Internet Appendix: The Determinants and Impact of Executive-Firm Matches

April 12, 2015

Appendix A: Sample Construction and Variable Definitions

The senior executive labor market for publicly traded companies is not only an important segment of the labor market, but it also provides an excellent framework to study the two-sided, many-to-one matching game with transfer. Unlike private companies where the transfer data are usually not observable to econometricians, public firms must report the details of the compensation packages for top executives in their annual proxy statements (DEF14A SEC form).

Information on executives including both compensation and the time in office comes from ExecuComp. Financial data on firms come from Compustat and CRSP. In addition, I collect executive background information from various sources: Dun & Bradstreet and Standard & Poor's publish annual directories of biographical data on officers of major firms. Corporate disclosure statements (10Ks and proxy statements) profile officers' backgrounds. Generally, firms also report executives' biographies in press announcements of turnovers. Supplemental educational background data are manually collected from various sources such as Marquis' Who's Who, Forbes Tearsheet, Zoominfo, and company websites.

Executive-firm-year observations are included in the sample when there is a succession event (i.e., a job match) in terms of either promoting internal candidates or hiring external candidates for the top five executive positions of a company.[1] I use both ExecuComp data and company press releases to identify the date and succession origin of turnover events. For each case, I collect information on the company and the manager before the appointment. I construct a sample of job matches for executives with tractable previous career paths, from 1993 to 2006. This process yields a total of 2,335 job matches before any other data restrictions. Table 1, Panel A in the main text lists the number of executive-firm matches by year.

I use the reported first full-year annual compensation after turnover, *tdc1*, to measure the endogenous transfer to the manager ($p(c_i, m_i)$ in the model). Because the managerial talent was recently acquired or retained, first-year compensation is less likely to be contaminated by entrenchment costs and is more informative about the market price of an observable managerial skill set to

[1] Because I condition on the set of occurred turnovers, my model is silent on why a manager experiences a turnover.

a company. Summary statistics of executive compensation, as well as firm and manager attributes, are provided in Table 1, Panel B, in the main text. There is a change in the variable *tdc1* after 2006, especially in the short-term compensation (bonus, non-equity incentives) part, since the SEC moved from a focus on timing (annual vs. long-term) to a focus on discretionary versus formula-based (regardless of whether annual or long term) compensation. Still, I collect data on managerial attributes such as education for a subsample of 609 turnovers between 2007 and 2012, from Capital IQ's *People Intelligence*, for robustness checks and out of sample predictions.

The firm attributes relevant for the three matching dimensions are size, diversification degree, and research intensity. To capture the real operating aspect of the size measure, I use the value of assets (*Assets*) as the main measure for firm size. I use other size measures such as revenue, number of employees, market value of equity, and firm value (market value of equity + book value of debt) in robustness checks. All size variables are in natural logarithms. Because I assume there is one national senior executive job market every year, an assumption that will be relaxed later, I convert all the variables to their CDFs in each year. Companies' and managers' ranks based on these characteristics thus reflect their market positions in various dimensions.

Following the diversification literature (e.g., Jacquemin and Berry 1979; Hund, Monk, and Tice 2010), the main measure I use for the degree of diversification is the entropy based on segment sales (Div_{sales}). Alternative measures for diversification used in this paper are the number of segments and entropy based on segment assets. The total entropy measure of diversification for firm i at year t is determined by:

$$Ent_{i,t} = \sum_{s=1}^{n} P_{s,i,t} ln \frac{1}{P_{s,i,t}}, \tag{1}$$

where n is the number of segments at the two-digit SIC level, and $P_{s,i,t}$ is the proportion of sales for segment s in firm i at year t. *Ent* is a continuous measure of diversification, and takes a higher value for firms operating in more business domains.[2] The degree of diversification captures the business scope of the firm and serves as a measure for the administrative, coordinative, and structural complexity of the firm.

A company's innovation intensity is measured by its R&D intensity (*RD*), which is R&D expenditure scaled by sales. It emphasizes the search for new opportunities, captures the innovation tendency of the company, and is strongly associated with its growth prospect.

With a competitive assignment model, this paper studies general, transferable managerial ability, which gives rise to the assortative matching on the executive labor market, not the firm-specific skills.[3] It also does not study the on-job learning component of the skill set, which is likely to be small: executives in this sample of S&P 1500 firms are mature (average age 54 when taking office) and experienced. I use management abilities and demographic information that are observable to

[2]The entropy-based measure is considered superior to the HHI measure because it is not influenced by how large the dominant business of a firm is. Total entropy changes as the distribution of sales across segments changes, even if the number of segments is held constant.

[3]With firm-specific skills, there is no "better" or "worse" manager but rather different styles that are best suited to different firms.

the labor market to capture relevant manager characteristics. "Economies of scale" implies that firm size is complementary to some general, execution-related skills. In this spirit, Kaplan et al. (2012) propose a talent measure from factor analysis, using interview scores. The first principal component correlates most strongly with the manager attribute "efficiency."

I construct two measures for managerial efficiency. The estimation method for the main efficiency measure, $Talent_{dea}$, uses input-oriented data envelopment analysis (DEA), based on an optimization algorithm in Demerjian et al. (2012).[4] It captures how far a company is from the industry-year-level operational efficiency frontier as a consequence of managerial decisions. It is available for a large sample and is mainly under management's control (unlike, e.g., media mentions).

I then calculate the efficiency scores at the executive-firm-year level by weighting firm-year-level raw scores with the executive's share of pay received for the whole management team. For example, if an executive receives total annual compensation of $2 million and the top management team receives $10 million, the raw efficiency score is weighted by 20% for the manager in this year. Following Li and Ueda (2006), who argue there is no clear relation between internal experience and innate managerial talent, I average an executive's efficiency scores ($Talent_{dea}$) only across the firm-years for which she worked before the current employer. Because of the restriction imposed by $Talent_{dea}$, all of the executives in my sample had experience as a senior manager at public firms before becoming the CEO for the current company. An alternative talent measure is constructed based on the accounting performance ROA (see Kuhnen and Zwiebel 2009) using the same procedure as for constructing $Talent_{dea}$.

In addition to efficiency, the second managerial attribute is her past experience in conglomerates. An executive's career experience helps shape the lens through which she views current opportunities and problems. Of great relevance is the diversity and complexity of the organizations she has managed in the past. I measure the diversity of the manager's industry expertise and the complexity of her past employers using the average entropy of the companies the executive had worked for, based on segment sales, Exp_{sales}. In this dimension, the low-type managers have focused experience; high-type managers have conglomerate experience.

The third managerial attribute is technical expertise and innovation propensity. The consistent finding in the management psychology literature is that the level of formal education of central decision makers is positively related to their receptivity to innovation (see, e.g., Kimberly and Evanisko 1981; Rogers and Shoemaker 1971; Wiersema and Bantel 1992). Also, science-related training is more likely to be associated with greater propensity for innovation, though formal professional education (e.g., the MBA degree) is not. I measure the technical expertise and innovation propensity of executives by the discrete variable *Education*. The *Education* score is 5 for executives who earned a doctoral degree (3%),[5] 4 for a master's degree in science (5%), 3 for a bachelor's degree in science (12%), 2 for a master's degree outside of science (i.e. MBA or M.A.) (27%), and

[4] http://faculty.washington.edu/smcvay/abilitydata.html

[5] The majority of CEOs with PhD degrees majored in science.

1 for a bachelor's degree outside of science (52%, including those who attended college but did not graduate).

Appendix B: Maximum Score Estimation

B.1 Algorithm

The algorithm for the maximum score estimation is as follows:

Step 1. Define markets. All firms with turnover events in one calendar year, and the managers hired at those events are grouped into one market.

Step 2. Construct independent variables. In each market, the independent variables (value-relevant interactions of attributes of each firm-manager combination) are constructed from the data set.

Step 3. Construct exchange pairs. Within each market, a comparing pair is formed by any two executive-firm matches with the same firm. A pair thus consists of one company and two managers working for different firms. I denote the pair with company c_i (actually) matching with manager m_i and company c_i (counterfactually) matching with manager m_j as $(c_i, m_i; c_i, m_j)$. Variations among independent variables on these two matches allow the identification of the coefficients on the interaction terms.

Step 4. Calculate interaction terms. I divide the interaction variables into two parts: proposed complementarities/substitutions and the "off-diagonal" interactions. Both the reduced-form and maximum score estimators show the ad hoc "off-diagonal" interactive terms are insignificant. This confirms the validity of the proxies I construct for the company and manager attributes. I therefore include only three interaction terms in the main analysis to test the proposed hypothesis.

Step 5. Estimation: Set the initial values for the parameters ω, compute the production functions for both the actual and counterfactual matches, and apply the differential evolution method (Storn and Price 1997) to search the global optimum for those parameters. For a guess of the production function parameter vector ω, the objective function asks whether the two inequalities are simultaneously satisfied. If so, the maximum score objective function increases by 1. The parameter value that brings the highest Q score in (9) is the maximum score estimator for ω. Programming the objective function involves evaluating the unknown production function and checking inequalities. Note that the consistency of the estimators does not require the inclusion of all inequalities. This allows subsampling, reduces the computational burden, and makes computation of confidence intervals feasible. also, not all inequalities will be satisfied, even at the maximizer $\hat{\omega}$ and even at the probability limit of the objective function. As the objective function is a step function, there will be more than one local maximum; finding one is sufficient for the estimation based on Fox (2010a). This is a semi-parametric estimation because I do not need to impose any structural assumption on the error terms, only specify the match production function up to a positive monotone transformation. Fox (2010a) also presents Monte Carlo evidence that the difference between having several small markets and one large market evaporates when the variance of the matching error term increases.

Step 6. Calculate confidence intervals. Maximum score estimation does not assume a distribution for the error terms. Therefore, I use subsampling to calculate the confidence intervals for the estimators. Specifically, I follow the procedure proposed in Fox's toolkit for matching maximum score estimation. For inference, I use subsample size equal to one-fourth of the matching markets and 100 replications (artificial data sets) in subsampling to obtain 100 sets of parameter estimates. Finally, for each parameter, I use these 100 estimates as its empirical distribution to calculate its confidence intervals.

The beauty of maximum score estimation compared to other techniques, like maximum likelihood estimation, is that it does not require distributional assumptions about the error terms.[6] Another advantage is that maximum score estimation does not suffer from the curse of dimensionality.[7] The advantage of maximum score estimation over propensity score matching is that it avoids solving for an equilibrium, which is prohibitive because of combinatorics in a matching game. Maximum score estimation works directly with F instead of the conditional probability of the assignment outcome.

B.2 Comparative Statics

Sections III.2 and III.3 in the main text describe structural identification of the matching model. The parameter ω describes the primitives governing the executive labor market. As Strebulaev and Whited (2011) suggest, one way to explain the intuition behind the features of the data that identify ω is to provide detailed comparative statics. Given a particular choice of the production function F and values for $X = C \cup M$, the characteristics of all firms and all managers in a matching market, one can compute the (unique) equilibrium assignment. Likewise, comparative statics in F can be undertaken by choosing production functions with different ω and examining how the assignments and pay distribution change. Alternatively, different values for the market characteristics X could be chosen, and the equilibrium computed for each choice of X. In Table A1, I carry out a few Monte Carlo exercises in this spirit.

For simplicity, I use a deterministic model here. Economic randomness (ϵ_{c_i,m_i}) will be addressed in the next subsection. I start with one-dimensional matching and specify the production function to be:

$$F_\omega(c_i, m_i) = \omega_1 C_{1i} M_{1i}, \tag{2}$$

where the firm attribute is drawn from a lognormal distribution $C_1 \sim 10logN(0,1)$.[8] I set the

[6]The error terms are not included in the objective function. Fox (2010a) provides Monte Carlo evidence that the estimates may perform well if the data are generated from a model with matching error. Not all inequalities are satisfied with the maximum score estimators, even at the probability limit of the objective function.

[7]To illustrate the large number of dimensions, and therefore the large number of error terms being integrated, take the example of a one-to-one matching game with n agents on each side of the market: the number of possible matches are n^2 and the number of possible assignments are $n!$.

[8]The results are robust to the choice of alternative distributions. I use lognormal because the empirical distribution

true parameter value to be $\omega_1 = 1$ and the number of firms (managers) to be 1,000, that is, 100 pairs in each of the 10 markets. In experiment (1.1), I impose perfectly assortative matching by setting M_1 (say, managerial talent) equal to C_1 . I assume that managerial compensation is $p(c_i, n_i) = 0.5F(c_i, n_i)$. Intuitively, given that the ranks of attributes of matched firms and managers perfectly line up, this compensation schedule along with assortative matching sustain a pairwise stable equilibrium. Indeed, the maximum score estimator recovers the true productivity parameter ($\hat{\omega_1} = 1$), and the matching model has a 100% fit – all I.R. conditions are satisfied.

In experiment (1.2), the production function F, the distributions of C_1 and M_1, and the sharing rule remain the same, but firms and managers are randomly matched. I estimate this model 100 times and get 100 estimates of $\hat{\omega_1}$ from the maximum score estimation. The average of the maximum score estimators across the 100 runs is 1, and the (5%, 95%) interval for $\hat{\omega_1}$ is (0.95, 1.06). However, the poor fit of the maximum score estimation suggests that an assignment model does not explain the data well. On average, only 26% of the I.R. conditions are satisfied. Such a market is not stable: if firm and manager attributes indeed complement each other to generate productivity (as suggested by F), one would observe an assortative pattern in firm-manager pairing in equilibrium, as better firms will outbid firms that are not as good, for desirable managerial skills.

In experiment (1.3), firms and managers are perfectly assortatively matched, but the compensation for each manager is randomly drawn from a uniform distribution. Maximum score estimation does not correctly recover ω_1 and has a poor fit. In experiment (1.4), both the firm-manager pairing and the managerial compensation are random. The matching model has the poorest fit; 6% of the I.R. conditions are satisfied. These exercises show that the matching model used in this paper evaluates the extent to which both observed firm-executive pairs *and* the cross-sectional distribution of executive compensation sustain a marketwide, pairwise stable equilibrium.

Next, I consider two matching dimensions:

$$F_\omega(c_i, m_i) = \omega_1 C_{1i} M_{1i} + \omega_2 C_{1i} M_{2i} \tag{3}$$

In experiments (2.1), (2.2), and (2.3), I fix X, the distribution of firm and manager attributes, but change ω and the sharing rule. In (2.1), I impose perfectly assortative matching on the first dimension by setting M_1 equal to C_1. The second managerial atttribute is positively, but not perfectly, correlated with C_1: $M_2 \sim 10U(0,1) + 0.1C_1$, where $U(0,1)$ is a uniform distribution between 0 and 1. This case is more realistic, as real firm and manager attributes do not always perfectly line up. Suppose the firm-manager pairs (c_i, m_i) and (c_j, m_j) are two adjacent pairs when one orders the matched firm-manager pairs in a market by F. Based on equations (4) and (5), manager j's compensation $p(c_j, m_j)$ could lie anywhere between $p(c_i, m_i) + F(c_i, m_j) - F(c_i, m_i)$ and $p(c_i, m_i) + F(c_j, m_j) - F(c_j, m_i)$, to be part of a pairwise stable equilibrium. In this exercise, I take the upper bound of pairwise stable compensation and normalize the reservation utility for the manager associated with the lowest F in each market to be 0.[9]

of firm size is fat-tailed (see Gabaix and Landier 2008).

[9]The results are similar as long as the managerial compensation lies in the interval constrained by the inequalities.

The positive estimates $(\hat{\omega_1}, \hat{\omega_2})$ suggest there is complementarity in each dimension: between C_1 and M_1, and between C_1 and M_2. Maximum score estimation recovers the true productivity parameters. In contrast, the relative strength of the complementarities cannot be correctly inferred from regressing compensation on the interactions of firm and manager attributes ($C_1 M_1$ and $C_1 M_2$). In fact, the regression coefficient on $C_1 M_2$ is even negative in this case, despite the true productivity parameter being positive. The average goodness of fit for the matching model is 44%, in terms of the percentage of satisfied I.R. conditions.[10]

The only difference from (2.1) to (2.2) is that now there is more assortative matching in the dimension that contributes *less* to F. As in (1.2), the matching model does not explain the observed labor market outcome well, because such a market is not pairwise stable. In a stable equilibrium, given the stronger complementarity between C_1 and M_2, firms and managers will match more on the second dimension. In experiment (2.3), firms and managers are more assortatively matched on the dimension with stronger complementarity, but the level of managerial compensation is randomly drawn from an uniform distribution. The model has a poor fit due to the random payoffs.

In sum, these exercises show that one can use the maximum score estimation only to estimate an assignment model as it considers the I.R. conditions that underlie the pairwise stable equilibrium. The marketwide firm-manager pairing as well as the cross-sectional distribution of managerial compensation jointly determine the equilibrium outcome, and thus, they are the inputs in the estimation. The sign of the maximum score estimators reveals whether there is assortative or anti-assortative matching in a particular dimension. The relative magnitude of the estimates reflects the relative strength in complementarity (or substitution) between firm and manager attributes. The overall goodness of the fit shows whether a matching model is a good model for the observed labor market outcome in terms of both assignments and pay distribution. In general, with real firm and manager attributes, perfectly assortative matching in multiple dimensions is hard to obtain. pairwise stable equilibrium requires that assortative matching being satisfied first in dimensions with stronger complementarity.

B.3 Rank Order Property

Following random utility models, unobserved heterogeneity is introduced as information that is observable to, and taken into account by, the players but is not observable to the econometricians. Statistical consistency of the maximum score estimators rests on the assumption that the error terms follow the rank-order property (Manksi 1975, 1985): the deterministic matching game under assignment A_2 violates more I.R. conditions than under assignment A_1 if and only if from an ex ante perspective A_1 is more likely to occur than A_2 after the errors are drawn. The rank order property thus relates the inequalities from pairwise stability to the probabilities of different equilibrium assignments, conditional on $X = C \cup M$. Given X, neither assignment A_1 nor A_2 may be a stable

[10] The fit is not 100%, as perfectly assortative matching in both dimensions cannot be achieved given the exogenously specified X in this case. Therefore, the model fit is likely to decrease with increased dimensionality, unlike OLS regressions, even if there is assortative matching in dimensions with stronger complementarity.

assignment to the deterministic matching model. But A_1 might dominate A_2 in the deterministic model in that at least one firm and one manager in A_2, say c_i and m_j, would prefer to match with each other instead of with their actual partners, leading to A_1. In a model with error terms, both A_1 and A_2 could be pairwise stable assignments to some realizations of the unobserved heterogeneity. The assumption says that A_1 is more likely to be a pairwise stable assignment to some realized model than is A_2.[11]

To my knowledge, no empirical matching paper has parametrically estimated distributions of unobserved characteristics in matching games.[12] Conceptually, however, if data are available on unmatched players, match-specific unobservables can be identified. An alternative is to impose a stronger functional form for F, for example, using random coefficients on additional observed characteristics or coefficients that are fixed across markets but heterogeneous within a market on the match-specific characteristics (see Fox and Yang 2012).[13]

Note that firm-specific unobserved heterogeneity drops out in equilibrium, due to the comparison between actual and counterfactual matches in the I.R. conditions. Manager-specific unobserved heterogeneity drops out in the model without transfer data (Fox 2010b), which I also estimate as a robustness check. Using Monte Carlo experiments, Fox (2010b) shows that if the match-specific (i.e., firm-manager-specific) errors are i.i.d., the bias is low. Type I extreme value errors also satisfy the rank order property, although maximum score estimation does not impose this parametric assumption. Fox (2010b) shows that the bias decreases with both the number of players in each market and the number of markets, for maximum score estimation without transfer.

Using similar Monte Carlo experiments, Akkus et al. (2014) show that maximum score estimation with transfer is similar in spirit to the estimation without transfer, but using transfer data can significantly improve the performance of the estimation, especially under the presence of unobserved, match-specific heterogeneity.

B.4 Robustness Checks

Table A2, Panel A reports maximum score estimates for the same production function using the approach in Fox (2010b). Fox (2010b) uses a less restrictive concept of "local production maximization." The idea is that for any two observed matching pairs, the sum of the value created by these matches exceeds the sum of the value that could be created if these pairs exchange their partners: $[F(c_i, m_i) + F(c_j, m_j)] \geq [F(c_i, m_j) + F(c_j, m_i)]$. Clearly, conditions (2) and (3) together imply Fox's (2010b) condition. Thus, my model imposes more stringent equilibrium conditions as it requires the data to match both assignments and pay observed in the data. The level of total matching output is not identified in Fox (2010b). The sign of ω_1 is superconsistently estimable, so I

[11] See Fox (2010a) for more details.

[12] For example, in Sorensen (2007), the unobserved ability of each VC is not measured.

[13] This approach relaxes the assumption that the match-specific characteristics entering the production function in the same manner for each match. Fixing coefficients across markets but not within markets only makes sense in a context where firm and manager indices such as i and j have a consistent meaning across markets. For example, the same set of firms and managers may participate in multiple matching markets in different years.

set it to the true value of +1, following Fox (2010b). The term "superconsistent" is commonly used in the time-series literature to describe estimators that converge faster than normal (e.g., based on sample average). The results are qualitatively similar, though the explanatory power of the matching model increases to 63%.

I also perform various robustness checks with different elements of the assignment model: additional ad hoc matching dimensions, different specifications of the production function, alternative definitions of the executive labor market, and alternative measures of firm and manager attributes.

First, I add ad hoc interaction terms ($Assets * Exp_{sale}$, $Assets * Education$, $Div_{sale} * Education$) to capture some of the potential correlations between "off-diagonal" firm and manager attributes. Results show that most of these terms are insignificant (except for $Assets * Exp_{sale}$, which is marginally significant) in the multivariate structural estimation and do not substantially increase the explanatory power of the model.

With multidimensional firm and manager attributes, assortative matching cannot be satisfied in all dimensions. Al alternative specification of the production function is used to test the effect of "imperfect matching":[14]

$$
\begin{aligned}
F_\omega(c_i, m_i) = \quad & -\omega_1 \mathbf{1}[C_{1i} M_{1i}](C_{1i} - M_{1i})^2 - \omega_4 \mathbf{1}[C_{1i} < M_{1i}](C_{1i} - M_{1i})^2 \\
& -\omega_2 \mathbf{1}[C_{2i} M_{2i}](C_{2i} - M_{2i})^2 - \omega_5 \mathbf{1}[C_{2i} < M_{2i}](C_{2i} - M_{2i})^2 \\
& -\omega_3 \mathbf{1}[C_{3i} M_{3i}](C_{3i} - M_{3i})^2 - \omega_6 \mathbf{1}[C_{3i} < M_{3i}](C_{3i} - M_{3i})^2 \\
& +\xi covariates_{c_i, m_i} + \varepsilon_{c_i, m_i}. \qquad (4)
\end{aligned}
$$

where $\mathbf{1}[C_{ni} > M_{ni}]$ is an indicator function that equals 1 if the company attribute is ranked higher than the corresponding manager attribute, and 0 otherwise. ω is also identified by the observed equilibrium pattern in terms of job assignments and cross-sectional distribution of pay. If managerial underqualification (i.e., manager ranked lower than the firm) reduces F, the estimated productivity parameters associated with underqualification indicators–$\hat{\omega}_1$, $\hat{\omega}_2$, $\hat{\omega}_3$–should be positive. If managerial overqualification reduces F, the estimated productivity parameters associated with overqualification indicators–$\hat{\omega}_4$, $\hat{\omega}_5$, $\hat{\omega}_6$–should be positive.

Table A2, Panel B reports the estimated productivity parameters for the alternative specification of this production function. Results show that there are penalties to both under- and overqualification for each matching dimension. The estimation results support the matching theory: the optimal allocation is a "bliss point" and compensation is non-monotone in firm type for a given manager. The positive estimates show that the level of joint output is reduced when firm and manager attributes in any dimension do not "fit." For the size/efficiency and R&D intensity/education dimensions, the matched pair is relatively equally punished regardless of whether executives are over- or underqualified. For the diversification/cross-industry experience dimension, the penalty is more severe for underqualified management: the cross-industry experience is a

[14] Empirically, this imperfection could also be due to either the economic randomness that is added to the deterministic model or the limited number of agents with indivisible, exogenous attributes in observed data.

particularly important yet scarce resource.

With maximum score estimation, correctly capturing the rank order of player attributes is important. One natural question is whether using alternative proxies for relevant firm and manager attributes changes the ranking. Section IV reports various alternative measures I use to capture firm and manager characteristics in unreported analysis. The results are similar, for example, whether using total book assets or total firm value to capture firm size, or using DEA- or ROA-based efficiency measures to capture managerial talent. Furthermore, the fact that the ad hoc interaction terms are mostly insignificant in the maximum score estimation suggests that the explanatory power of the size/efficiency complementarity will not increase much using alternative talent measures related to career and educational credentials (e.g., Falato et al. 2014).

Next, using the principal component analysis, I decompose size and diversification into a common component (PC) and residual components (ε for $Assets$ and ζ for Div_{sale}). I find that ε comoves with $Talent_{dea}$, ζ comoves with Exp_{sale}, and the principal component PC, which explains 62% of the variations in both size and the degree of diversification, comoves only with Exp_{sale}. This explains (1) why there is assortative matching on both dimensions, (2) why diversification/experience dominates size/talent, and (3) why controlling for the proposed dimensions, the off-diagonal interaction terms $Div_{sale} * Talent_{dea}$ and $Assets * Exp_{sale}$ are not relevant. Furthermore, to change the order of the importance of economies of scale versus economies of scope, the increase in correlation between firm size and managerial talent needs to mainly come from the increase in correlation between PC and $Talent_{dea}$, which is currently insignificant.

In the main analysis, I assume that the executive labor markets in different years are independent of each other. This is similar to adding year fixed effects in a regression to reflect the market conditions, such as business cycle. To relax the assumption of historical independence, I pool together agents in two consecutive years and perform the same estimation. The point estimates are robust, but the statistical fit of the estimation deteriorates because of the increased number of inequality conditions. In the alternative, one could envision several distinctive markets for one year, which requires imposing more structure on the data in terms of market segmentation. In this spirit, I restrict the executive labor market for each firm to a limited set of relevant industries only, based on the manager migration matrix in Cremers and Grinstein (2011). The estimates are robust, as most of the inequalities for counterfactuals from irrelevant industries are trivially satisfied. The results are similar if I focus on the CEO labor market only. Finally, to capture firm-specific matching or inertia of the manager, I construct an indicator variable that equals 1 for internal promotions. However, compared to industry-specific skills, firm-specific skills are less important in explaining the executive labor market outcome.

B.5 Many-to-one Matching

Less than 10% of the firm-years in the sample are associated with multiple hires. Therefore, hiring a management team rather than hiring a single manager does not appear to be a big concern with this sample. Still, I estimate a many-to-one matching model, following Fox and Bajari (2013), in this subsection. In particular, I relax the third assumption that is required to make many-to-one matching isomorphic to one-to-one matching – additive separability condition. The revised model is the following:

There are a=1, ... , K firms and m =1, ... , L managers. With many-to-one matching, firms may hire a number of new executives. Let H_a denote such a hire of new management team. An allocation of management teams to firms can be written as $A = H_1, H_2, ..., H_K$. Let p^L denote the pay distribution for the L managers.

The outcome $(p^L, A) = (p^L, H_1, H_2, ..., H_K)$ satisfying $\bigcup_{a\in K} H_a \subseteq M$ and $H_a \bigcap H_b = \varnothing$ for all firms a and b is a pairwise stable outcome in both job assignments and executive pay if, for each firm a=1, ... , K, corresponding new management team $H_a \subset L$, and executives $i \in H_a$ and $j \notin H_a$, $j \in L$,

$$[f_a(H_a) - p_i] \geqslant [f_a(H_a \setminus \{i\} \cup \{j\}) - p_j]$$

and

$$[f_b(H_b) - p_j] \geqslant [f_b(H_b \setminus \{j\} \cup \{i\}) - p_i]$$

The production function is specified as:

$$\begin{aligned} f_a(H_a) \quad = \quad & \omega_1 Assets_a \Sigma_{H_a}(Talent_{dea}) + \omega_2 Div_a \Sigma_{H_a}(Exp_{sale}) + \omega_3 RD_a \Sigma_{H_a}(Edu_{tech}) \\ & + \gamma_1 \delta_{H_a}(Talent_{dea}) + \gamma_2 \delta_{H_a}(Exp_{sale}) + \gamma_3 \delta_{H_a}(Edu_{tech}) \\ & + \xi covariates_{a,H_a} + \sum_{m\in H_a} \varepsilon_{a,H_a} \end{aligned}$$

where Σ_{H_a}'s are the sums of managerial efficiency, conglomerate experience, and technical education across all newly hired executives in a firm-year; and δ_{H_a}'s are the standard deviation in the corresponding variables across all newly hired executives in a firm-year. $covariates_{i,Q}$ measures the percentage of newly hired executives with managerial experience in the same industry in a firm-year. If there is only one new hire in a given firm-year, the production function reduces to equation (4) in the main text.[15] If there are multiple new hires in a given firm-year, the δ terms could pick up the interactions (e.g., complementarities) among the new executives. Because of these terms, the production function is not additively separable across multiple matches for the same firm-year.

The objective function for the many-to-one matching is:

[15] This is validated using the sample of firm-manager-years with only one-to-one matching.

$$\sum_{y=1}^{Y}\sum_{a=1}^{M_y-1}\sum_{b=a+1}^{M_y}\sum_{i=1}^{|H_a|}\sum_{j=1}^{|H_b|}\mathbf{1}\{[f_a(H_a)-p_i] \geqslant [f_a(H_a\setminus\{i\}\cup\{j\})-p_j]\} \cap\{[f_b(H_b)-p_j] \geqslant [f_b(H_b\setminus\{j\}\cup\{i\})-p_i]\}$$

where M_y is the number of matches observed in market (year) y.

Table A3 reports the results. The estimated ω's are all positive, with similar relative magnitudes as in one-to-one matching (Table 4, Panel B in the main text). The estimated γ's are all negative. Only the estimate on dispersion within new hires in terms of technical education is significant. The estimates on talent and conglomerate experience indicate no significant penalty for dispersions in managerial characteristics among the newly hired executives for a firm. Taken together, at least with this sample, firms do not appear to be concerned about hiring a new team rather than a new manager.

Appendix C: Additional Reduced-form Analysis

In addition to the efficiency measure $Talent_{dea}$, in Table A4, I use experience in conglomerates and technical education of the manager as alternative measures of managerial talent, and interact them with firm size. None of these additional interactions is significant. Interestingly, a principal component analysis on $Talent_{dea}$, Div_{sale} and the generalist ability index suggests that the dominant principal component is mostly correlated with $Talent_{dea}$ (at 0.72), confirming $Talent_{dea}$ as an important measure for managerial talent.

Appendix D: Matching Quality and Firm Performance

In addition to the test on announcement abnormal returns, I also test whether superior performance achieved by certain firm-manager pairs is indeed associated with higher estimated complementarities, measured by $(\sum_{n=1,2,3}\hat{\omega_n}C_nM_n)$. Because most of the important corporate decisions are made collectively, firm-year-level matching quality $M_{quality}$ is the average of firm-executive-year-level complementarities within the senior executive team of a company.

The corporate performance measures I use are annual Tobin's Q, ROA, and ROE. Q is constructed as market-to-book ratio of firm assets. The numerator equals the market value of equity plus book assets minus the sum of the book value of common equity and deferred taxes. The denominator is assets at book value. ROA (in percentage) is constructed as net income before extraordinary items and discontinued operations divided by total assets. ROE (in percentage) is constructed as the net income before extraordinary items and discontinued operations divided by common equity.

I run regressions of Tobin's Q, ROA, and ROE on lagged matching quality and other firm financial characteristics and growth opportunities. Average Tobin's Q for firms in the sample is 1.81, with a standard deviation of 1.5. Average ROA in the sample is 2.39%, with a standard deviation of 9.62%. Average ROE in the sample is 8.43%, with a standard deviation of 27.47%.

Table A5 presents the results. In specification [1], I investigate the association between firm performance and $M_{quality}$, controlling for industry and year fixed effects. In specification [2], I control for firm fixed effects. In specification [3], I include managerial ownership and its square term, following Morck et al. (1988), as well as a set of control variables widely used in the literature on the determinants of firm performance.[16] Results show a positive and significant relation between $M_{quality}$ and all three performance variables, both cross-sectionally and within firms. A one-standard-deviation change in $M_{quality}$ is associated with a change in Tobin's Q by 0.13, which can be translated into a $260 million change in market value for an average-size firm. A one-standard-deviation change in $M_{quality}$ changes ROA by 0.6% and ROE by 1.92%. It is important to note that the goal of this subsection is not to establish causality, but to provide external validation of the estimated match-specific productivity. The positive and significant association between postmatching firm performances and $M_{quality}$ confirms that the complementarities in size/talent, diversification/cross-industry experience, and R&D/education are important sources of the match-specific productivity.

[16] In the ROA and ROE regressions, *Free Cash Flow* was dropped as a control to avoid collinearity with the dependent variable.

References

[1] Demerjian, Peter R., Baruch Itamar Lev, and Sarah McVay, 2012, "Quantifying managerial ability: A new measure and validity tests," *Management Science* 58(7), 1229-1248.

[2] Falato, Antonio, Dan Li, and Todd Milbourn, 2014,"Which skills matter in the market for CEOs? Evidence from pay for CEO credentials," *Management Science*, Forthcoming.

[3] Fox, Jeremy, and Patrick Bajari, 2013, "Measuring the efficiency of an FCC spectrum auction," American Economic Journal: Microeconomics, 5(1), 100-146.

[4] Hund, John, Don Monk, and Sheri Tice, 2010, "Uncertainty About Average Profitability and the Diversification Discount," *Journal of Financial Economics* 96(3), 463-484.

[5] Jacquemin, A.P. and C.H. Berry, 1979, "Entropy measure of diversification and corporate growth," *Journal of Industrial Economics* 27, 359-369.

[6] Kaplan, Steven N., Mark Klebanov and Morten Sorensen, 2012, "Which CEO Characteristics and Abilities Matter?" *Journal of Finance* 67(3), 973-1007.

[7] Kimberly, J. R., and M. J. Evanisko, 1981, "Organizational innovation: The influence of individual, organizational and contextual factors on hospital adoption of technological and administrative innovation," *Academy of Management Journal* 24, 689-713.

[8] Kuhnen, Camelia M., and Jeffrey Zwiebel, 2009, "Executive pay, hidden compensation and managerial entrenchment," working paper.

[9] Li, Fei, and Masako Ueda, 2006, "CEO-firm match and principal-agent theory," working paper.

[10] Morck, Randall, Andrei Shleifer, and Robert W. Vishny, 1988, "Management ownership and market valuation," *Journal of Financial Economics* 20, 293-315.

[11] Rogers, Everett and F.F. Shoemaker, 1971, *Communication of innovations: A cross cultural approach*, The Free Press.

[12] Storn, R., and Price, K., 1997, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization* 11, 341-359.

[13] Strebulaev, Ilya and Toni Whited, 2012, "Dynamic models and structural estimation in corporate finance," *Foundations and Trends in Finance* 6, 1-163.

[14] Wiersema, Margarethe F., and Karen A. Bantel, 1992, "Top management team demography and corporate strategic change," *Academy of Management Journal* 25, 91-121.

Table A1: Comparative Statics

In this table, I carry out a few Monte Carlo exercises in the spirit of comparative statics explained in Section III.4. $logN(0,1)$ is a lognormal distribution with mean 0 and standard deviation of 1. $U(0,1)$ is a uniform distribution between 0 and 1. In each experiment, I specify X, F, the matching pattern and the sharing rule (i.e., compensation p). I run each model 100 times, and report the average maximum score estimator, the (5%, 95%) confidence internal of the estimates, and the model fit in the last column. Experiment (1.1) provides a case of perfectly assortative matching between firms and managers, where managerial compensation also sustains a pairwise stable equilibrium. Experiment (1.2) provides a case where managers are randomly assigned to firms, and they evenly split F. Experiment (1.3) provides a case where firms and managers are assortatively matched, but managerial payoff is a random draw from the uniform distribution. Experiment (1.4) provides a case where both assignments and pay are random. Matching models in experiments (2.1) to (2.3) are multidimensional. In experiments (2.1) and (2.3), firms and managers are more assortatively matched on the dimension with higher complementarity. They differ in that managerial payoff in (2.3) is random. Experiment (2.2) provides a case where firms and managers are more assortatively matched on the dimension with less complementarity.

	First Dim. C_1M_1	Second Dim. C_1M_2	F	p	Maximum Score Estimator
(1.1)Perfectly assortative, pairwise stable	$C_1 \sim 10logN(0,1)$ $M_1 = C_1$		C_1M_1	$0.5F$	$\hat{\omega_1} = 1$ 100% I.R. satisfied
(1.2)Random pairing, 50/50 split	$C_1 \sim 10logN(0,1)$ $M_1 \sim 10logN(0,1)$		C_1M_1	$0.5F$	$\hat{\omega_1} = 1$ (0.95, 1.06) 26% I.R. satisfied
(1.3)Assortative matching, random payoff	$C_1 \sim 10logN(0,1)$ $M_1 = C_1$		C_1M_1	$p \sim$ $10U(0,1)$	$\hat{\omega_1} = 0.02$ (0.02, 0.03) 14% I.R. satisfied
(1.4)Random pairing, random payoff	$C_1 \sim 10logN(0,1)$ $M_1 \sim 10logN(0,1)$		C_1M_1	$p \sim$ $10U(0,1)$	$\hat{\omega_1} = 0.03$ (-0.03, 0.04) 6% I.R. satisfied
(2.1)More assort. in the dim. with higher complem.	$C_1 \sim 10logN(0,1)$ $M_1 = C_1$	$M_2 \sim$ $10U(0,1)$ $+0.1C_1$	$10C_1M_1$ $+C_1M_2$	$p(c_j,n_j) =$ $F(c_j,n_j)-$ $F(c_j,n_i)+p(c_i,n_i)$	$\hat{\omega_1} : \hat{\omega_2} = 10.0$ (9.8, 10.3) 44% I.R. satisfied
(2.2)More assort. in the dim. with lower complem.	$C_1 \sim 10logN(0,1)$ $M_1 = C_1$	$M_2 \sim$ $10U(0,1)$ $+0.1C_1$	C_1M_1 $+10C_1M_2$	$p(c_j,n_j) =$ $F(c_j,n_j)-$ $F(c_j,n_i)+p(c_i,n_i)$	$\hat{\omega_1} : \hat{\omega_2} =$ -1.6 (-9.9, 0.9) 4% I.R. satisfied
(2.3)Random payoff	$C_1 \sim 10logN(0,1)$ $M_1 = C_1$	$M_2 \sim$ $10U(0,1)$ $+0.1C_1$	$10C_1M_1$ $+C_1M_2$	$p \sim$ $100U(0,1)$	$\hat{\omega_1} : \hat{\omega_2} = 0.2$ (-35.0, 11.8) 13% I.R. satisfied

Table A2: Maximum Score Estimates in Specification (5)–Without Transfer Data

In this table, I report results using conditions in Fox (2010b), without using the transfer data, and the results for the production function (8). $Assets$, Div_{sale}, RD, $Talent_{dea}$, Exp_{sale}, Edu_{tech} are firm and manager characteristics in their annual CDFs.

Panel A. Maximum Score Estimates without Using Transfer

	Production Function Estimates		Production Function Estimates	
Interaction Terms	Point Estimates	95% CI	Point Estimates	95% CI
$Assets * Talent_{dea}$	1	Superconsistent	1	Superconsistent
$Div_{sale} * Exp_{sale}$	8	(3.4, 9.2)	7.6	(5.0, 9.4)
$RD * Edu_{tech}$	11.3	(4.2, 12.8)	10.7	(6.1, 12.8)
$Assets * Talent_{dea}$			1.2	(0.9,2.2)
$Assets * Edu_{tech}$			0.3	(-0.7,0.6)
$Div_{sale} * Edu_{tech}$			0.5	(-0.3, 0.7)
No. of I.R.	171,764		171,764	
% satisfied	63.30%		67.10%	

Panel B. Maximum Score Estimates in Specification (8)

Interaction Terms	Point Estimates	95% CI
-$1[Assets \geq Talent_{dea}] * (Assets - Talent_{dea})^2$	2.8	(0.9, 3.1)
-$1[Assets < Talent_{dea}] * (Assets - Talent_{dea})^2$	2.6	(1.2, 3.5)
-$1[Div_{sale} \geq tExp_{sale}] * (Div_{sale} - Exp_{sale})^2$	3.2	(1.7, 4.8)
-$1[Div_{sale} < Exp_{sale}] * (Div_{sale} - Exp_{sale})^2$	1.0	(0.5, 1.3)
-$1[RD \geq Edu_{tech}] * (RD - Edu_{tech})^2$	3.0	(1.8, 4.2)
-$1[RD < Edu_{tech}] * (RD - Edu_{tech})^2$	2.6	(1.7, 3.8)
1(same industry)	53	(28.1, 70.5)
No. of I.R.	343,528	
% satisfied	54.50%	

Table A3: Many-to-one Matching

This table reports the results for the many-to-one matching model outlined in section A.4. Firm characteristics $Assets$, Div_{sale}, and RD, all in their annual CDFs. Manager characteristics $Talent_{dea}$, Exp_{sale}, and Edu_{tech} are all in their annual CDFs as well. The Σ variables sum over each manager attribute across the new hires in each firm-year. The δ variables are standard deviations in each manager attribute across the new hires in each firm-year. 1[same ind] is an indicator variable that equals one if the executive had managerial experience at firms in the same (two digit SIC) industry as the current firm before becoming its top executive. The percentage of inequalities satisfied with the estimates serves as a measure of the statistical fit.

	Production Function Estimates	
Interaction	PointEst	95%CI
$Assets * \Sigma(Talent_{dea})$	0.4	(0.1,0.8)
$Div_{sale} * \Sigma(Exp_{sale})$	4.8	(3.0,6.1)
$RD * \Sigma(Edu_{tech})$	2.8	(1.2, 3.8)
$Assets * \delta(Talent_{dea})$	-19.0	[-27.8, -4.3]
$Div_{sale} * \delta(Exp_{sale})$	-7.3	[-14.3, -2.6]
$RD * \delta(Edu_{tech})$	-12.6	[-20.1, -9.5]
1[same ind]	76.8	(67.3, 89.2)
No. of I.R.	343,528	
% satisfied	53.80%	

Table A4: Additional Reduced-form Analysis

This table reports the regression results Using Exp_{sale} and Edu_{tech} as talent measures.

	$Tcomp$	
	(1)	(2)
$Assets$	0.30*	0.55***
	(0.17)	(0.16)
Div_{sale}	0.03	0.01
	(0.13)	(0.10)
RD	-0.47	-0.43
	(0.30)	(0.29)
Exp_{sale}	-0.13	-0.01
	(0.14)	(0.09)
Edu_{tech}	-0.15	0.0001
	(0.20)	(0.20)
$Assets * Exp_{sale}$	0.26	
	(0.19)	
$Assets * Edu_{tech}$		-0.19
		(0.15)
$Div_{sale} * Exp_{sale}$	-0.0004	0.04
	(0.18)	(0.14)
$RD * Edu_{tech}$	0.52*	0.45*
	(0.31)	(0.27)
Firm FE	x	x
Observations	2,025	2,025
Adj. R^2	0.360	0.358

Table A5: Matching Quality and Firm Performance

This table reports regression results for Tobin's Q, ROA, and ROE (definitions provided in section IV) on matching quality with control variables. Panel [1] includes the industry and year F.E. Panels [2] and [3] include firm and year F.E. $M_{quality}$ is measured by $\sum_{n=1,2,3} \hat{\omega_n} C_n M_n$ with $\hat{\omega_n}$ estimated in Table 4, Panel B, Column [3]. *RD* and *Advertising* are the company's R&D and advertising expenses scaled by the stock of property, plant, and equipment (PPE). *Leverage* is calculated as long-term debt divided by total assets. The managerial ownership variable includes the direct stock ownership, restricted stock holdings, shares acquired by option exercise, and shares acquired on vesting. $log(Sales)$ is the logarithm of sales. $Tangible/S$ is the ratio of PPE to sales. *CapEx* is the ratio of capital expenditures to PPE. *FreeCF* is calculated as operating cash flow divided by total assets. *Volatility* is calculated as the standard deviation of daily stock returns for the fiscal year. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	[1]			[2]			[3]		
	Q	ROA	ROE	Q	ROA	ROE	Q	ROA	ROE
$M_{quality}$	0.04*	0.21**	0.67*	0.07**	0.13*	0.84**	0.05**	0.22*	0.74*
	(0.02)	(0.08)	(0.35)	(0.03)	(0.07)	(0.43)	(0.02)	(0.12)	(0.38)
RD							0.44	-5.80	-18.17**
							(0.55)	(3.80)	(7.99)
$Advertising$							1.40	0.46	-5.93
							(1.65)	(5.88)	(20.19)
$Leverage$							-0.54*	-12.12***	-25.98**
							(0.28)	(3.06)	(11.50)
$Ownership$							0.10	0.49	1.10
							(0.06)	(0.39)	(0.93)
$Ownership^2$							-0.39*	-3.88***	-7.61**
							(0.22)	(1.23)	(3.10)
$log(Sales)$							-1.55***	7.68*	11.81
							(0.55)	(3.93)	(9.56)
$log(Sales)^2$							0.06**	-0.44*	-0.65
							(0.03)	(0.24)	(0.59)
$Tangible/S$							-1.76**	-24.02***	-46.36***
							(0.65)	(5.49)	(12.37)
$(Tangible/S)^2$							0.30**	5.61***	11.24***
							(0.13)	(1.74)	(3.81)
$CapExp$							3.21***	14.12**	46.16***
							(1.03)	(5.49)	(13.42)
$FreeCF$							3.74***		
							(0.66)		
$Volatility$							0.01	-2.79***	-7.56***
							(0.02)	(0.48)	(1.53)
Ind & Yr FE	x	x	x						
Firm & Yr FE				x	x	x	x	x	x
Observations	3,548	3,527	3,415	3,548	3,527	3,415	3,223	3,210	3,108
Adj. R^2	0.08	0.09	0.08	0.41	0.49	0.46	0.52	0.53	0.53