VI) |
|---|---|---|---|---|
| VI | 0.002 | 0.006 | 0.011 | 0.028 |
| R&D/sales | 0.106 | 0.065 | 0.048 | 0.027 |
| #Patents/assets | 0.007 | 0.008 | 0.008 | 0.007 |
| log(1+#Patents) | 0.444 | 0.531 | 0.654 | 0.848 |

*Note*: This table displays averages by (annually sorted) quartiles based on text-based vertical integration ($VI$). The sample includes 61,463 observations. All variables are defined in the Appendix.

Table IA.11: Examples of Vertically Integrated firms: Top 30 in 2008

| Company | Rank | #Segments | $VI$ | Perc.($VI$) | Perc.($VI(Segment)$) |
|---|---|---|---|---|---|
| HANDY & HARMAN LTD | 1 | 5 | 0.091 | 1 | 0.969 |
| PARKER-HANNIFIN CORP | 2 | 2 | 0.079 | 0.999 | 0.000 |
| EATON CORP | 3 | 5 | 0.076 | 0.999 | 0.966 |
| EMERSON ELECTRIC CO | 4 | 6 | 0.074 | 0.999 | 0.991 |
| FRANKLIN ELECTRIC CO INC | 5 | 1 | 0.073 | 0.998 | 0.717 |
| COMMERCIAL VEHICLE GROUP INC | 6 | 1 | 0.069 | 0.998 | 0.000 |
| ROCKWOOD HOLDINGS INC | 7 | 5 | 0.069 | 0.997 | 0.959 |
| SCHNITZER STEEL INDS -CL A | 8 | 3 | 0.064 | 0.997 | 0.000 |
| LEGGETT & PLATT INC | 9 | 3 | 0.062 | 0.997 | 0.710 |
| DOVER CORP | 10 | 4 | 0.058 | 0.996 | 0.641 |
| SIFCO INDUSTRIES | 11 | 2 | 0.055 | 0.996 | 0.994 |
| MYERS INDUSTRIES INC | 12 | 1 | 0.053 | 0.996 | 0.000 |
| AMPCO-PITTSBURGH CORP | 13 | 2 | 0.053 | 0.995 | 0.681 |
| SONOCO PRODUCTS CO | 14 | 3 | 0.052 | 0.995 | 0.000 |
| LKQ CORP | 15 | 1 | 0.052 | 0.995 | 0.000 |
| P & F INDUSTRIES -CL A | 16 | 2 | 0.052 | 0.994 | 0.760 |
| BERKSHIRE HATHAWAY | 17 | 9 | 0.051 | 0.994 | 0.000 |
| PRECISION CASTPARTS CORP | 18 | 2 | 0.051 | 0.993 | 0.790 |
| MATTHEWS INTL CORP -CL A | 19 | 6 | 0.051 | 0.993 | 0.884 |
| RELIANCE STEEL & ALUMINUM CO | 20 | 1 | 0.050 | 0.993 | 0.000 |
| CARLISLE COS INC | 21 | 6 | 0.050 | 0.992 | 0.962 |
| UNVL STAINLESS & ALLOY PRODS | 22 | 1 | 0.050 | 0.992 | 0.000 |
| AMERICAN AXLE & MFG HOLDINGS | 23 | 1 | 0.049 | 0.992 | 0.000 |
| ENCORE WIRE CORP | 24 | 1 | 0.049 | 0.991 | 0.000 |
| HAWK CORP | 25 | 1 | 0.049 | 0.991 | 0.000 |
| KANSAS CITY SOUTHERN | 26 | 1 | 0.049 | 0.991 | 0.000 |
| AMERICAN ELECTRIC TECH INC | 27 | 3 | 0.049 | 0.990 | 0.885 |
| DREW INDUSTRIES INC | 28 | 1 | 0.049 | 0.990 | 0.000 |
| CHINA PRECISION STEEL INC | 29 | 1 | 0.048 | 0.989 | 0.000 |
| COLEMAN CABLE INC | 30 | 1 | 0.048 | 0.989 | 0.000 |

*Note*: The table displays the 30 most vertically integrated firms in 2008 based on our text-based measure of vertical integration ($VI$). The table also presents the number of Compustat segments, the $VI$ score, the firm's percentile $VI$ ranking, and the firm's percentile $VI(Segment)$ ranking.

Figure 1: Bootstrapped Models. This we performed a bootstrap analysis in which we re-estimate our baseline probit specification 1,000 times on sub-samples composed of 3,000 randomly selected firms. The dependent variable is a dummy indicating whether the given firm is a target in a vertical transaction in a given year. Vertical transactions are identified using the Vertical Text-10% network. All estimations include control variables similar to our baseline model, defined in the Appendix of the paper. The independent variables are standardized for convenience. All estimations include year fixed effects. Standard errors are clustered by FIC-300 industry and year. We present the distribution of the estimated coefficients on firms' R&D and patenting intensity, as well as the corresponding $t$-statistics.