Learning Production Process Heterogeneity: Implications of Machine Learning for Corporate M&A Decisions

Jongsub Lee*

Hayong Yun†

July 2024

*Seoul National University (SNU) Business School. E-mail: jongsub.lee@snu.ac.kr

†Eli Broad College of Business, Michigan State University. E-mail: yunhayon@msu.edu

The authors greatly appreciate helpful comments from discussions with Mei Danqing (discussant), Xin Deng (discussant), Jung Hoon Han, Gerard Hoberg, Hyoung-Goo Kang (discussant), Srinivasan Krishnamurthy (discussant), Gordon Phillips, Nagpurnanand Prabhala (discussant), Tao Shu (discussant), Chris James, Mike Ryngaert, Andy Naranjo, Charles Hadlock, Evan Dudley, Alejandro Lopez Lira, and Mauricio Praximer (discussant). We thank seminar and conference participants at University of Florida, Seoul National University (SNU) Business School, Asian Bureau of Finance and Economic Research (ABFER), China International Conference in Finance (CICF), Joint Conference with the Allied Korea Finance Associations (AKFA), Financial Management Association (FMA) Annual Meeting, Capital Market Research in the Era of AI Conference at HKUST, EUROFIDAI-ESSEC Paris December Finance Meeting, MSU AI Networking Event for valuable comments, and Gayoung Koo, Doyeon Kim, and Jaehyung Cho for data assistance. Lee gratefully acknowledges financial support from the Institute of Finance and Banking and the Institute of Management Research at Seoul National University.

Abstract

We introduce novel metrics to evaluate production process heterogeneity using both machine learning (ML) and traditional kernels. ML kernels, particularly through economically motivated transfer learning models, enhance M&A forecasting accuracy. A wider gap in firms' production processes predicts fewer M&As, lower success rates, reduced returns, diminished post-M&A growth, and increased divestiture. Dynamic learning among repeat acquirors mitigates adverse effects of production process dissimilarity on post-M&A growth. The adoption of Right-to-Work laws, reducing employees' bargaining power, significantly alleviates detrimental effects of heterogeneous production processes. Our findings underscore the pivotal role of technology heterogeneity in shaping integration synergy and firm boundary decisions.

JEL Classifications: G3, G34, L2, M1

Keywords: Machine learning, production process heterogeneity, merger and acquisitions (M&A's), integration synergy, firm boundaries

1 Introduction

Merger decisions have long been examined in corporate finance, and prior work offers several explanations for why firms undertake these transactions. One prevailing perspective is the neoclassical theory, which posits that mergers are orchestrated with the primary objective of maximizing shareholder value. Behavioral theories introduce nuanced dimensions, attributing corporate merger and acquisition (M&A) decisions to factors such as misvaluation of equities and CEO overconfidence.

Building upon the foundations laid by Jovanovic and Rousseau (2001, 2002), we begin from a neoclassical viewpoint of mergers, honing in on the synergy motive. According to this rationale, an acquiring company (hereafter referred to as an "acquiror") envisions that the combined entities will surpass their individual worth due to asset complementarity. This perspective, further influenced by industry shocks and cross-industry linkages, contends that structural changes in underlying production technologies play a pivotal role in shaping the expected value of synergy (Harford, 2005; Ahern and Harford, 2014).

Empirical evidence, such as the findings of Devos, Kadapakkam, and Krishnamurthy (2009), reveals that synergy constitutes a substantial percentage (10.03%) of the merged company's equity value. Unpacking this, operating synergy, which encompasses elements like cost reduction and margin improvement through economies of scale and pricing power, constitutes 8.38%, while tax savings account for the remaining 1.65%. Notably, the distribution of these gains varies with the overlap in production processes; operating synergies take precedence in focused mergers, while tax savings play a pivotal role in diversifying deals.

Recent advances in deep learning show that these models can approximate flexible, non-parametric functions. This has been useful in fields where researchers want to measure concepts that people recognize but find hard to describe formally. One example is artistic or musical style, where a model learns how an artist or composer transforms inputs into outputs. The resulting functions can then be compared to evaluate how similar two styles are. This illustrates that even a black-box, correlation-based approach can provide informative measures for objects that do not have clear analytical definitions.

A similar idea applies to economics. Management practices, organizational culture, and policy choices can be viewed as functions that map information into decisions. Prior work, including Bertrand and Schoar (2003), shows that individual managers systematically influence firm policies. Under a functional view, these patterns reflect differences in decision style. Quantifying such styles can help address counterfactual questions such as how a different central bank chair would have responded to a macroeconomic shock or how a firm might have evolved under a different CEO.

Our paper applies this functional perspective to production style, defined as the mapping from a firm's production inputs to its economic output. Deep learning provides a natural method for estimating these functions without strong parametric assumptions. Comparing the learned functions for acquirors and targets yields a measure of production-style distance that reflects technological, organizational, and cultural differences that matter for integration. This helps explain why production-process heterogeneity is difficult to measure with standard tools and why it plays a central role in M&A outcomes.

However, a challenge arises in quantifying the degree of heterogeneity in the production processes of two merging firms, considering factors like technological distances, cultural dissimilarity, and organizational capital heterogeneity—all of which are complex and latent by nature. This inherent heterogeneity is a critical factor in understanding the cost of integration and the resulting value of synergy (Sher, 2012; Christensen et al., 2011; Barbin, 2017). While existing studies have attempted to capture these differences using text-based product characterization (Hoberg and Phillips, 2010), we contribute to this discourse by employing novel machine learning (ML) techniques to quantify underlying production process heterogeneity. This endeavor aims to shed light on the intricate decision-making processes behind corporate M&A transactions and enhance our understanding of the synergy they bring to the involved entities.

Measuring the distance between two firms' production processes (or, more broadly, corporate value-creation kernels) poses a challenge due to the non-linear nature of these kernels, which are not simply vector representations of inputs. Existing methodologies, like the Hoberg-Phillips TNIC similarity score and other text-based analyses, utilize cosine similarity between word vectors of corporate disclosures to assess the divergence in the product focuses of two firms (Figure 1a). However, as depicted in Figure 1b, production processes are intricate functions of inputs, making it unclear how to accurately measure these non-linear kernel distances. Moreover, the literature lacks exploration into how such distances impact business integration decisions.

[Insert Figure 1 Here]

To address these challenges, we propose a two-step log-mean squared error (MSE) ratio approach designed to quantify the disparity between the production processes of acquirors and targets. In the first step (Figure 2a), we examine peer groups of firms to derive the production function of the acquiror ($f_A(\cdot)$) and the target ($f_T(\cdot)$). This involves analyzing the production inputs and outputs of peer firms to discover how a vector of neoclassical production inputs, x, correlates with the valuation output, y (Tobin's Q). The vector of production inputs encompasses total assets, capital expenditure, financial leverage, labor, asset tangibility, intangible assets, and advertising expenses.

While our approach is general, we focus on five specific kernels: Cobb-Douglas, linear, XGBoost, fully connected layer neural network, and transfer learning-based neural network, with the latter three being ML-based. Although deep learning uses typical observables as production inputs, the incorporation of weights, which indicate how to combine these inputs across multiple decision layers, enables us to capture the latency and complexity of the production decision-making process within a firm. Notably, transfer learning-based neural networks have yet to be explored in the finance literature. We use this novel deep learning technique to demonstrate the critical role of "economic fitting" in comprehending the strategic decision-making process within an organization.[1]

[Insert Figure 2 Here]

In the second step, as depicted in Figure 2b, we compute the distance in production decision processes between any pair of acquiror (A) and target (T), either at the industry level using industry-based peer groups or at the firm level using Hoberg-Phillips TNIC similarity score peers. We calculate the mean squared error ($||y_T - \hat{f}_A(x_T; w_A)||$) by optimizing the weights, w_A, for inputs using the acquiror's production function, $\hat{f}_A(.)$ and the target's input and output data (x_T and y_T, respectively).[2] This is then compared to the mean squared error using the target's own production function, $\hat{f}_T(.)$ under the newly optimized weights, w_T, estimated for the target's input and output data (i.e., $||y_T - \hat{f}_T(x_T; w_T)||$). The former, constrained optimization, employs the acquiror's production kernel with inputs from the target, introducing a loss of fit in the first measurement $||y_T - \hat{f}_{\boldsymbol{A}}(x_T; \boldsymbol{w_A})||$ compared to the latter, which is without such restriction, $||y_T - \hat{f}_T(x_T; w_T)||$.[3] A large loss of fit ratio could, therefore, quantify the degree of heterogeneity between the acquiror's and the target's underlying production processes.

For industry-level distances, we estimate the loss of fit ratio for all pairs of Hoberg-Phillips

[1] While both simple deep learning, utilizing fully connected layers, and transfer learning can aptly model production functions, transfer learning excels in this context. Unlike simple deep learning, transfer learning places greater emphasis on transforming inputs into intermediate production factors during the initial learning stage. This approach tends to capture more economically significant elements of a firm's decision-making process, making it notably more robust in empirically explaining merger outcomes compared to simple deep learning methods. For more technical details of our machine learning-based kernels, please refer to Section 3 and Appendix A.

[2] $||.||$ indicates L^2-norm.

[3] The weight in bold font indicates that we are fixing the parameters from the acquiror's production function estimation when learning the target's input-output relationship.

25 Fixed Industry Classifications (FIC) from 1988 to 2021.[4] Beyond inter-industry distances, we also gauge functional distances between acquiror and target firms using those with a high TNIC similarity score as their instantaneous peers. This novel concept of functional distance allows us to quantify production process heterogeneity across industries and firms for the first time in the literature.

Using this novel metric, we hypothesize that the greater the functional distance between the acquiror's and the target's production kernels, the less synergistic value will be created in M&As between them. Business integration is expected to be more costly and less synergetic when two organizations with distinct production kernels are combined into a single organization (Damodaran, 2005; Devos, Kadapakkam, and Krishnamurthy, 2009; Hoberg and Phillips, 2010; Deng, Kang, and Low, 2013). For instance, merging a labor-intensive firm with a capital-intensive target is likely to result in limited synergy, partly due to the high costs associated with adjusting the two distinct factors during integration. Even when firms share the same production factors, those with different weights on those factors tend to experience more expensive business integration than firms with identical production factors and weights.[5] In essence, both weights and factor structure heterogeneity significantly impact the expected value of synergy. We further argue that adjusting weights is less expensive than modifying factor structures. Altering weights might be feasible through the transfer of processing knowledge from the acquiror to the target, while addressing differences in factor structures necessitates substantial financial resources and, if required, additional investments and divestitures.

We test these hypotheses using comprehensive M&A transaction data from SDC Platinum covering the period from 1988 to 2021. Before delving into the impact of the functional

[4]Our industry-level results are also robust to an alternative industry classification, such as the Fama-French 12 industry codes. We report the results in the Internet Appendix, which is available upon request.

[5]For example, a capital-dependent firm with a positive weight on the capital factor may face high integration costs when merging with a firm with a negative weight on the same capital factor.

distance between acquiror-target production processes on M&A outcomes, we first assess the performance of each kernel in fitting the underlying production functions of the two firms. Our findings reveal that flexible ML-based kernels (XGBoost, fully connected layer neural network, and transfer learning-based neural network) outperform conventional Cobb-Douglas or linear kernels, as evidenced by the estimated mean squared error, at both the industry and firm levels.

Moving to industry-pair-level analysis, we investigate whether industries with substantial functional distance have a lower number of M&A transactions between them. Utilizing our five production kernel distances (Cobb-Douglas, linear, XGBoost, fully connected layer neural network, and transfer learning-based neural network) as the primary explanatory variables, we find robust support for the ML-based production kernels. Among them, the transfer learning-based neural network model demonstrates the most accurate predictions regarding the aggregate M&A activities between two industries. Specifically, a one standard deviation increase in the transfer learning-based kernel distance results in a 4.6% reduction in log(Number of M&A Deals) from its sample standard deviation, with statistical significance at the 1% level. These results hold consistently across pooled panel regressions and year-by-year cross-sectional regressions. In contrast, distances calculated using the linear kernel fail to explain the merger intensity between two industries.

In the deal-level analysis using firm-level distances, we construct instantaneous peer groups with high TNIC similarity scores for acquiring and target firms. We evaluate the completion rate of each deal using a dummy variable that indicates a successfully completed deal (Completed dummy) as our main dependent variable. We control for TNIC similarity score and various M&A and firm characteristics. We find that all our ML-based production process distances exhibit significantly negative coefficients, indicating that a greater functional distance is associated with a lower likelihood of successfully completed deals. Once again, the transfer learning-based neural network outperforms other kernels in predicting deal-level outcomes. The results persist across different specifications, including year and industry fixed effects, exclusion of partial mergers and divestitures, and even vertical mergers.[6]

We also explore M&A announcement returns, long-term survival rates, and divestitures for the combined organizations post-M&A. Employing specifications similar to Deng, Kang, and Low (2013), we find that a one standard deviation increase in transfer learning-based distance corresponds to a -14% poorer stock market reaction around the deal announcement date, with statistical significance at the 1% level. For long-term synergy effects, the survival rate of merged organizations over ten to fifteen years following deal completion decreases by 55% for a marginal increase in our transfer learning-based functional distance. These findings collectively suggest that heterogeneous production processes between the acquiror and the target are strongly associated with unfavorable merger outcomes in both short and long runs. This suggests elevated integration costs and diminished synergy in M&A transactions characterized by disparate production processes.

Finally, we explore dynamic learning effects among repeated acquirors. We identify an adverse impact of production process dissimilarity on firm growth—a trend that is mitigated for repeated acquirors. These acquirors tend to target firms with smaller transfer learning-based kernel distances, particularly following an incomplete prior deal marked by significant production process heterogeneity. Furthermore, we conduct supplementary quasi-natural experiments to refine the inclusion restrictions of our innovative distances utilizing the adoption of Right-to-Work laws, which law weaken employees' bargaining power against employers. We find that the exogenously reduced integration cost associated with the labor factor significantly mitigates the adverse effects of heterogeneous production processes on merger outcomes.[7]

[6] Complementary production processes between the acquiror and the target are often desired in vertical mergers.

[7] Our results also validate the findings of Deng, Kang, and Low (2013), who emphasize the role of orga-

We make several important contributions to the literature on machine learning and corporate finance. Notably, we introduce a conceptual framework that, to the best of our knowledge, marks the first attempt to quantify production process heterogeneity across industries and firms utilizing cutting-edge ML techniques. Production processes are inherently complex, involving not only technological aspects but also corporate culture, organizational capital, and other stakeholder-related characteristics (Sah and Stiglitz, 1984; Dessein, 2002; Dessein and Santos, 2006; Deng, Kang, and Low, 2013). These elements are latent by nature, making their estimation challenging. ML-based models, especially deep neural networks, are particularly suitable for unraveling the latent and intricate nature of the underlying production processes within a firm. Unlike ordinary least squares (OLS), which may introduce non-linearity through the inclusion of various polynomials, deep neural networks autonomously identify the most effective functional form of non-linear factors through multi-layer learning weights without relying on ex-ante parametric assumptions. This distinction enhances the model's capacity to capture the complexity and nuances of production processes, rendering it a valuable tool in quantifying production process heterogeneity.

We also provide economic intuitions behind the estimation outcomes. Jovanovic and Rousseau (2001, 2002) propose the Q-theory of mergers, which predicts that high-Q industries acquire low-Q industries due to a valuation advantage. Rhodes-Kropf and Robinson (2008) propose an alternative theory, like-buys-like, in which firms in similar valuation cohorts are more likely to be integrated due to asset complementarity. Another stream of studies in the literature focuses on the industry merger wave and cross-industry merger dynamics (Harford, 2005; Hoberg and Phillips, 2010, 2016; Hoberg, Phillips, and Prabhala, 2014; Ahern, 2012; Ahern and Harford, 2014). All these studies jointly emphasize the importance of industrial organization dynamics, product market characteristics, and their networks as the

nizational capital, such as corporate social responsibility (CSR), in smoothing business integration process post-mergers.

main drivers of the observed patterns in M&As. Using a novel metric, we quantify production process heterogeneity across industries, which turns out to be an important determinant of cross-industry M&A activities.

Our findings extend beyond the simple dichotomous merger classifications in the literature, such as focused versus diversifying mergers. We propose a novel proxy for merger synergy even within diversifying mergers, thereby meaningfully expanding the merger classification and synergy literature. While Devos, Kadapakkam, and Krishnamurthy (2009) highlight the tax benefits of diversifying mergers, we demonstrate that mergers between distinct industries could also exhibit significant operational synergy, contingent upon the overlap in underlying production processes between the merged companies. Our measures complement the text-based identification of firms' end product markets pioneered by Hoberg and Phillips (2010). We establish a close connection between our functional distances and their product classifications.[8]

Finally, we highlight the importance of "economic fitting" when applying the latest deep learning techniques to corporate finance data. Prior literature has applied various novel machine learning methods to advance our understanding in finance, such as corporate governance (Erel et al., 2021), venture capital (Bonelli, 2023; Hu and Ma, 2021; Lyonnet and Stern, 2022), corporate finance (Jha et al., 2024), term structure (Van Binsbergen, Han, and Lopez-Lira, 2023), and asset pricing (Gu, Kelly, and Xiu, 2020, 2021). Our findings indicate varying effectiveness among ML-based kernels, emphasizing the significance of an approach capable of embedding hierarchical decision-making processes within a firm, such as transfer learning-based neural networks. When we evaluate the similarity and dissimilarity between two firms' production processes, the ability to segment decision layers and identify transferable portions plays a crucial role in understanding the underlying production processes. This

[8] Hoberg and Phillips (2016) mention that SIC codes are created based on production processes. The approach reflected in the SIC codes can also be complemented by our machine-based classifications of underlying production processes.

underscores the importance of economically motivated model selection in the application of cutting-edge ML techniques to corporate finance.

2 Motivating Example

To illustrate the idea of our novel metric, we start with a common Cobb-Douglas type CES production kernel for firm A, as often employed in Q theory literature:

$$q_A = CK_A^{\alpha}L_A^{1-\alpha}, \tag{1}$$

where Tobin's Q (q_A) represents the output, and inputs are denoted as x_A=$\{K_A$,$L_A\}$, referring to firm A's capital (K_A) and labor (L_A). The weights associated with capital and labor factors are α and $1-\alpha$, respectively. The parameters to be estimated are encapsulated in $w_A = \{C, \alpha\}$.

Consider an acquiror in the manufacturing sector, where the stochastic production output relies solely on the capital input, expressed as $y_A = CK_A^1L_A^0 + \epsilon_A$, with $||\epsilon_A|| = 0$. Suppose that this firm acquires a target company in the service sector, where production output depends solely on the labor input, i.e., $y_T = CK_T^0L_T^1 + \epsilon_T$, with $||\epsilon_T|| = 0$. A viable approach to assessing the divergence between the acquiror and target's production processes is to examine how the target's output differs when the target's input is applied to the acquiror's capital-intensive production kernel, i.e., $||y_T - CK_T^1||$. We normalize this distance by a baseline scenario where the target's inputs are applied to its own production function, i.e., $||y_T - CL_T^1||$. Our production function distance is then defined as the natural logarithm of the following ratio:

$$d(y_A, y_T) = log\left(\frac{||y_T - CK_T^1||}{||y_T - CL_T^1||}\right) \gg 0. \tag{2}$$

This ratio yields a positive value because for the labor-intensive target, $||y_T - CL_T^1|| \approx 0$, while $||y_T - CK_T^1|| > 0$. Essentially, this metric adeptly encapsulates dissimilarity in both factor structures (i.e., K vs. L) and the associated weights with the factors $[(\alpha, 1-\alpha) = (1, 0)$ vs. $(0, 1)]$ between the acquiror and target's production technologies.

As we will demonstrate in the following sections, this Cobb-Douglas production kernel proves too rigid for effectively fitting empirical production outputs. Consequently, we explore more adaptable production kernels that better capture the complexities of real-world production processes. Although we consider a linear production function with additional inputs as one versatile form, our analysis reveals that even a linear function with more inputs falls short of adequately fitting the data. As a result, we turn to more flexible fitting algorithms commonly utilized in contemporary machine learning and deep neural network literature. Among these, we find that a transfer learning-based neural network rooted in economic principles performs exceptionally well in explaining the integration cost of a merger.

3 Construction of Distance Measure

We link an acquiror's input vector, x_A, to the valuation output, y_A, using the firm's production kernel, $f_A(\cdot)$, each year according to the following expression:

$$f_A(x_A; w_A) = \hat{y}_A, \tag{3}$$

where w_A represents the parameter vector used to tailor the production kernel, $f_A(\cdot)$, to the real-world data. We explore five production kernels: Cobb-Douglas, linear, XGBoost, and two others based on deep neural networks.

XGBoost operates as a tree-based bootstrap regression method, generating estimates by averaging results from multiple trees. As implied by its full name—eXtreme Gradient Boosting— it represents a refined version of the gradient boosting algorithm. It builds a series of decision trees sequentially, with each tree improving upon the errors of the last, thereby enhancing accuracy and efficiency through gradient descent. This method outperforms Random Forest, another ensemble approach, in terms of computational power and flexibility. Unlike Random Forest's independent tree construction, XGBoost's interconnected tree-building strategy often yields more accurate results for various datasets. Given its widespread application in the finance literature (Gu, Kelly, and Xiu, 2020; Erel et al., 2021), we opted for XGBoost as a benchmark against our primary method, deep neural networks.[9]

We explore two neural network-based kernels: one utilizes fully connected layer neural networks (Figure 3), while the other employs a novel weight-transfer mechanism, referred to as transfer learning-based neural networks (Figure 4). For the former fully connected layer deep learning, we implement five-layer neural networks ($L = 5$ in Figure 3). These networks consist of an input layer with 8 nodes (representing eight production inputs; see below) and ReLU activation, followed by three hidden layers with 100 nodes each and ReLU activation in the middle layer. The network concludes with an output layer featuring a single node under linear activation.

The production input, x_A, contains the following eight nodes: the natural logarithm of total assets, capital expenditure to total assets ratio, short-term debt divided by total assets, long-term debt divided by total assets, employment divided by total assets, property plant and equipment (PPENT) divided by total assets, advertisement expenses divided by

[9]Basic descriptions of this method are provided in Gu, Kelly, and Xiu (2020). For more in-depth insights into tree-based regressions, consult Chapter 15 of Hastie et al. (2009).

total assets, and R&D expenditure divided by total assets. When we use the Cobb-Douglas kernel, these inputs are log-transformed to align with the natural logarithm of Tobin's Q, the common production output, y_A, for all production kernels examined in this study.[10]

We remove industry-specific effects in each year by demeaning all the variables using their industry-year average values.[11] For instance, an input factor, x_{ijt}, of firm j in industry i in year t, is demeaned by $\bar{x}_{it} = \frac{1}{N_{it}} \sum_{k=1}^{N_{it}} x_{ijt}$, where N_{it} is the number of firms in industry i in year t. The adjustment is made as follows:

$$\tilde{x}_{ijt} = x_{ijt} - \bar{x}_{it}. \tag{4}$$

In the initial step of measuring distance, we estimate the production function of the acquiror based on its peer group. Two approaches are employed for forming peer groups to estimate the production function of the acquiror or target company. First, for inter-industry distance measurement, we use the 25 Hoberg-Phillips Fixed Industry Classifications (FIC) as a peer group.[12] In the second approach, for inter-firm distance measurement, we use instantaneous peer groups based on Hoberg-Phillips TNIC score similarity. Specifically, we select firms with a similarity score greater than 0.09 (approximately the 95th percentile in each year's full TNIC score data) as instantaneous peers for a firm of interest.

Given that conventional R^2 is not suitable for nonlinear regressions like XGBoost or deep learning, we rely on the mean squared error (MSE) to evaluate the output of the production

[10] When we replace the valuation output, Tobin's Q, with the operational profitability variable, such as return on assets (ROA), we obtain consistent results. The results are reported in the Internet Appendix. They are available upon request.

[11] We standardize all our independent and dependent variables using their industry average values for each year. Although consistent results are obtained using raw values, we focus on these deviations from the annual industry mean. This approach accounts for potential cross-industry shifts and shocks in industrial production characteristics, ensuring a more meaningful comparison across industries.

[12] In the Internet Appendix, we also consider Fama-French 12 industry classification as a peer group for measuring industry-to-industry distance; results are qualitatively similar to those reported in the main text. These results are available on request.

function ($\hat{y}_A$) and the actual observed production outcome (y_A):

$$MSE(\hat{y}_A) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{A,j} - \hat{y}_{A,j})^2}, \tag{5}$$

where N is the number of peers, in the case of inter-industry distance estimation, representing the firms in the acquiror's FIC 25 industry.

In each fitting, we search model parameters, w_A. For instance, when fitting a fully layered deep neural network kernel (see Figure 3), we aim to minimize the mean squared errors by solving:

$$\min_{w_A} \mathcal{L}_A(x_A; w_A) = \min_{w_A} MSE(\hat{y}_A) = \min_{w_A} \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{A,j} - \hat{y}_{A,j})^2}. \tag{6}$$

In the second step, we retain the resulting trained model parameters ($\boldsymbol{w_A}$), such as the neural weights of each layer, ($\boldsymbol{w_A}^{\boldsymbol{(0)}}\boldsymbol{, \ldots, w_A}^{\boldsymbol{(L)}}$), which fit the production functions of the firms in the acquiror's peer group (e.g., acquiror's FIC 25 industry peers in the case of inter-industry distance estimation, or acquiror's instantaneous peers based on TNIC similarity score for the inter-firm distance estimation). Once the acquiror's production function is estimated, we assess the goodness of fit for the target by computing the mean squared error for the target's production function. This is done while **fixing** its model parameters to be the same as those estimated for the acquiror's peer group:

$$MSE(y_T, x_T; \boldsymbol{w_A}) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{T,j} - \hat{y}_{T,j}(x_T))^2}, \tag{7}$$

where $y_{T,j}$ is the actual production outcome of firm j in the target industry, and $\hat{y}_{T,j}$ is the model estimate of firm j's production function using the trained weights from the acquiror's industry (w_A), i.e., $f_A(x_T; \boldsymbol{w_A}) = \hat{y}_T(x_T)$.

Finally, we define the distance between the acquiror and the target as the natural logarithm of the normalized mean squared error obtained above by the mean squared error of the target's peer group without the weight constraint from the acquiror's peer group:

$$d(y_A, y_T) = log\left(\frac{||y_T - \hat{f}_A(x_T; \boldsymbol{w_A})||}{||y_T - \hat{f}_T(x_T; w_T)||}\right) = log\left(\frac{MSE(y_T, x_T; \boldsymbol{w_A})}{MSE(y_T, x_T; w_T)}\right). \tag{8}$$

Figure 3 explains the process of estimating industry-to-industry distance using this fully connected layer neural network kernel.[13]

The aforementioned full-layer-based deep learning approach, although effective in capturing variations between acquirors' and targets' production functions, might be overly sensitive. This approach could exaggerate minor production process differences that are easily accommodated in practice. Neural network kernels, which are structured in layers, allow for the retention of a subset of neural weights from selected layers. The retained weights remain fixed, while new layers can be added downstream (adjacent to the output layer) and trained with additional data. This capability, known as transfer learning (Goodfellow, Bengio, and Courville, 2016), enables us to leverage the neural network's transfer-learning feature to identify only the significant differences in production kernels between the acquiror and target.

In practical terms, the distance is estimated using a similar approach, with the key distinction being that we utilize the trained peer group's weights from the acquiror's production function only up to the second-to-last layer of the deep neural network of the target's peer group. In simpler terms, we retain $\boldsymbol{w_A}^{(\boldsymbol{0})}, \ldots, \boldsymbol{w_A}^{(\boldsymbol{L-1})}$, while allowing the weight in the last layer, $w_T{}^{(L)}$, to be freely adjustable.

[13]Note that this estimation is based on out-of-sample predictions utilizing 80% of the data for training and 20% for testing. This process is repeated 10 times and the average distance value is used in the end. It is important to recognize that these distances are asymmetric, reflecting the perspective of acquirors who apply their production function to the target's production process. In simpler terms, a complex firm acquiring a target with a straightforward task may have a different industry distance than a scenario where a simple firm acquiring a complex target company. Our measure captures such differences.

Under this relaxed-weights constraint (see Figure 4), we minimize the mean squared errors of the production outputs of the target's peer group:

$$\min_{w_T{}^{(L)}} \mathcal{L}_T(x_T; \boldsymbol{w_A}^{(0)}, \ldots, \boldsymbol{w_A}^{(L-1)}, w_T^{(L)}) = \min_{w_T{}^{(L)}} \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{T,j} - \hat{y}_{TF,T,j})^2} \tag{9}$$

The resulting estimated production function is denoted as:

$$f_A(x_T; \boldsymbol{w_A^{(0)}}, \ldots, \boldsymbol{w_A^{(L-1)}}, w_T^{(L)}) = \hat{y}_{TF,T}. \tag{10}$$

We normalize the resulting mean squared error by the mean squared error of the production function of the target industry itself, without any weights restriction, denoted as $MSE(y_T, x_T; w_T{}^{(0)}, \ldots, w_T{}^{(L)})$. This normalization leads to the following transfer learning-based distance:

$$d_{\boldsymbol{TF}}(y_A, y_T) = log\left(\frac{MSE(y_T, x_T; \boldsymbol{w_A}^{(0)}, \ldots, \boldsymbol{w_A}^{(L-1)}, w_T{}^{(L)})}{MSE(y_T, x_T; w_T{}^{(0)}, \ldots, w_T{}^{(L)})}\right). \tag{11}$$

It is important to note that, while the transfer learning framework could allow for the addition of multiple layers to train the target's peer groups, this paper, for illustrative purposes, incorporates only a single layer, $w_T{}^{(L)}$, just before the output layer.[14] This specific setup depicts a situation in which only a minor modification at the final stage of the production decision is permitted. A significant transfer learning-based distance in production processes implies that a mere adjustment of the target's final production process is inadequate for smooth integration into the acquiror's production technology. Consequently, in such cases, the merging firms would display considerable heterogeneity in their production processes.[15]

[14] Additionally, while transfer learning is applicable to neural networks due to their layered structure, it is neither feasible nor practical for other types of kernels, such as linear, Cobb-Douglas, and XGBoost, which lack such layered networks.

[15] See Appendix B for a theoretical illustration of this intuition.

4 Simple Simulation

This section provides a simple numerical example that connects the closed-form two-factor linear model in Appendix B with a neural network implementation. The goal is to clarify why deep learning is well suited for measuring latent economic objects such as production distance. The value of deep learning in this setting is not only its ability to capture nonlinearities, but also its modular structure. In particular, transfer learning allows us to impose economically meaningful restrictions on which parts of the production process are easy to adapt and which are costly to change.

4.1 Problem Setup

We consider three stylized industries with simple nonlinear production technologies. The baseline industry is capital-intensive, with output given by $y_1 = K^3$. The second industry relies on the same underlying factor but differs in the final transformation, with output $y_2 = -K^3$. The third industry is based on a different production factor altogether, with output $y_3 = L^3$.

Economically, these cases correspond to three distinct acquisition scenarios. When an acquiring firm faces a target that shares the same production factor but differs in a minor way, integration should be relatively inexpensive. Retooling a capital-intensive plant to reverse the sign of output is costly but feasible. By contrast, acquiring a firm whose production depends on a different factor requires rebuilding core organizational and technological structures, implying much higher integration costs.

A neural network trained from scratch does not distinguish between these cases. It treats K^3 and $-K^3$ as fundamentally different functions and therefore assigns a large production distance between y_1 and y_2. In contrast, a transfer learning design freezes the layers that map inputs into latent production factors and re-estimates only the final output layer. Under this restriction, K^3 and $-K^3$ are recognized as sharing the same underlying technology, with differences that can be resolved through limited reweighting at the final stage. At the same time, the distance between K^3 and $-K^3$ remains large, reflecting the need to replace the factor structure itself.

This mirrors the economic logic in Appendix B. Adjusting weights at the final stage of production is relatively cheap, while changing the underlying factor structure is costly. The example anticipates our MSE-ratio distance and clarifies why the transfer learning kernel provides a tighter link between statistical fit and economic integration cost.

4.2 Network Structure

We generate synthetic production data for each industry using noisy nonlinear processes. Specifically, we draw 1,000 observations per industry according to $y_1 = K^3 + 0.05\varepsilon$, $y_2 = -K^3 + 0.05\varepsilon$, and $y_3 = L^3 + 0.05\varepsilon$, where $\varepsilon \sim N(0,1)$. Capital and labor are independent inputs.

The production function is approximated using a fully connected neural network with six layers. The input layer has two nodes for capital and labor, followed by four hidden layers with 8, 8, 8 and 4 nodes, respectively. ReLU activations are used in all hidden layers and the output layer is linear.

We compare two training approaches that correspond to different integration assumptions. In the vanilla approach, the network is trained on data from the capital-intensive industry and then directly applied to data from the target industry. The resulting mean squared error captures the loss from imposing the acquiror's full production process on the target.

In the transfer learning approach, the network is first trained on the capital-intensive industry in the same way. When evaluating fit on the target industry, we retain the trained weights in the first five layers and allow only the final layer to be re-estimated. This restriction reflects an integration process in which latent production factors are preserved, while the final mapping from factors to output is adjusted. As a result, the model can accommodate sign reversals within the same factor structure, but not a shift from capital-based to labor-based production.

4.3 Result of Learning

4.3.1 Production Process Differences: Fully Connected Method

Figure 5 illustrates how the vanilla neural network measures production process differences across industries using the mean squared error (MSE) between actual and predicted output.

Panel A considers the benchmark case in which both the acquiror and the target are capital-intensive industries. The figure (a) shows the true production data for the K-intensive industry, while the figure (b) shows the predictions of the model trained in the same industry. The predicted output closely matches the true data, and (c) shows that the resulting MSE is close to zero. This confirms that when the acquiror and target share the same production process, the estimated production distance is small, as expected.

Panel B considers a more subtle case in which the acquiror and target rely on the same production factor but differ in the sign of output. The figure (a) shows the true production data for the inverted K-intensive industry, while (b) shows predictions from the model trained on the original K-intensive industry. The model continues to generate a positive cubic relationship in capital, which does not match the inverted structure of the target. As a result, the figure (c) shows a large MSE, indicating a substantial production distance despite the shared underlying factor.

Panel C examines the case in which the acquiror and target rely on different production factors. The figure (a) shows that output in the L-intensive industry depends on labor rather than capital. Because the model was trained using only capital variation, its predictions in (b) do not meaningfully track the true production process and largely reflect arbitrary extrapolation.

Consequently, the figure (c) shows a large production distance. The appendix confirms that this result is robust across repeated simulations and reflects structural mismatch rather than noise.

Taken together, these figures show that the vanilla neural network is highly sensitive to both sign changes and factor changes. It treats minor differences within the same factor structure and fundamental changes in production technology in a similar way, motivating the transfer learning approach introduced next.

[Insert Figure 5 Here]

4.3.2 Production Process Differences: Transfer Learning

Figure 6 reports results from the transfer learning approach, which allows limited adjustment of the production process when moving from the acquiror to the target industry. Unlike the vanilla model, transfer learning permits adaptation only at the final stage of production, while preserving the underlying factor structure learned from the acquiror.

Panel A considers the case in which the acquiror is capital-intensive and the target is an inverted capital-intensive industry. The figure (a) shows the true production data for the target, while (b) shows predictions from the transfer learning model. The predicted output closely matches the true inverted cubic relationship. As a result, (c) shows that the mean squared error is close to zero. This indicates a small production distance, consistent with the fact that the two industries share the same production factor and differ only in a minor way that can be accommodated through limited reweighting.

Panel B considers a more fundamental difference in production technology, where the acquiror is capital-intensive and the target is labor-intensive. The figure (b) shows that the transfer learning model fails to capture the target's production process, and (c) shows a large prediction error. In this case, allowing adjustment only at the final layer is insufficient because the identity of the dominant production factor differs across industries. Together, these figures illustrate a key distinction between changes in factor weights and changes in factor identity. Transfer learning can accommodate sign reversals within the same factor structure, but it cannot bridge differences that require rebuilding the underlying production technology. This behavior aligns closely with economic intuition about integration costs in mergers.

To better understand how these differences arise, we next examine how production inputs are transformed into latent representations across the layers of the network.

[Insert Figure 6 here]

4.4 Process of Learning: Flow of Layers

[Insert Figure 7 Here]

Figure 7 illustrates how production differences are progressively encoded as inputs propagate through the inherited layers of the network. We begin with the Vanilla Method (Fully Connected Neural Network) trained on the capital-intensive industry and evaluate transfer learning as we vary the depth of the inherited (frozen) representation while restricting adjustment to the final aggregation layer.

Each row corresponds to a different number of inherited hidden layers (H1–H4). Passing the target data through a deeper set of fixed layers applies progressively richer nonlinear transformations of the original inputs before the final layer is re-estimated. Within each row, the left column reports the Vanilla fit for the capital-intensive case, while the right columns report transfer-learning fits for the sign-reversal case ($-K^3$) and the factor-switch case (L^3), respectively.

With only a shallow inherited representation (H1), the model yields a coarse piecewise-linear approximation with limited segmentation of the input space. Economically, this corresponds to capturing an average marginal relationship that summarizes how capital maps into production. When applied to the inverted capital-intensive case, the final layer can partially compensate by reversing the sign, but the fit remains coarse because the inherited representation provides only limited nonlinear structure.

As the inherited depth increases (H2 and H3), the network partitions the input space into more regions, each with its own local slope. In economic terms, deeper inherited layers allow the model to approximate production as a sequence of regimes, where different ranges of capital intensity are governed by different marginal responses. This produces additional kinks in the fitted function and substantially improves fit for the sign-reversal case. Although the marginal effect flips sign, the underlying factor remains capital, so the model can realign local slopes without changing factor identity.

By contrast, the labor-intensive case remains poorly fitted across depths when only the final aggregation layer is re-estimated. Even with the deepest inherited representation (H4), predictions remain anchored on capital-based latent features. Additional depth increases the number of nonlinear segments, but these segments are still constructed from the wrong underlying factor. As a result, greater flexibility along the capital dimension does not translate into meaningful improvements when output depends primarily on labor.

Overall, Figure 7 clarifies the distinction between factor weights and factor identity. Transfer learning succeeds when integration requires reweighting within a shared factor structure (e.g., sign reversals in a capital-intensive technology), but it fails when integration requires replacing the factor structure itself (capital to labor). This mechanism underlies why the transfer learning–based distance captures economically meaningful integration costs that can be obscured by fully flexible neural network distances.

5 Data

5.1 Sample Construction

Our sample consists of the annual panel of U.S. firms in 625 Hoberg-Phillips 25 Fixed Industry Classification (FIC) pairs from 1988 to 2021. There are a total of 21,250 industry pairs in this database. Production process distances are constructed using the financial information from the Compustat annual database for all U.S. firms in our sample.

We require a firm to have total assets greater than $10 million. We also require no missing information for our production input and output variables, including total assets, Tobin's Q (total shares outstanding, year-end stock price, market value of common equity, deferred taxes), book leverage (debt in current liabilities, long-term debt), capital expenditure/assets, labor (number of employees), asset tangibility (PPENT), intangible assets (advertising expenses, R&D expenses), and free cash flow.[16] This data requirement results in 162,415 firm-year observations from annual Compustat, which we use for training and testing production functions.

To inspect the relationship between our industry-level functional distances and Hoberg and Phillips (2010, 2016)'s firm rival scores, we merge the average TNIC scores (full data version) for each of our 625 Hoberg-Phillips 25 FIC pairs. TNIC scores are at the firm level, and thus, we compute the average scores for each of the 25 FIC pair in each year.[17]

We obtain deal-level M&A information from the SDC Platinum database and merge it with acquirors' and targets' firm financial information from Compustat. For inclusion in our sample, deals have to have the following information: whether acquiror and target were in a high technology industry, whether the merger was hostile, whether it used tender

[16]We use free cash flow as a control variable in our regression analyses rather than a production input in our kernel estimation.

[17]TNIC scores are available from 1988 to 2021 (21,250 observations). We obtain the scores from Hoberg and Phillips's data library (https://hobergphillips.tuck.dartmouth.edu/).

offers, whether the deal was financed partly by stocks, deal size, and acquiror's equity value. We require the transaction value to be greater than \$100 million.[18] We exclude partial acquisitions and divestitures. These are all standard data filters employed in the literature.

The most restrictive filter requires the SDC Platinum data to match with full TNIC scores (for which both acquiror and target need to have GVKEYs, i.e., both should be publicly traded firms). This filter results in 785 deals from 1988 to 2021 for the main sample with the \$100 million transaction value filter.[19] Later in our tests, we also consider a subsample that restricts the sample to horizontal mergers whose acquiror-target pairs have vertscores (Vertical TNIC data) smaller than 0.02.

5.2 Variable Description

Our main explanatory variables are production function distances using Cobb-Douglas, linear, XGBoost, and two neural network kernels (fully-connected and transfer-learning types), denoted as CD Distance, Linear Distance, XGB Distance, FC Distance, and TF Distance, respectively. These distances are computed at two distinct levels: the acquiror-target industry level (utilizing the Hoberg-Phillips 25 Fixed Industry Classifications) and the acquiror-target firm level (employing firms with a high Hoberg-Phillips TNIC similarity score as instantaneous peers).

As a proxy for M&A intensity between a pair of Hoberg-Phillips 25 FIC pairs (i.e., 625 pairs in each year), we compute the Number of M&A Deals between each Hoberg-Phillips 25 FIC pair reported in the SDC Platinum data. Log(Number of M&A Deals) is the natural logarithm of this variable. For the M&A deal-level analyses, we construct an indicator for M&A deal completion (Completed) and withdrawal (Withdrawn). As controls in our deal-level analyses, we consider the following deal and acquiror/target characteristics widely used

[18] We also conduct tests for deals with values exceeding \$10 million and obtain robust results.

[19] For the transaction value greater than \$10 million sample, we are left with 922 deals.

in the literature (see Deng, Kang, and Low, 2013, among many others). Diversify is an indicator for a diversifying merger (the acquiror's 25 FIC differs from that of the target's). Hostile is an indicator for a hostile merger, which takes a value of one if SDC Platinum records the deal as a hostile merger and zero otherwise. High Tech is an indicator that takes a value of one if both acquiror and target are in high technology sectors and zero otherwise (i.e., SDC Platinum records them as "Primary Business not Hi-Tech"). Tender Offer is an indicator for tender offer deals. Stock Deal is an indicator for deals that are (partly) financed by stock according to the report in the SDC database. Relative Deal Size is the deal value reported in the SDC normalized by the acquiror's market capitalization.

Additional firm fundamentals compiled from Compustat are defined as follows. Firm size is the natural logarithm of total assets, log(Asset). Tobin's Q is the ratio of total assets plus the market value of equity (total shares outstanding times year-end stock price) minus the total value of common equity minus deferred taxes to total assets. Book Leverage is debt in current liabilities plus total long-term debt divided by total assets. Cash Flow to Assets is operating income before depreciation minus total interest-related expenses minus total income taxes minus capital expenditures divided by total assets.

To study the announcement effect, we compute acquiror-target equal-weighted and value-weighted cumulative abnormal returns for three event windows, CAR[t-1,t], CAR[t-1,t+1], and CAR[t-1,t+2], respectively. For post-merger real effects, we use a post-merger acquiror survival indicator, Survive Within 10 Years, which equals one if the acquiror exists and is not acquired as a target in the SDC database within 10 years after the merger. We also consider a Divestiture Within 10 Years dummy, which is one for a firm that appears as a target for divestiture within 10 years after the merger, and Asset Growth [t+1,t+10], the log difference in total assets between one and 10 years after the merger. Market value growth [t+1,t+10] is the log difference in market capitalization between one and 10 years after the merger. Prior Deals Within 10 Years, Prior Completed Deals Within 10 Years, and Prior

Failed Deals Within 10 Years are, respectively, the number of deals, completed deals, and failed deals within the 10 years before the current announcement date. Most Recent Deal Within 10 Years is an indicator of the latest merger within the past 10 years. Most Recent Deal Within 10 Years: Completed (Most Recent Deal Within 10 Years: Incomplete) are subindicators for the latest complete (incomplete) merger transaction within the past 10 years.

5.3 Summary Statistics

Figure 8 illustrates the fitting performance, represented by MSE for each kernel. We evaluate how effectively each kernel fits the production process of each Hoberg-Phillips 25 FIC in each year. In Figure 8a, we show the log(MSE) of the Cobb-Douglas kernel compared to the deep neural network; the Cobb-Douglas kernel's values are much larger than those of the neural network. Specifically, the log(MSE) of the Cobb-Douglas kernel typically ranges from -2 to 2, whereas that of the neural network ranges from -7 to -4. Despite the frequent use of structural models like Cobb-Douglas CES kernels in the theoretical literature, it appears less flexible in fitting the empirical data than ML-based kernels.

[Insert Figure 8 Here]

Figure 8b compares the log(MSE) of the linear kernel to the neural network kernel. While the linear kernel's fitting performance is better than that of Cobb-Douglas except for a couple of outliers, it still falls significantly short of the neural network's performance (i.e., the difference of log(MSE) greater than 25% of the neural network's log(MSE)) in 5% of the industry-years. In contrast, the neural network's log(MSE) performs significantly worse than the linear kernel in fewer than 0.5% of industry-years.

In Figure 8c, we compare the log(MSE) of XGBoost to the neural network. The log(MSE) values for both kernels are in a comparable range, indicating that both XGBoost and neural networks fit production functions similarly well.

Based on the fitting performance of the production function in Figure 8, we conclude that ML-based kernels (XGBoost and neural network kernels) better capture the difference in production processes between two industries than other traditional kernels.

[Insert Table 1 Here]

Table 1 reports the summary statistics of the key variables used in this paper. In the top panel, which shows the mean, standard deviation, and 1st/50th/99th percentile values of the Hoberg-Phillips 25 FIC pairs by year (625 pairs each year), we find that the means of acquiror-target industry-level production process MSE ratios are typically close to one (though, due to random initial training conditions, the ratio can be smaller than one), but exhibit substantial variation.[20] Production function distances, i.e., the log-transformed MSE ratios, are typically close to zero for all five kernels with substantial variation. The annual 25 FIC pair average TNIC score has a mean value of 0.012 with a standard deviation of 0.014. The mean of the log number of M&A deals is 0.033 with a standard deviation of 0.191.

The bottom panel shows the same statistics at the firm level.[21] The panel also reports the mean value of the firm-level TNIC score (0.189) with its standard deviation (0.109). About 24 pecent of our sample comprises diversifying mergers (acquiror and target have different FIC), while 34 percent are in the high-tech sector, and 4.5 percent are hostile mergers. Acquirors tend to be larger (mean log assets are 8.194) than targets (mean log

[20] In untabulated results, the MSE ratio of the fully connected layer-based neural network kernel, for example, has a mean value of 1.196 with a standard deviation of 0.382, while the 99th percentile value is 2.594. The MSE ratio for the transfer learning neural network kernel is lower because it does not take into account small differences in the production processes of the acquiror and the target. This means that the MSE ratio is usually close to one and only gets big when the differences in the production processes cannot be fixed at the neural network's last layer. The mean value is 0.904 with a standard deviation of 0.085, but the 99th percentile is 1.132.

[21] The firm-level production process MSE ratios have mean values closer to one than industry-level MSE ratios. For example, the mean value of neural network-based MSE ratios is 1.046 for the fully connected layer model, and 1.009 for the transfer learning-based model. However, there is substantial variation in MSE ratios ranging from 0.124 (transfer learning-based neural network) to 4.788 (linear kernel).

assets are 6.625) but have similar book leverage (mean value is 0.248 for acquiror and 0.252 for target). Cumulative Abnormal Returns have means ranging from 1.8 percent (t-1 to t) to 2.5 percent (t-1 to t+1) with substantial variation (standard deviations are 0.064 for t-1 to t event window and 0.078 for the t-1 to t+2 event window). The 10-year survival indicator has a mean value of 0.079 with a large standard deviation (0.269). The indicator for divestiture within the past 10 years has a mean value of 0.014 with a standard deviation of 0.119.

[Insert Table 2 Here]

Table 2 reports the correlation among production process distances and TNIC similarity score. In Panel A, Hoberg-Phillips 25 FIC level variables are presented. The 25 FIC level TNIC scores are negatively correlated with all production process distances. The most substantial negative correlation is -0.5264 (XGB Distance), while the smallest negative correlation is -0.1346 (TF Distance). Notably, the transfer learning-based neural network measure shows a small correlation with traditional kernels (Cobb-Douglas and linear) and a large correlation with other nonlinear ML-based kernels (XGBoost and fully connected neural network kernels). Panel B displays similar correlations using firm-level production process distances.

6 Results

6.1 M&A Activities and Production Process Heterogeneity Across Industries

We now examine the relationship between differences in production processes (CD, Linear, XGB, FC, and TF Distances) and industry-level M&A activities. We conduct industry-pair-level panel regressions using a total number of observations during our sample period from 1988 to 2021 of 21,250 ($= 25 \times 25 \times 34$). Table 3 reports the results of the pooled panel regressions of the log number of M&A deals between two industries on our industry-level production function distances (i.e., CD, Linear, XGB, FC, and TF Distances).

[Insert Table 3 Here]

We hypothesize that M&A activities are negatively correlated with acquiror-target industry-level distances. In Table 3, we find results that are largely consistent with our hypotheses: M&A activities are significantly and negatively correlated with XGBoost, FC, and TF distances, both economically and statistically. For example, in Column V, the point estimate for TF Distance is -0.095, indicating that for a one standard deviation increase in TF Distance (0.093), there is a 0.88 percent reduction in log(Number of M&A Deals) between two industries. In the same table, the annual average TNIC score is positively correlated with cross-industry merger intensity, implying that acquiror-target industries with similar end products tend to engage in a greater number of M&As. However, it should be highlighted that our acquiror-target industry-level production process distances are economically and statistically significant even after accounting for these TNIC effects. Unlike the three ML-based production functional distances, conventional CD and Linear Distances fail to explain cross-industry merger activities. Their point estimates are either in the wrong sign (CD Distance) or statistically insignificant (Linear Distance). These results underscore the utility and significance of non-parametric ML approaches for measuring the highly complex and latent underlying production processes in each industry.

The negative relation between M&A activity and industry-level production process heterogeneity remains robust when using alternative production outputs, such as the natural logarithm of sales or return on assets, rather than Tobin's Q. The result holds as well when we employ alternative distances measured with a positive weight restriction on the first layer of the deep neural networks. Excluding the leverage variables in the estimation, which could potentially be endogenous to production outputs, does not alter our results either.

[Insert Figure 9 Here]

We further examine the negative relation between M&A activities and acquiror-target production process distances year by year. Figure 9 reports the results from these cross-sectional regressions of log(Number of M&A Deals) on our five functional distances for each year from 1988 to 2021. Each regression includes the annual average TNIC score for acquiror-target industry pairs (625 FIC industry pairs in each year), acquiror and target 25 FIC industry fixed effects, and an intercept. Figure 9 displays the point estimates of log functional distances (along with their 95-percent confidence intervals) for each year. Additionally, we compare the point estimate for TF Distance with those for our four other kernels. As shown in Figure 9a (Cobb-Douglas kernel) and Figure 9b (linear kernel), the point estimates of linear kernels are much smaller in magnitude and often statistically insignificant, while all other non-linear ML-based kernels demonstrate statistical significance throughout the entire sample period. This once again highlights the importance of utilizing machine learning techniques in corporate finance.

[Insert Figure 10 Here]

Finally, Figure 10 presents the results from bootstrapping based on the same specification as in Column V of Table 3, with randomly re-shuffled acquiror-target industry distances used 500 times. As depicted in the figure, the distribution is centered around zero, ranging from -0.04 to 0.04. Given that our previous point estimate for the correct acquiror-target production process distance was -0.095, this negatively significant impact on cross-industry M&A activities is unlikely to be attributable to random measurement errors.

6.2 Deal-Level M&A Activities and Firm-Level Production Process Heterogeneity

In this section, we conduct the deal-level analysis. We examine the relationship between the likelihood of M&A deal completion to firm-level acquiror-target production process dissimilarity.[22] We employ probit regression to analyze the impact of acquiror-target production process heterogeneity on the likelihood of their M&A deal completing. Specifically, we use an indicator for M&A deal completion as our main dependent variable. As explained earlier, firm-level acquiror-target production process distances are estimated using their instantaneous peer groups formed by firms with a high TNIC similarity score to the firms of interest. Control variables include TNIC score, deal characteristics (Diversify, Hostile, High Tech, Tender Offer, Stock Deal, Relative Deal Size), and fundamental characteristics of acquirors and targets, such as the logarithm of assets, Tobin's Q, book leverage, and the cash flow to assets ratio. In all regressions, we include acquirors' and targets' 25 FIC industry fixed effects as well as year fixed effects. We cluster the standard errors at the year level. Panel A reports the point estimates of production kernel distances, while Panel B of the same table reports the marginal effect on the probit propensity of each explanatory variable.

[Insert Table 4 Here]

Columns I to V of Table 4, Panel A present estimates of production process distances without control variables. In these regressions, we include only industry and year fixed effects. We limit our sample to deals with transaction amounts greater than $100 million. Consistent with our hypothesis that the M&A completion rate is lower for acquirors and targets with a wide gap in their production processes, all point estimates are negative, except for the linear kernel-based distance. These point estimates are statistically significant, generally falling

[22] In our sample, it is observed that the size of acquiring firms is similar to that of their respective peers, and similarly, the size of target firms is comparable to that of their own peers.

within the 1 to 5 percent confidence levels.

In Columns VI to X, we incorporate the acquiror-target TNIC score into the regression. While our point estimates remain negative, the economic magnitude of acquiror-target production process distances becomes smaller compared to those without this additional control. Only the two neural network-based distances are still statistically significant. Among them, our transfer learning-based distance, TF Distance, continues to significantly explain the negative consequence of production process heterogeneity between two firms on their likelihood of deal completion. This result holds at the 5 percent statistical significance level.

We consider several possible specifications in Columns XI through XVI. Given that TF Distance showed the most significant influence on the transaction completion rate analysis, we decided to adopt it as our main distance measure. We then include the TNIC score along with all deal- and firm-specific information. In this analysis, the magnitude of the TF Distance effect increases to -1.897 with controls, compared to -1.323 without them.

In Column XII of the same table, we exclude vertical mergers where the acquiror-target Hoberg-Phillips vertscore is more than 0.02. In that column, we discover that the estimation for the production process distance becomes economically more significant (-2.148) than those in earlier analyses. A possible explanation for this result is that companies may purchase targets with radically dissimilar production processes to vertically integrate downstream production processes. Although the production processes of the acquiror and the target differ significantly in these vertical mergers, they may still work well together. Eliminating these cases may emphasize the increasing integration cost associated with significant functional gaps between targets and acquirors.

In Column XIII, we use the alternative linear probability model and find that the previous results remain valid. In Columns XIV and XV, we further consider different M&A samples. Specifically, relatively smaller deals (with a transaction value greater than $10 million) are considered in Column XIV, while very large deals (with a transaction value greater than

$900 million) are considered in Column XV. Our results remain robust across these various test samples.

Finally in Column XVI, we substitute the deal withdrawal indicator (Withdrawn) for the deal completion dummy (Completed). At the one percent statistical significance level, we obtain a positive point estimate for TF Distance, our main proxy for production process heterogeneity between acquiror and target.

Overall, the results reported in Tables 3 and 4 strongly support our hypotheses, indicating that heterogeneity in the acquiror-target production processes could reduce M&A activities between the two firms. These findings are consistently supported at both the industry and company pair levels.

6.3 M&A Performance Consequences: Announcement Effects

In this section, we further explore whether the cost of integration arising from heterogeneous production processes between the acquiror and target would be reflected in lower announcement returns for M&A deals. We posit that the expected synergy of a merger is likely to be overshadowed by the cost of integration when the production processes between the acquiror and target are very different. Stock return reactions around the deal announcement date are likely to capture such negative effects of the rising integration cost.

For each announced deal, we construct cumulative abnormal returns (CARs) for (-1, 0), (-1, +1), and (-1, +2)-day windows using a simple market model. Abnormal returns from the market model are accumulated over the three event windows, serving as the main dependent variables in our announcement effect analyses. We present the results in Table 5. Utilizing a linear panel regression specification with TF Distance as the primary explanatory variable, we control for the same list of deal- and firm-level characteristics for the acquiror and target as we did in Table 4.

[Insert Table 5 Here]

Table 5 reports the announcement effects for all deals in Columns I through VI. The announcement effects for horizontal mergers alone are further examined in Columns VII through XII. We consider both value-weighted and equal-weighted announcement effects. In Columns IV to VI and Columns X to XII, the CAR is weighted by the firm sizes of the acquiror and target. In contrast, in Columns I to III and Columns VII to IX, the CAR is equally weighted for the acquiror and target.

First, in a sample that includes both horizontal and vertical mergers, the value-weighted CARs for all three event windows (Columns IV to VI) are significantly negative, whereas the equal-weighted CARs for the same event windows (Columns I to III) are not statistically significant. For the event window from t-1 to t, the value-weighted CAR is marginally significant, with a value of -0.105 (as shown in Column IV). For the longer event windows, the effect becomes stronger, with CAR values of -0.148 and -0.143 in Columns V and VI, respectively. The significantly negative value-weighted CARs, in contrast to the insignificant equal-weighted CARs, indicate that acquirors tend to bear the integration costs of a merger rather than the targets.

To vertically integrate manufacturing processes not already present in the acquiring firm, companies occasionally purchase targets in other industries. In such cases, while there may be a large difference between the acquiror's and target's production processes, these differences can offer value, making integration costs less significant. We examine the effect of the production functional distances between the acquiror and the target exclusively for horizontal mergers in Columns VII through XII. The value-weighted CARs are all negative and statistically significant at the one percent level, while the announcement effects of the equal-weighted CARs (Columns VII to IX) are insignificant. Furthermore, the point estimates for these horizontal merger samples in Columns X to XII are more pronounced than those for Columns IV to VI, where our sample consists of both horizontal and vertical mergers.

Overall, investors thoroughly evaluate the potential integration costs associated with a merger in the lead-up to the announcement date. The stock prices during various event windows surrounding the announcement date effectively mirror investors' perceptions of the merger's costs.

6.4 M&A Performance Consequences: Post-Merger Survival

In this section, we examine the impact of integration costs on the long-term success of mergers. We anticipate a higher likelihood of long-term survival for the combined organization if the acquiror smoothly integrates the target's main production technologies, thereby realizing the expected synergy following the merger. To capture this long-term effect on M&As, we conduct probit regressions using the acquiror's survival for ten (or fifteen) years post-M&A as the key dependent variable. The survival indicator takes a value of one if the acquiror has a positive amount of assets and has not appeared in SDC Platinum as a target (i.e., the acquiror is not for sale) during the 10 years (or 15 years) post-merger. We also consider acquiror divestiture within the 10 years after the merger as an alternative dependent variable. Control variables are the same as those employed in our deal-level analyses (i.e., Tables 4 and 5). The key independent variable is TF Distance.

[Insert Table 6 Here]

In Column I, we consider the 10-year survival dummy for the acquiror using all deals with a transaction value greater than $10 million. We find that acquirors are less likely to survive when the acquiror-target production process distance (TF Distance) is large. The point estimate of -3.762 for TF Distance is statistically significant at the five percent level. Column II considers a subsample with larger deals (transaction value greater than $50 million). There, we observe slightly weaker effects; however, TF Distance still exhibits a strong negative correlation with the acquiror's survival over a 10-year horizon post-merger, with a point estimate of -2.847. Column III repeats the same exercise as in Column I, using a longer post-merger horizon of 15 years. We find a much stronger effect of TF Distance (-6.528). Finally, in Column IV, we analyze the pure divestiture of the acquiror, defined as the acquiror becoming a target with divestiture in the SDC Platinum data, within the 10 years post-merger. We find that acquirors are more likely to experience divestiture within 10 years when the acquiror-target production process distance (TF Distance) is large.

6.5 M&A Performance Consequences: Post-Merger Firm Growth

In this section, we examine the relationship between production process distances and post-merger firm growth. For firm growth, we measure 10-year asset growth as the log difference in total assets 10 years after the merger compared to the merger year, and market value growth as the log difference in market value 10 years after the merger relative to the merger year. Market value is calculated as the sum of total assets and the market value of equity, the latter of which is derived from the total shares outstanding multiplied by the year-end stock price.

[Insert Table 7 Here]

Table 7 shows results on post-merger growth. We analyze two samples: one considering all types of merger (Columns I and III) and another excluding vertical mergers (Hoberg-Phillips vertscore<0.02) (Columns II and IV). As shown in Column I, mergers between firms with a large difference in production processes (TF Distance) tend to have lower growth post-merger. The estimate on TF Distance is -2.178, resulting in 24% lower asset growth for a one standard deviation increase in the distance. In Column II, we repeat the same exercise for a subsample without vertical mergers and find stronger results: a one-standard-deviation increase in TF Distance leads to 27% lower asset growth post-merger. Similar findings are observed for market value growth. In Column III, a one-standard-deviation increase in TF

Distance results in 26% lower 10-year market value growth, and in Column IV, which focuses solely on the horizontal merger subsample, the decrease is 29%.

[Insert Table 8 Here]

Table 8 presents the dynamic learning effects of repeated acquisitions. We use 10-year asset growth (Columns I and VI) and 10-year market value growth (as shown in Columns VII and IX) as a proxy for firm growth. The key independent variables capturing repeat acquirors' learning effects are Prior Deals Within 10 Years, Completed Deals Within 10 Years, Prior Failed Deals Within 10 Years, and their interactions with TF Distance.

In Table 8, we observe that post-merger firm growth rates decrease when there is a large production process distance between the acquiror and the target. Specifically, a one-standard-deviation increase in TF Distance results in 34% (-3.085×0.11) lower 10-year asset growth and a 31% decrease in 10-year market value growth (-2.811×0.11) following the merger announcement.

However, the adverse effects of a large production process distance are mitigated when acquirors have more experience, as evidenced by their past deals. In Column I, the interaction between TF Distance and Prior Deals Within 10 Years is 0.063. This indicates that each additional past deal by the acquiror reduces the adverse effects on firm growth of a large TF Distance by 0.7% (calculated as 0.063×0.11×1). In Column II, we observe similar effects when vertical mergers are excluded.

Columns III and IV quantify acquirors' past experiences through the number of deals completed in the past 10 years. In these columns, we observe a similar economic magnitude in the mitigating effect of a large TF Distance. Specifically, for all mergers (Column III), the point estimate for the interaction between TF Distance and the dummy variable for Prior Completed Deals Within 10 Years is 0.064, translating to a 0.7% mitigating effect. For horizontal mergers in Column IV, similar effects are observed, with a point estimate of

0.060, corresponding to a mitigating effect of 0.66%.

When analyzing the number of incomplete deals in the past 10 years (Prior Failed Deals Within 10 Years), we find much stronger mitigating effects. An additional failed deal in the past 10 years is associated with a 3.8% mitigating effect on the adverse impacts of production process distance, as shown by a calculation of $0.349 \times 0.11 \times 1$, and a 3.8% mitigating effect, calculated as $0.344 \times 0.11 \times 1$, for horizontal mergers (Column VI).

Comparing results from analyzing past completed deals (Columns III and IV) against past incomplete deals (Columns V and VI) shows that repeated acquirors reflect on and learn significantly more from past failures (incomplete deals) than past successes (completed deals). In Columns VII to XII, the analysis is repeated, utilizing 10-year post-merger market value growth as the measure for firm growth instead of asset growth. Qualitatively similar results are obtained, albeit slightly weaker.

[Insert Table 9 Here]

Finally, Table 9 illustrates the impact of repeated acquisition on TF Distance. The dependent variable, TF Distance, measures the distance between acquirors and targets in terms of their production processes. The key independent variables include indicators for having a recent merger deal within 10 years (Most Recent Deal Within 10 Years), for the most recent deal within 10 years being completed (Most Recent Deal Within 10 Years: Completed), and for the most recent deal within 10 years being incomplete (Most Recent Deal Within 10 Years: Incomplete).

As depicted in Table 9, TF Distance is smaller when acquirors had a deal within the last 10 years for all types of mergers (Column I) and specifically for horizontal mergers (Column II). However, no significant learning effects are observed when the most recent merger within 10 years was completed, as shown in Columns III and IV. In contrast, a stronger learning effect is observed from failed mergers in the past 10 years (Most Recent Deal Within 10

Years: Incomplete). In Column VI, an incomplete most recent merger within the past 10 years is associated with a TF Distance reduction of -0.011. These findings in Table 9 suggest that repeated acquirors tend to acquire targets with similar production processes, indicating a learned strategy to mitigate the burdens of costly integration post-merger.

Overall, heterogeneous production processes between acquirors and targets are strongly correlated with poor valuation outcomes reflected in both short-term (Table 5) and long-term effects (Table 6). Collectively, these findings suggest that merger deals between industries and firms with heterogeneous production functions are associated with higher integration costs and reduced expected synergy.

6.6 Impact of Employee Bargaining Power: State Adoption of Right-To-Work Laws

In this section, as a supplementary test to identify the main channels of the effect we document, we consider the adoption of Right-to-Work laws across states. Over the past several decades, numerous U.S. states have adopted Right-to-Work laws. These laws have weakened labor rights and employees' bargaining power with employers through various provisions, such as the prohibition of mandatory union membership.

[Insert Table 10 Here]

Panel A of Table 10 indicates that more than half of the 50 U.S. states adopted Right-to-Work laws. In the context of mergers, weakening employee bargaining power may reduce integration costs by facilitating relocating or downsizing the labor force more easily. Hence, we expect the adverse effects of large differences in production processes between acquiror and target to be mitigated when states adopt Right-to-Work laws.

As indicated in Panel B of Table 10, the interaction between production process distance and the indicator of post Right-to-Work is negative. The point estimate for the interaction term in Column V is -1.725, significantly mitigating the negative effect of TF Distance, which is -2.045. Column VI shows that the mitigating effect of Right-to-Work laws becomes more pronounced in larger deals, with the interaction estimate being -2.216. This effectively neutralizes the negative effect of TF Distance (-2.191) when focusing on mergers with deal sizes greater than $200 million. Finally, as demonstrated in Column VII, the mitigating effect of Right-to-Work laws persists when focusing on only horizontal mergers.

In summary, the facilitation of labor restructuring post-merger, achieved by weakening employees' bargaining power through the adoption of Right-to-Work laws, significantly reduces the integration costs arising from heterogeneous production processes between acquirors and targets.

7 Conclusion

We analyze the impact of production process heterogeneity between the acquiror and the target on corporate M&A decisions. Using machine learning (ML) techniques, we estimate each firm's underlying production decision-making process at both the industry and firm levels. In particular, we consider tree-based XGBoost and two other neural network kernels that are novel to the literature (fully-connected layers and transfer learning-based neural networks). Our analysis incorporates various neoclassical production inputs, including firm size, capital expenditure, financial leverage, labor, asset tangibility, intangible assets, and advertising expenses. We then map these inputs to production outcomes, such as valuation and sales. After estimating these kernels non-parametrically, we compute the functional distances between acquirors and targets.

Drawing on an intuition similar to that of a Lagrange multiplier, we introduce novel production function distances. These are constructed based on the ratio of the mean squared error (MSE) from constrained production function optimization to the MSE from the unconstrained optimization. This MSE ratio captures the difficulty an acquiring firm faces when integrating the target's production process with theirs. As mentioned earlier, our approach extends beyond applying a naive ML algorithm. We also employ advanced deep learning transfer learning algorithms capable of ”extracting layers” from the trained set (i.e., the acquiror's peer groups) for application to the target's data. By analyzing the functional distances estimated through both conventional and ML-based kernels (including Cobb-Douglas, linear, XGBoost, fully connected layers, and transfer learning-based distances), we examine key hypotheses about the impact of production functional distances on corporate merger activities and outcomes across industries and firms.

We examine whether acquirors and targets from industries with a significant disparity in their production decision processes are less likely to undergo business integration via M&As. Additionally, we investigate whether deals involving firms with heterogeneous production processes are less likely to be completed and/or whether they exhibit weaker stock market reactions upon deal announcement to investors. Furthermore, we explore whether firms involved in a merger with a greater functional distance are less likely to survive and more likely to divest in the long run, specifically over periods of 10 and 15 years following the merger completion.

Utilizing more than 30 years of comprehensive M&A history for U.S. firms from the SDC Platinum database, our empirical findings largely align with our predictions. We observe that transfer learning-based production process dissimilarity most effectively captures the critical differences in production process heterogeneity. This approach provides a more accurate prediction of merger outcomes than conventional Cobb-Douglas and linear kernels. Furthermore, we find that M&As between firms with heterogeneous production factors, decision-making rules, cultures, and organizational capitals are more likely to be incomplete, leading to lower announcement returns and poorer long-term growth and survival rates. Stock market reactions to deal announcements are approximately 10% lower when there is significant dissimilarity in the production processes between the acquiror and the target. Such mergers also exhibit an approximately 55% lower likelihood of survival within the fifteen years following the announcement. Collectively, these findings suggest that expected synergy is lower in a merger between two firms with heterogeneous production technologies that attempt to unite their business operations into a single new entity.

Interestingly, we also observe dynamic learning in repeat acquirors who glean insights from past failed deals. We identify a negative effect of production process dissimilarity on firm growth that tends to be mitigated in the case of repeated acquisitions. Following an incomplete prior attempt, these repeat acquirors shift their focus to new targets with less production functional heterogeneity. Through a supplementary quasi-natural experiment designed to refine the inclusion restrictions of our innovative distances, we show that the adoption of Right-to-Work laws can alleviate the detrimental effects of production functional dissimilarity between the acquiror and the target. The adoption of these laws weakens employees' bargaining power with employers, thus facilitating a smoother post-merger integration process.

In summary, our results suggest that novel machine learning approaches in corporate finance can effectively capture the intricate and latent production decision-making process within a firm. Using deep neural network-based kernels, we uncover production factors and weights that mirror the underlying production technology and other intangible organizational capitals, such as corporate culture and hierarchies. Our production process estimates offer insights into many facets of corporate financial decision-making. We anticipate that these innovative techniques can address broader aspects of corporate finance issues extending beyond mergers and acquisitions. This exploration could shed light on managerial decision-making preferences, CEO turnover patterns (such as whether new CEOs are internally promoted or externally hired); it could also help determine which type of CEO is better suited to lead a given organization, and assess how quickly a new CEO can adapt to a new business environment. We aim to continue our research in subsequent studies on these topics.

References

Ahern, K. R. 2012. Bargaining power and industry dependence in mergers. *Journal of Financial Economics* 103:530–50.

Ahern, K. R., and J. Harford. 2014. The importance of industry links in merger waves. *The Journal of Finance* 69:527–76.

Barbin, C. 2017. The three reasons why tech m&a deals fail to deliver value. In *Forbes, October 19*.

Bonelli, M. 2023. Data-driven investors. In *Data-driven Investors: Bonelli, Maxime*. [Sl]: SSRN.

Christensen, C. M., R. Alton, C. Rising, and A. Waldeck. 2011. The new m&a playbook. *Harvard business review* 89:48–57.

Damodaran, A. 2005. The value of synergy, new york university-stern school of business.

Deng, X., J.-k. Kang, and B. S. Low. 2013. Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of financial Economics* 110:87–109.

Dertat, A. 2017. Applied deep learning-part 4: Convolutional neural networks. *Towards Data Science* 26.

Dessein, W. 2002. Authority and communication in organizations. *The Review of Economic Studies* 69:811–38.

Dessein, W., and T. Santos. 2006. Adaptive organizations. *Journal of political Economy* 114:956–95.

Devos, E., P.-R. Kadapakkam, and S. Krishnamurthy. 2009. How do mergers create value? a comparison of taxes, market power, and efficiency improvements as explanations for synergies. *The review of financial studies* 22:1179–211.

Erel, I., L. H. Stern, C. Tan, and M. S. Weisbach. 2021. Selecting Directors Using Machine Learning. *The Review of Financial Studies* 34:3226–64. ISSN 0893-9454. doi:10.1093/rfs/hhab050.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep learning*. MIT press.

Gu, S., B. Kelly, and D. Xiu. 2020. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies* 33:2223–73. ISSN 0893-9454. doi:10.1093/rfs/hhaa009.

———. 2021. Autoencoder asset pricing models. *Journal of Econometrics* 222:429–50.

Harford, J. 2005. What drives merger waves? *Journal of financial economics* 77:529–60.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer.

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *The Review of Financial Studies* 23:3773–811.

———. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124:1423–65.

Hoberg, G., G. Phillips, and N. Prabhala. 2014. Product market threats, payouts, and financial flexibility. *The Journal of Finance* 69:293–324.

Hu, A., and S. Ma. 2021. Persuading investors: A video-based study. Working Paper, National Bureau of Economic Research.

Jacot, A., F. Gabriel, and C. Hongler. 2018. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems* 31.

Jha, M., J. Qian, M. Weber, and B. Yang. 2024. Chatgpt and corporate policies. Working Paper, National Bureau of Economic Research.

Jovanovic, B., and P. L. Rousseau. 2001. Why wait? a century of life before ipo. *American Economic Review* 91:336–41.

———. 2002. The q-theory of mergers. *American Economic Review* 92:198–204.

Lyonnet, V., and L. H. Stern. 2022. Venture capital (mis) allocation in the age of ai. *Fisher College of Business Working Paper* 002.

Pan, S. J., and Q. Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22:1345–59.

Rhodes-Kropf, M., and D. T. Robinson. 2008. The market for mergers and the boundaries of the firm. *The Journal of Finance* 63:1169–211.

Roberts, D. A., S. Yaida, and B. Hanin. 2022. *The principles of deep learning theory.* Cambridge University Press Cambridge, MA, USA.

Sah, R. K., and J. E. Stiglitz. 1984. The architecture of economic systems: Hierarchies and polyarchies. Working Paper, National Bureau of Economic Research.

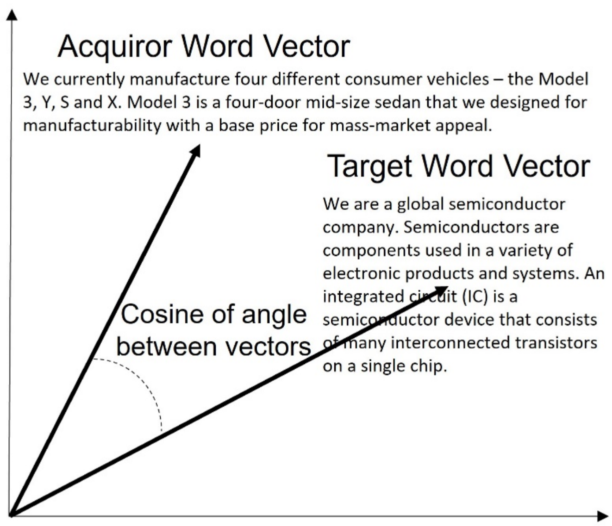
Sher, R. 2012. Why half of all m&a deals fail, and what you can do about it. In *Forbes Leadership Forum.*

Van Binsbergen, J. H., X. Han, and A. Lopez-Lira. 2023. Man versus machine learning: The term structure of earnings expectations and conditional biases. *The Review of Financial Studies* 36:2361–96.

Figure 1. Vector Distance vs. Function Distance

This figure graphically illustrates the differences in measuring distance between vectors and functions. Figure 1a shows that the distance between two vectors is measured by the cosine of theangle between them. Figure 1b demonstrates that functions cannot be measured by the cosine method and require a new definition. In this context, x_A represents the acquiror's production input, y_A is the acquiror's production output, and $f_A(\cdot)$ is the acquiror's production function ($y_A = f_A(x_A)$). Similarly, x_T is the target's production input, y_T is the target's production output, and $f_T(\cdot)$ is the target's production function ($y_T = f_T(x_T)$).

(a) Distance between vectors



(b) How do we measure distance between functions?

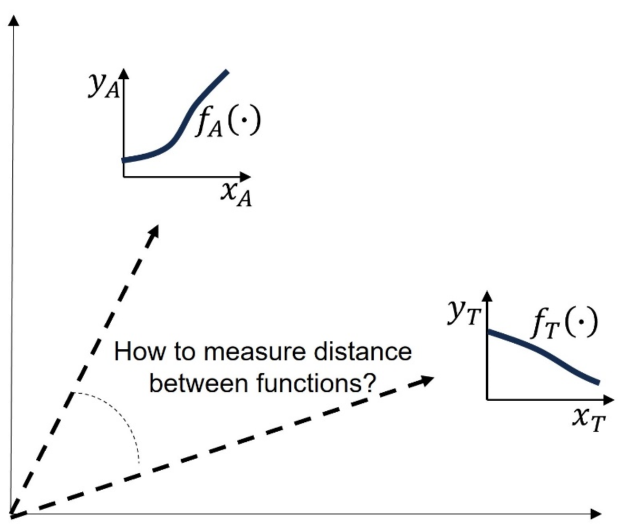
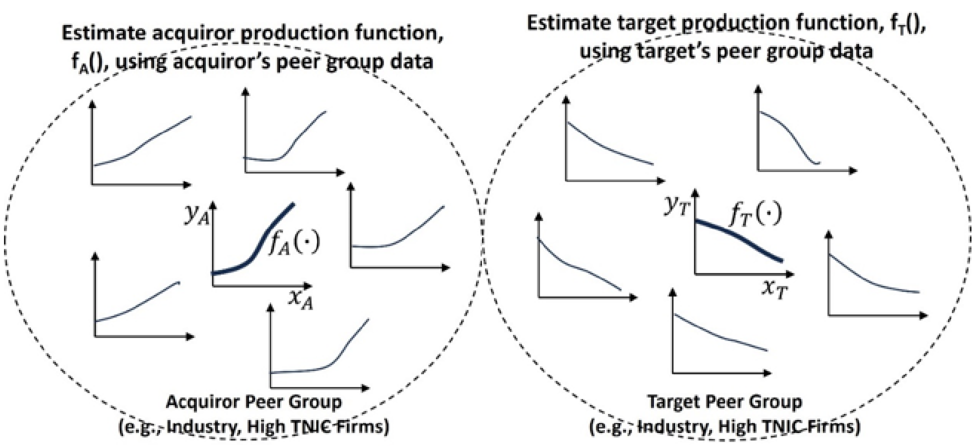


Figure 2. Estimation of Production Function Distance

This figure graphically illustrates our method of estimating production functional distances. Figure 2a depicts the first step, where the production functions of the acquiror and target are estimated using a peer group. This group may consist of firms in the same industry or firms with a high TNIC similarity score. Figure 2b illustrates the second step, which involves computing the mean squared errors (MSE) from applying target data to the acquiror's production function, denoted as $||y_T - \hat{f}_A(x_T)||$, and to the target's own production function, represented as $||y_T - \hat{f}_T(x_T)||$. The natural logarithm of the ratio of these two MSE values represents the production function distance between the two firms:

$$d(f_A(\cdot), f_T(\cdot)) = log\frac{||y_T - \hat{f}_A(x_T; w_A)||}{||y_T - \hat{f}_T(x_T; w_T)||} = log\frac{MSE(y_T, x_T; w_A)}{MSE(y_T, x_T; w_T)} \tag{12}$$

(a) Estimating acquiror's and target's production functions.



(b) Computing mean squared errors (MSE).

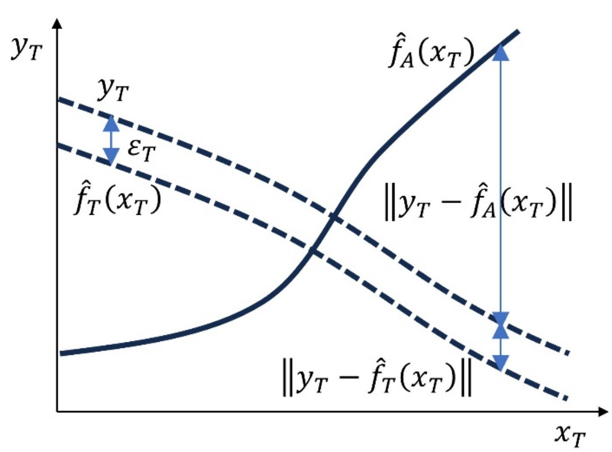


Figure 3. The Network Structure for Estimating the Acquiror-Target Production Function Distance

This figure illustrates the estimation of acquiror-target distance, whether industry-to-industry or firm-to-firm, using a fully connected neural network kernel (denoted as Neural Network: Fully Connected). The top panel depicts the training of a production function for the acquiror's FIC 25 industry. The acquiror's industry's input-outcome data is used to fit the production function using a deep neural network, formulated as $f_A(x_A; {w_A}^{(0)}, \ldots, {w_A}^{(L)}) = \hat{y}_A$. The fitting criteria aim to find neural weights that minimize the mean squared error (MSE) between the neural network-represented production kernel and the actual production outcome:

$$\min_{{w_A}^{(0)},\ldots,{w_A}^{(L)}} \mathcal{L}_A(x_A; {w_A}^{(0)}, \ldots, {w_A}^{(L)}) = \min_{{w_A}^{(0)},\ldots,{w_A}^{(L)}} MSE(\hat{y}_A) = \min_{{w_A}^{(0)},\ldots,{w_A}^{(L)}} \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{A,j} - \hat{y}_{A,j})^2} \tag{13}$$

We retain the resulting neural weights, $\boldsymbol{{w_A}^{(0)}), \ldots, {w_A}^{(L)}}$, which have been optimized to fit the production function of the firms within the acquiror's industry. The bottom panel illustrates the process of obtaining the MSE for production functions using data from the target's industry:

$$MSE(y_T, x_T; \boldsymbol{{w_A}^{(0)}, \ldots, {w_A}^{(L)}}) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{T,j} - \hat{y}_{T,j}(x_T))^2}. \tag{14}$$

In this context, $y_{T,j}$ represents the actual production outcome of firm j in the target's industry, while $\hat{y}_{T,j}$ denotes the model's estimate of firm j's production function. This estimate employs weights that have been trained and optimized based on firms in the acquiror's industry:

$$f_A(x_T; \boldsymbol{{w_A}^{(0)}, \ldots, {w_A}^{(L)}}) = \hat{y}_T(x_T) \tag{15}$$

The distance between the acquiror's and target's industries is calculated as the ratio of the mean squared error, $MSE(y_T, x_T; \boldsymbol{{w_A}^{(0)}, \ldots, {w_A}^{(L)}})$, obtained from the model using weights optimized for the acquiror's industry, to the mean squared error, $MSE(y_T, x_T; {w_T}^{(0)}, \ldots, {w_T}^{(L)})$, when the model is fitted using unconstrained weights for the firms in the target's industry:

$$d(y_A, y_T) = log\left(\frac{MSE(y_T, x_T; \boldsymbol{{w_A}^{(0)}, \ldots, {w_A}^{(L)}})}{MSE(y_T, x_T; {w_T}^{(0)}, \ldots, {w_T}^{(L)})}\right) \tag{16}$$

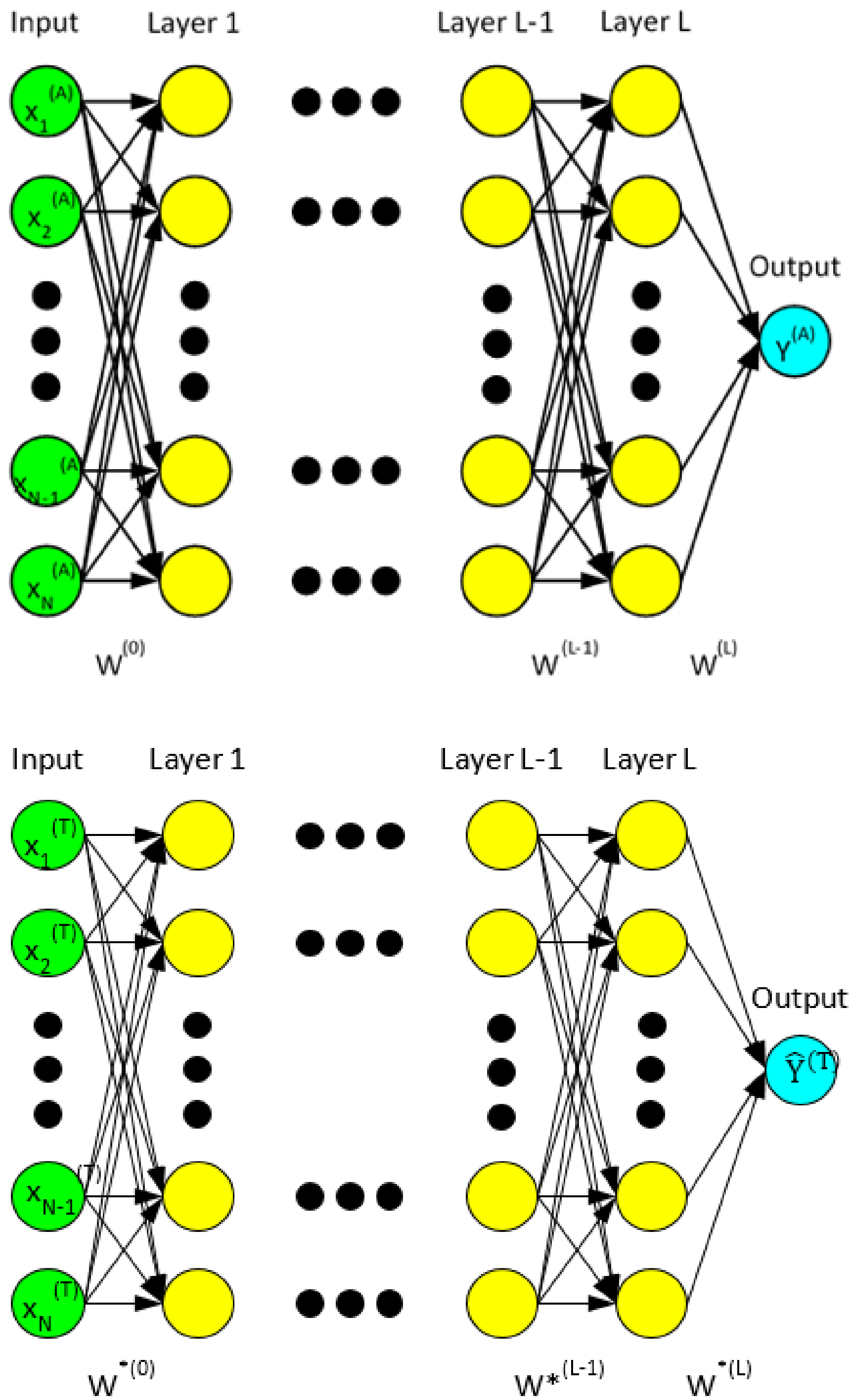

Input
Layer 1
Layer L-1
Layer L
Output
$X_1^{(A)}$
$X_2^{(A)}$
$X_{N-1}^{(A)}$
$X_N^{(A)}$
$Y^{(A)}$
$W^{(0)}$
$W^{(L-1)}$
$W^{(L)}$
Input
Layer 1
Layer L-1
Layer L
Output
$X_1^{(T)}$
$X_2^{(T)}$
$X_{N-1}^{(T)}$
$X_N^{(T)}$
$\hat{Y}^{(T)}$
$W^{*(0)}$
$W^{*(L-1)}$
$W^{*(L)}$

Figure 4. The Network Structure for a Transfer Learning-based Distance Estimation

This figure illustrates the estimation of acquiror-target distance (either industry-to-industry or firm-to-firm) using a transfer learning-based approach. The top panel shows the process of training the acquiror's production function using its peer group's input and output data. This estimation process follows the same methodology as depicted in Figure 3. The bottom panel depicts the process of obtaining the mean squared error (MSE) for the production function using the target's peer group data, with a partial transfer of the acquiring firm's production weights. In this transfer learning-based (TF) Distance, we apply the weights trained in Step 1 up to the second-to-last layer of deep neural networks, denoted as ${w_A}^{(0)}, \ldots, {w_A}^{(L-1)}$. The weight of the last layer, ${w_T}^{(L)}$, is then adjusted to minimize the MSE based on the target's industry production data.

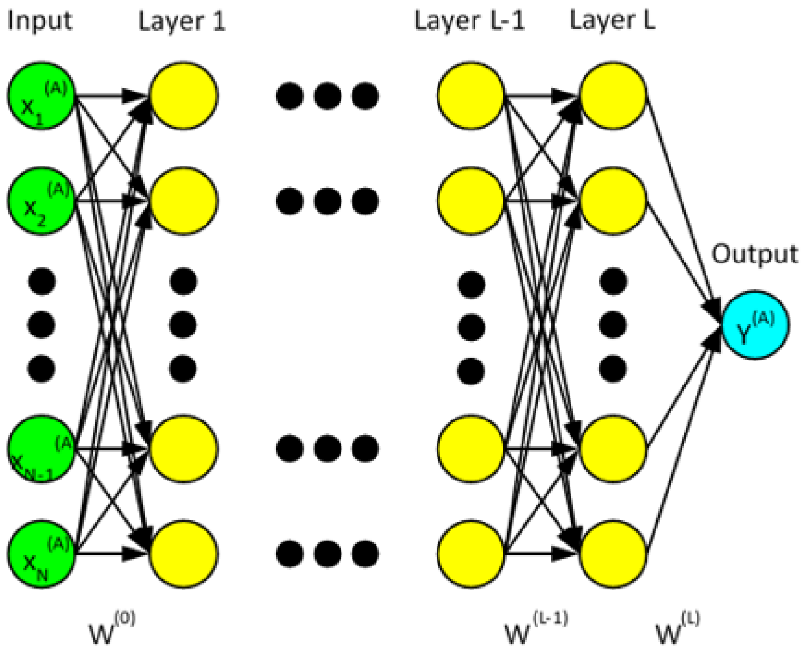
$$\min_{w_T^{(L)}} \mathcal{L}_T(x_T; \boldsymbol{w_A}^{\boldsymbol{(0)}}, \boldsymbol{\ldots}, \boldsymbol{w_A}^{\boldsymbol{(L-1)}}, w_T^{(L)}) = \min_{w_T^{(L)}} \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{T,j} - \hat{y}_{TF,T,j})^2} \tag{17}$$

The resulting estimated production function is denoted as,

$$f_A(x_T; \boldsymbol{w_A^{(0)}}, \boldsymbol{\ldots}, \boldsymbol{w_A^{(L-1)}}, w_T^{(L)}) = \hat{y}_{TF,T}. \tag{18}$$

Using the resulting mean squared error with respect to $MSE(\hat{y}_T, y_T, x_T; \boldsymbol{w_T}^{\boldsymbol{(0)}}, \boldsymbol{\ldots}, \boldsymbol{w_A}^{\boldsymbol{(L-1)}}, {w_T}^{(L)})$ gives the transfer learning-based acquiror-target distance

$$d_{TF}(y_A, y_T) = log\left(\frac{MSE(y_T, x_T; \boldsymbol{w_A}^{\boldsymbol{(0)}}, \boldsymbol{\ldots}, \boldsymbol{w_A}^{\boldsymbol{(L-1)}}, {w_T}^{(L)})}{MSE(y_T, x_T; {w_T}^{(0)}, \ldots, {w_T}^{(L)})}\right). \tag{19}$$


Input
Layer 1
Layer L-1
Layer L
Output
$X_1^{(A)}$
$X_2^{(A)}$
$X_{N-1}^{(A)}$
$X_N^{(A)}$
$Y^{(A)}$
$W^{(0)}$
$W^{(L-1)}$
$W^{(L)}$

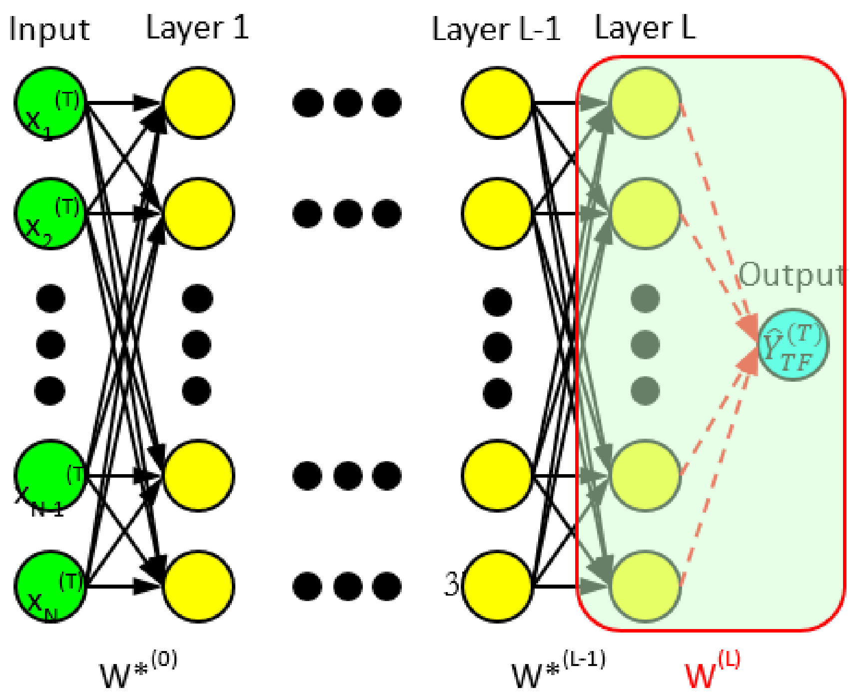
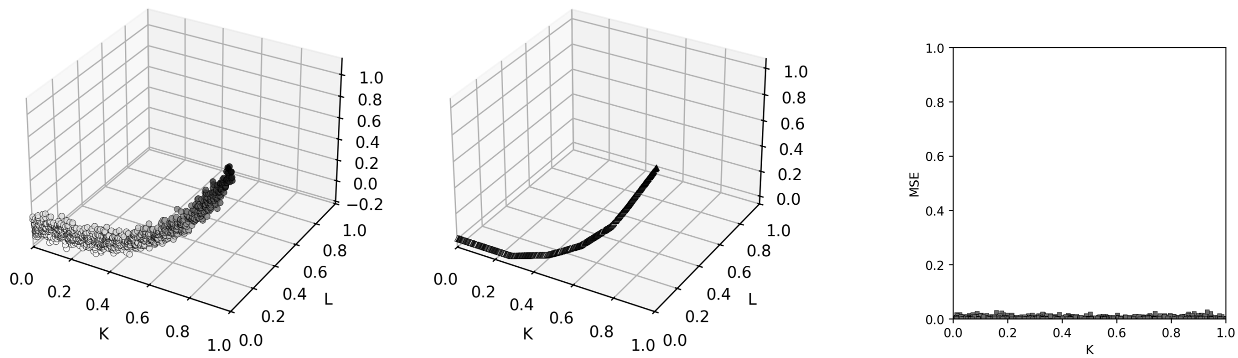

Input
Layer 1
Layer L-1
Layer L
Output
$X_1^{(T)}$
$X_2^{(T)}$
$X_{N-1}^{(T)}$
$X_N^{(T)}$
$\hat{Y}_{TF}^{(T)}$
$W*^{(0)}$
$W*^{(L-1)}$
$W^{(L)}$

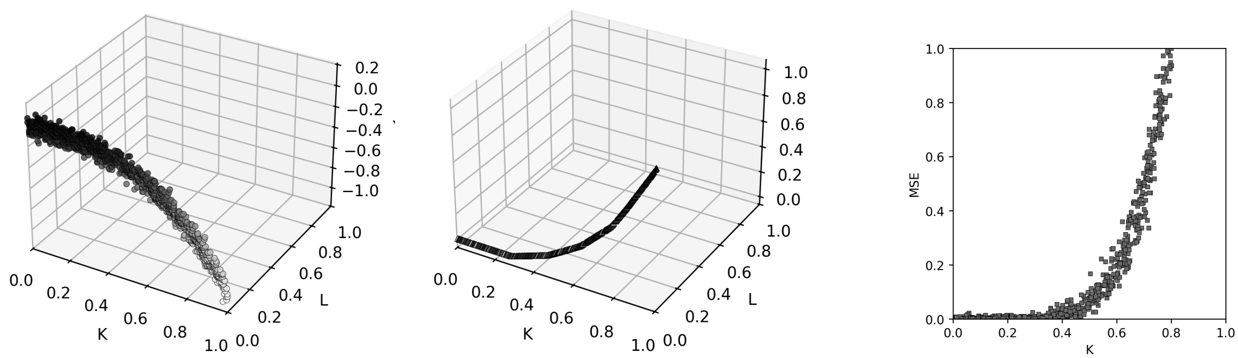
Figure 5. Result of Fully Connected Method

These figures evaluate how fully connected neural network trained on capital-intensive structure performs when applied across different industries. Each row corresponds to a different mapping, and each column reports the true outcome, the model's prediction, and the resulting mean squared error.



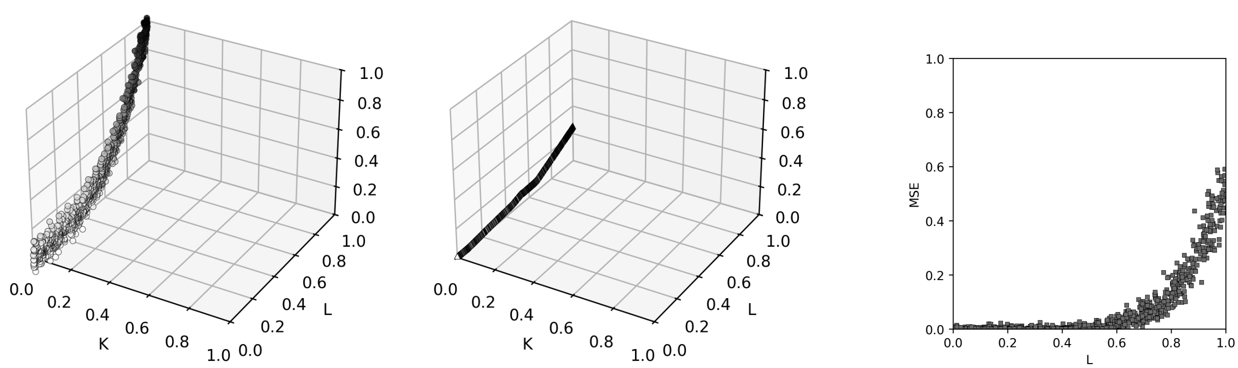
(a) True **(b)** Prediction **(c)** MSE

Panel A: Prediction for Capital Intensive Industry



(a) True **(b)** Prediction **(c)** MSE

Panel B: Prediction for Inverted Capital Intensive Industry

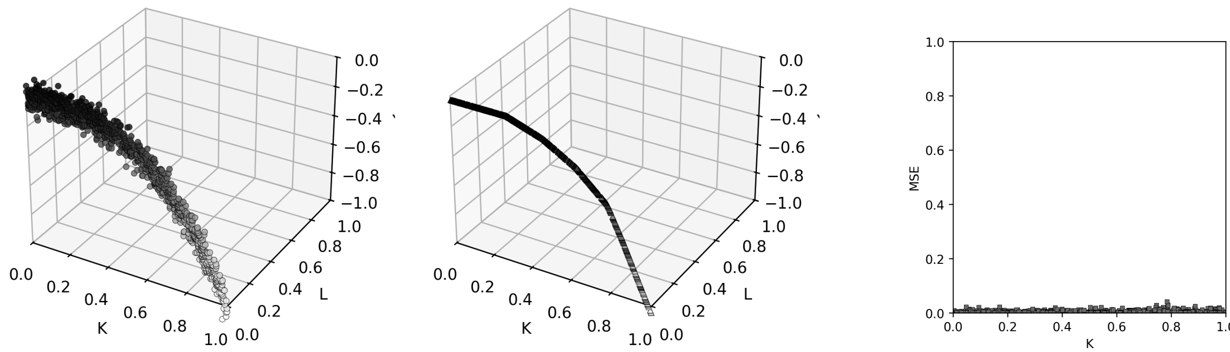


(a) True **(b)** Prediction **(c)** MSE

Panel C: Prediction for Labor Intensive Industry

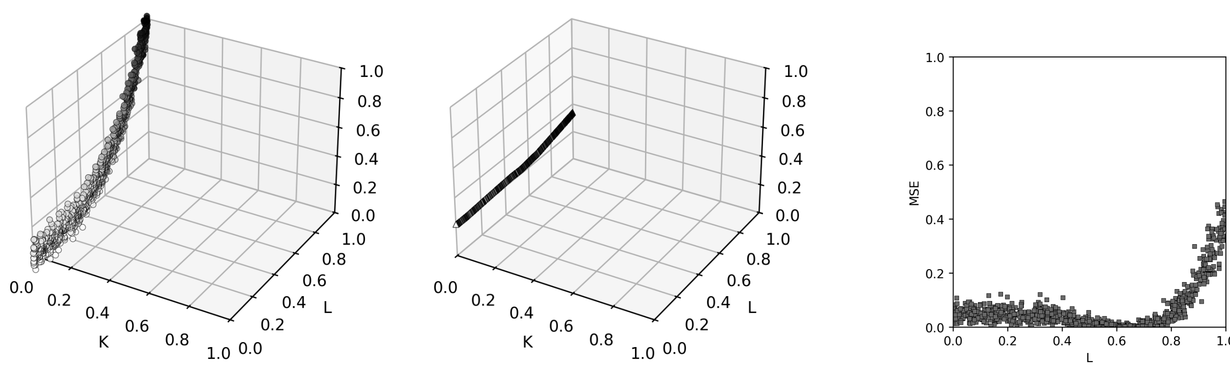
Figure 6. Result of Transfer Learning Method

These figures evaluate how the neural network transfer learning across production regimes, when applied post-learned industry. Each row corresponds to a different mapping, and each column reports the true outcome, the model's prediction, and the resulting mean squared error.



(a) True **(b)** Prediction **(c)** MSE

Panel A: Transfer Learning on Inverted Capital Intensive Industry



(a) True **(b)** Prediction **(c)** MSE

Panel B: Transfer Learning on Labor Intensive Industry

Figure 7. Process of Learning across layers

Starting from the Vanilla Method trained on the capital-intensive setting (K^3), the table compares layer-wise predictions for transfer learning under a sign reversal ($-K^3$) and a factor switch (L^3).

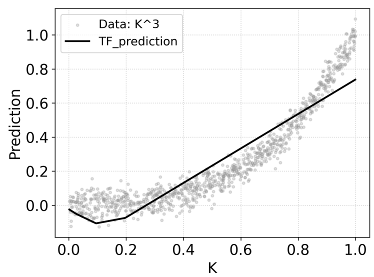
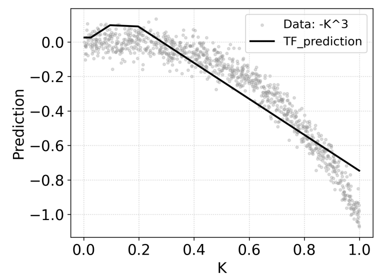
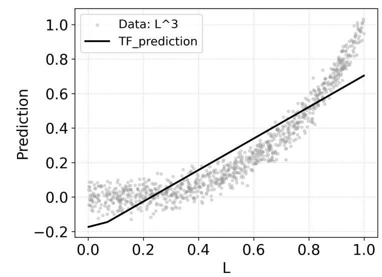
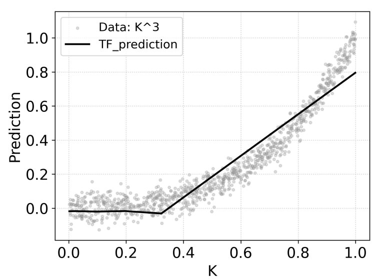
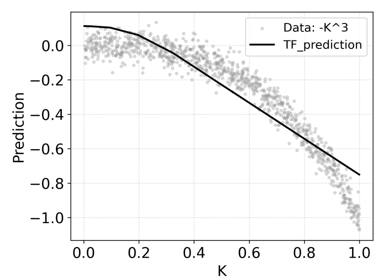
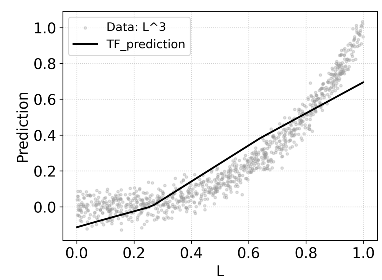
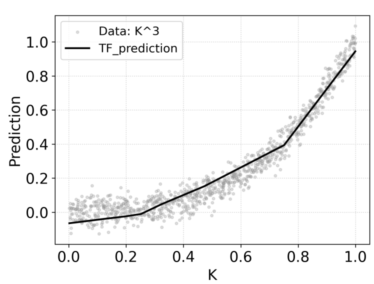
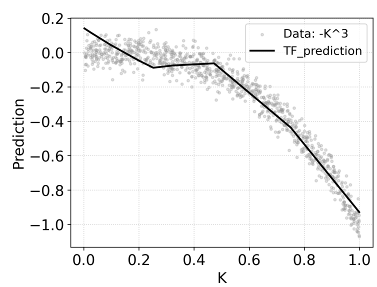
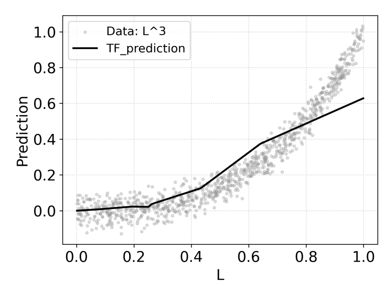
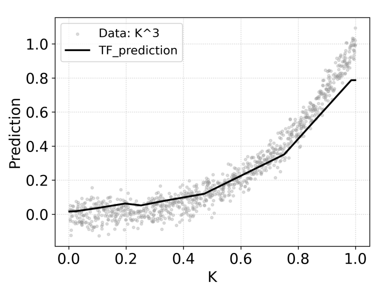
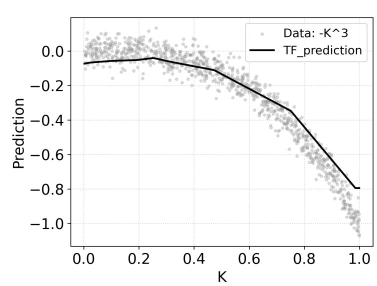
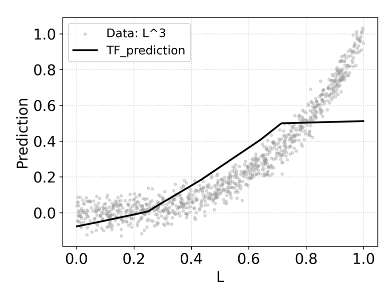
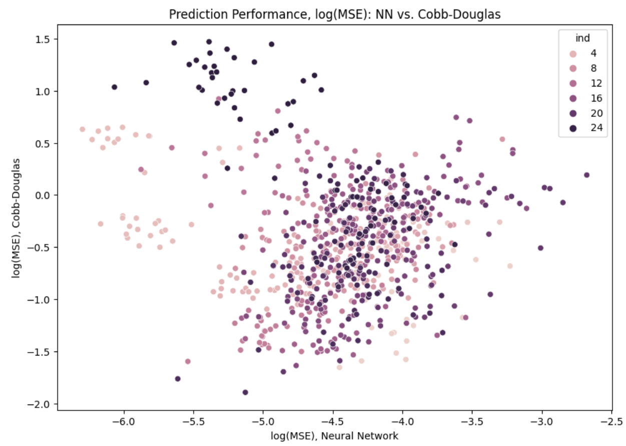
Hidden Layer	Vanilla Method (K^3)	TF: Different Sign ($-K^3$)	TF: Different Factor (L^3)
1st			
2nd			
3rd			
4th			

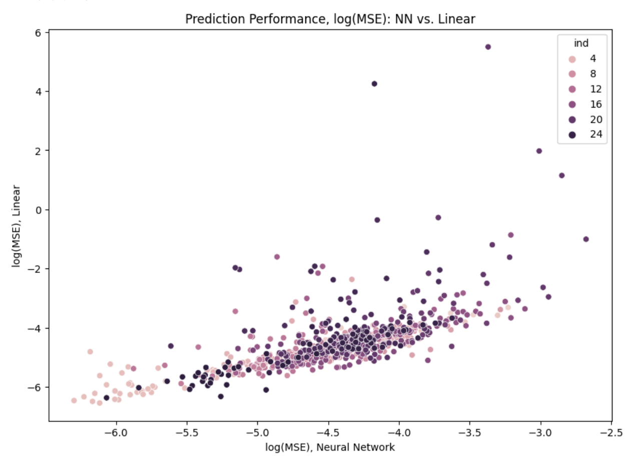
Figure 8. Fitting the Value Creation (Production) Function

In this figure, we demonstrate the creation of production functions for each group of firms (Hoberg-Phillips 25 Fixed Industry Classification) in each year (from 1988 to 2021). For kernel functions, we experiment with Cobb-Douglas, linear, XGBoost, and neural network functions. The inputs in production functions for linear, XGBoost, and neural network kernels include the natural logarithm of total assets, capital expenditures divided by total assets, short-term debt divided by total assets, long-term debt divided by total assets, employment divided by total assets, property, plant, and equipment divided by total assets, advertisement expenses divided by total assets, and R&D expenditure divided by total assets. For the Cobb-Douglas kernel, we use capital expenditures divided by total assets and employment divided by total assets as inputs. The output is represented by the natural logarithm of Tobin's Q. Fitting performance is measured by the natural logarithm of mean squared error (MSE). Figure 8a shows log(MSE) of neural network vs. Cobb-Douglas kernels. Figure 8b shows log(MSE) of neural network vs. linear kernels. Figure 8c shows log(MSE) of neural network vs. XGBoost kernels.

(a) Cobb-Douglas vs. Neural Network



(b) Linear vs. Neural Network



(c) XGBoost vs. Neural Network

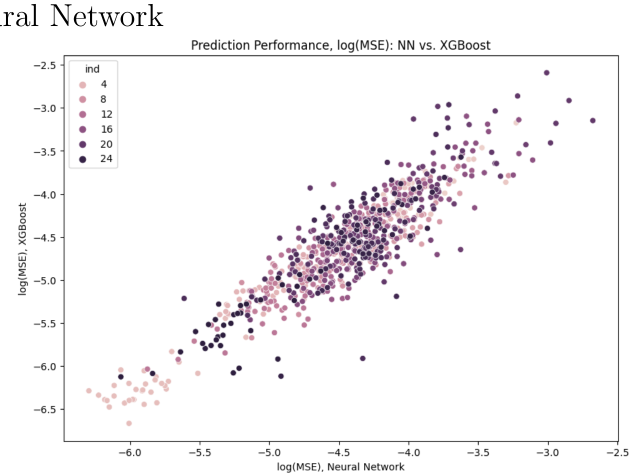


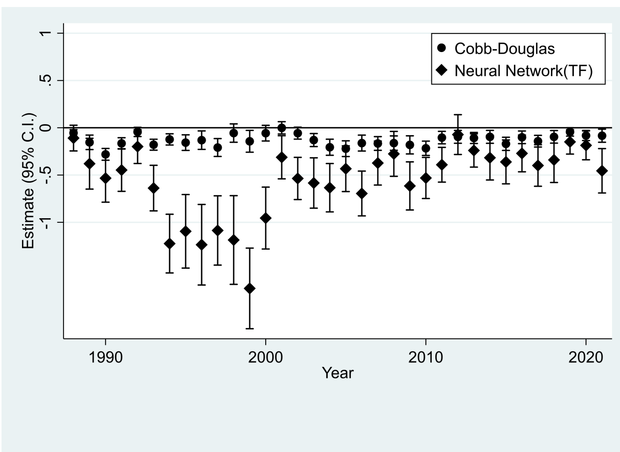
Figure 9. M&A Activities (Year-by-Year: 1988-2021)

This figure shows the year-by-year relationship between M&A activities and industry distance measures from 1988 to 2021. We report the point estimate of the industry-to-industry production process distance for each year, where the dependent variable is log(Number of M&A Deals) between the acquiror's and target's industries.

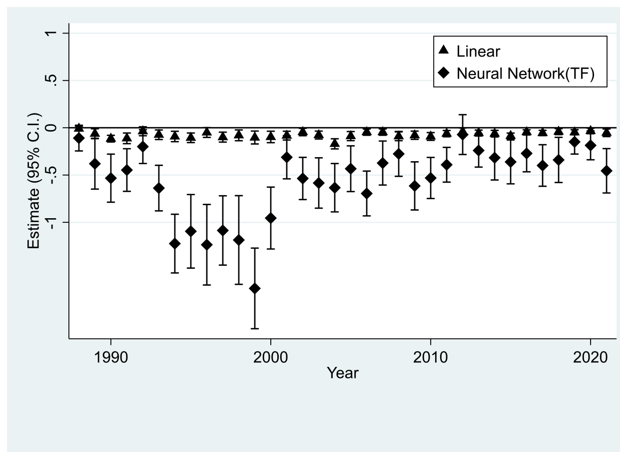
$$ln(\#M\&A)_{i,j,t} = \alpha_t + \beta_i + \gamma_j + log(Distance)_{i,j,t} + \epsilon_t \tag{20}$$

Each regression has 625 observations (Hoberg-Phillips 25 Fixed Industry Classifications pairs) and includes industry fixed effects and intercepts (not shown in the table). Figure 9a shows regression estimates of log(distance) for neural network (transfer learning) vs. Cobb-Douglas kernels. Figure 9b shows regression estimates of log(distance) for neural network (transfer learning) vs. linear kernels. Figure 9c shows regression estimates of log(distance) for neural network (transfer learning) vs. XGBoost kernels. Figure 9d shows regression estimates of log(distance) for neural network (transfer learning) vs. neural network (fully connected) kernels.

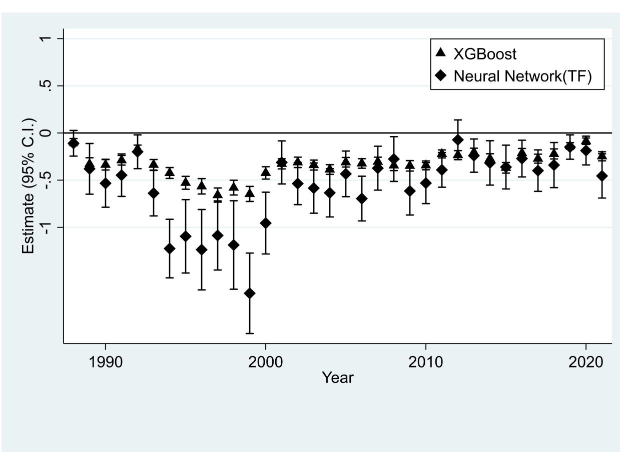
(a) Cobb-Douglas vs. Neural Network (TF)



(b) Linear vs. Neural Network (TF)



(c) XGBoost vs. Neural Network (TF)



(d) Neural Network (FC) vs. Neural Network (TF)

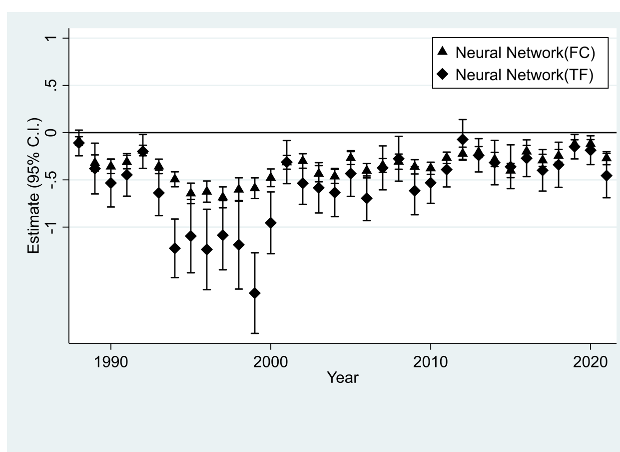


Figure 10. Robustness: Distribution of Bootstrap Estimates

This figure presents regression estimates from Column (V) of Table 3, with the exception that the industry-to-industry distances of acquiror-target industry pairs have been randomly shuffled. We conduct this placebo regression 500 times, reporting the frequency of parameter estimates related to industry distance. The dotted vertical line represents the value of the TF Distance parameter estimate from Column (V) of Table 3.

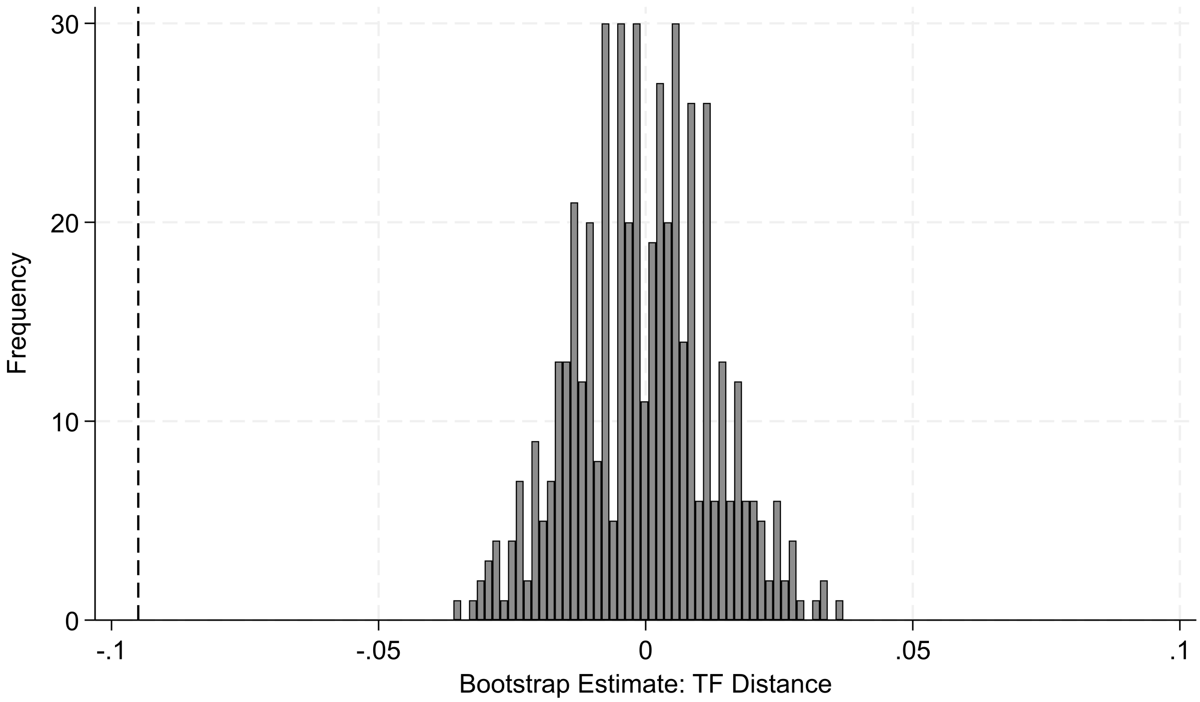


Table 1. Summary Statistics

This table shows summary statistics (the mean, standard deviation, and 1/50/99 percentiles) for the variables used in this paper. The top panel shows Hoberg-Phillips 25 fixed industry classification (FIC) pair-level variables. Log MSE ratios (Cobb-Douglas, linear, XGboost, Neural Network (Fully Connected), and Neural Network (Transfer Learning) kernels) are CD Distance, Linear Distance, XGB Distance, FC Distance, and TF Distance, respectively. Log(Number of M&A Deals) is the natural logarithm of the number of merger deals (excluding partial acquisitions and divestitures) for each acquiror-target Hoberg-Phillips 25 FIC pair. Firm-level log production MSE ratios are similarly defined as CD, linear, XGBoost, FC, and TF Distances, except using instantaneous acquiror/target peers based on TNIC similarity scores (similarity scores greater than 0.09, or the 95th-percentile of full TNIC data). TNIC score (firm level) is the Hoberg-Phillips TNIC score from the full data. They are reported in the bottom panel (Firm Level Variables). For the M&A deal-level analyses, we construct an indicator for M&A deal completion (Deal Completed indicator) and withdrawal (Withdrawal indicator). As controls in our deal-level analyses, we consider the following deal and acquiror/target company characteristics widely used in the literature (Deng, Kang, and Low, 2013, among many others): Diversify is an indicator for a diversifying merger (the acquiror's FIC differs from that of the target's). Hostile is an indicator for a hostile merger, which takes a value of one if SDC Platinum records the deal as a hostile merger and zero otherwise. High Tech is an indicator that takes a value of one if both acquiror and target are in high technology sectors and zero otherwise (i.e., SDC Platinum records them as "Primary Business not Hi-Tech"). Tender Offer is an indicator for tender offer deals. Stock Deal is an indicator for deals that are (partly) financed by stock, according to the report in SDC. Relative Deal Size is the deal value reported in SDC normalized by the acquiror's market capitalization. Firm size is the natural logarithm of total assets. Tobin's Q is the ratio of total assets plus market value of equity (total shares outstanding times year-end stock price) minus the total value of common equity minus deferred taxes to total assets. Book Leverage is debt in current liabilities plus total long-term debt divided by total assets. Cash Flow to Assets is operating income before depreciation minus total interest-related expenses minus total income taxes minus capital expenditures divided by total assets. TNIC scores (Hoberg and Phillips, 2010, 2016) are at the firm level, while our industry distances are measured at the Hoberg-Phillips 25 FIC pair level. Hence, we compute the average scores for each of the 625 acquiror-target Hoberg-Phillips 25 FIC pair in each year. Acquiror-target value-weighted cumulative abnormal returns are computed for three event windows, t-1 to t (CAR[t-1,t]), t-1 to t+1 (CAR[t-1,t+1]), t-1 to t+2 (CAR[t-1,t+2]). Post-merger acquiror survival indicators equal one if the acquiror exists and is not acquired as a target in the SDC database within 10 years after merger; we also use an indicator for divestiture, which is one if the acquiror appears as a target for divestiture within 10 years after a merger. Asset Growth [t+1,t+10] is the log difference of total assets 10 years after the merger. Market value growth [t+1,t+10] is the log difference of market value 10 years after the merger. Prior Deals Within 10 Years, Prior Completed Deals Within 10 Years, and Prior Failed Deals Within 10 Years are the number of deals, completed deals, and failed deals within the past 10 years as of the current deal's announcement date. The Most Recent Deal Within 10 years is an indicator of having a recent merger deal within 10 years. Most Recent Deal Within 10 Years: Completed is an indicator for having completed the most recent deal within 10 years. Most Recent Deal Within 10 Years: Incomplete is an indicator for having the most recent deal within 10 years be incomplete.

Hoberg-Phillips FIC 25 Level Variables	N	Mean	Std.Dev.	p1	p50	p99
CD Distance (Cobb-Douglas Kernel)	21250	0.255	0.404	-0.047	0.082	1.156
Linear Distance	21250	0.362	0.730	-0.231	0.160	1.775
XGB Distance (XGBoost Kernel)	21250	0.123	0.300	-0.246	0.120	0.563
FC Distance (Neural Network: Fully Connected)	21250	0.136	0.290	-0.344	0.116	0.630
TF Distance (Neural Network: Transfer Learning)	21250	-0.106	0.093	-0.257	-0.102	0.034
TNIC Score (Mean)	21250	0.012	0.014	0.002	0.008	0.035
log(Number of M&A Deals)	21250	0.033	0.191	0.000	0.000	0.000

Firm Level Variables	N	Mean	Std.Dev.	p1	p50	p99
CD Distance (Cobb-Douglas Kernel)	760	0.064	0.208	-0.080	0.006	0.414
Linear Distance	785	0.130	0.587	-0.406	-0.003	1.156
XGB Distance (XGBoost Kernel)	785	-0.242	0.379	-0.814	-0.241	0.347
FC Distance (Neural Network: Fully Connected)	785	1.046	0.371	0.685	0.993	2.109
TF Distance (Neural Network: Transfer Learning)	785	1.009	0.124	0.790	0.992	1.363
TNIC Score	785	0.189	0.109	0.037	0.179	0.357
Completed	785	0.703	0.457	0.000	1.000	1.000
Withdrawn	785	0.209	0.407	0.000	0.000	1.000
Diversify	785	0.243	0.429	0.000	0.000	1.000
Hostile	785	0.045	0.207	0.000	0.000	0.000
High Tech	785	0.341	0.474	0.000	0.000	1.000
Tender Offer	785	0.110	0.313	0.000	0.000	1.000
Stock Issue	785	0.020	0.141	0.000	0.000	0.000
Relative Deal Size	683	1.205	1.303	0.640	1.000	1.903
Acquiror log(Asset)	783	8.194	2.264	4.570	8.338	11.615
Acquiror Tobin's Q	783	2.063	2.098	0.971	1.398	5.520
Acquiror Book Leverage	778	0.248	0.201	0.000	0.221	0.619
Acquiror CF/Assets	783	-0.015	0.180	-0.001	0.000	0.000
Target log(Asset)	784	6.625	1.885	3.730	6.685	9.706
Target Tobin's Q	784	1.958	2.907	0.889	1.226	4.708
Target Book Leverage	774	0.252	0.233	0.000	0.206	0.692
Target CF/Assets	784	-0.002	0.022	-0.003	0.000	0.001
CAR[t-1, t]	629	0.018	0.064	-0.079	0.007	0.128
CAR[t-1, t+1]	629	0.025	0.075	-0.083	0.015	0.161
CAR[t-1, t+2]	628	0.024	0.078	-0.095	0.015	0.161
Survive Within 10 Years	978	0.079	0.269	0.000	0.000	1.000
Divestiture Within 10 Years	978	0.014	0.119	0.000	0.000	0.000
Asset Growth [t+1,t+10]	407	0.592	0.720	-1.404	0.582	2.084
Market Value Growth [t+1,t+10]	404	0.556	0.764	-1.665	0.553	2.342
Prior Deals Within 10 Years	978	11.100	13.262	0.000	7.000	76.000
Prior Completed Deal Within 10 Years	978	8.209	11.009	0.000	5.000	66.000
Prior Failed Deals Within 10 Years	978	2.892	3.461	0.000	2.000	18.000
Most Recent Deal Within 10 Years	978	0.944	0.230	0.000	1.000	1.000
Most Recent Deal Within 10 Years: Completed	978	0.319	0.466	0.000	0.000	1.000
Most Recent Deal Within 10 Years: Incomplete	978	0.138	0.345	0.000	0.000	1.000

Table 2. Correlations

This table displays correlations among the TNIC score and production process distances. Panel A presents correlations among Hoberg-Phillips 25 FIC industry-to-industry distancesin Cobb-Douglas (CD Distance), linear (Linear Distance), XGBoost (XGB Distance), neural network (fully connected) (FC Distance), neural network (transfer learning) (TF Distance) kernels. Panel B shows correlations among TNIC score and firm-to-firm distances based on instantaneous peer groups with high Hoberg-Phillips TNIC score firms (similarity score above 0.09, which is approximately the 95th percentile of the full TNIC database each year). All correlations are significant at the 1% level.

Panel A. Industry-To-Industry Distance

	TNIC Score (Mean) (I)	CD Distance (II)	Linear Distance (III)	XGB Distance (IV)	FC Distance (V)
CD Distance	-0.1618				
Linear Distance	-0.1658	0.1296			
XGB Distance	-0.5264	0.1645	0.2375		
FC Distance	-0.3254	0.1905	0.2654	0.6896	
TF Distance	-0.1346	0.0723	0.0557	0.2579	0.4414

Panel B. Firm-To-Firm Distance

	TNIC Score (I)	CD Distance (II)	Linear Distance (III)	XGB Distance (IV)	FC Distance (V)
CD Distance	-0.3189				
Linear Distance	-0.1829	0.2479			
XGB Distance	-0.4095	0.3898	0.4485		
FC Distance	-0.3156	0.3381	0.3507	0.5257	
TF Distance	-0.1578	0.0791	0.1501	0.2332	0.4831

Table 3. M&A Activity: Industry-to-Industry Distance

This table shows the relationship between M&A activities and industry distance measures in the Hoberg-Phillips 25 Fixed Industry Classification pairs by year from 1988 to 2021. The dependent variable is log(Number of M&A Deals). The key independent variables are CD Distance, Linear Distance, XGB Distance, FC Distance, and TF Distance (production process distances using Cobb-Douglas, linear, XGBoost, neural network (fully connected), and neural network (transfer learning) kernels, respectively). Each regression has 21,250 observations (625 industry pairs times 34 years). Both year and industry fixed effects are controlled, but their point estimates are not reported in this table. The t-statistics are shown in square brackets, and standard errors are clustered at the year level. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	log(Number of M&A Deals)				
	(I)	(II)	(III)	(IV)	(V)
CD Distance	0.020***				
	[4.33]				
Linear Distance		-0.002			
		[-1.00]			
XGB Distance			-0.139***		
			[-8.95]		
FC Distance				-0.074***	
				[-7.76]	
TF Distance					-0.095***
					[-5.85]
TNIC Score (FIC-Pair Annual Mean)	5.803***	5.657***	3.894***	5.060***	5.571***
	[18.37]	[18.34]	[20.88]	[18.35]	[18.74]
Intercept	-0.152***	-0.147***	-0.086***	-0.118***	-0.154***
	[-14.23]	[-14.33]	[-13.06]	[-13.89]	[-14.29]
Model	Linear	Linear	Linear	Linear	Linear
FIC FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	21,250	21,250	21,250	21,250	21,250
R-squared	0.199	0.199	0.223	0.205	0.201

Table 4. M&A Completion: Firm-to-Firm Distance

This table shows the relationship between the likelihood of M&A deal completion and acquiror-target production function distances from 1988 to 2021. Panel A shows their point estimates, while Panel B shows the marginal effects of the distances (except Column XIII, which uses a linear probability model). The dependent variable is an indicator for M&A deal completion (Columns I to XV), which is one if the deal is completed and zero otherwise, or an indicator for withdrawal (Column XVI). The key independent variables are firm-to-firm CD Distance (Columns I, VI), Linear Distance (Columns II, VII), XGB Distance (Columns III, VIII), FC Distance (Columns IV, IX), and TF Distance (Columns V, X, and XVIV to XI) kernels. Firm-to-firm distances are estimated using neighboring firms with a Hoberg-Phillips similarity score greater than 0.09. Column XII restricts the sample to horizontal mergers whose acquiror-target pairs have a vertscore (Vertical TNIC data) smaller than 0.02. Column XIII uses a linear probability model. Column XIV includes small deals (deal size greater than $10 million). Column XV restricts to very large deals (deal size greater than $900 million). In Columns XI to XVI, we control for Diversify, Hostile, High Tech, Tender Offer, Stock Deal, Relative Deal Size, acquirors' and targets' Firm Size, Tobin's Q, Book Leverage, and Cash Flow to Assets Ratio, as well as year and industry fixed effects. The t-statistics are shown in square brackets, and standard errors are clustered at the year level. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Panel A.

Dependent Variables	Completed									
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
CD Distance	-0.593**					-0.391				
	[-2.35]					[-1.40]				
Linear Distance		-0.111					-0.038			
		[-0.99]					[-0.34]			
XGB Distance			-0.249**					-0.021		
			[-2.09]					[-0.16]		
FC Distance				-0.789***					-0.461*	
				[-3.36]					[-1.81]	
TF Distance					-1.555***					-1.323**
					[-3.00]					[-2.44]
TNIC Score						1.715***	2.267***	2.280***	1.930***	2.063***
						[3.14]	[3.77]	[3.55]	[3.06]	[3.42]
Diversify										
Hostile										
High Tech										
Tender Offer										
Stock Issue										
Relative Deal Size										
Acquiror log(Asset)										
Acquiror Tobin's Q										
Acquiror Book Leverage										
Acquiror CF/Assets										
Target log(Asset)										
Target Tobin's Q										
Target Book Leverage										
Target CF/Assets										
Intercept	3.808***	3.532***	3.617***	3.390***	3.431***	3.602***	3.388***	3.433***	3.312***	3.152***
	[5.40]	[5.31]	[5.23]	[4.87]	[4.58]	[5.08]	[5.18]	[5.09]	[4.87]	[4.40]
Sample	100m+	100m+	100m+	100m+	100m+	100m+	100m+	100m+	100m+	100m+
Model	Probit	Probit	Probit	Probit	Probit	Probit	Probit	Probit	Probit	Probit
FIC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	754	779	779	779	779	748	773	773	773	773
R-squared	0.1987	0.1964	0.1978	0.2033	0.2037	0.2093	0.212	0.2119	0.2143	0.2178

Dependent Variables	Completed					Withdrawn
	(XI)	(XII)	(XIII)	(XIV)	(XV)	(XVI)
CD Distance						
Linear Distance						
XGB Distance						
FC Distance						
TF Distance	-1.897***	-2.148***	-0.349**	-1.042**	-5.658**	1.959***
	[-3.00]	[-3.00]	[-2.59]	[-2.05]	[-2.39]	[2.94]
TNIC Score	2.400***	2.076**	0.447**	2.328***	0.889	-2.916***
	[2.95]	[2.57]	[2.70]	[3.77]	[0.62]	[-3.08]
Diversify	0.095	0.137	0.004	0.014	1.010**	-0.502**
	[0.49]	[0.61]	[0.10]	[0.08]	[2.19]	[-2.51]
Hostile	-1.520***	-1.967***	-0.430***	-1.569***	-1.248	1.513***
	[-4.06]	[-4.76]	[-4.38]	[-5.88]	[-1.60]	[3.97]
High Tech	-0.009	0.166	-0.011	-0.112	-1.042**	0.140
	[-0.03]	[0.64]	[-0.14]	[-0.64]	[-2.15]	[0.46]
Tender Offer	0.651**	0.810***	0.096*	0.655***	0.332	-0.649**
	[2.43]	[3.01]	[1.72]	[3.24]	[0.47]	[-2.27]
Stock Issue	-0.541	-0.270	-0.100	-0.170	0.141	0.522
	[-1.20]	[-0.51]	[-0.77]	[-0.38]	[0.13]	[1.20]
Relative Deal Size	-0.120**	-0.159**	-0.027*	-0.088*	0.110	0.145***
	[-2.21]	[-2.45]	[-1.91]	[-1.71]	[1.34]	[2.82]
Acquiror log(Asset)	0.287***	0.303***	0.058***	0.166***	0.257*	-0.369***
	[4.32]	[4.31]	[3.51]	[3.17]	[1.69]	[-4.65]
Acquiror Tobin's Q	-0.002	-0.003	-0.001	-0.022	-0.197	0.023
	[-0.03]	[-0.06]	[-0.09]	[-0.67]	[-1.57]	[0.46]
Acquiror Book Leverage	0.121	0.245	-0.005	-0.406*	-1.600*	0.343
	[0.32]	[0.64]	[-0.06]	[-1.67]	[-1.74]	[0.99]
Acquiror CF/Assets	13.236	14.426	3.702	4.740**	-2,611.855**	-3.507
	[0.62]	[0.65]	[0.95]	[2.49]	[-2.33]	[-0.16]
Target log(Asset)	-0.290***	-0.283***	-0.055***	-0.139**	-0.445**	0.328***
	[-3.75]	[-3.49]	[-2.88]	[-2.17]	[-2.29]	[3.63]
Target Tobin's Q	-0.062	-0.063	-0.015	-0.036***	0.234	0.069*
	[-1.62]	[-1.60]	[-1.59]	[-2.69]	[1.63]	[1.69]
Target Book Leverage	0.557	0.357	0.135*	0.394	-0.162	-0.583
	[1.57]	[1.09]	[1.81]	[1.05]	[-0.15]	[-1.40]
Target CF/Assets	-0.224	-0.257	-0.581	3.678	221.708**	-0.597
	[-0.01]	[-0.02]	[-0.14]	[1.23]	[2.50]	[-0.04]
Intercept	-7.173***	-7.357***	0.616***	-1.328	0.091	7.531***
	[-10.86]	[-10.28]	[4.52]	[-1.20]	[0.05]	[12.54]
Sample	100m+	H-MA	100m+	10m+	900m+	100m+
Model	Probit	Probit	Linear	Probit	Probit	Probit
FIC FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	613	587	669	922	211	613
R-squared	0.2693	0.2819	0.272	0.2399	0.3648	0.3099

Panel B. Marginal Effects

Dependent Variables	Completed									
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
CD Distance	-0.160**					-0.104				
	[-2.38]					[-1.42]				
Linear Distance		-0.031					-0.010			
		[-0.99]					[-0.34]			
XGB Distance			-0.068**					-0.006		
			[-2.07]					[-0.16]		
FC Distance				-0.214***					-0.124*	
				[-3.38]					[-1.81]	
TF Distance					-0.422***					-0.353**
					[-3.03]					[-2.44]
TNIC Score						0.458***	0.610***	0.614***	0.518***	0.551***
						[3.12]	[3.80]	[3.60]	[3.07]	[3.42]
Diversify										
Hostile										
High Tech										
Tender Offer										
Stock Issue										
Relative Deal Size										
Acquiror log(Asset)										
Acquiror Tobin's Q										
Acquiror Book Leverage										
Acquiror CF/Assets										
Target log(Asset)										
Target Tobin's Q										
Target Book Leverage										
Target CF/Assets										
Sample	100m+	100m+	100m+	100m+	100m+	100m+	100m+	100m+	100m+	100m+
Model	Probit	Probit	Probit	Probit	Probit	Probit	Probit	Probit	Probit	Probit
FIC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	754	779	779	779	779	748	773	773	773	773
R-squared	0.1987	0.1964	0.1978	0.2033	0.2037	0.2093	0.212	0.2119	0.2143	0.2178

Dependent Variables	Completed					Withdrawn
	(XI)	(XII)	(XIII)	(XIV)	(XV)	(XVI)
CD Distance						
Linear Distance						
XGB Distance						
FC Distance						
TF Distance	-0.422***	-0.469***	-0.349**	-0.243**	-1.185**	0.390***
	[-3.03]	[-3.07]	[-2.59]	[-2.08]	[-2.52]	[2.97]
TNIC Score	0.534***	0.453***	0.447**	0.543***	0.186	-0.580***
	[2.93]	[2.60]	[2.70]	[3.64]	[0.62]	[-3.09]
Diversify	0.021	0.030	0.004	0.003	0.212**	-0.100**
	[0.49]	[0.61]	[0.10]	[0.08]	[2.19]	[-2.56]
Hostile	-0.338***	-0.429***	-0.430***	-0.366***	-0.261	0.301***
	[-4.13]	[-5.01]	[-4.38]	[-6.12]	[-1.62]	[3.98]
High Tech	-0.002	0.036	-0.011	-0.026	-0.218**	0.028
	[-0.03]	[0.64]	[-0.14]	[-0.64]	[-2.26]	[0.46]
Tender Offer	0.145**	0.177***	0.096*	0.153***	0.070	-0.129**
	[2.40]	[2.92]	[1.72]	[3.27]	[0.47]	[-2.21]
Stock Issue	-0.120	-0.059	-0.100	-0.040	0.030	0.104
	[-1.19]	[-0.51]	[-0.77]	[-0.38]	[0.13]	[1.19]
Relative Deal Size	-0.027**	-0.035**	-0.027*	-0.021*	0.023	0.029***
	[-2.27]	[-2.53]	[-1.91]	[-1.75]	[1.31]	[2.95]
Acquiror log(Asset)	0.064***	0.066***	0.058***	0.039***	0.054*	-0.073***
	[4.55]	[4.67]	[3.51]	[3.20]	[1.70]	[-5.18]
Acquiror Tobin's Q	-0.000	-0.001	-0.001	-0.005	-0.041	0.005
	[-0.03]	[-0.06]	[-0.09]	[-0.67]	[-1.62]	[0.46]
Acquiror Book Leverage	0.027	0.053	-0.005	-0.095*	-0.335*	0.068
	[0.32]	[0.64]	[-0.06]	[-1.66]	[-1.74]	[0.98]
Acquiror CF/Assets	2.945	3.150	3.702	1.106**	-547.150**	-0.697
	[0.61]	[0.65]	[0.95]	[2.48]	[-2.19]	[-0.15]
Target log(Asset)	-0.065***	-0.062***	-0.055***	-0.033**	-0.093**	0.065***
	[-3.80]	[-3.58]	[-2.88]	[-2.18]	[-2.41]	[3.74]
Target Tobin's Q	-0.014	-0.014	-0.015	-0.008***	0.049*	0.014*
	[-1.64]	[-1.61]	[-1.59]	[-2.69]	[1.66]	[1.74]
Target Book Leverage	0.124	0.078	0.135*	0.092	-0.034	-0.116
	[1.57]	[1.09]	[1.81]	[1.06]	[-0.15]	[-1.39]
Target CF/Assets	-0.050	-0.056	-0.581	0.858	46.445***	-0.119
	[-0.01]	[-0.02]	[-0.14]	[1.22]	[2.63]	[-0.04]
Sample	100m+	H-MA	100m+	10m+	900m+	100m+
Model	Probit	Probit	Linear	Probit	Probit	Probit
FIC FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	613	587	669	922	211	613
R-squared	0.2693	0.2819	0.272	0.2399	0.3648	0.3099

Table 5. Announcement Effect

This table shows the relationship between M&A deal announcement returns and the acquiror-target production functional distance based on the transfer learning algorithm (i.e., TF Distance) from 1988 to 2021. We use linear specification. The dependent variable is daily cumulative abnormal returns around announcement dates from t-1 to t (Column I, IV, VII, X), t-1 to t+1 (Column II, V, VII, XI), and t-1 to t+2 (Column III, VI, IX, XII). Columns I to VI include both horizontal and vertical mergers. Columns VII to XII exclude mergers that are likely to be vertical mergers (exclude deals with a Hoberg-Phillips vertscore greater than 0.02) All regressions use the value weighted acquiror and target cumulative abnormal returns. The key independent variable is TF Distance. In all regressions, we control for TNIC Score, Merger Completed, Diversify, Hostile, High Tech, Tender Offer, Stock Deal, Relative Deal Size, acquirors' & targets' firm size, Tobin's Q, Book Leverage, Cash Flow to Asset Ratio, and year- and industry fixed effects. The t-statistics are shown in square brackets, and standard errors are double-clustered at the acquiror-target industry level. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Dependent Variables	CAR [t-1, t] (I)	CAR [t-1, t+1] (II)	CAR [t-1, t+2] (III)	CAR [t-1, t] (IV)	CAR [t-1, t+1] (V)	CAR [t-1, t+2] (VI)
TF Distance	0.054	-0.083	-0.092	-0.105*	-0.148***	-0.143***
	[0.22]	[-0.35]	[-0.40]	[-2.01]	[-3.39]	[-2.84]
TNIC Score	-0.080	-0.053	-0.063	-0.009	-0.012	-0.007
	[-1.43]	[-0.53]	[-0.62]	[-0.28]	[-0.24]	[-0.14]
Merger Completed	0.067**	0.066***	0.081***	0.023***	0.026***	0.033***
	[2.19]	[3.88]	[4.52]	[3.03]	[3.13]	[3.57]
Diversify	-0.060**	-0.053	-0.051	-0.002	-0.004	-0.001
	[-2.07]	[-1.60]	[-1.61]	[-0.26]	[-0.29]	[-0.07]
Hostile	0.097**	0.133***	0.137**	0.049*	0.059**	0.061**
	[2.21]	[2.87]	[2.77]	[1.93]	[2.42]	[2.32]
High Tech	-0.006	-0.017	-0.022	-0.001	0.001	-0.001
	[-0.21]	[-0.35]	[-0.45]	[-0.13]	[0.14]	[-0.10]
Tender Offer	0.025	0.037	0.040	0.010	0.014	0.017
	[0.72]	[1.30]	[1.42]	[0.78]	[1.14]	[1.45]
Stock Issue	0.042	-0.028	-0.013	0.018	0.008	0.014
	[0.49]	[-0.32]	[-0.14]	[0.67]	[0.26]	[0.41]
Relative Deal Size	-0.001	0.001	0.002	0.000	-0.003	-0.002
	[-0.05]	[0.07]	[0.27]	[0.09]	[-0.73]	[-0.71]
Acquiror log(Asset)	0.030***	0.035***	0.035***	-0.007***	-0.010***	-0.009***
	[5.52]	[4.13]	[3.85]	[-3.26]	[-3.58]	[-3.11]
Acquiror Tobin's Q	0.000	0.003	0.001	0.001	0.002	0.001
	[0.06]	[0.74]	[0.22]	[0.83]	[1.13]	[0.41]
Acquiror Book Leverage	-0.022	-0.035	-0.041	-0.010	-0.003	-0.014
	[-0.33]	[-0.56]	[-0.68]	[-0.54]	[-0.16]	[-0.76]
Acquiror CF/Assets	10.253*	6.820	8.928	-1.852	-1.670	-0.457
	[1.83]	[1.00]	[1.22]	[-0.53]	[-0.45]	[-0.11]
Target log(Asset)	-0.035***	-0.049***	-0.049***	0.007**	0.007**	0.007*
	[-4.22]	[-5.31]	[-4.97]	[2.71]	[2.46]	[2.05]
Target Tobin's Q	-0.004	-0.012	-0.010	-0.001	-0.002	-0.001
	[-0.58]	[-1.71]	[-1.31]	[-0.86]	[-1.64]	[-0.51]
Target Book Leverage	-0.053	-0.025	-0.020	0.003	-0.010	-0.009
	[-1.36]	[-0.45]	[-0.33]	[0.21]	[-0.70]	[-0.51]
Target CF/Assets	-2.887	-5.275	-3.987	-0.756	-1.444	-0.619
	[-0.79]	[-1.59]	[-1.27]	[-0.74]	[-1.51]	[-0.66]
Intercept	0.250*	0.256*	0.245*	-0.016	-0.011	-0.018
	[1.86]	[2.05]	[1.90]	[-0.43]	[-0.49]	[-1.45]
Model	Linear	Linear	Linear	Linear	Linear	Linear
Exclude Vertical Merger	No	No	No	No	No	No
Acquiror-Target Weighting	Equal	Equal	Equal	Value	Value	Value
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	629	629	628	629	629	628
R-squared	0.296	0.311	0.303	0.244	0.279	0.260

Dependent Variables	CAR [t-1, t] (VII)	CAR [t-1, t+1] (VIII)	CAR [t-1, t+2] (IX)	CAR [t-1, t] (X)	CAR [t-1, t+1] (XI)	CAR [t-1, t+2] (XII)
TF Distance	-0.062	-0.193	-0.203	-0.139***	-0.176***	-0.169***
	[-0.26]	[-0.82]	[-0.89]	[-3.08]	[-4.05]	[-3.17]
TNIC Score	-0.112*	-0.089	-0.092	-0.012	-0.017	-0.009
	[-2.02]	[-0.86]	[-0.90]	[-0.35]	[-0.34]	[-0.17]
Merger Completed	0.073**	0.073***	0.089***	0.024***	0.027***	0.035***
	[2.35]	[4.20]	[5.05]	[3.01]	[3.63]	[4.43]
Diversify	-0.053*	-0.044	-0.044	0.005	0.000	0.002
	[-1.97]	[-1.37]	[-1.40]	[0.59]	[0.02]	[0.13]
Hostile	0.092*	0.145**	0.149**	0.047	0.062**	0.064**
	[1.85]	[2.80]	[2.77]	[1.65]	[2.26]	[2.24]
High Tech	-0.011	-0.028	-0.033	-0.002	-0.001	-0.004
	[-0.35]	[-0.56]	[-0.65]	[-0.20]	[-0.15]	[-0.41]
Tender Offer	0.030	0.024	0.026	0.012	0.013	0.015
	[0.83]	[0.78]	[0.85]	[0.82]	[0.97]	[1.23]
Stock Issue	0.000	-0.065	-0.048	0.010	0.000	0.008
	[0.00]	[-0.77]	[-0.50]	[0.36]	[0.01]	[0.21]
Relative Deal Size	0.005	-0.005	-0.003	0.002	-0.003	-0.002
	[0.40]	[-0.47]	[-0.37]	[0.30]	[-0.64]	[-0.56]
Acquiror log(Asset)	0.028***	0.032***	0.032***	-0.007***	-0.011***	-0.010***
	[5.63]	[3.92]	[3.68]	[-3.38]	[-3.72]	[-3.27]
Acquiror Tobin's Q	0.001	0.004	0.002	0.001	0.002	0.001
	[0.13]	[0.92]	[0.37]	[0.94]	[1.22]	[0.50]
Acquiror Book Leverage	-0.047	-0.050	-0.055	-0.015	-0.006	-0.016
	[-0.72]	[-0.77]	[-0.88]	[-0.74]	[-0.32]	[-0.86]
Acquiror CF/Assets	13.635*	10.255	13.375	-1.802	-1.810	-0.274
	[1.99]	[1.31]	[1.65]	[-0.53]	[-0.53]	[-0.07]
Target log(Asset)	-0.033***	-0.049***	-0.049***	0.007**	0.008**	0.008**
	[-3.91]	[-5.18]	[-4.86]	[2.80]	[2.51]	[2.08]
Target Tobin's Q	-0.003	-0.011	-0.009	-0.001	-0.002	-0.001
	[-0.47]	[-1.61]	[-1.18]	[-0.86]	[-1.58]	[-0.39]
Target Book Leverage	-0.039	-0.024	-0.017	0.003	-0.011	-0.009
	[-0.88]	[-0.45]	[-0.31]	[0.18]	[-0.74]	[-0.49]
Target CF/Assets	-2.474	-4.771	-3.477	-0.738	-1.379	-0.553
	[-0.74]	[-1.54]	[-1.18]	[-0.73]	[-1.48]	[-0.61]
Intercept	0.219*	0.257**	0.247*	-0.029	-0.018	-0.024**
	[1.83]	[2.09]	[1.97]	[-0.97]	[-1.06]	[-2.80]
Model	Linear	Linear	Linear	Linear	Linear	Linear
Exclude Vertical Merger	Yes	Yes	Yes	Yes	Yes	Yes
Acquiror-Target Weighting	Equal	Equal	Equal	Value	Value	Value
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	611	611	610	611	611	610
R-squared	0.304	0.317	0.309	0.255	0.288	0.268

Table 6. Divestiture and Survival

This table shows the relationship between the likelihood of post-merger acquiror survival and the acquiror-target transfer learning-based distance (i.e., TF Distance) from 1988 to 2021 using the probit model. The dependent variables are indicator for post-merger acquiror survival (Columns I to III) that is one if the acquiror exists and has not been acquired as a target (a target in the SDC database) for at least 10 years after a merger (Columns I and II) or at least 15 years after a merger (Column III), and an indicator for divestiture, which is one if the acquiror appears as target for divestiture within 10 years after a merger (Column IV). The key independent variable is TF Distance. In all regressions, we control for TNIC Score, Merger Completed, Diversify, Hostile, High Tech, Tender Offer, Stock Deal, Relative Deal Size, acquirors' and targets' Firm Size, Tobin's Q, Book Leverage, Cash Flow to Assets Ratio, and year- and industry fixed effects. The t-statistics are shown in square brackets, and standard errors are clustered at the year level. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Panel A.

Dependent Variables	Survive 10 Years	Survive 10 Years	Survive 15 Years	Divestiture Within 10 Years
	(I)	(II)	(III)	(IV)
TF Distance	-3.762**	-2.847**	-6.528*	8.988**
	[-1.97]	[-2.08]	[-1.79]	[2.02]
TNIC Score	-0.341	-0.946	1.561	-3.750
	[-0.26]	[-0.59]	[0.66]	[-1.37]
Merger Completed	0.059	-0.059	-0.256	1.349*
	[0.30]	[-0.27]	[-0.56]	[1.75]
Diversify	-0.741*	-0.298	-1.303*	-6.087***
	[-1.83]	[-0.76]	[-1.88]	[-9.36]
Hostile	0.396	0.343	1.185	
	[0.73]	[0.59]	[1.22]	
High Tech	-0.916*	-1.290	-1.576	-1.528*
	[-1.81]	[-1.63]	[-1.54]	[-1.96]
Tender Offer	0.492	0.861***	0.058	-1.164**
	[1.63]	[2.95]	[0.09]	[-2.35]
Stock Issue	0.388	0.371		
	[0.63]	[0.65]		
Relative Deal Size	0.069	0.056	0.183**	-2.398**
	[1.30]	[1.14]	[2.36]	[-2.05]
Acquiror log(Asset)	-0.004	-0.048	0.321*	0.008
	[-0.06]	[-0.57]	[1.94]	[0.05]
Acquiror Tobin's Q	-0.026	-0.023	-0.059	-1.373**
	[-0.66]	[-0.57]	[-0.87]	[-2.33]
Acquiror Book Leverage	-0.425	-0.150	-1.597	0.184
	[-0.58]	[-0.17]	[-1.35]	[0.11]
Acquiror CF/Assets	-131.126	-101.470	-277.306	-93.236
	[-1.48]	[-0.53]	[-1.42]	[-0.64]
Target log(Asset)	0.038	0.103	-0.142	0.132
	[0.40]	[1.01]	[-0.86]	[0.70]
Target Tobin's Q	0.071***	0.125***	-0.165	0.234
	[2.78]	[2.86]	[-1.13]	[1.27]
Target Book Leverage	0.088	-0.342	0.114	-3.305*
	[0.16]	[-0.39]	[0.10]	[-1.85]
Target CF/Assets	-4.426***	10.995	-5.376***	31.006*
	[-3.96]	[0.64]	[-3.98]	[1.74]
Intercept	-7.404***	-7.485***	-10.460***	1.506
	[-4.46]	[-3.97]	[-6.67]	[0.33]
Model	Probit	Probit	Probit	Probit
Sample	Deal size10m	Deal size50m	Deal size10m	Deal size10m
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	589	439	367	192
Pseudo R2	0.3136	0.3315	0.4619	0.3984

Panel B. Marginal Effects

Dependent Variables	Survive 10 Years	Survive 10 Years	Survive 15 Years	Divestiture Within 10 Years
	(I)	(II)	(III)	(IV)
TF Distance	-0.549**	-0.411**	-0.530*	0.755**
	[-1.98]	[-2.05]	[-1.76]	[2.02]
TNIC Score	-0.050	-0.137	0.127	-0.315
	[-0.26]	[-0.59]	[0.68]	[-1.37]
Merger Completed	0.009	-0.009	-0.021	0.113*
	[0.30]	[-0.27]	[-0.57]	[1.75]
Diversify	-0.108*	-0.043	-0.106*	-0.512***
	[-1.89]	[-0.77]	[-1.82]	[-9.36]
Hostile	0.058	0.049	0.096	
	[0.73]	[0.59]	[1.23]	
High Tech	-0.134*	-0.186*	-0.128	-0.128*
	[-1.83]	[-1.71]	[-1.61]	[-1.96]
Tender Offer	0.072	0.124***	0.005	-0.098**
	[1.64]	[3.14]	[0.09]	[-2.35]
Stock Issue	0.057	0.054		
	[0.64]	[0.65]		
Relative Deal Size	0.010	0.008	0.015**	-0.202**
	[1.28]	[1.14]	[2.47]	[-2.05]
Acquiror log(Asset)	-0.001	-0.007	0.026**	0.001
	[-0.06]	[-0.58]	[2.14]	[0.05]
Acquiror Tobin's Q	-0.004	-0.003	-0.005	-0.115**
	[-0.66]	[-0.57]	[-0.89]	[-2.33]
Acquiror Book Leverage	-0.062	-0.022	-0.130	0.015
	[-0.58]	[-0.17]	[-1.31]	[0.11]
Acquiror CF/Assets	-19.145	-14.641	-22.501	-7.836
	[-1.48]	[-0.53]	[-1.37]	[-0.64]
Target log(Asset)	0.006	0.015	-0.012	0.011
	[0.40]	[1.00]	[-0.88]	[0.70]
Target Tobin's Q	0.010***	0.018***	-0.013	0.020
	[2.80]	[2.96]	[-1.14]	[1.27]
Target Book Leverage	0.013	-0.049	0.009	-0.278*
	[0.16]	[-0.39]	[0.10]	[-1.85]
Target CF/Assets	-0.646***	1.586	-0.436***	2.606*
	[-4.02]	[0.64]	[-3.51]	[1.74]
Model	Probit	Probit	Probit	Probit
Sample	Deal size10m	Deal size50m	Deal size10m	Deal size10m
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	589	439	367	192
Pseudo R2	0.3136	0.3315	0.4619	0.3984

Table 7. Post-Merger Growth of the Combined Firm

This table shows the relationship between firm growth and the acquiror-target transfer learning-based distance (i.e., TF Distance) from 1988 to 2021. The dependent variables are 10-year asset growth (Columns I to II) and market value growth (Columns III and IV). Asset growth (market value growth) is the log difference of total assets (market value) 10 years after the merger and one year after the merger. Market value is total assets plus the market value of equity (total shares outstanding times year-end stock price). The key independent variable is TF Distance. We require the deal transaction value to be greater than $10 million. In all regressions, we control for TNIC Score, Merger Completed, Diversify, Hostile, High Tech, Tender Offer, Stock Deal, Relative Deal Size, acquirors' and targets' Firm Size, Tobin's Q, Book Leverage, Cash Flow to Assets Ratio, and year- and industry fixed effects. The t-statistics are shown in square brackets, and standard errors are clustered at the year level. ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Dependent Variables	Asset Growth		Market Value Growth	
Post-Merger Time Horizon	10 Years	10 Years	10 Years	10 Years
	(I)	(II)	(III)	(IV)
TF Distance	-2.178***	-2.332***	-2.345***	-2.599***
	[-3.06]	[-3.22]	[-3.02]	[-3.15]
TNIC Score	0.041	-0.018	0.061	-0.095
	[0.09]	[-0.04]	[0.10]	[-0.15]
Merger Completed	0.154	0.170	0.177	0.171
	[0.91]	[1.00]	[1.31]	[1.22]
Diversify	0.070	0.117	0.028	0.068
	[0.43]	[0.66]	[0.19]	[0.42]
Hostile	-0.036	0.027	-0.143	-0.101
	[-0.19]	[0.12]	[-0.94]	[-0.60]
High Tech	0.081	0.043	-0.028	-0.031
	[0.66]	[0.33]	[-0.21]	[-0.22]
Tender Offer	-0.058	-0.071	0.080	0.072
	[-0.57]	[-0.64]	[0.65]	[0.58]
Stock Issue	0.134	0.080	0.154	0.102
	[0.39]	[0.23]	[0.39]	[0.25]
Relative Deal Size	-0.047	-0.051	-0.011	-0.010
	[-1.24]	[-1.23]	[-0.30]	[-0.27]
Acquiror log(Asset)	0.002	-0.001	-0.039	-0.040
	[0.06]	[-0.03]	[-1.40]	[-1.42]
Acquiror Tobin's Q	0.013	0.013	-0.035**	-0.032**
	[0.71]	[0.78]	[-2.39]	[-2.39]
Acquiror Book Leverage	0.137	-0.003	-0.134	-0.256
	[0.51]	[-0.01]	[-0.52]	[-0.90]
Acquiror CF/Assets	3.101**	2.965**	3.186**	3.198**
	[2.37]	[2.29]	[2.57]	[2.51]
Target log(Asset)	-0.034	-0.036	-0.013	-0.007
	[-0.97]	[-1.01]	[-0.43]	[-0.21]
Target Tobin's Q	-0.008	-0.001	-0.012**	-0.017
	[-1.14]	[-0.09]	[-2.12]	[-0.90]
Target Book Leverage	0.037	0.067	0.028	0.029
	[0.21]	[0.38]	[0.13]	[0.14]
Target CF/Assets	1.070***	5.338	1.875***	-1.983
	[3.54]	[0.70]	[5.68]	[-0.17]
Intercept	0.755	0.778	0.889	0.816
	[1.45]	[1.45]	[1.51]	[1.28]
Sample	All	Horizontal	All	Horizontal
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	407	393	404	390
R-squared	0.306	0.310	0.342	0.333

Table 8. Learning by Repeat Acquirors: Post-Merger Growth

This table shows the relationship between the learning effect of repeated acquirors for firm growth and the acquiror-target transfer learning-based distance (i.e., TF Distance) from 1988 to 2021. The dependent variables are 10-year asset growth (Columns I to VI) and market value growth (Columns VII and XII). Asset growth (market value growth) is the log difference of total assets (market value) 10 years after the merger and one year after the merger. Market value is total assets plus the market value of equity (total shares outstanding times year-end stock price). The key independent variables are Prior Deals Within 10 Years, Completed Deals Within 10 Years, Prior Failed Deals Within 10 Years, and their interaction with TF Distance. Prior Deals Within 10 Years, Completed Deals Within 10 Years, and Prior Failed Deals Within 10 Years are the number of deals, completed deals, and failed deals within the past 10 years as of the current merger's announcement date. Columns I, III, V, VII, IX, XI consider all types of mergers, and Columns II, IV, VI, VIII, X, XII consider horizontal mergers (Hoberg-Phillips vertscore < 0.02). We require the deal transaction value to be greater than $10 million. In all regressions, we control for TNIC Score, Merger Completed, Diversify, Hostile, High Tech, Tender Offer, Stock Deal, Relative Deal Size, acquirors' and targets' Firm Size, Tobin's Q, Book Leverage, Cash Flow to Assets Ratio, and year- and industry fixed effects; The t-statistics are shown in square brackets, and standard errors are clustered at the year level. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Dependent Variables	10-Year Asset Growth					
	(I)	(II)	(III)	(IV)	(V)	(VI)
Prior Deals Within 10 Years	0.063**	0.060**				
x TF Distance	[2.46]	[2.19]				
Prior Completed Deals Within 10 Years			0.064**	0.060*		
x TF Distance			[2.28]	[2.04]		
Prior Failed Deals Within 10 Years					0.349***	0.344***
x TF Distance					[2.98]	[2.86]
Prior Deals Within 10 Years	0.009	0.008				
	[1.69]	[1.35]				
Prior Completeds Deal Within 10 Years			0.010*	0.009		
			[1.84]	[1.47]		
Prior Failed Deals Within 10 Years					0.039*	0.039
					[1.87]	[1.64]
TF Distance	-3.085***	-3.146***	-2.888***	-2.968***	-3.086***	-3.179***
	[-3.53]	[-3.54]	[-3.39]	[-3.44]	[-3.68]	[-3.66]
TNIC Score	0.007	-0.053	0.024	-0.039	0.008	-0.039
	[0.01]	[-0.11]	[0.05]	[-0.08]	[0.02]	[-0.08]
Merger Completed	0.175	0.187	0.172	0.185	0.158	0.168
	[1.03]	[1.10]	[1.01]	[1.08]	[0.94]	[1.00]
Diversify	0.049	0.093	0.055	0.098	0.026	0.074
	[0.30]	[0.53]	[0.34]	[0.56]	[0.16]	[0.42]
Hostile	-0.017	0.045	-0.020	0.040	-0.039	0.020
	[-0.09]	[0.21]	[-0.11]	[0.19]	[-0.22]	[0.10]
High Tech	0.090	0.055	0.088	0.051	0.102	0.072
	[0.74]	[0.42]	[0.72]	[0.39]	[0.84]	[0.56]
Tender Offer	-0.070	-0.085	-0.069	-0.082	-0.067	-0.079
	[-0.70]	[-0.76]	[-0.68]	[-0.74]	[-0.67]	[-0.71]
Stock Issue	0.133	0.081	0.131	0.079	0.106	0.051
	[0.40]	[0.24]	[0.39]	[0.24]	[0.32]	[0.15]
Relative Deal Size	-0.039	-0.043	-0.040	-0.044	-0.043	-0.047
	[-0.99]	[-1.00]	[-1.02]	[-1.03]	[-1.14]	[-1.13]
Acquiror log(Asset)	0.011	0.009	0.005	0.002	0.022	0.020
	[0.32]	[0.25]	[0.14]	[0.06]	[0.64]	[0.57]
Acquiror Tobin's Q	0.020	0.020	0.018	0.017	0.023	0.022
	[1.01]	[1.05]	[0.90]	[0.96]	[1.13]	[1.16]
Acquiror Book Leverage	0.120	-0.011	0.126	-0.007	0.116	-0.009
	[0.45]	[-0.04]	[0.47]	[-0.02]	[0.44]	[-0.03]
Acquiror CF/Assets	3.292**	3.110**	3.260**	3.096**	3.564**	3.354**
	[2.60]	[2.50]	[2.56]	[2.46]	[2.69]	[2.59]
Target log(Asset)	-0.036	-0.039	-0.034	-0.037	-0.041	-0.042
	[-1.06]	[-1.12]	[-0.99]	[-1.05]	[-1.21]	[-1.22]
Target Tobin's Q	-0.012	-0.003	-0.011	-0.003	-0.011	-0.002
	[-1.43]	[-0.22]	[-1.37]	[-0.20]	[-1.29]	[-0.13]
Target Book Leverage	0.008	0.033	0.017	0.044	0.013	0.036
	[0.04]	[0.16]	[0.09]	[0.22]	[0.07]	[0.19]
Target CF/Assets	0.955***	6.772	0.982***	6.302	0.944***	7.126
	[3.11]	[0.86]	[3.19]	[0.80]	[3.08]	[0.96]
Intercept	0.617	0.662	0.677	0.718	0.508	0.543
	[1.11]	[1.15]	[1.24]	[1.26]	[0.91]	[0.94]
Exclude Vertical Merger	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	407	393	407	393	407	393
R-squared	0.317	0.320	0.314	0.317	0.325	0.328

Dependent Variables	10-Year Market Value Growth					
	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)
Prior Deals Within 10 Years	0.037**	0.033*				
x TF Distance	[2.30]	[2.01]				
Prior Completed Deals Within 10 Years			0.034*	0.029		
x TF Distance			[1.90]	[1.68]		
Prior Failed Deals Within 10 Years					0.271*	0.274
x TF Distance					[1.92]	[1.62]
Prior Deals Within 10 Years	0.007	0.007				
	[1.36]	[1.21]				
Prior Completeds Deal Within 10 Years			0.008	0.007		
			[1.31]	[1.18]		
Prior Failed Deals Within 10 Years					0.041	0.043
					[1.48]	[1.27]
TF Distance	-2.811***	-2.991***	-2.656***	-2.844***	-3.048***	-3.278***
	[-3.12]	[-3.34]	[-2.90]	[-3.15]	[-3.52]	[-3.56]
TNIC Score	0.041	-0.112	0.061	-0.095	-0.008	-0.154
	[0.07]	[-0.17]	[0.10]	[-0.15]	[-0.01]	[-0.23]
Merger Completed	0.184	0.175	0.181	0.172	0.179	0.169
	[1.36]	[1.24]	[1.33]	[1.21]	[1.34]	[1.21]
Diversify	0.011	0.050	0.016	0.055	-0.010	0.030
	[0.07]	[0.31]	[0.11]	[0.34]	[-0.07]	[0.20]
Hostile	-0.137	-0.099	-0.140	-0.104	-0.141	-0.102
	[-0.91]	[-0.58]	[-0.92]	[-0.60]	[-0.93]	[-0.61]
High Tech	-0.023	-0.024	-0.023	-0.025	-0.019	-0.013
	[-0.18]	[-0.17]	[-0.18]	[-0.18]	[-0.15]	[-0.10]
Tender Offer	0.072	0.064	0.074	0.067	0.071	0.062
	[0.60]	[0.52]	[0.61]	[0.54]	[0.59]	[0.51]
Stock Issue	0.146	0.096	0.143	0.092	0.140	0.089
	[0.37]	[0.24]	[0.36]	[0.23]	[0.36]	[0.22]
Relative Deal Size	-0.007	-0.007	-0.009	-0.008	-0.007	-0.006
	[-0.18]	[-0.16]	[-0.22]	[-0.21]	[-0.17]	[-0.14]
Acquiror log(Asset)	-0.043	-0.045	-0.045	-0.047	-0.034	-0.035
	[-1.17]	[-1.22]	[-1.27]	[-1.33]	[-1.06]	[-1.09]
Acquiror Tobin's Q	-0.032*	-0.031*	-0.034**	-0.032**	-0.028	-0.026
	[-1.96]	[-2.06]	[-2.13]	[-2.25]	[-1.66]	[-1.68]
Acquiror Book Leverage	-0.135	-0.252	-0.133	-0.251	-0.141	-0.252
	[-0.51]	[-0.86]	[-0.50]	[-0.86]	[-0.53]	[-0.88]
Acquiror CF/Assets	3.371**	3.372**	3.347**	3.360**	3.524***	3.496***
	[2.79]	[2.68]	[2.74]	[2.63]	[3.05]	[2.92]
Target log(Asset)	-0.011	-0.005	-0.011	-0.004	-0.015	-0.008
	[-0.40]	[-0.17]	[-0.37]	[-0.15]	[-0.53]	[-0.28]
Target Tobin's Q	-0.013**	-0.017	-0.013**	-0.017	-0.013**	-0.017
	[-2.28]	[-0.96]	[-2.23]	[-0.95]	[-2.36]	[-0.94]
Target Book Leverage	0.020	0.022	0.026	0.029	0.012	0.007
	[0.09]	[0.10]	[0.12]	[0.13]	[0.05]	[0.03]
Target CF/Assets	1.827***	-1.758	1.843***	-1.968	1.800***	-1.193
	[5.51]	[-0.14]	[5.58]	[-0.16]	[5.37]	[-0.10]
Intercept	0.862	0.804	0.889	0.829	0.758	0.688
	[1.32]	[1.16]	[1.38]	[1.21]	[1.22]	[1.03]
Exclude Vertical Merger	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	404	390	404	390	404	390
R-squared	0.345	0.335	0.344	0.334	0.349	0.339

Table 9. Learning by Repeat Acquirors: TF Distance

This table shows the relationship between the learning effect of repeated acquirors for current TF Distance and recent deal (within 10 years) from 1988 to 2021. The dependent variable is TF Distance. The key independent variables are an indicator for having a past merger deal within the most recent 10 years (Most Recent Deal Within 10 Years), an indicator for the most recent deal within 10 years that is completed (Most Recent Deal Within 10 Years: Completed), and an indicator for the most recent deal within 10 years that is incomplete (Most Recent Deal Within 10 Years: Incomplete). Columns I, III, and V consider all types of mergers, and Columns II, IV, and VI consider horizontal mergers (Hoberg-Phillips vertscore < 0.02). We require the deal transaction value to be greater than $10 million. All regressions are controlled for TNIC Score, Merger Completed, Diversify, Hostile, High Tech, Tender Offer, Stock Deal, Relative Deal Size, acquirors' and targets' Firm Size, Tobin's Q, Book Leverage, Cash Flow to Assets Ratio, and year- and industry fixed effects; The t-statistics are shown in square brackets, and standard errors are clustered at the year level. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	TF Distance					
	(I)	(II)	(III)	(IV)	(V)	(VI)
Most Recent Deal Within 10 Years	-0.017**	-0.015*				
	[-2.05]	[-1.82]				
Most Recent Deal Within 10 Years: Completed			0.001	0.002		
			[0.39]	[0.47]		
Most Recent Deal Within 10 Years: Incomplete					-0.011**	-0.010*
					[-2.00]	[-1.87]
TNIC Score	-0.005	-0.011	-0.005	-0.011	-0.007	-0.013
	[-0.26]	[-0.55]	[-0.25]	[-0.51]	[-0.35]	[-0.60]
Merger Completed	0.003	0.002	0.003	0.002	0.003	0.002
	[0.63]	[0.40]	[0.59]	[0.35]	[0.57]	[0.34]
Diversify	0.078***	0.082***	0.078***	0.082***	0.077***	0.081***
	[9.60]	[10.51]	[9.43]	[10.32]	[9.49]	[10.41]
Hostile	0.009	0.003	0.010	0.004	0.010	0.004
	[1.17]	[0.39]	[1.24]	[0.45]	[1.29]	[0.51]
High Tech	0.005	0.004	0.005	0.004	0.005	0.004
	[0.59]	[0.57]	[0.57]	[0.54]	[0.60]	[0.58]
Tender Offer	0.000	0.000	0.000	0.000	0.000	0.000
	[0.07]	[0.05]	[0.04]	[0.04]	[0.02]	[0.03]
Stock Issue	-0.003	-0.011	-0.002	-0.010	-0.003	-0.011
	[-0.30]	[-1.01]	[-0.20]	[-0.91]	[-0.29]	[-0.97]
Relative Deal Size	0.002	0.002	0.002	0.002	0.002	0.002
	[1.57]	[1.21]	[1.47]	[1.14]	[1.48]	[1.16]
Acquiror log(Asset)	-0.000	-0.001	-0.000	-0.001	-0.001	-0.001
	[-0.14]	[-0.62]	[-0.36]	[-0.79]	[-0.56]	[-0.99]
Acquiror Tobin's Q	0.001	0.002	0.001	0.002	0.001	0.002
	[1.34]	[1.57]	[1.39]	[1.64]	[1.27]	[1.52]
Acquiror Book Leverage	-0.019	-0.022*	-0.020	-0.022*	-0.021*	-0.023*
	[-1.65]	[-1.80]	[-1.69]	[-1.82]	[-1.77]	[-1.90]
Acquiror CF/Assets	0.044	0.063	-0.004	0.021	0.006	0.030
	[0.49]	[0.79]	[-0.04]	[0.30]	[0.07]	[0.44]
Target log(Asset)	0.001	0.001	0.001	0.001	0.001	0.001
	[0.61]	[0.70]	[0.64]	[0.72]	[0.65]	[0.74]
Target Tobin's Q	-0.000	-0.000	-0.000	-0.000	-0.000	0.000
	[-0.44]	[-0.28]	[-0.15]	[-0.01]	[-0.15]	[0.00]
Target Book Leverage	0.002	0.006	0.002	0.006	0.002	0.006
	[0.29]	[0.88]	[0.21]	[0.81]	[0.25]	[0.87]
Target CF/Assets	-0.137***	-0.064	-0.135***	-0.057	-0.134***	-0.050
	[-6.57]	[-0.46]	[-6.20]	[-0.40]	[-6.31]	[-0.36]
Intercept	-0.250***	-0.245***	-0.265***	-0.258***	-0.261***	-0.254***
	[-9.56]	[-8.94]	[-10.47]	[-9.76]	[-10.57]	[-9.86]
Exclude Vertical Mergers	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	983	958	983	958	983	958
R-squared	0.512	0.523	0.510	0.522	0.512	0.524

Table 10. M&A Completion: Impact of Employee Bargaining Power (Right-to-Work)

This table shows the impact of states' adoption of Right-to-Work laws on the relationship between the likelihood of M&A deal completion and various firm-to-firm production process heterogeneity measures from 1988 to 2021. Panel A shows the years when various states adopted Right-to-Work law. Panel B shows the regression results. The dependent variable is an indicator for M&A deal completion (Columns I to VII), which is one if the deal is completed and zero otherwise. The key independent variables are firm-to-firm CD Distance (Column I), Linear Distance (Column II), XGB Distance (Column III), FC Distance (Column IV), TF Distance (Columns V to VII), and interactions of these distances with the post-Right-to-Work indicator. Firm-to-firm distances are estimated using neighboring firms with a Hoberg-Phillips similarity score greater than 0.09. Columns (I) to (V) consider deal sizes greater than $100 million. Columns (VI) and (VII) consider deal sizes greater than $200 million. Column (VII) considers horizontal mergers only (Hoberg-Phillips vertscore less than 0.02). Year and industry fixed effects are controlled but not reported in this table. The t-statistics are shown in square brackets, and standard errors are clustered at the year level. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Panel A. Right-to-Work Adoption Years

State	Right-to-Work Year	State	Right-to-Work Year
Alabama	1953	Montana	
Alaska		Nebraska	1946
Arizona	1946	Nevada	1952
Arkansas	1944	New Hampshire	
California		New Jersey	
Colorado		New Mexico	
Conneticut		New York	
Delaware		North Carolina	1947
District of Columbia		North Dakota	1947
Florida	1944	Ohio	
Georgia	1947	Oklahoma	2001
Hawaii		Oregon	
Idaho	1985	Pennsylvania	
Illinois		Rhode Island	
Indiana	2012	South Carolina	1954
Iowa	1947	South Dakota	1947
Kansas	1958	Tennessee	1947
Kentucky	2017	Texas	1947
Louisiana	1976	Utah	1955
Maine		Vermont	
Maryland		Virginia	1947
Massachusetts		Washington	
Michigan	2012 (repealed 2023)	West Virginia	2016
Minnesota		Wisconsin	2015
Mississippi	1954	Wyoming	1963
Missouri	2017 (repealed 2018)		

Panel B. Results

Dependent Variable	Completed						
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
CD Distance	-0.252						
	[-0.65]						
CD Distance x Post-Right-to-Work	-0.351						
	[-0.71]						
Linear Distance		-0.115					
		[-0.85]					
Linear Distance x Post-Right-to-Work		0.200					
		[1.25]					
XGB Distance			-0.318*				
			[-1.85]				
XGB Distance x Post-Right-to-Work			0.842***				
			[2.63]				
FC Distance				-0.302			
				[-0.96]			
FC Distance x Post-Right-to-Work				-0.467			
				[-0.75]			
TF Distance					-2.045***	-2.191**	-2.106**
					[-3.02]	[-2.46]	[-2.35]
TF Distance x Post-Right-to-Work					1.725*	2.216**	2.081*
					[1.87]	[1.99]	[1.82]
Post-Right-to-Work	0.159	0.085	0.315*	0.111	0.091	-0.003	-0.008
	[1.06]	[0.60]	[1.91]	[0.79]	[0.65]	[-0.02]	[-0.05]
TNIC Score	1.731***	2.282***	2.330***	1.940***	2.077***	1.842***	1.721***
	[3.05]	[3.69]	[3.74]	[2.94]	[3.42]	[2.81]	[2.86]
Intercept	3.546***	3.284***	3.268***	3.250***	3.435***	3.238***	4.602***
	[4.96]	[5.10]	[4.87]	[4.64]	[5.07]	[4.14]	[5.54]
Model	Probit	Probit	Probit	Probit	Probit	Probit	Probit
Deal Size Cutoff	100M+	100M+	100M+	100M+	100M+	200M+	200M+
Merger Type	All	All	All	All	All	All	Horizontal
FIC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	748	773	773	773	773	626	609
R-Squared	0.2109	0.214	0.2202	0.2157	0.2215	0.2787	0.2743

Appendix

Appendix A. Background Information on Transfer Learning

Transfer learning and domain adaptation are processes in which knowledge acquired in one setting is employed to improve generalization in another (Goodfellow, Bengio, and Courville, 2016). This technique is particularly beneficial in situations where a model developed for one task is adapted for another, especially when the second task has limited data or when training for it is time-consuming.

The essence of transfer learning lies in its efficiency: a model trained on a large dataset (like ImageNet for computer vision or extensive text corpora for natural language processing) is then adjusted for a specific, data-scarce task. The procedure typically involves first training a base network on a primary dataset and task, then repurposing the learned features for a second target network, tailored to a different dataset and task. This not only saves time but also leverages the extensive knowledge already captured by the model.

In fields such as computer vision and natural language processing, the impact of transfer learning has been substantial. It has enabled the adaptation of models for more specialized tasks, even with limited data. Recent applications extend to Language Learning Models like Google's word2vec, BERT, and OpenAI's GPT series, showcasing its versatility and effectiveness across different domains. For further details refer to Pan and Yang (2009), a classic review article.

Our method leverages transfer learning by concentrating on the transformed production factors of both acquiror and target, while disregarding the final production stage. This is achieved by examining the deep learning network's differences up to the penultimate layers, allowing the final layer to adapt (fine-tuning) to the target's specific input/output data (Figure 4). Recent studies on neural tangent kernels, such as those by Jacot, Gabriel, and Hongler (2018) and Roberts, Yaida, and Hanin (2022), describe the supervised learning of deep neural networks as a diffusion process. In this process, initially random model parameters (neural weights) are progressively updated from input to downstream layers through stochastic gradient optimization. To illustrate, Dertat (2017) provides images showing several representative layers of the VGG16 Image Classification network (see Figure A1). These images demonstrate how the original image is incrementally simplified until it reaches the classification (softmax) layer.

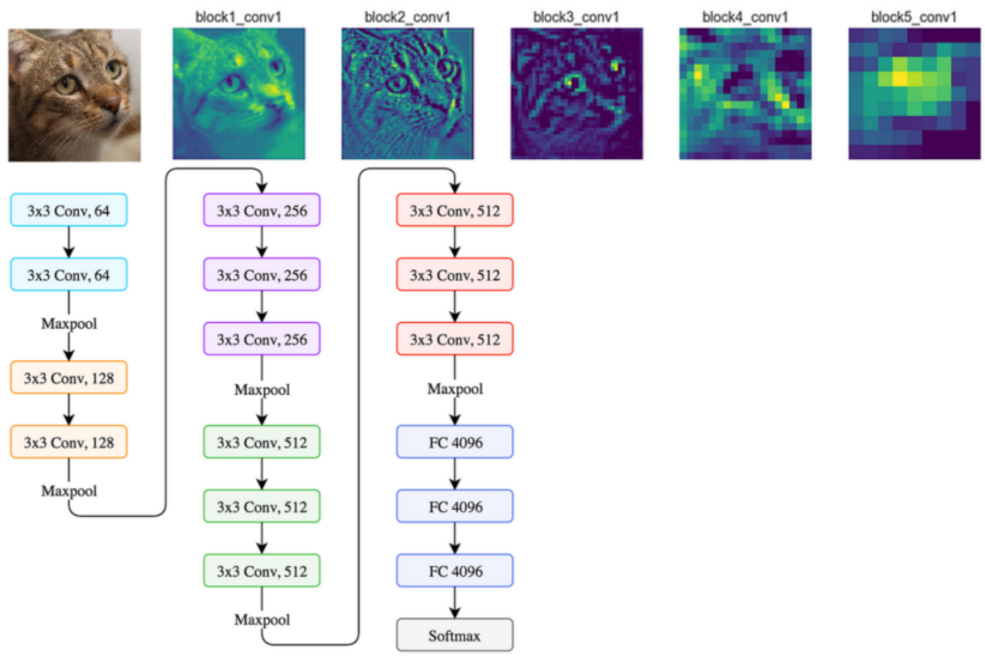
Our use of transfer learning, aimed at isolating transformed production factors from the input stage up to the penultimate layers, is driven by the concept of simplifying input data into more basic elements. Take, for example, a truck company case: here, the input data comprises factors such as capital and labor, whereas the production factors in the intermediate stages are elements like frames, doors, and wheels. The focus of our transfer learning-based approach on production process distance is to compare the differences in how the acquiror and target transform these input factors. Notably, it overlooks the variations in the final transformation from production factors to output, as these differences are generally easier to align during the integration phase of mergers.

As we demonstrate in subsequent sections of our main text, incorporating this economic perspective into our transfer learning approach significantly enhances the prediction accuracy and robustness of M&A outcomes, outperforming traditional methods like fully connected deep neural networks or other flexible machine learning techniques such as XGBoost.

Appendix B. The Fully Connected Layer Neural Network vs. the Transfer Learning-Based Neural Network

In this Appendix, we use a simplified two-factor linear production function to examine production decision differences between acquirors and targets across industries. Our analysis covers three industries: a baseline industry, one with similar latent production factors to the baseline but different factor weights, and another with distinct production factors. This

Figure A1.

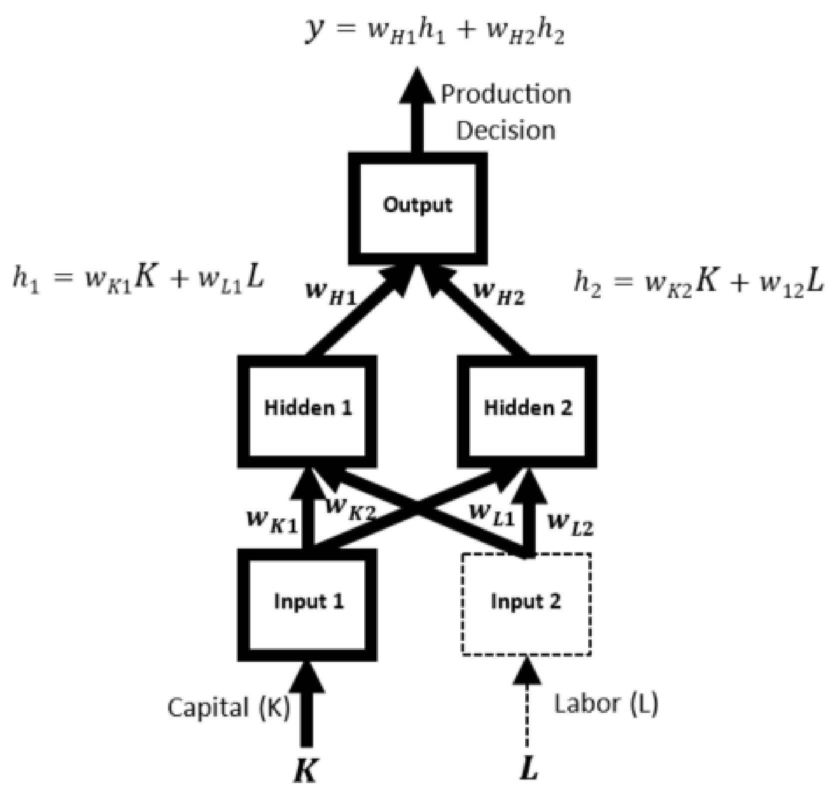


example shows that when acquirors and targets share common production factors, there may be discrepancies in fully connected neural network distance, but not significantly in transfer learning distance. However, when they differ in production factors, both fully connected neural networks and transfer learning ones reveal significant distances. Empirical results in the main text demonstrate that transfer learning-based distances are strong predictors of M&A likelihood. A large distance indicates major differences in production processes between acquirors and targets, which may be challenging to integrate without major adjustments to organizational rules, decision-making processes, and corporate culture, as noted by Sah and Stiglitz (1984) and Dessein and Santos (2006).

Model Setup

Consider an economy with firms that have two-factor, three-layer hierarchy linear production functions as shown in Figure B1.

Figure B1. Two-Factor Linear Production Functions



This figure shows the two-factor, three-layer hierarchy linear production function example. These firms take capital (K) and labor (L) as input, where these factors are assumed to be uniformly distributed, i.e., K∼ U(0,1) and L∼ U(0,1). Linear combinations of these factors are fed into the middle-level hierarchy (we refer to h_1 and h_2 as latent factors):

$$h_1 = w_{K1}K + w_{L1}L$$

$$h_2 = w_{K2}K + w_{L2}L$$

In turn, linear combinations of these middle-level hierarchical outputs are fed to the top hierarchy to make the final production output:

$$y(K,L) = w_{H1}h_1 + w_{H2}h_2,$$

which can be explicitly expressed in terms of input factors of capital and labor as

$$y(K,L) = w_K K + w_L L,$$

where

$$w_K \equiv w_{H1}w_{K1} + w_{H2}w_{K2}$$
$$w_L \equiv w_{H1}w_{L1} + w_{H2}w_{L2}.$$

In this economy, we consider three groups of firms (i.e., industries) based on the actual realization of outputs using input factors (K and L). The first group is the baseline industry whose output depends on capital input and random noise:

$$y_1(K,L) = K + \epsilon,$$

where

$$\epsilon \sim N(0,\sigma^2).$$

The second group's output also depends on capital input and random noise but the relationship between output and capital input differs from that of the baseline industry:

$$y_2(K, L) = -K + \epsilon.$$

The second industry reflects the group of firms that share production factors with the baseline industry, but differ in how these factors are reflected in the production output. The last group of firms has output that depends on the distinct factor from the baseline industry, i.e., the labor factor, and random noise:

$$y_3(K, L) = L + \epsilon.$$

The third industry represents firms that have different production factors and weights from those in the baseline industry.

Production Process Distance

Now we compute the production process distance using two different estimation approaches: one without accommodation (from latent input factors to final output), referred to as the fully connected neural network-based distance (FC Distance), and one with accommodation (from latent input factors to final output), referred to as the transfer learning-based distance (TF Distance).

The objective of optimizing a business organization is to adjust the weights of each factor, $w \equiv \{w_{K1}, w_{K2}, w_{L1}, w_{L2}, w_{H1}, w_{H2}\}$, such that the expected mean squared errors (MSE) between the model prediction, $y(K, L)$, and the actual observed outcome of the production, $y_i(K, L)$, are minimized

$$MSE_i = E\left[\int_0^1 \int_0^1 (y(K,L) - y_i(K,L))^2 dK dL\right].$$

It is straightforward to see that the MSE for the baseline industry is minimized when $w_1^* \equiv \{1,0,0,0,1,0\}$ with a resulting mean squared error of σ^2. Proofs are provided in the footnote.[23]

A particular interest arises in cross-industry M&As where acquirors and targets belong to different industries. It is straightforward to show that the mean squared errors of a firm in the baseline industry (group 1) and the other two industries (groups 2 and 3) are σ^2, $\frac{4}{3}+\sigma^2$, and $\frac{13}{6}+\sigma^2$, respectively. [24] That is, the integration cost is minimal for within-industry mergers ($MSE_{FC,11} = \sigma^2$). The next lowest integration cost is between industries that share the same production factors, i.e., between groups 1 and 2 ($MSE_{FC,12} = \frac{4}{3}+\sigma^2$). Finally, the integration cost is the highest between the firms in industries with different input factors, i.e., groups 1 and 3 ($MSE_{FC,13} = \frac{13}{6}+\sigma^2$).

In practice, an acquiring firm can reasonably accommodate a target firm's production process

[23] To see this, $MSE_1 = E\left[\int_0^1 \int_0^1 (y(K,L;w_1^*) - y_1(K,L))^2 dK dL\right] = E\left[\int_0^1 \int_0^1 (K-(K+\epsilon))^2 dK dL\right] = E\left[\int_0^1 \int_0^1 \epsilon^2 dK dL\right] = E\left[\epsilon^2\right] = \sigma^2$.

i.e., it cancels out the input factor, K, and is left with only the contribution from random noise. Note that this is not the only mean squared error minimizing solutions. Other solutions are $w_1^{**} \equiv \{\eta, 1-\eta, 0, 0, 1, 1\}$, $w_1^{***} \equiv \{1, 1, 0, 0, \eta, 1-\eta\}$, $w_1^{****} \equiv \{\alpha, \beta, 0, 0, \frac{1}{\alpha}, \frac{1}{\beta}\}$, where $0 \leq \eta \leq 1$, $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$. In this example, we will focus on w_1^* to simplify exposition.

[24] The MSE of group 1 model applied to group 1 input-output data is
$MSE_{FC,11} = E\left[\int_0^1 \int_0^1 (K-(K+\epsilon))^2 dK dL\right] = E\left[\int_0^1 \int_0^1 \epsilon^2 dK dL\right] = E\left[\epsilon^2\right] = \sigma^2$.

The MSE of the group 1 model applied to group 2 input-output data is
$MSE_{FC,12} = E\left[\int_0^1 \int_0^1 (4K^2 - 4K\epsilon + \epsilon^2) dK dL\right] = E\left[\frac{4}{3} - 2\epsilon + \epsilon^2\right] = \frac{4}{3}\sigma^2$.

The MSE of group 1 model applied to group 3 input-output data is
$MSE_{FC,13} = E\left[\int_0^1 \int_0^1 (K-(L+\epsilon))^2 dK dL\right] = E\left[\int_0^1 (\frac{4}{3} + L^2 + \epsilon^2 - L + \epsilon - 2L\epsilon) dL\right] = \frac{13}{6} + \sigma^2$.

when the underlying factors (latent factors in this model) significantly overlap. To reflect such flexibility in accommodating the target firm's production process, the second measure preserves organizational structure from the trained industry up to the middle-hierarchy (i.e., fix $w_{K1}, w_{K2}, w_{L1}, w_{L2}$) but allows accommodation (fine-tuning) at the top-hierarchy of the decision tree in the second industry (i.e., adjust w_{H1}, w_{H2}). It is clearly seen that the mean squared errors of a firm in the baseline industry and the other two groups with this additional layer extraction are σ^2, σ^2, $\frac{4}{3} + \sigma^2$, respectively.[25]

When layer adjustments (fine-tuning) are allowed, integration costs are generally lower than without such accommodations. The resulting integration costs are lowest under the layer adjustment for same industry mergers ($MSE_{TF,11} = \sigma^2$) and cross-industry mergers with the same production factors ($MSE_{TF,12} = \sigma^2$). The integration cost is highest between firms from industries with completely different production factor structures, i.e., groups 1 and 3 ($MSE_{TF,13} = \frac{13}{6} + \sigma^2$).

The above model further highlights the sensitivity of business integration cost with and without layer adjustment. The integration cost without layer adjustment is more sensitive to the difference in input factors across industries than those with layer extraction. The

25

The MSE of the group 1 model applied to group 1 input-output data is transfer learning w_{1*} to fit $y_1(K,L) = K + \epsilon$ using model $y(K,L;w_{(TF,1)} = w_1^*) = w_K K + w_L L = K$, which is the same as the no transfer learning case because the weights are already optimal, and no transfer learning is needed, i.e., $MSE_{TF,11} = \sigma^2$.

The MSE of the group 1 model applied to group 2 input-output data is transfer learning w_1^* to fit $y_2(K,L) = -K + \epsilon$ using model $y(K,L) = w_K K + w_L L$. The optimal weights for $y_2(K,L)$ are $w_2^* \equiv \{1,0,0,0,-1,0\}$, which can be achieved by altering the last layers' weights of $w_1^* \equiv \{1,0,0,0,1,0\}$ to $w_{TF,2} = \{1,0,0,0,-1,0\}$, i.e., change to $w_{H1} = -1$. Hence, $y(K,L;w_{TF,2}) = w_K K + w_L L = -K$. As a result, $MSE_{TF,12} = E\left[\int_0^1 \int_0^1 (-K - (K+\epsilon))^2 dK dL\right] = E\left[\epsilon^2\right] = \sigma^2$.

The MSE of the group 1 model applied to group 3 input-output data is transfer learning w_1^* to fit $y_3(K,L) = L + \epsilon$ using model $y(K,L;w_{TF,3}) = w_K K + w_L L$. Since the true output, $y_3(K,L)$, only depends on labor and not on capital, the second best (because the first layer weights from labor are zero and there is no way to incorporate labor input effects) way to minimize mean squared error loss is to suppress the capital input effects which only adds noise irrelevant to the true output factor (labor), $w_{TF,3} = \{1,0,0,0,0,0\}$, i.e., change to $w_{H1} = 0$, and we get $y(K,L;w_{TF,3}) = 0$.

$MSE_{TF,13} = E\left[\int_0^1 \int_0^1 (0 - (L+\epsilon))^2 dK dL\right] = \frac{4}{3} + \sigma^2$.

latter approach is, therefore, more likely to capture salient rather than minor differences in the underlying production processes between two industries.

Appendix C. Variable Definitions

Variables	Definition	Sources
Hoberg-Phillips FIC 25 Level Variables		
CD Distance	Acquiror industry-to-target industry production process distance as measured by log MSE ratios using Cobb-Douglas kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target industries are based on Hoberg-Phillips 10-K Text-based 25 Industry Classification.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
Linear Distance	Acquiror industry-to-target industry production process distance as measured by log MSE ratios using linear kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target industries are based on Hoberg-Phillips 10-K Text-based 25 Industry Classification.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
XGB Distance	Acquiror industry-to-target industry production process distance as measured by log MSE ratios using XGBoost kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target industries are based on Hoberg-Phillips 10-K Text-based 25 Industry Classification.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications

FC Distance	Acquiror industry-to-target industry production process distance as measured by log MSE ratios using neural network (fully connected layers) kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target industries are based on Hoberg-Phillips 10-K Text-based 25 Industry Classification.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
TF Distance	Acquiror industry-to-target industry production process distance as measured by log MSE ratios using neural network (fully connected layers) kernel for fitting production process, but using transfer learning to allow for fine-tuning the last layer to accommodate target industry. Detail procedures for computing distance is described in Section 3. Acquiror and target industries are based on Hoberg-Phillips 10-K Text-based 25 Industry Classification.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
TNIC Score (FIC-Pair Annual Mean)	Annual industry mean of Hoberg-Phillips Text-based Network Industry Classification (TNIC) data. industries are based on Hoberg-Phillips 10-K Text-based 25 Industry Classification.	Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
log(number of MA deals)	Natural logarithm of annual number of deals in SDC Platinum database for each acquiror-target industry pair.	SDC Platinum

Firm Level Variables

CD Distance	Acquiror-to-target firm-level production process distance as measured by log MSE ratios using Cobb-Douglas kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target instantaneous peer groups are based on Hoberg-Phillips TNIC score similarity. Specifically, we select firms with a similarity score greater than 0.09 (approximately the 95th percentile in each year's full TNIC score data) as an instantaneous peer group for a corresponding firm.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
Linear Distance	Acquiror-to-target firm-level production process distance as measured by log MSE ratios using linear kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target instantaneous peer groups are based on Hoberg-Phillips TNIC score similarity. Specifically, we select firms with a similarity score greater than 0.09 (approximately the 95th percentile in each year's full TNIC score data) as an instantaneous peer group for a corresponding firm.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications

XGB Distance	Acquiror-to-target firm-level production process distance as measured by log MSE ratios using XGBoost kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target instantaneous peer groups are based on Hoberg-Phillips TNIC score similarity. Specifically, we select firms with a similarity score greater than 0.09 (approximately the 95th percentile in each year's full TNIC score data) as an instantaneous peer group for a corresponding firm.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
FC Distance	Acquiror-to-target firm-level production process distance as measured by log MSE ratios using neural network (fully connected layers) kernel for fitting production process. Detail procedures for computing distance is described in Section 3. Acquiror and target instantaneous peer groups are based on Hoberg-Phillips TNIC score similarity. Specifically, we select firms with a similarity score greater than 0.09 (approximately the 95th percentile in each year's full TNIC score data) as an instantaneous peer group for a corresponding firm.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications

TF Distance	Acquiror-to-target firm-level production process distance as measured by log MSE ratios using neural network (fully connected layers) kernel for fitting production process, but using transfer learning to allow for fine-tuning the last layer to accommodate target industry. Detail procedures for computing distance is described in Section 3. Acquiror and target instantaneous peer groups are based on Hoberg-Phillips TNIC score similarity. Specifically, we select firms with a similarity score greater than 0.09 (approximately the 95th percentile in each year's full TNIC score data) as an instantaneous peer group for a corresponding firm.	Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
TNIC Score	Hoberg-Phillips Complete 10-K Text-based Network Industry Classification (TNIC) Industry Data.	Hoberg-Phillips 10-K Text-based Fixed Industry Classifications
Completed	Deal completion indicator which one if Status is "Completed".	SDC Platinum
Withdrawn	Deal incomplete indicator which one if Status is "Withdrawn".	SDC Platinum
Diversify	Indicator for a diversifying merger, i.e., the acquiror's 25 FIC differs from that of the target's (variable construction follows Deng, Kang, and Low (2013)).	SDC Platinum, Compustat, Hoberg-Phillips 10-K Text-based Fixed Industry Classifications

Hostile	Indicator for a hostile merger, which takes a value of one if SDC Platinum records the deal as a hostile merger (Attitude is "Hostile") and zero otherwise (variable construction follows Deng, Kang, and Low (2013)).	SDC Platinum
High Tech	Indicator that takes a value of one if both acquiror and target are in high technology sectors and zero otherwise, i.e., SDC Platinum records them as "Primary Business not Hi-Tech" (variable construction follows Deng, Kang, and Low (2013)).	SDC Platinum
Tender Offer	Indicator for tender offer deals, i.e., TenderOffer is"Yes" in SDC (variable construction follows Deng, Kang, and Low (2013)).	SDC Platinum
Stock Issue	Indicator for deals that are (partly) financed by stock according to the report in SDC, i.e., SourceofFundsPreferred_StockIssu is "Yes" or SourceofFundsCommonStockIssue is "Yes" (variable construction follows Deng, Kang, and Low (2013)).	SDC Platinum
Relative Deal Size	The deal value reported in SDC normalized by the acquiror's market capitalization (variable construction follows Deng, Kang, and Low (2013)).	SDC Platinum
Acquiror log(Asset)	Acquiror's natural logarithm of total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat

Acquiror Tobin's Q	Acquiror's ratio of total assets plus market value of equity (total shares outstanding times year-end stock price) minus the total value of common equity minus deferred taxes to total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat
Acquiror Book Leverage	Acquiror's debt in current liabilities plus total long-term debt divided by total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat
Acquiror CF/Assets	Acquiror's operating income before depreciation minus total interest related expenses minus total income taxes minus capital expenditures divided by total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat
Target log(Asset)	Target's natural logarithm of total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat
Target Tobin's Q	Target's ratio of total assets plus market value of equity (total shares outstanding times year-end stock price) minus the total value of common equity minus deferred taxes to total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat
Target Book Leverage	Target's debt in current liabilities plus total long-term debt divided by total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat

Target CF/Assets	Target's operating income before depreciation minus total interest related expenses minus total income taxes minus capital expenditures divided by total assets (variable construction follows Deng, Kang, and Low (2013)).	Compustat
CAR[t-1, t]	Acquiror-target value-weighted cumulative abnormal return for event window one day prior to announcement date to announcement date.	CRSP, Fama-French 5 Factors
CAR[t-1, t+1]	Acquiror-target value-weighted cumulative abnormal return for event window one day prior to announcement date to one day after announcement date.	CRSP, Fama-French 5 Factors
CAR[t-1, t+2]	Acquiror-target value-weighted cumulative abnormal return for event window one day prior to announcement date to two days after announcement date.	CRSP, Fama-French 5 Factors
Survive 10 Years	Indicator that is one if the acquiror exists and not acquired as target (a target in SDC database) within 10 years after merger.	Compustat, SDC Platinum
Survive 15 Years	Indicator that is one if the acquiror exists and not acquired as target (a target in SDC database) within 15 years after merger.	Compustat, SDC Platinum
Divestiture Within 10 Years	Indicator which is one if acquire appear as target for divestiture at within 10 years after merger.	SDC Platinum
Asset Growth 10 Years	Log difference of total assets 10 years after to one year after merger.	Compustat

Market Value Growth 10 Years	Log difference of market value 10 years after to one year after merger. Market value is total assets plus year end share price times shares outstanding minus common equity and deferred taxes.	Compustat
Prior Deals Within 10 Years	Acquiror's number of deals within prior 10 years to the announcement of current deal.	SDC Platinum
Prior Completed Deals Within 10 Years	Acquiror's number of completed deals within prior 10 years to the announcement of current deal.	SDC Platinum
Prior Failed Deals Within 10 Years	Acquiror's number of incomplete deals within prior 10 years to the announcement of current deal.	SDC Platinum
Most Recent Deal Within 10 Years	Indicator for acquiror having a recent merger deal within 10 years	SDC Platinum
Most Recent Deal Within 10 Years: Completed	Indicator for acquiror's most recent deal within 10 years is completed.	SDC Platinum
Most Recent Deal Within 10 Years: Incomplete	Indicator for acquiror's most recent deal within 10 years is incomplete.	SDC Platinum
Post-Right-to-Work	Indicator for corresponding state already adopted Right-to-Work law in the current deal announcement year.	National Conference of State Legislators