This is the final working paper version of an article that appeared in the American Economic Review, 2008, 98(3), p. 642-668.

The Difference That CEOs Make: An Assignment Model Approach

Marko Terviö*
Haas School of Business
University of California, Berkeley

July 30, 2007

Abstract

This paper presents an assignment model of CEOs and firms. The distributions of CEO pay levels and firms' market values are analyzed as the competitive equilibrium of a matching market where talents, as well as CEO positions, are scarce. It is shown how the observed joint distribution of CEO pay and market value can then be used to infer the economic value of underlying ability differences. The variation in CEO pay is found to be mostly due to variation in firm characteristics, whereas implied differences in managerial ability are small and make relatively little difference to shareholder value. The value-added of scarce CEO ability within the 1000 largest firms in the US was about $21-25 billion in 2004, of which the CEOs received about $4 billion as ability rents while the rest was capitalized into market values.

*This paper is based on Chapter 2 of my Ph.D. Thesis. I thank David Autor, Abhijit Banerjee, Ernesto Dal Bó, Glenn Ellison, Ben Hermalin, Bengt Holmström, Jonathan Leonard, Michael Katz, and Thomas Philippon for helpful comments, and the Yrjö Jahnsson Foundation and the Finnish Cultural Foundation for financial support.

1 Introduction

The levels of CEO pay are controversial and widely resented in the popular press. The academic literature on executive compensation has mainly focused on the incentive structure of CEO pay, while its levels have received much less attention. The average pay of a CEO in the largest 100 firms has hovered between \$16 and \$30 million in recent years. Could anything resembling competition for talent be able to explain such pay levels—or are they prima facie evidence of wrongdoing or at least market imperfections?

This paper argues that purely competitive models have not been taken as far as they could, and that observed pay levels are consistent with competition for small differences in talent. We develop a simple assignment model and show how it can be used to infer the unobserved distribution of ability from the observed joint distribution of CEO pay and market value, assuming it is the competitive equilibrium of a market where heterogeneous firms and individuals match. The model is then calibrated to measure the social value of scarce executive ability and to gauge the extent to which observed levels of CEO pay can be explained by differences in talent and to what extent by variation of firm size.

The predominant fact about the distribution of executive pay is that large firms pay their CEOs more than small firms.[1] Intuition suggests that the economic impact of a manager's decisions depends on the amount of resources under his control, so that the observed strong relation of firm size and CEO pay levels is a reflection of scarce executive ability being worth more to larger firms. That this relation should result in high levels and a skewed distribution of income for CEOs was proposed by Mayer (1960), who termed this the "scale-of-operations" effect. In a similar spirit, Manne (1965) argued that a major benefit of corporate mergers and takeovers is to allocate the control of resources according to managerial abilities. Lucas (1978) invoked Manne's suggestion to devise a theory of firm size distribution based on the allocation of capital to a population of potential managers of heterogeneous ability. Rosen (1982) presented a related model with a focus on the division of labor into managers and workers and the allocation

[1] The elasticity of CEO pay to firm size has been estimated at about 0.3 using various measures of firm size. See the survey by Murphy (1999), and Kostiuk (1990) whose data goes back to 1930's.

of subordinate labor between managers. In all these models, firms are inherently homogeneous: In equilibrium, all size differences between firms, as well as differences in CEO pay, arise from the heterogeneity of managerial ability. Any differences among firms are merely a manifestation of the mechanism by which the variation in talent is magnified into higher variation in CEO pay. As a result, all differences in CEO pay are then necessarily explained by differences in talent, either directly or via the scale of operations effect.

In the spirit of Rosen (1982), we develop a model where "the distributions of firm size and managerial reward are the joint outcomes of the same underlying problem." However, we will argue that not just individuals, but also firms are differentiated by important indivisible characteristics that cannot easily be shuffled among firms. In other words, there is also an exogenous firm-specific component behind the cross-sectional variation in firm size. This simple feature has far-reaching implications for the understanding of CEO pay. It means that an assignment model is needed to understand the determination of the levels of CEO pay.[2] In an assignment model, different types of indivisible units of production—here managers and firms—are matched in fixed proportions, and the equilibrium distributions of income to both factors depend in a non-obvious way on the full distributions of the qualities of both factors. In particular, the competitive price of ability does not reflect its marginal productivity in the usual sense of the term.

The assignment model of CEOs and firms to be applied in this paper was first presented in Terviö (2003). It builds on the "differential rents" model of Sattinger (1979) by adding adjustable capital that is endogenously allocated between the matched pairs of firms and managers.[3] The basic simplifying assumption is that there is a competitive and frictionless labor market for executive ability, which is equally applicable in all companies, but is more productive at larger companies. Even though all firms would rather hire the most able individual for the job, it is the companies where ability is at its most productive that will pay the most for it and therefore

[2] The seminal assignment models are Koopmans and Beckmann (1957) and Tinbergen (1956, 1957).

[3] See also the survey of assignment models by Sattinger (1993), which includes a detailed exposition of the "differential rents" model, and a related general equilibrium approach in Teulings (1995).

attract the best individuals. In equilibrium, each firm must prefer hiring its CEO at his equilibrium pay level to hiring any other company's CEO at his or her pay level. The setup has a continuous distribution of workers and firms, which rules out match-specific rents and, therefore, any need to model bargaining, and a complementary production function which generates positive assortative matching (here meaning the matching of the best managers with the largest firms). In this setup, the pay levels of individuals depend on the distributions of firm size and CEO ability in the economy in a relatively straightforward way.

The basic assignment model shows how the economic surplus produced by matched pairs of CEOs and firms gets divided into incomes. However, the firms' share of this surplus is not directly observable in the data: Market values are equilibrium outcomes, into which the effects of both current and future CEOs are capitalized. Also, part of firms' capital stock may be adjustable even in the short run; however, any income accruing to such adjustable assets is not determined according to the assignment model (instead, adjustable assets should just earn their marginal product). We extend the basic assignment model to resolve these issues, and then show how the model can be used to infer the unobserved distributions of ability and firm characteristics (up to undetermined constants) from the observed joint distribution of CEO pay and market value. The model can then be used to answer quantitative questions about the effects of CEO ability on profits and CEO pay. However, these questions necessarily take the form of counterfactuals about the distributions of ability or firm size.

In the empirical part we use CompuStat data on the 1000 largest publicly traded companies in the US in 1994–2004. First we quantify the relative importance of heterogeneity in ability and firm size toward explaining the cross-sectional variation in CEO pay. As will be explained, the value of existing ability can only be measured relative to some counterfactual distribution of replacement CEOs. Our main counterfactual is the difference that CEOs make to total economic surplus, compared to if they were all replaced by the lowest type CEO in the sample. In 2004, this added value was about \$21–25 billion, of which the top CEOs received \$4.4 billion in total as a rent to their scarce ability. The remainder is capitalized in market values, and is quite small compared to the total market value of \$12.6 trillion. Similarly, the additional value if all CEOs became as good as the current highest type would be worth less than \$3.5 billion. Such an increase in ability levels would be associated with a decrease of over \$1 billion in total CEO pay (as increased competition would reduce the Ricardian component in the rents to top CEOs), leaving a net gain of about \$4.5 billion for the shareholders.

By contrast, if abilities remained unchanged but the existing firms were replaced by the 1000th largest type, the effects on CEO pay would be much more dramatic. Under status quo, the CEO at the largest firm is expected to earn over \$15 million more a year than the CEO at the 1000th largest; this pay difference would be cut by a factor of 5. In total, the rents to ability would be reduced from \$4.4 billion to less than \$2 billion. If, on the other hand, all firms were as large as the actual biggest firm, then total CEO rents would increase by about a factor of 100, with the highest types expecting to earn over \$700 million per year. We conclude that the observed high levels of top CEOs are mainly due to firm scale rather than the scarce ability of CEOs.

We then investigate how well the assignment model can explain the recent fluctuations in the levels of CEO pay and market values, purely based on time-variation of a single scaling parameter. We interpret this parameter as capturing the (size-neutral) variation in productivity, which interacts with current CEO ability to generate future profits. By and large, the model provides a reasonable fit for the coevolution of CEO pay and market values during the sample years. However, the boom years 2000–2001 fit less well; specifically the unusually high highest levels of CEO pay cannot be generated from the same model that fits in other years.

Unlike most of the literature, in this paper the structure of pay is not considered, only the level. Differences in required effort or risk-bearing are presumed to have insignificant explanatory power for the variation in pay levels compared to individual ability and firm-specific usefulness for that ability. Systemic failures and agency problems, such as, for example, the skimming explanation of Bertrand and Mullainathan (2001), empire-building (Jensen, 1986), and ratcheting (Ang and Nagel, 2006), are likewise ignored. Our model takes a reduced-form approach to all incentive problems; the expected cost of a CEO's compensation is interpreted as the market price of the effective managerial ability that can be bought with the existing contracting technology. The upside of ignoring the structure of pay is that we are able to analyze the determination of the whole distribution of CEO pay levels as an equilibrium outcome.

A paper within the incentive literature, that is somewhat similarly motivated to this one, is Baker and Hall (2004). They explore the relation of incentives and firm size while assuming away differences in ability. In their model, effort and firm size are allowed to be complementary, so the optimal level of effort and sensitivity of compensation to market value depend on firm size. Using cross-sectional data on the structure of CEO pay and firm size, they find evidence for a substantial complementarity.

The recent literature on the levels of CEO pay in the competitive framework includes Murphy and Zabojnik (2004) and Frydman (2005) who attribute the growth to increasing generality (as opposed to firm-specificity) of the required managerial skills. Cuñat and Guadalupe (2006) find evidence that increased foreign competition has raised the productivity of managerial talent. Gayle and Miller (2005) present a calibration method and find that the increase in firm size has increased the cost of deterring moral hazard. Recently also Gabaix and Landier (2007) study CEO pay with an assignment model, which differs from Terviö (2003) by assuming specific forms for the unobserved distributions (with talent following the extreme value distribution), and introduces several extensions. Using market value as an exogenous measure of firm size, they find that the six-fold increase of CEO pay in the US between 1980 and 2003 can be fully explained by the six-fold increase in market capitalization during that period.

The body of the paper is divided into four parts. Section 2 introduces the basic assignment model, and derives the equilibrium distributions of pay and profits using a method similar to that of screening models. This section can be skimmed by those familiar with assignment models, although the solution method and the discussion of comparative statics may be of independent theoretical interest. In Section 3 the assignment model is modified to take into account the specific features of the CEO-firm setup. It is shown how the model can be used to back out the differences in ability from the joint distribution of CEO pay and market value. The meaning and interpretation of the exogenous component if firm size is discussed in Section 4. The empirical results are presented in Section 5.

2 An Assignment Model of Pay

The distinctive feature of assignment models is that productive resources are embedded in indivisible units and these units must be combined in fixed numbers to produce output. Here the units will be individual managers and firms, and they are matched one to one. A production function describes the output resulting from matching any individual with any firm as a function of their fixed characteristics. A particularly tractable assignment model results from three simplifying assumptions: one-dimensional qualities, continuity, and complementarity. (This is the "differential rents" setup of Sattinger (1979).) The first two assumptions are made for analytical convenience, but the complementarity assumption is central to the analysis. Other simplifying assumptions are symmetric information and risk neutrality.

The first assumption means that individual and firm characteristics affecting output can both be summed up by one number; these factor qualities will be referred to simply as "ability" and "firm size," denoted by a and b respectively. Note that one-dimensional ability does not preclude different individuals having different strengths contributing to their ability to affect output.

Second, it is assumed that the production function is continuous and strictly increasing in both of its arguments, and that there is a unit mass of individuals and firms with "smoothly" distributed characteristics. The distributions of a and b have continuous finite supports without gaps; the resulting distributions of output and factor incomes will inherit these properties. Dispensing with this assumption would only complicate the notation without bringing more insights.

The substantive assumption about technology is the complementarity between ability and firm size, i.e., that the production function has a positive cross-partial. In this case, efficiency requires positive assortative matching: the best individual must be matched with the largest firm, the second best with the second largest etc. If the sorting were not perfect, then total output could be increased by shuffling some individuals between firms.[4] The

[4] Positive assortative matching ("positive sorting") maximizes the output from matching $a_1 \leq a_2$ and $b_1 \leq b_2$ if $Y(a_1, b_1) + Y(a_2, b_2) \geq Y(a_1, b_2) + Y(a_2, b_1)$. Rearranging this inequality to $Y(a_2, b_2) - Y(a_1, b_2) \geq Y(a_2, b_1) - Y(a_1, b_1)$ illustrates that complementarity can also be defined as "increasing differences" in the production function.

equilibrium matching of individuals and firms is thus very simple as is the determination of equilibrium output. It is the division of output into factor incomes (wages and profits) that requires further analysis.

It will be convenient to refer to distributions by their inverse distribution functions or "profiles." Think of the individuals as ordered by their ability on the unit interval, so that $a[i]$ is the ability of an i quantile individual and $a'[i] > 0$. Denoting the distribution function by F_a, the profile of a is defined by

$$a[i] = a \text{ st. } F_a(a) = i. \tag{1}$$

If there were atoms in the distribution of a they would correspond to flat parts in the profile, while gaps in the support of a would appear as jumps.

Using the quantile i as the variable (instead of the arbitrary units of ability) the distributions of factor incomes can then intuitively be solved in a manner analogous to the standard method of solving screening models. We believe this quantile approach to be more intuitive and tractable than the traditional method of working with density functions, especially when considering empirical applications.

2.1 Equilibrium: Determination of Pay Levels and Profits

In competitive equilibrium, the profiles of factor incomes must support the efficient matching of individuals and firms, which we know involves perfect sorting by quality. Two types of conditions must hold in competitive equilibrium. First, there are the sorting constraints: every firm must prefer hiring its efficient match at the equilibrium wage to hiring any other individual at their equilibrium wage. Second, there are the participation constraints: All firms and individuals must be earning at least their outside income.

$$\begin{array}{rlll} Y(a[i], b[i]) - & w[i] \geq Y(a[j], b[i]) - w[j] & \forall i, j \in [0,1] & \text{SC}(i,j) \\ Y(a[i], b[i]) - & w[i] \geq \pi^0 & \forall i \in [0,1] & \text{PC-}b[i] \\ & w[i] \geq w^0 & \forall i \in [0,1] & \text{PC-}a[i] \end{array} \tag{2}$$

The outside opportunities (w^0, π^0) are assumed to be the same for all units.[5] The unit mass should be thought of as a normalization of the mass of pairs of individuals and firms that are active in equilibrium. The lowest active

[5] A weaker assumption would do here, namely that the outside opportunities increase slower along the profile than the equilibrium incomes.

firm-individual pair ($i = 0$) is the one that just breaks even with the outside opportunity:

$$Y(a[0], b[0]) = \pi^0 + w^0. \tag{3}$$

The firms are not residual claimants in any sense: The equilibrium conditions could equivalently be stated in terms of individuals hiring firms.

The sorting constraints in (2) are mathematically analogous to the incentive compatibility conditions in a typical nonlinear pricing problem.[6] As in nonlinear pricing problems, the amount of constraints can be reduced drastically by noticing that, for any $i \geq j \geq k$, the sum of two adjacent sorting conditions SC(i, j)+SC(j, k) implies SC(i, k). The binding constraints are the marginal sorting constraints that keep firms from wanting to hire the next best individual, and the participation constraints of the lowest types. Regrouping the sorting constraint SC$(i, i - \varepsilon)$ and dividing it by ε gives

$$\frac{Y\left(a[i], b[i]\right) - Y(a[i - \varepsilon], b[i])}{\varepsilon} \geq \frac{w[i] - w[i - \varepsilon]}{\varepsilon}. \tag{4}$$

This becomes an equality as $\varepsilon \to 0$ and, via the definition of the (partial) derivative, yields the slope of the wage profile.

$$w'[i] = Y_a(a[i], b[i])a'[i] \tag{5}$$

The wage profile itself is then obtained by integrating the slope and adding in the binding participation constraint $w[0] = w^0$.

$$w[i] = w^0 + \int_0^i Y_a(a[j], b[j])a'[j]\mathrm{d}j \tag{6}$$

where Y_a denotes the partial derivative. Analogously, or as the remainder from $y = \pi + w$, the profile of profits satisfies

$$\begin{aligned} \pi'[i] &= Y_b(a[i], b[i])b'[i] && (7) \\ \pi[i] &= \pi^0 + \int_0^i Y_b(a[j], b[j])b'[j]\mathrm{d}j. && (8) \end{aligned}$$

All inframarginal pairs produce a surplus over the sum of their outside opportunities, and the division of this surplus depends on the distributions of factor quality. At any given point in the profile the increase in surplus

[6] Note that here one-to-one matching precludes "bunching" that is common in screening models.

is shared between the factors in proportion to their contributions to the increase at that quantile.[7]

Due to the continuity assumptions, the factor owners do not earn rents over their next best opportunity within the industry. In a discrete model there would be a match-specific rent left for bargaining, as the difference in the pay of two "neighboring" individuals could be anywhere between the differences of their firms' valuation for the ability difference between them. In a continuous model there is nothing to be bargained over because all units have arbitrarily close competitors. If one of the profiles has a jump at some quantile, then all of the increase in surplus at that point goes to the factor with a jump because the other side is still perfectly competitive. (There would be match-specific rents only if both of the exogenous factor profiles had jumps at the exact same quantile.)

One striking feature of this model industry is that factor owners are only affected by changes in the quality of those below them in the rankings. Mathematically, this is clear from the fact that the equations for factor income profiles take the form of integrals over the profiles below. Intuitively, the binding constraint on any factor owner is the quality and price of their next best competitor. For example, if an individual's next best competitor becomes less productive, then she can raise her price by a fixed amount, and this price increase spills upwards along the whole profile by shifting the division of surplus to individuals' favor by that same fixed amount at every firm.

A central feature to understand about the assignment model is that the unobserved productivity characteristics a and b are essentially ordinal. Any increasing transformation of "the scale of measurement" for a factor quality, combined with the inverse change in the functional form of the production function, changes nothing of substance in the model. This means, for example, that using a Cobb-Douglas form $Y(a,b) = Aa^{\gamma}b^{1-\gamma}$, as opposed to a simple multiplicative $y = ab$, would be superfluous, or even misleading if it causes one to believe that the income shares should have any tendency to

[7] The complementary two-factor model generalizes into more factors in the natural way. For example, if firms have many tasks, with workers for each task t drawn from a separate ability distribution $a_t[i]$, then the equilibrium income of the position-t worker at quantile i is $w_t[i] = w_t^0 + \int_0^i Y_{a_t}(b[j], a_0[j], a_1[j], \ldots, a_T[i]) a_t'[j] \mathrm{d}j$. This generalization will be invoked later with Y as the present value of surplus and a_t as the ability of the period$-t$ CEO.

be related to the exponents. This is a special case of a more general mistake of assuming that factors are paid their marginal products, in a situation where the amounts of two matching factors cannot be shifted across different units of production. This transferability of factors of production between firms is what pins down the linear scale of measurement for factor qualities in the usual case, and only the total quantity of a factor of production in the economy must adhere to some budget constraint. In an assignment setup there is much less flexibility. The "division" of productive characteristics between the units is what it is, and the economic problem is how to match these factor units into units of production.[8]

It would be incorrect to say that factors earn their marginal productivity by the usual definition of marginal productivity, because the increase in output if the individual of ability $a[i]$ were to increase in ability is proportional to $b[i]$, which is not the message of the wage equation (6). But if she were to increase in ability, then, in equilibrium, she would also move up in the ranking and be matched with a higher b—and other individuals would have to move down and experience a decrease in productivity.[9] Here we note that the "differential rents" assignment models (including our model) satisfy "the No-Surplus Condition" of Ostroy (1980, 1984), which is an alternative definition for a perfectly competitive equilibrium. This means that individuals in fact do receive their marginal product, once the margin is defined correctly. As ability cannot conceivably be extracted from one individual and poured into another, the relevant margin here is whether an individual will participate in the industry or not—and if not, then the effect of the resulting rearrangement of remaining individuals is part of the marginal product.

[8]While it is not sensible to make predictions about the effects of taxation in a model where effort is supplied inelastically, it is worth pointing out as a curiosity that any level of progressivity in income taxation would not reduce efficiency here, as long as the equilibrium matching is not disturbed, i.e., as long as after-tax income is increasing in pre-tax income.

[9]An alternative method for deriving the wage equation (6), as the properly defined marginal product of individual ability, is to consider the decrease in industry surplus if a vanishingly small mass of individuals at quantile i were to leave the industry. See Section 2.4 in Terviö (2003).

2.2 Comparative Statics

Uniform productivity growth Consider a change by which the production function Y is multiplied by some constant G but the distributions of factor qualities a and b remain unchanged. By inspection of (6) and (8), it is clear that the rents earned over the outside opportunities must then change by that same multiple. After all, such productivity growth is mathematically equivalent to changing the units of measurement for output. If also the outside opportunities $(w^0,\ \pi^0)$ change by the same multiple, then factor incomes adhere to the same scaling.

Scaling Lemma. If $Y_t\left(a,b\right) = GY\left(a,b\right), w_t^0 = Gw^0$ and $\pi_t^0 = G\pi^0$ then $w_t[i] = Gw[i]$ and $\pi_t[i] = G\pi[i]$ for all $i \in [0,1]$.

Notice that the scaling of factor incomes holds for any production function and regardless of the shapes of the distributions of a and b. However, if the outside opportunities do not move in lockstep with productivity, then the break-even level output does not scale with productivity and the size of the industry would change through activation or inactivation of some potential firms. This in turn could change the division of surplus at all firms.

When the production function is multiplicative, $Y(a,b) = ab$, then any change in overall productivity is observationally equivalent to the same change having affected either all ability levels or all firm sizes. When all incomes double there is no way of telling from income data whether it is due to a doubling of abilities or firm sizes.

Change in the shape of a distribution The multiplicatively separable production function, which will be used in the empirical part of this paper, lends itself to a simple graphical depiction of the equilibrium and the comparative statics of the model. Figure 1 depicts an example of a matching graph that arises from two particular distributions of factor qualities a and b. The graphical convenience of multiplicativity comes from the fact that the level of output from matching an individual of type a and a firm of type b is the rectangle between the point $\{b,a\}$ and the origin. The matching graph $a = \varphi(b)$, defined by $a[F_b(b)] = \{(a,b) \text{ st. } F_a(a) = F_b(b)\}$, is a strictly increasing curve, with slope

$$\varphi'(b) = a'[F_b(b)]f_b(b) = \left.\frac{a'[i]}{b'[i]}\right|_{i=F_b(b)}. \tag{9}$$

Changes in the distribution of either factor quality appear as a change in the shape of the matching graph. (By contrast, uniform productivity growth would merely change the levels of output associated with each isoquant, but have no effect on their shapes or the shape of the matching graph.) For example, if individuals become more able, in the sense of first-order stochastic dominance, then the matching graph shifts up. Note, however, that the matching graph alone does not tell how the mass of matched pairs is distributed on top of it, only that there is a positive mass.

The area of the smaller rectangle in Figure 1 is the break-even level of output, $y[0] = a[0]b[0]$, that just covers the reservation prices of the factors. The division of this minimum output is exogenous; the area of the shaded triangle represents the reservation wage of individuals, $w[0]$. Inframarginal pairs $i > 0$ create an additional surplus $y[i]-y[0]$, whose division depends on the distributions of a and b in a simple way: the surplus going to individual i is represented by the area between $a[i]$ and $a[0]$ and to the left of the matching graph.[10] While moving up the matching graph, the size of the rectangle representing output increases. The contribution of a higher ability a to this increase is proportional to the horizontal side of the rectangle, which is $b = \varphi^{-1}(a)$. Conversely, the marginal productivity of b is $\varphi(b)$. The wage of any type of an individual can be read off the graph in a similar matter; the entire shaded region represents the wage of the highest type $a[1]$. Note again that an individual's wage is not merely a function of ability and the size of her own firm, but depends on the shapes of the distributions of a and b below, summarized in the matching graph.

To illustrate the model's nonstandard implications, it is useful to do a comparative statics exercise with the distribution of abilities as the variable. Suppose that the ability of individuals between quantiles i^* and 1 is increased, while the qualities of firms and lower-ranked individuals are unchanged. The new matching graph is shown by the dashed line above the small dark shaded region in Figure 1. The distribution of b is not changed, so the quantiles move vertically: the ith quantile is matched exactly above

[10] To see why this is the case, change the variable of integration in the wage equation (6) from quantile j to ability a. Then $j(a) = F_a(a)$, $\frac{\mathrm{d}j}{\mathrm{d}a} = \frac{1}{a'[j]}$ and $b[j(a)] = \varphi^{-1}(a)$. This results in

$$w[i] - w^0 = \int_0^i a'[j]b[j]\mathrm{d}j = \int_{a[0]}^{a[i]} \varphi^{-1}(a)\mathrm{d}a, \tag{10}$$

which is indeed the area of the shaded region between $a[i]$ and $a[0]$.

where it used to. It can be seen that the pay levels of the highest types of individuals must go down, even though the output is up at every firm in $(i^*, 1)$ and unchanged elsewhere. The loss in the pay of the very highest type $a[1]$ is the entire dark shaded region.

Individuals gain from their increased productivity, but also lose due to tougher competition from other individuals below. The income of the lower range of the improved individuals necessarily goes up as a result of the change, as the competition effect is not very strong. For example, for the individual at quantile i, whose ability increases from $a[i]$ to $a^+[i]$, the effect of tougher competition from below is the the income loss represented by the area of the dark shaded region below $a[i]$. The gain from increased ability for the same individual is the area of the light shaded region between $a[i]$ and $a^+[i]$. The highest types must be made worse off since all they get is the loss. Of course, if everyone's ability were to increase sufficiently, then all individuals can be better off. For this it would be necessary for the highest level of ability to increase enough to retain a sufficient relative advantage over its lower-ability competitors. Inspection of the wage equation (6) reveals that a sufficient condition for everyone's pay to increase is that the slope of the ability profile should increase at every quantile.

From the point of view of the unchanging factor, the gains are unambiguous: all firms of types $b[i^*]$ or higher are better off than before. The dark region to the left of $b[i]$ is the resulting gain for firm i. The converse results hold if a section of firms became "more productive," i.e., experienced an increase in b. Individuals of lower ability would feel no "trickle-down" effect from increased productivity at the higher-level firms, but instead there would be a trickle-up effect and the ablest individuals would gain the most. High-ability individuals gain whether the level of output at their firm is increased or not, because the value of ability at lower-ranked firms has been increased, shifting the division of surplus to individual's favor.

These comparative statics results can be summed up in terms of first-order stochastic dominance of an interval that excludes the maximum. If the new distribution of ability dominates its old distribution and the distribution of firm size is held fixed, then the new distribution of profits dominates its old distribution, but the new distribution of pay levels does *not* dominate its old distribution (vice versa for a change in firm size distribution).

This example also illustrates why studying an earnings function can be misleading. Equilibrium relations such as $w(a)$, or equivalently $w(b)$, depend on the full distributions of a and b. Even if ability and the earnings function were observed directly, it would give the wrong predictions about (even the signs of) the changes in earnings, if (more than a zero measure of) individuals were to change in ability.

3 An Assignment Model for CEOs

3.1 The Setup

To adapt the assignment model to the CEO market, two complications must be dealt with. The first is that the economic surplus created as a result of the interaction between the current CEO and the firm is not directly observable. The market value of a firm is affected not just by the current CEO but also by the expectations concerning all future CEOs. The second complication is that part of market value may reflect the value of capital that could easily be transferred between firms. The income to such adjustable factors is not determined within the equilibrium of the assignment problem: The marginal product of adjustable capital is defined in the traditional way so every unit of it should expect to earn the same return at every firm. In this section we show how, under simplifying assumptions, both of these complications affect the empirical interpretation of the assignment model.

Our basic functional form assumption is that ability and firm size interact multiplicatively. Surplus generated in one period is assumed to be $y(a, b) \propto ab$, where b is firm size and a is management ability. This form covers, without loss of generality, all multiplicatively separable production functions $y(a, b) = f(a)h(b)$, where f and h are strictly increasing functions. The motivation for selecting the multiplicative form is to use the simplest form that exhibits the complementarity necessary to generate assortative matching.[11] The exogenous component of firm size, b, describes the potential for surplus that is specific to the firm—we will return to discuss its interpretation in Section 4.[12]

[11] In the absence of assortative matching, a more general assignment model can still be relevant, but is likely to yield fewer clear insights. The seminal assignment model of Koopmans and Beckmann (1957) considers a general problem for matching plants and locations in a linear programming framework.

[12] Note that the distribution of market values is necessarily an equilibrium outcome—it

Market Value and the Effect of CEOs

The relevant management ability a that impacts the current flow of surplus may depend on both past CEOs and the current CEO. We assume that firms are infinitely lived and that the effective management ability for surplus in period t at a firm with a history of CEO abilities of $a_t, a_{t-1}, a_{t-2}, \ldots$ is a weighted average of current and past abilities

$$\begin{aligned} A_t &= A(a_t, a_{t-1}, a_{t-2}, \ldots) \\ &= \sum_{\tau=0}^{\infty} \alpha_\tau a_{t-\tau}. \end{aligned} \tag{11}$$

The impact weight α_τ gives the share of currently effective management ability that comes from the ability of the CEO τ periods ago. (By the same token, α_τ is also the fraction of the current CEO's total impact that occurs τ periods into the future.) Assuming that the CEO impact fades at a constraint rate λ, i.e., $\alpha_{\tau+1} = \alpha_\tau / (1+\lambda)$, and utilizing the normalization $A(a, a, a, \ldots) = a$, the impact weights are

$$\alpha_\tau = \frac{\lambda}{(1+\lambda)^{\tau+1}}. \tag{12}$$

In the limit of $\lambda \to \infty$ CEOs impact only contemporaneous earnings.[13]

To keep the model tractable, we make a strong stationarity assumption. It consists of three parts: 1) The distributions of ability $a[i]$ and firm size $b[i]$ are constant over time, 2) productivity grows deterministically at rate g at every firm, and 3) the values of outside opportunities w^0 and π^0 grow at rate g. The crucial simplification resulting from this assumption is that all firms can expect to stay at their current quantile in the distribution of firm size, and therefore to keep matching with a CEO of the same ability. The surplus generated at a firm of size b and a current CEO of type a, t periods from now, is therefore

$$y_t(a, b) = (1+g)^t ab \tag{13}$$

depends on the quality of CEOs that the firms are able to match with—so it cannot be used as the firm size variable b in the model.

[13] In the calibrations, λ will vary between 0.1 and ∞, implying a half-life of 7.3 years or less for the CEO impact. The lower bound for λ is admittedly "from the hat;" we hope to see empirical work on intertemporal substitutability of CEO ability in the future.

in year-t dollars. The present value of all surplus can now be stated as

$$Y(a,b) = \sum_{t=0}^{\infty} B^t ab = \frac{ab}{1-B}, \tag{14}$$

where $B = (1+g)/(1+r)$ is the growth-adjusted discount factor. The part of this surplus that goes to the firm is capitalized into market value, but most of the part that goes to CEO pay consists of the pay of future CEOs and is therefore inherently unobservable.

However, we know from the Scaling Lemma in Section 2.2 that, if the change in productivity is proportional across all units and if the outside opportunities of the factors also increase by the same proportion, then the share of the surplus going to each factor at each quantile i stays the same. In other words, under the strong stationarity assumption, the CEO pay at each firm is expected to grow at rate g. It follows that the ratio of the current CEO pay to the present value of all (current and future) CEO pay has to equal the "price-earnings" ratio of $1/(1-B)$. This is how the stationarity assumption allows us to translate the observed flow of CEO pay and the observed stock of market value into common units of measurement that can later be used to calibrate the assignment model. The breakdown of the surplus (14) into factor incomes at firm i can now be expressed as

$$\frac{w[i]}{1-B} + v[i] = \frac{a[i]b[i]}{1-B} \tag{15}$$

where the first term is the present value of pay to all CEOs that firm i will ever employ, and $v[i]$ is the market value of the firm. The division of this surplus into factor incomes is then determined from the distributions of b and a and their outside opportunities as seen in Section 2.1.

Adjustable Capital

In terms of the theory, the effects of adjustable factors of production have already been "partialled out" of the assignment model: The surplus is defined as the maximized surplus, net of the cost of all adjustable inputs. However, when some adjustable inputs are physically or legally embedded in one of the matching parties then they become a potential confounding factor in the empirical analysis. One such input could be the effort of the CEO, but we believe that the variation in CEO effort levels could only explain a trivial fraction of the variation in CEO pay and therefore abstract away from effort choice completely. However, the role of adjustable capital is potentially a significant issue here, so we now explore how the assignment model can be reconciled with the presence of adjustable capital.

The observed market value, to be denoted by v^*, is in fact a sum of the capitalized economic profits and the value of optimally chosen adjustable capital, k^*.

$$v^* = v + k^* \tag{16}$$

Only the "net" market value v is determined from the division of surplus in accordance of the assignment model, whereas adjustable capital must earn the market rate of return r. In order to model the determination of k^*, we assume that the gross surplus has constant elasticity θ with respect to adjustable capital. This means that the (net) surplus, that is available to be divided between the firm and the CEOs, consists of

$$y_t(a,b) = \max_{k_t} \left\{ \left(ab\zeta\,(1+g)^t\right)^{1-\theta} k_t^{\theta} - rk_t \right\} \tag{17}$$

where the strong stationarity assumption was used for $A_t = a$. By a convenient choice of the constant ζ,[14] the optimal level of adjustable capital can be written as

$$k_t^* = \frac{\theta}{r\,(1-\theta)}\,(1+g)^t\,ab. \tag{18}$$

The maximized surplus is then still

$$y_t(a,b) = (1+g)^t\,ab \tag{19}$$

which means that (15) continues to hold, despite the presence of adjustable capital. The value of adjustable capital can be removed from observed market value v^* by combining (15), (16), and (18), and solving for v. This results (in current period $t = 0$) in

$$\begin{aligned} v_0 &= \xi v_0^* - (1-\xi)\,\frac{w_0}{1-B}, \text{ where} \\ \xi &\equiv \frac{1-\theta}{1-\theta+\frac{\theta}{r}\,(1-B)}. \end{aligned} \tag{20}$$

Note that, at $\theta = 0$, all capital is sunk and $v_0 = v_0^*$.

[14] The convenient choice is $\zeta \equiv (1-\theta)^{-1}\,(r/\theta)^{\theta/(1-\theta)}$; this can be interpreted as the units of measurement for b. The placement of a and b inside the exponent $(1-\theta)$ is likewise a convenient choice of units, resulting in exponent-free units for a and b in (19).

The bottom line is that we remove the confounding effects of adjustable capital by transforming the observed market value into a part that reflects the capitalized income to the fixed factor b, which is determined in the equilibrium of the matching market between firms and CEOs. The transformation is done via equation (20), so it depends on the assumed values of three parameters: The discount rate r, the growth rate g (embedded in the discount factor B), and the elasticity θ. The calibrations will be conducted under a wide range of parameter values. Unfortunately there is no good way to pin down the value of θ. The data imposes an upper bound on the logically possible values of θ, as the value (20) cannot be negative at any quantile; this upper bound is $0.92 - 0.95$ in the data.[15] In addition, Interbrand's estimates of brand value for worlds' largest companies are between 5–51% of market capitalization. (Brand capital is perhaps the most clear-cut example of firm-specific sunk capital that affects the potential for profits.) For these reasons we believe that 0.8 is a loose upper bound for the share of adjustable capital for our calibrations. Fortunately, a number of interesting questions about ability and CEO pay turn out to not be overly sensitive to the assumed value of θ.

3.2 Empirical Inference of Distributions of Factor Quality

The basic idea for the inference of unobserved abilities comes from the observation that the slopes of the equilibrium factor income profiles (5) and (7) form a system of two differential equations, while the break-even condition (3) provides a boundary condition. Using the observed factor income profiles, the profiles of fixed factor qualities $a[i]$ and $b[i]$ can be solved from this system, up to an unknown constant of integration. In general, this system consists of a pair of nonlinear differential equations, and only a numerical solution is available. However, with the multiplicatively separable production function, the distributions of a and b can be solved in closed form (i.e., as functions of the data), up to multiplicative constants. The conditions for the income profiles must take into account the specific assumptions introduced in Section 3.1.

Use $Y(a_0, a, b)$ to denote the present value of surplus at a firm of size b that is in equilibrium matched with ability a, but where the current period

[15] These bounds obtain under the range of values that will be assumed for r and g.

CEO is of ability a_0. Combining (11) and (19), the present value is

$$\begin{aligned} Y(a_0, a, b) &= \sum_{t=0}^{\infty} B^t (\alpha_t a_0 + (1-\alpha_t)\, a) b \\ &= \frac{\lambda}{\lambda + 1 - B} (a_0 - a)\, b + \frac{ab}{1-B}, \end{aligned} \tag{21}$$

where (12) was used in the second step. To apply the equilibrium condition (5) we need the derivative of $Y(a_0, a, b)$ with respect to a_0. Considering that, in equilibrium $a_0 = a$, the specific form of the wage profile is now

$$w'[i] = \frac{\lambda}{\lambda + 1 - B} a'[i] b[i]. \tag{22}$$

Since pay levels grow at rate g, the present value of all CEO pay at firm i is $w[i]/(1-B)$; The rest of the present value of surplus must be capitalized into market value. Differentiating (15) with respect to i and combining it with (22) yields

$$\begin{aligned} v'[i] &= \frac{\partial}{\partial i} Y(a[i], b[i]) - \frac{w'[i]}{1-B} \\ &= \frac{a[i] b'[i]}{1-B} + \frac{a'[i] b[i]}{\lambda + 1 - B} \end{aligned} \tag{23}$$

The last term (obtained after some simplification) reflects the lingering impact of past CEOs. Since the CEOs get paid a flow income, the firm in effect "lends" the CEOs some of the present value of their impact at the discount rate r.

Finally, the pair of differential equations that describes the equilibrium is given by (22) and (23), and can be solved for

$$\frac{a[i]}{a[0]} = \exp\left\{ \frac{\lambda}{\lambda + 1 - B} \int_0^i \frac{w'[j]}{w[j] + v[j]\,(1-B)} \mathrm{d}j \right\}, \tag{24}$$

$$\frac{b[i]}{b[0]} = \exp\left\{ \int_0^i \frac{v'[j] - w'[j]/\lambda}{w[j] + v[j]\,(1-B)} \mathrm{d}j \right\}. \tag{25}$$

The indeterminacy of the constants $a[0]$ and $b[0]$ is due to the fact that we have no way of inferring anything about the relative contributions of ability and firm size to the surplus created at and below the smallest firm in the sample. Inference is only possible about the contributions of the "excess" of ability $a[i]/a[0]$ and firm size $b[i]/b[0]$ relative to the lowest types into the excess of surplus relative to the lowest pair.

The economic questions that can be answered based on the inferred distributions (24) and (25) will be built up of counterfactuals of the following type: What would be the impact on welfare (present value of surplus) by the counterfactual employment, for the duration of one period, of an individual of the type found at quantile I at a firm in quantile i? Utilizing the border condition $v[0] + w[0]/(1 - B) = a[0]b[0]/(1 - B)$ and (21), we get

$$Y(a[I], a[i], b[i]) - Y(a[i], b[i]) = \frac{\lambda}{\lambda + 1 - B}\left(v[0] + \frac{w[0]}{1 - B}\right)\frac{(a[I] - a[i])\, b[i]}{a[0]b[0]}. \tag{26}$$

These impacts can be calculated from the data because they include only the relative factor qualities: The undetermined constants wash out of the predicted economic effects of hypothetical rearrangements of individuals and firms.[16]

It would also be possible, in principle, to analyze the value of the exogenous component in firm size by considering changes in firm size while holding fixed abilities fixed. However, such results are very sensitive to the assumed model parameters, and will therefore only be mentioned in passing.

4 What is the Exogenous Component of Firm Size?

The fixed factor embedded in individuals has a very natural interpretation as ability, which obviously can not be removed from one person and grafted onto another. But what are the firm-specific characteristics that can not be chopped into pieces and shuffled between firms? For there to be exogenous (as opposed to CEO-induced) heterogeneity behind the cross-sectional variation in market values, there must be some fixed firm-specific qualities.

Different firms occupy different niches of the economy. The size of the niche includes all exogenous determinants of the scope of a firm's operations, everything that is inherent in technology and consumer preferences. Under this interpretation, b could be dubbed "the natural scale" of a firm. It seems clear that there is wide variation in the natural scale of firms: Even if all

[16] Note that, if the model is to be taken seriously, then observed rearrangment of CEOs between firms only reflects changes in information about their ability and can not help identify the value of ability; although inference from such movements can be sensible within other models (e.g., Hayes and Schaefer, 1999, and Parrino, 1997, whose findings support the presence of assortative matching.)

managers were exactly equal in ability, there would still be vast size differences between firms. For example, the manufacturing of wide-body aircraft is always going to be a bigger business than building yachts, and probably managed separately from it under most circumstances. The assignment model shows how such heterogeneity of firms interacts with heterogeneity of managerial ability to generate the joint distribution of profits and pay.[17]

An economy-wide increase in productivity or demand results in growth of exogenous firm size. A bull market could thus be interpreted as an across-the-board increase in the natural scale of firms. CEO pay levels should then be procyclical (and in apparent defiance of relative performance evaluation) because the marginal impact of management ability is then procyclical, as has been pointed out by Himmelberg and Hubbard (2000). As much as the decisions of the current period CEO interact with the future surplus potential of firms, then expectations of future scale of firms should affect the pay of the current CEO as well as the market value of the firm, although not necessarily current earnings.[18]

Role of firm heterogeneity If firms were inherently identical (i.e., if $b[i] = \bar{b}$ constant), then the matching between firms and CEOs would be irrelevant and the only reason for companies to differ by market value would be the scale-of-operations effect. As ability is complementary with capital it is still efficient to allocate more capital to the management of more able CEOs, but there is no "assignment" anymore. This is how firm size differences are determined in the classic models of Lucas (1978) and Rosen (1982). (Similar to these models, the assignment model also requires decreasing returns to combining several firms into the management of a single individual, as otherwise all firms should be merged to be under the command of the

[17] According to our story, the results in Rose and Shepard (1997) could be interpreted as there being another dimension of firm heterogeneity with which CEO ability is complementary, namely the natural level of diversification of the firm's operations. Similarly, from Palia (2000), lack of regulation appears to be complementary with CEO ability.

Montgomery and Wernerfelt (1988) discuss a number of further reasons for the existence of fixed firm-specific factors.

[18] Bebchuk and Grinstein (2005) use regression analysis to find that one fifth of the recent increases in executive pay can be attributed to the increase in firm size as measured by current year sales. We are tempted to speculate that a larger fraction could be attributed to increase in expected future sales.

most able CEO.) There the size distribution of firms (by earnings) is solely a reflection of the economy's solution to allocating productive resources to different managers: if all managers were equally apt, then all firms would have the same market value. By contrast, in an assignment model, both sides of the match are inherently heterogeneous. This two-sided heterogeneity implies that both sides must earn rents relative to their outside opportunity (except for the lowest types that match). There is a contemporaneous rent to being a firm in a large niche—partly as the simple Ricardian rent due to occupying an inframarginal niche, and partly because being highly ranked allows a firm to be matched with a more able CEO. Some of these rents merely compensate for the capital that has been sunk in the past and now show up as part of the firm's natural scale b.[19]

5 Empirical Application

5.1 Data

The sample comprises the 1000 publicly traded US companies with the largest market value in each year from 1994 to 2004 in the CompuStat database.[20] The variable for CEO pay is total compensation, with options valued using the standard Black-Scholes formula. We measure CEO pay as a firm-specific variable: For firms that had several CEOs during the same year, their compensation is summed up to measure that year's CEO pay. We do not consider the incentive structure of CEO pay; what matters for our model is the expected cost of compensation.

In practice, the magnitude of the potential impact of CEO ability in any given firm depends on many factors, and, even in the absence of stochastic factors, could not be expected to have a perfect rank correlation with market value. However, to calibrate our assignment model, the input data of CEO pay and market value needs to exhibit perfect rank correlation, so the

[19] By no means does the existence of rents to firms imply that their owners earn excess returns—these rents should have been capitalized into market value all along. The rent from occupying a lucrative niche should have been dissipated back when it was decided who got to occupy that niche (perhaps in a patent race, or through premature entry, or as a rent to talented or lucky founders).

[20] While the data on CEO pay is available from 1992, we drop the first two years because they seem to miss many of the smaller companies that should have been in the top 1000.

observed noisy relation of CEO pay and market value needs to be smoothed into a strictly monotonic relation. For this purpose we performed a Lowess smoothing of the relation of the levels of CEO pay and the rank by market value (separately for each year).[21] The relationship of CEO compensation and firm ranking by market value is shown in Figure 2 for 2004.[22] It bears out the well known fact that larger companies pay their CEOs more. The cross-sectional Spearman rank correlation of market value and CEO pay varies between 0.47 and 0.57 during the sample period; this can be considered a "measure of fit" for the key assumption of assortative matching.

In principle, the smoothing could be done the other way around: by fitting the market values against rank by CEO pay. We use the rank by market value as "the true ordering," because as a stock variable it is likely to be a better measure of a firm's value then a year's flow of CEO pay is a measure of CEO compensation at that firm. CEO pay is more volatile partly because compensation granted for a given year does not in practice necessarily compensate for the services in that year alone, due to deferred pay and bonuses.[23]

From now on, except where explicitly stated, observed or actual distribution of CEO pay refers to the pay levels that have been smoothed in their relation to market value. Since rank by market value is used to order the observations, there is no need to do any smoothing of market values themselves: A simple connect-the-dots interpolation suffices to create continuous distributions.

5.2 Cross-Section Results

Role of CEO Ability for Economic Welfare The value of existing ability can only reasonably be measured relative to some hypothetical level of ability that would replace it. In this section we measure the aggregate value of CEO ability in the sample by considering three replacement types

[21] Stata implementation was used with bandwidth 0.7 (the smallest to yield strictly increasing fit in every year) and adjustment of means to match the the sample means. For details on Lowess (LOcally WEighted Scatterplot Smoothing), see Cleveland (1979).

[22] We define rank such that largest firm has rank 1, so the relation of rank $n \in \{1, \dots, 1000\}$ and quantile $i \in [0, 1]$ is $n = 1000 - 999i$.

[23] The higher stability of market value is well captured by the fact that the correlation of current and lagged rank by market value is 0.95, while the correlation of current and lagged rank by the CEO pay by firm is only 0.73.

for the existing distribution of ability, while the existing distribution of firm size is kept in place. In each counterfactual, the replacement CEOs at every firm would be of the same type—the baseline type (from the 1000th largest firm), the highest type (from the largest firm), and the median sample type (from firm 500).[24]

Table 1 tabulates the results of these counterfactuals. The results are calculated under four different combinations of assumed parameter values.[25] In the first and the most important case we assume the replacement CEOs to be of the baseline type, i.e., of the type managing the smallest sample firm. Using data from 2004, the value of existing ability over the baseline level is between \$21.3 and \$25 billion, depending on the assumed model parameters. For comparison, the actual total pay of the top 1000 CEOs was \$7.1 billion. The baseline CEO earned in fact \$2.7 million per year: This would be the equilibrium pay of all top 1000 CEOs if they were all of that same type, so the counterfactual total pay of 1000 baseline types is \$2.7 billion.[26] The difference, \$4.4 billion, is the rent to the scarce ability of the actual top 1000 CEOs, namely the part of their ability that exceeds that of the baseline type. This amounts to between 17% and 21% of the economic value of ability. The rest of the value of ability is left to shareholders, so the implied effect that the CEOs had on shareholder value in 2004 was about \$16.9–20.6 billion, which was 0.13–0.16% of the total market capitalization of the largest 1000 firms. Recall that the value of ability is measured as a flow here: It is the contribution of the current-year CEO to the present value of economic surplus. The implied "stock" value of current and future CEO ability is obtained by dividing the above values by the appropriate discount rate.

The assessed value of ability relative to the baseline types is depicted in Figure 3 for all sample years, using the two parameter assumptions with

[24] The exact median rank among the top 1000 firms is 500.5; for the sake of brevity we admit a small "rounding error" in the discussion.

[25] The parameters are used in transforming the market values, via equation (20), into the flow income for firm's fixed factors. The parameters were chosen from the set $\lambda \geq 0.1$, $0 \leq \theta \leq 0.8, 0.05 \leq r \leq 0.1$, $g \geq 0.02$, with restriction $0.06 \geq r - g \geq 0.025$, giving implied "P/E" ratios between 17 and 44. First and last column (A and B) pick out the combinations that result in the most extreme dollar values of ability within this set of parameter values.

[26] In equation (6), there would be $a'[i] = 0$ at all i.

the most extreme results. The millennium boom/ bubble stands out in the figure. This could be expected: in our assignment model, thanks to the complementarity between managerial ability and firm size, increases in expected productivity increase the value of current CEO ability, even if the productivity is only expected to affect future earnings potential. Whether the expectations of growth in productivity are "irrational" or not is a different question: either way, in competitive equilibrium, more optimistic expectations induce the owners to immediately bid more for the CEOs. This is because current CEO ability interacts with future scale to generate future profits.

The second result in Table 1 shows the optimistic counterfactual: The increase in total surplus, if all top 1000 CEOs were to be replaced with CEOs of the current highest type. In this case the gains are much more modest than the losses in the previous grim scenario: This value is about \$3.2–3.4 billion.[27] The gains from improved ability are relatively small because CEO ability is increased the least at the largest firms where ability is at its most productive, as these managers are already close to the highest type. (By contrast, in the previous counterfactual, the largest decreases in ability took place at the largest firms, so the total effect was larger.) In this case there would be a jump in the ability profile at the baseline, and the resulting jump in surplus would accrue to CEOs so that every CEO in the top 1000 would get paid \$2.9–3.4 million more than the best CEO outside the sample. (This assumes that the distributions are continuous below the baseline, so that the best outside type is a baseline type.) Above the baseline managers are homogeneous and all additional surplus goes to firms. This improvement in overall ability would melt away most of the pay advantage of the CEOs at the very top, as they would now face tough competition in the matching market. (This would mean a decrease of about \$15.3 million for the highest pay level.) In total, the top 1000 CEOs would earn \$1–1.5 billion less under this (socially) optimistic scenario, leaving the shareholders a total windfall gain of \$4.4–4.7 billion. In the third reported scenario the median (500th) type is used as the replacement; This would decrease total surplus by \$6.6–

[27] This gain amounts to $0.025 - 0.027\%$ of total market capitalization. Using the same data but restricting the shape of the talent distribution, Gabaix and Landier (2007) find an analogous gain of 0.016% at the 250th largest firm. (Note that their discussion refers to Terviö 2003, which used data from 1999).

7.1 billion.[28]

Role of Firm Size for CEO Pay The difference in the pay of two CEOs at firms of different size reflects both the difference in their abilities and the difference in the size of the firms they manage. We can measure the economic value of differences in ability that arise solely from the differences in ability by hypothetically matching the CEOs with a firm of the same size. Due to the complementarity of ability and firm size, this dollar value is larger the larger the firm that is used as the reference point. To illustrate the role of firm size behind the levels of CEO pay, we pick three reference ranks and calculate the difference that other CEOs would have made to surplus compared to the actual CEO at a firm of the reference rank. For comparison, these plots also show the difference in actual CEO pay compared to the CEO managing the reference firm. The counterfactual difference that CEOs would have made to surplus is, by construction of the equilibrium, everywhere lower than the actual pay difference. (If it were not, then the reference firm would be better off hiring someone other than its own match even after paying her more than her equilibrium pay.)

The calculations in Figure 4 are based on a counterfactual where all firms become identical with the reference firm $b[I]$, while the distribution of ability remains unchanged. Counterfactual wage differences can be computed with equation (26) by replacing $b[i]$ with a constant $b[I]$. Now that firms are homogeneous, all of the difference that CEOs make to economic surplus accrues as simple Ricardian talent rents to CEOs, i.e., the difference to surplus made by a CEO would also be the difference in her equilibrium wage compared to the reference CEO. For example, in the middle panel of Figure 4, the CEO at the median firm in fact earned about \$3.2 million more than the CEO at the baseline, but \$12.1 million less than the CEO at the top. In the counterfactual, these differences would be about \$5.5–6.1 and \$3.5–3.8 million respectively. Compared to "status quo", the homogenization of firms to the median size increases the pay of the median CEO because firms

[28] It is also possible to calculate counterfactuals where the ability distribution is held fixed while firms are replaced, but the implied welfare effects are very sensitive to the assumed parameter values. For example, replacing all firms with the same type as the actual 1000th largest firm gives the mirror image of the "value of ability over baseline." This difference made by fixed firm specific factors—including sunk capital—is inferred to be between \$626 and \$91 billion.

below her in the ranking become larger and thus contribute to the (properly defined) marginal product of her ability. By contrast, the top ranked CEO makes about \$9–10 million less in the counterfactual, because the negative effect of the decrease in the size of the top half of firms dominates the positive effect coming from the increase in size below the median. (The comparative statics with respect to changes in a factor quality distribution were discussed in Section 2.2).

The top panel of Figure 4 shows the value of ability differences with the baseline firm as the reference, and the bottom panel with the largest firm as the reference. If all top 1000 firms were as small as the 1000th, then the CEO of the largest firm would be paid only about \$2.8–3.4 million more than the baseline CEO, a roughly five-fold reduction. If, instead, the firms became as large as the actual largest firm, then this pay difference would be about \$720–860 million. Note that in this case the only thing to change at the top firm is the equilibrium division of rents: The CEO gains from the increase in demand for the ability of all the lower types. The related aggregate values of CEO rent (pay minus pay of baseline type) are shown in Table 2. These results highlight the overwhelming role of firm scale in driving the differences in CEO pay. The pessimistic change cuts total rents to ability by more than half, to about \$1.7–1.9 billion, whereas the increase in firm size increases the rents by roughly hundred-fold. The results are fairly robust to the assumed parameter values, except as to their effect on counterfactual market values, which are not the focus here.

We saw before that replacing CEOs by other sample types makes a relatively small difference to total economic surplus. By contrast, a hypothetical variation in the exogenous component of firm size generates huge differences in pay (and, of course, in market value). In this sense, the high levels of CEO pay at the top can be said to be mostly due to the exogenous component in firm scale.

5.3 Time Series Results

In this section we investigate whether a simple assignment model with a time-invariant distribution of ability and firm size is sufficient to generate the recent fluctuations in the distribution of CEO pay.[29] For this purpose we now calibrate the model over the sample years 1994–2004 under the restriction that distributions of both fixed factors are constant over time.

More specifically, we now allow a multiplier in the surplus function to vary freely over time. Thus the (expected present value of) surplus in year t generated at a firm of type b that hires CEOs of type a is assumed to be

$$Y_t(a, b) = G_t ab, \tag{27}$$

where G_t is an economy-wide parameter capturing expected productivity that will also absorb the multiplicative constants in the inferred distributions of a and b. Changes in G are assumed to come as a surprise and expected to stay permanent. In terms of the model, a change in G is observationally equivalent to a proportional across-the-board change in firm size. Our favored interpretation is that changes in G capture changes in expected potential for surplus, which depends on expectations about productivity and demand. Reductions in G reflect lowered expectations.

Denote the abilities and firm sizes inferred from the data in year t by a_t and b_t, where distributions for each year are calculated using (24) and (25). We now take the average of the distributions over the sample period ($T = 11$):

$$\hat{a}[i] = \frac{1}{T}\sum_t \frac{a_t[i]}{a_t[0]} \quad \text{and} \quad \frac{\hat{b}[i]}{\hat{b}[0]} = \frac{1}{T}\sum_t \frac{b_t[i]}{b_t[0]}. \tag{28}$$

Figure 5 shows inferred relative abilities $a_t[i]/a_t[0]$ from selected quantiles over the sample years, as well as the corresponding averages. Based on a year-by-year inference, the implied relative abilities would be higher during the millennial boom before returning to the previous levels.

The time-varying scaling factor G_t is chosen so that predicted total surplus in each year will exactly fit the data, i.e., to satisfy

$$G_t \int_0^1 \hat{a}[i]\hat{b}[i]\mathrm{d}i = \int_0^1 \left(\frac{w_t[i] - w_t[0]}{1 - B} + v_t[i] - v_t[0]\right)\mathrm{d}i. \tag{29}$$

As market values dwarf CEO pay, the relative changes in the "year effects" G_t are mostly driven by changes in total market value.

[29] This is not to say that distribution of ability should stay constant in the long run: changes in the process of selection into CEO and in managerial learning-by-doing can result in changes in the distribution CEO ability, even if the underlying population distribution of talent is unchanged.

After obtaining the time invariant distributions $\hat{a}[i]$ and $\hat{b}[i]$, and the time-variant factors G_t, these are then plugged into the equilibrium conditions of the model, (6), (8), and (27), to generate predictions or "fitted values" for the factor income distributions in each year.

$$\hat{w}_t[i] = w_t[0] + G_t \frac{\lambda}{\lambda + 1 - B} \int_0^i \hat{a}'[j]\hat{b}[j]\mathrm{d}j \tag{30}$$

$$\hat{v}_t[i] = v_t[0] + G_t \int_0^i \left(\frac{\hat{a}[j]\hat{b}'[j]}{1 - B} + \frac{\hat{a}'[j]\hat{b}[j]}{\lambda + 1 - B} \right) \mathrm{d}j \tag{31}$$

The predicted market values are obtained by using the (inverted) equation (20) to add the implied value of adjustable capital to $\hat{v}_t$.

Figure 6 depicts the model's predicted wages over time at selected ranks.[30] The baseline pay levels are taken directly from the data, as the model cannot predict them, only the additional surplus relative to them. Therefore the predictions will necessarily fit exactly at the lowest ranked firm. (Notice that the graphs show the difference in pay over the baseline levels, so this "automatic" part of the fit does not contribute to the fit of the model.) By varying the scaling parameter, the model provides a reasonable fit to the fluctuations in the distribution of CEO pay.

While for the most part the variation of a single parameter is able to generate the changes in the distribution of observed CEO pay, the boom years 2000–2001 represent a particular problem for the model. The unusually high pay levels of the top CEOs in these years cannot be explained within the multiplicative assignment model by the variation of a single productivity parameter. At the same time, the predicted market values, shown in Figure 7, are too high at the smaller firms. This is because the very largest firms were unusually highly valued relative to the rest, while the model forces the same shape for the distribution of market values for all years. These discrepancies could of course mean that something unusual that is completely missed by the model was going on (of which anecdotal and, more recently, even court evidence is plentiful). Nevertheless the relation of CEO pay and market values has apparently since gotten "back to normal," inasmuch as an 11-year sample allows us to call a few years of it abnormal.[31] The higher pay

[30] The choice between parameter scenarios makes only a negligible difference for the predicted pay levels because these calculations do not involve any counterfactual matches.

[31] One theoretical possibility is that the shape of the distribution $b_t[i]$ was unusual in the

levels and market values in the last sample year (2004) relative to the first (1994) can be explained by a size-neutral increase in expected productivity.

As the impact of managerial ability on future profits depends on productivity, the economic value of ability has, by construction, also fluctuated with market values over the years. The aforementioned value of ability relative to the baseline has varied during the sample period between \$10–12 billion (1994) and \$31–37 billion (2000), as seen in Figure 3, with about a fifth of the value always accruing to the CEOs. The values calculated from cross sections are also depicted, and are naturally more volatile as a and b are allowed to vary over time.

Our model does not address the question of why much of the fluctuation in CEO pay has taken the form of stock options, but the results are consistent with the explanation of Oyer (2004). Since the equilibrium levels of CEO pay increase and decrease together with market values in the matching market, a simple way to automatically link the CEOs pay to the fluctuating competitive level would be to tie it to the market values. This way there would be no need to rewrite the contract every time the market values move up or down. The only purpose of such "incentive pay" would be retainment.

6 Conclusion

There is a time-honored tradition in economics to assume that prices are competitive and reflect all available information, at least as the first approach in analyzing price data. This paper has shown how assignment models can be used in this spirit to analyze income data from markets with assortative matching, and developed a theory-driven estimate for the economic impact of CEO ability. The role of ability was studied within the assignment model by evaluating the predicted effects of counterfactual distributions of ability and firm size on CEO pay and shareholder value.

This paper is merely the first attempt in applying an assignment model into the market for CEOs, and leaves in (and hopefully lays bare) the admittedly many strong assumptions utilized to take the model to the data. Firm size and CEO ability were assumed to interact in a multiplicatively separable fashion, while uncertainty, frictions, incentive problems, comple-

anomalous years, but in the absence of any theoretical justification for such an anomaly, we view this "explanation" as too flexible to be useful.

mentarities between different CEOs across time, and many other features were assumed away, to allow the model to stay in a reasonably tractable form. To resolve the problem that the observed market values depend on the expected course of the whole CEO market in the future, the world was assumed to be stationary in a very strong sense: Total factor productivity is assumed to grow at the same deterministic rate at every firm, and each firm to match with a CEO of current ability forever. Clearly the empirical results must be taken with a grain of salt: They are simply based on the assumption that the model is true.

We have shown that plausible (i.e., very small) differences in ability among CEOs and a tractable assignment model of CEOs and firms can together generate the high levels of observed CEO pay as part of a perfectly competitive equilibrium. There are certainly other sensible explanations for the size-pay relationship and the current setup does not allow us to test our model against them. By no means do we claim to have shown that CEOs are not "overpaid" as our assignment model does not include that possibility. (For example, a market-level version of the skimming view could be that there is a "stealing technology" that allows CEOs to steal more at bigger firms—hence the robust size-pay relation. Presumably this technology would come with some decreasing returns to scale, since the ratio of CEO pay to market value is decreasing in the latter).[32] We believe that the matching of CEOs with exogenously heterogeneous firms has a genuinely important role in driving the competitive levels of CEO pay, but not that competition is the sole force behind CEO pay. While much work remains to be done, we hope to have shown that assignment models have much to offer in helping to understand the determination of CEO pay levels.

References

Ang, James and Gregory L Nagel (2006): "The Influence of Peers on CEO Pay Since 1970." *Working paper.*

Bertrand, Marianne and Sendhil Mullainathan (2001): "Are CEOs Rewarded for Luck? The Ones Without Principles Are." *Quarterly Journal*

[32] Such "scalable" stealing would be consistent with the options backdating found in Yermack (1997).

of Economics, 116, pp. 901–932.

Baker, George P and Brian J Hall (2004): "CEO Incentives and Firm Size." *Journal of Labor Economics*, 22 , pp. 767–798 .

Bebchuk, Lucian and Yaniv Grinstein (2005): "The Growth of Executive Pay." *Oxford Review of Economic Policy*, 21, pp. 283–303.

Cleveland, William S (1979): "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association*, 74, pp. 829–836.

Cuñat, Vicente and Maria Guadalupe (2006): "Globalization and the Provision of Incentives inside the Firm: The Effect of Foreign Competition." *Columbia University working paper.*

Frydman, Carola (2005): "Rising Through the Ranks. The Evolution of the Market for Corporate Executives, 1936-2003." *MIT Sloan working paper.*

Gabaix, Xavier and Augustin Landier (2007): "Why has CEO Pay Increased do Much." *Forthcoming in the Quarterly Journal of Economics.*

Gayle, George-Levi and Robert A Miller (2005): "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *Carnegie Mellon University working paper.*

Hayes, Rachel M and Scott Schaefer (1999): "How Much Are Differences in Managerial Ability Worth?" *Journal of Accounting and Economics*, pp. 125–148.

Himmelberg, Charles P and R Glenn Hubbard (2000): "Incentive Pay and the Market for CEOs: An Analysis of Pay-for-Performance Sensitivity." *Columbia University working paper.*

Jensen, Michael C (1986): "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review Papers and Proceedings,* 76, pp. 323–329.

Koopmans, Tjalling C and Martin Beckmann (1957): "Assignment Problem and the Location of Economic Activities." *Econometrica*, 25, pp. 53–76.

Kostiuk, Peter F (1990): "Firm Size and Executive Compensation." *Journal of Human Resources*, 25, pp. 91–105.

Lucas, Robert E Jr (1978): "On the Size Distribution of Business Firms." *Bell Journal of Economics*, 9, pp. 508–523.

Manne, Henry G (1965): "Mergers and the Market for Corporate Control." *Journal of Political Economy*, 73, pp. 110–120.

Mayer, Thomas (1960): "Distribution of Ability and Earnings." *Review of Economics and Statistics*, pp. 189–195.

Montgomery, Cynthia A and Birger Wernerfelt (1988): "Diversification, Ricardian Rents, and Tobin's q." *RAND Journal of Economics*, 19, pp. 623–632.

Murphy, Kevin J (1999): "Executive Compensation." in *Handbook of Labor Economics Volume 3*, edited by Orley Ashenfelter and David Card, pp. 2485–2563.

Murphy, Kevin J and Zábojník, Ján (2004): "CEO Pay and Appointments: A Market-Based Explanation for Recent Trends." *American Economic Review Papers and Proceedings*, 94, pp. 192–196.

Ostroy, Joseph M (1980): "The No-Surplus Condition as a Characterization of Perfectly Competitive Equilibrium." *Journal of Economic Theory*, 22, pp. 183–207.

— (1984): "A Reformulation of the Marginal Productivity Theory of Distribution." *Econometrica*, 52, pp. 599–630.

Oyer, Paul (2004): "Why Do Firms Use Incentives That Have No Incentive Effects?" *Journal of Finance*, 59, pp. 1619–1649.

Palia, Darius (2000): "The Impact of Regulation on CEO Labor Markets." *RAND Journal of Economics*, 31, pp. 165–179.

Parrino, Robert (1997): "CEO Turnover and Outside Succession: a Cross-Sectional Analysis." *Journal of Financial Economics*, 46, pp. 165–197.

Rose, Nancy and Andrea Shepard (1997): "Diversification and Executive Compensation: Ability Premia or Managerial Entrenchment?" *RAND Journal of Economics*, 28, pp. 489–514.

Rosen, Sherwin (1982): "Authority, Control and the Distribution of Earnings." *Bell Journal of Economics*, 13, pp. 311–323.

Sattinger, Michael (1979): "Differential Rents and the Distribution of Earnings." *Oxford Economic Papers*, 31, pp. 60–71.

— (1993): "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature*, 31, pp. 831–880.

Teulings, Coen N (1995): "The Wage Distribution in a Model of the Assignment of Skills to Jobs." *Journal of Political Economy*, 102, pp. 280–315.

Terviö, Marko (2003): "The Difference that CEOs Make: An Assignment Model Approach." Chapter 2 in *Studies of Talent Markets*, Ph.D. Thesis, MIT.

Tinbergen, Jan (1956): "On the Theory of Income Distribution." *Weltwirtschaftliches Archiv*, 77, pp. 155–175.

— (1957): "Some Remarks on the Distribution of Labour Incomes." in *International Economic Papers, no. 1. Translations prepared for the International Economic Association*, edited by Peacock, Alan T et al., Macmillan, pp. 195–207.

Yermack, David (1997): "Good Timing: CEO Stock Option Awards and Company News Announcements." *Journal of Finance*, 52, pp. 449–476.

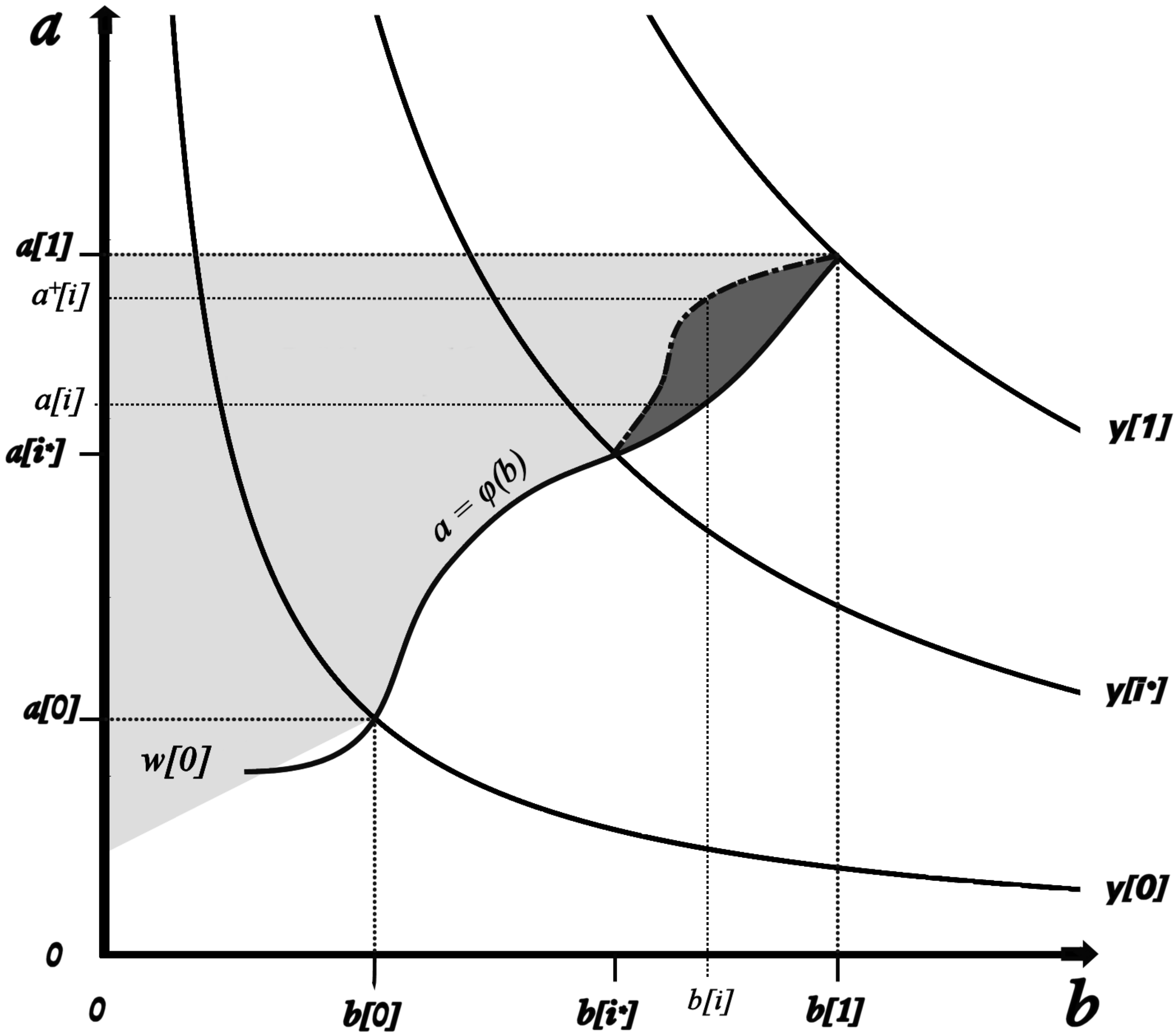

Figure 1. Comparative statics in the multiplicative case.
The increasing graph covers both the active matches {b, φ(b)} above b[0] and potential but inactive matches below b[0]. The three decreasing graphs are the isoquants for levels of output y[0], y[i*], and y[1]. The entire shaded region is the equilibrium wage of the highest ability type, a[1]. The dark shaded region is the decrease in wage and increase in profits for the highest types if the matching graph between quantiles i* and 1 were to shift up to the dashed line.

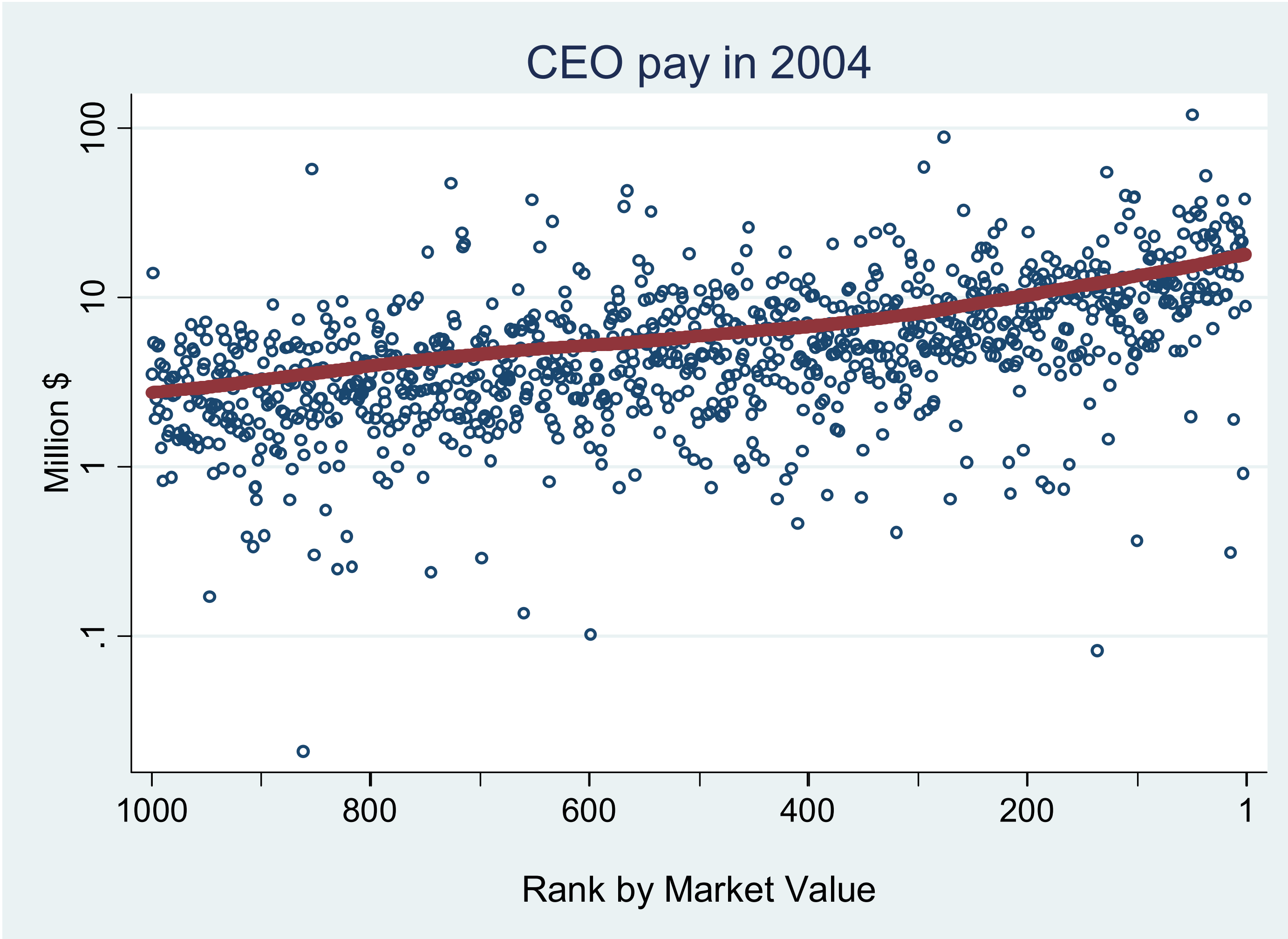

Figure 2. Relation of CEO pay and firm rank by market value in 2004. The smoothed relation (obtained with the Lowess method) appears upwards biased in the graph because the pay levels are depicted on log scale.

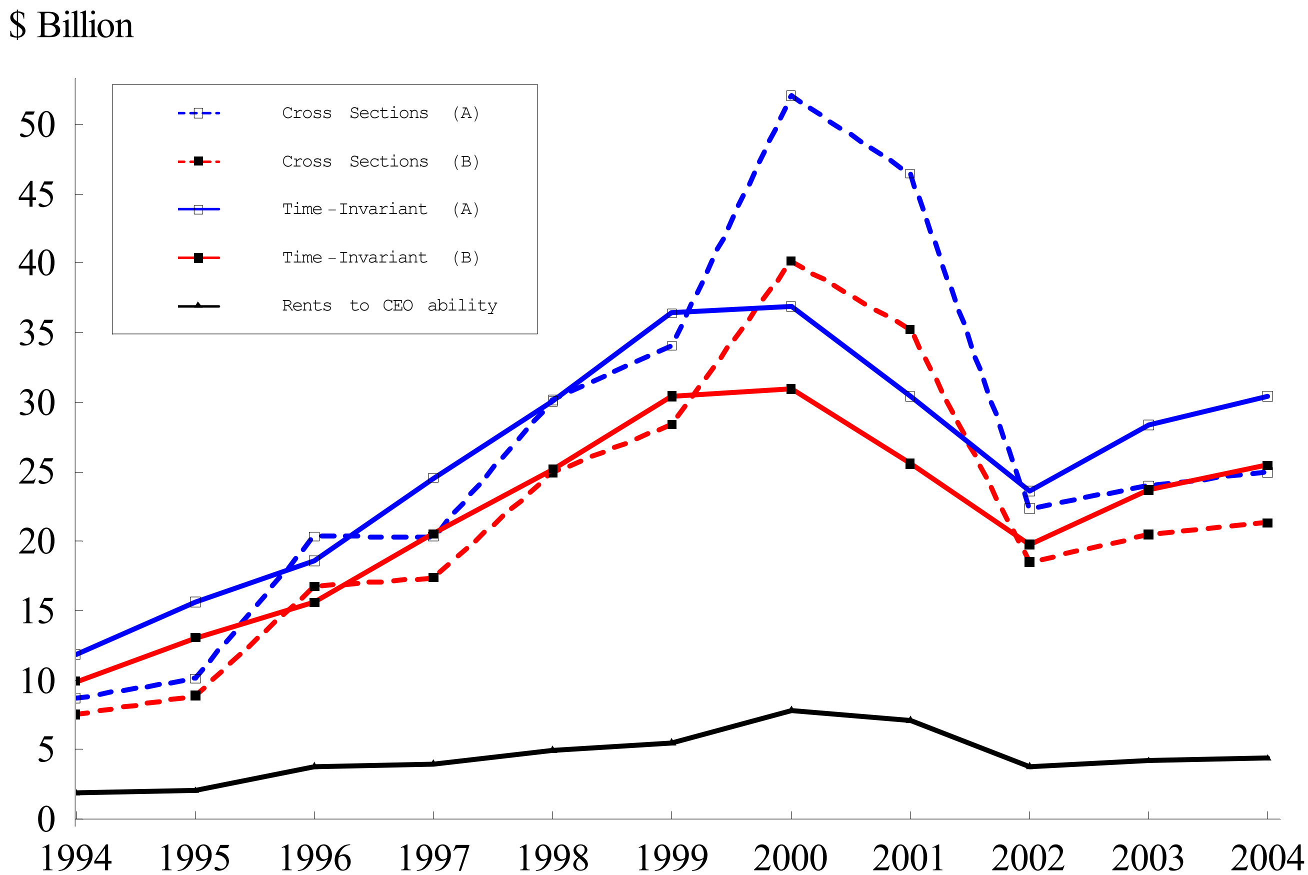

Figure 3. Value of CEO ability and rents to CEOs relative to baseline ability, at the 1000 largest firms. In "Cross Sections" the profiles of ability are inferred separately in each sample year, whereas in the "Time-invariant" specification the distribution of ability is forced to be constant over time. The parameter assumptions with the most extreme results are shown for both specifications (see Table 1 for details). All values are in 2004 dollars.

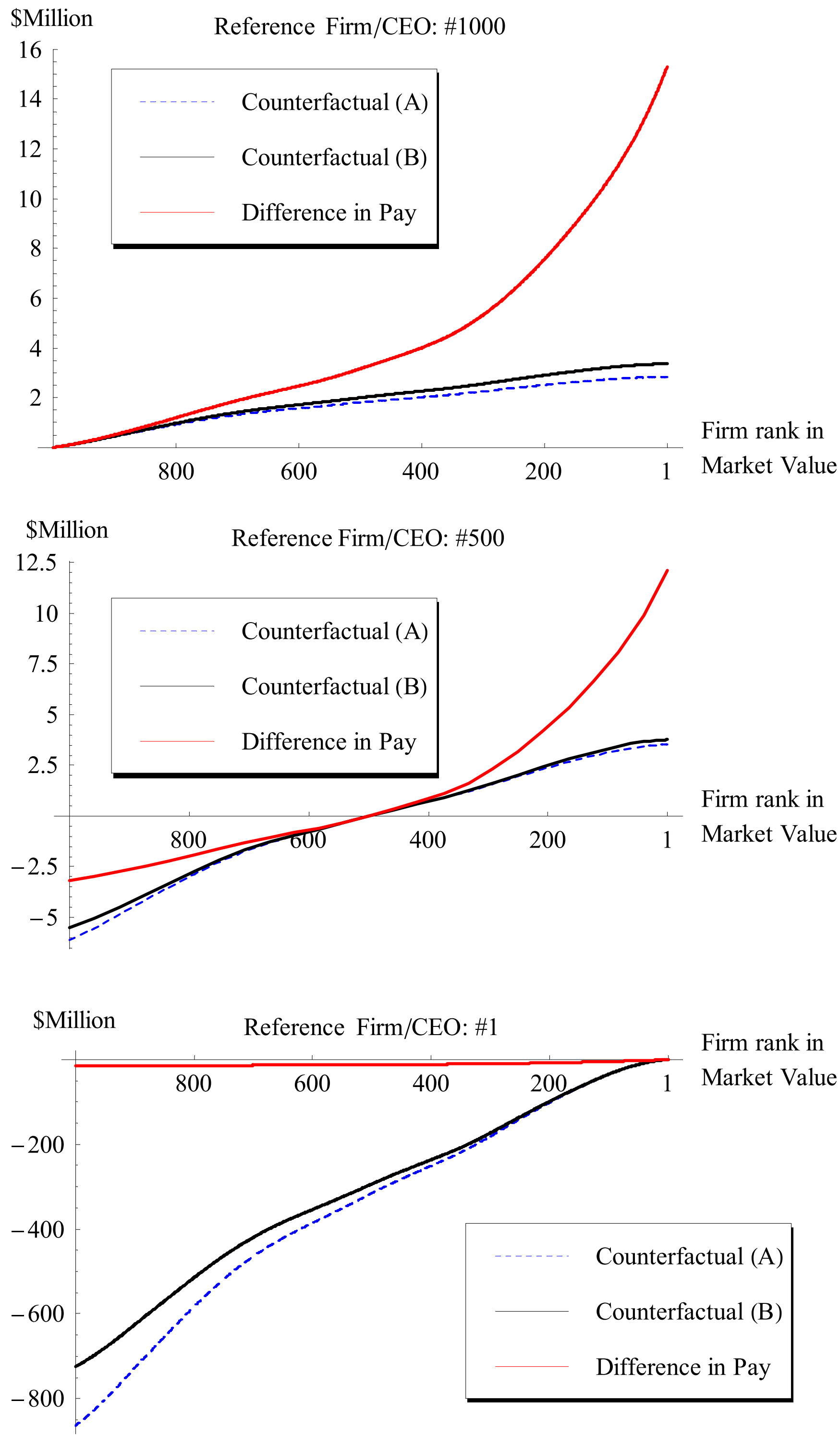

Figure 4. Impact of CEO ability by reference firm.[1]

[1] The counterfactual difference that CEOs would make to economic surplus created at the reference firm if they were to replace the actual CEO at the reference firm. The value is calculated under two assumptions of the model parameters (A and B, defined in Table 1). The red line depicts the difference in actual pay of the CEOs relative to that of the reference rank.

Figure 5. Inferred CEO abilities at 1st, 250th, 500th, and 750th largest firm (relative to 1000th) by year. Dashed lines give the average over this time period, used as the time-invariant distribution of ability in Section 5.3.[2]

[2] This figure is based on calibrations with parameters from column A in Table 1; the picture is very similar with other parameter scenarios, except for the scale of *a[i]/a[0]*, which get larger the higher the assumed share of adjustable capital θ.

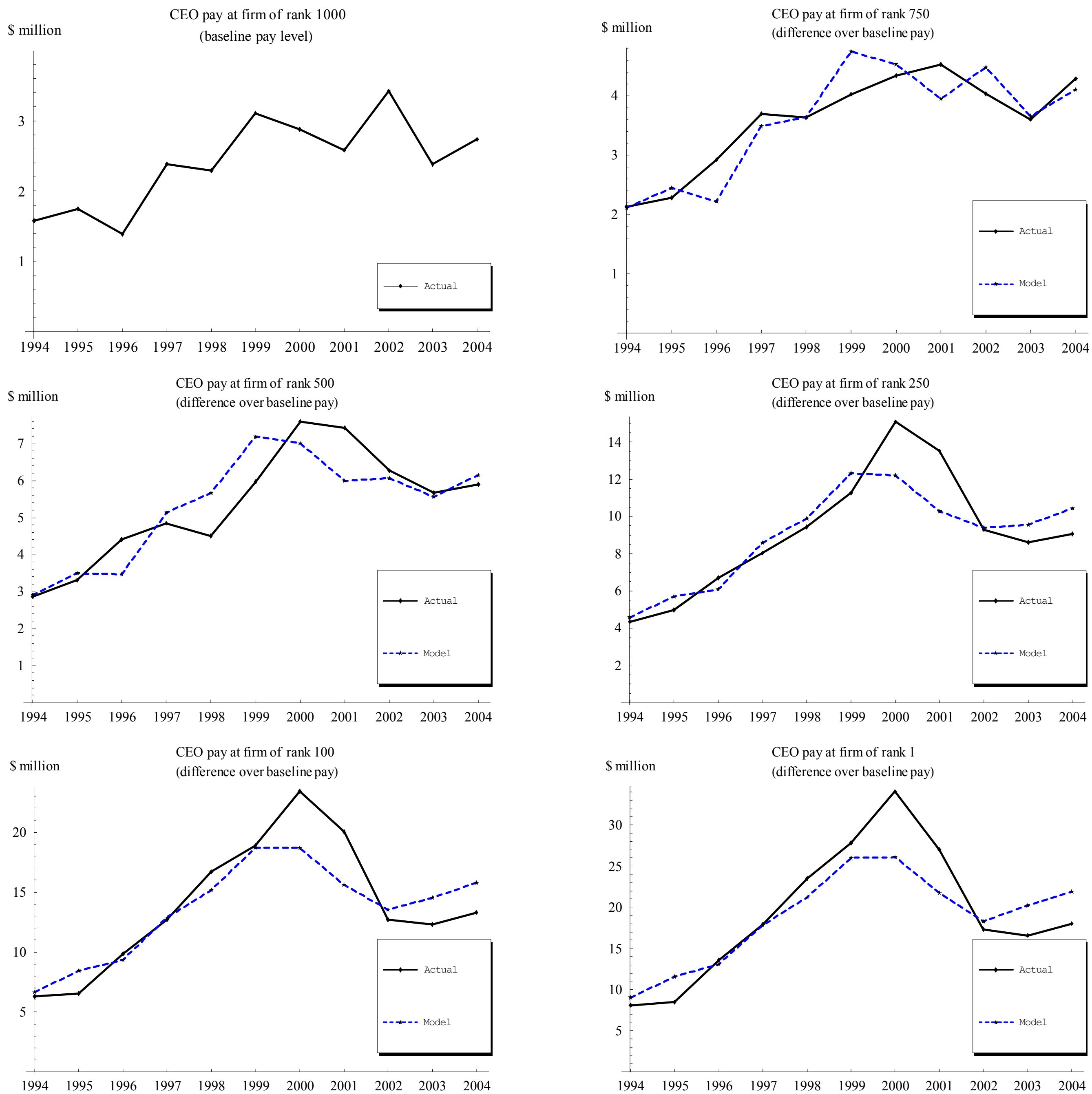

Figure 6. The difference in pay between the CEOs of selected ranks and the baseline (1000th) CEO. "Actual" pay refers to the smoothed CEO pay and "Modeled" pay refers to the CEO pay level generated by the assignment model while imposing time-invariant distributions of ability and firm size, and size-neutral productivity growth.[3] All values are in 2004 dollars.

[3] The parameters used in these calibrations are same as in column A, Table 1. The effect of assumed parameters on these figures is negligible. Parameter assumptions (within the assumed set) make only negligible (below 1%) difference to the fitted wages so only one fit is shown.

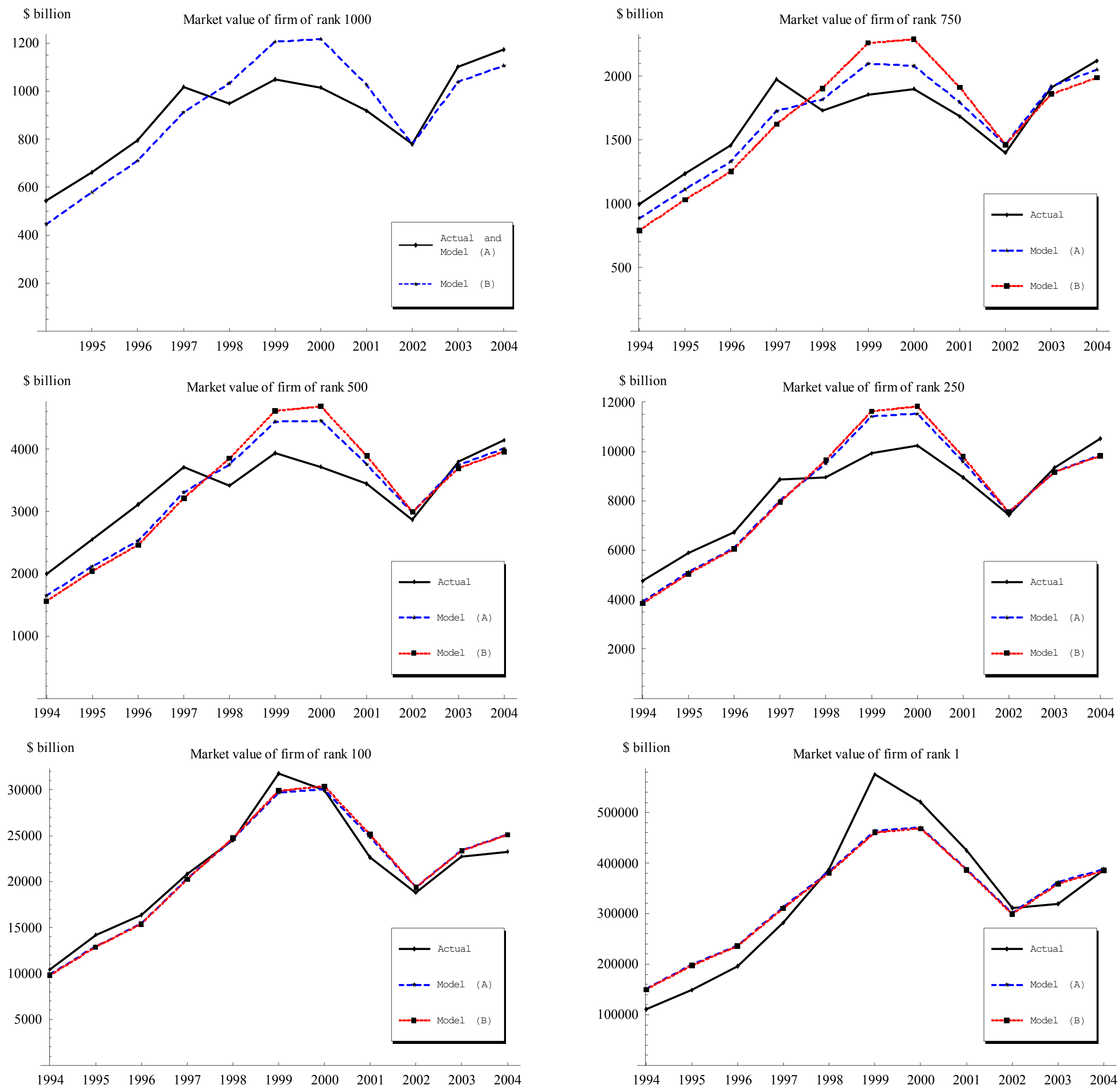

Figure 7. Predicted market values in the time-invariant calibration.[4] All values are in 2004 dollars.

[4] The parameters used in these calibrations are same as in columns A and B, Table 1. The market values are calculated under the two assumptions of the model parameters that generate the most extreme results (within the assumed set).

Table 1. Value of existing CEO ability at 1000 largest firms (2004).
Values in $Billions.

Assumptions	(A)	(a2)	(c)	(b2)	(B)		
Discount rate (r)	0.08	0.08	0.05	0.05	0.05		
Growth rate (g)	0.02	0.02	0.025	0.025	0.025		
Share of Adj. Capital (θ)	0	0.4	0.4	0.4	0.8		
Rate of impact fading (λ)	∞	∞	0.5	0.1	0.1		
						Counterfactual rent to CEOs	
Results	(A)	(a2)	(c)	(b2)	(B)	(A)	(B)
Value over Baseline	24.97	24.75	23.36	23.12	21.33	0	0
Value below Max	3.16	3.17	3.25	3.27	3.40	2.85	3.36
Value over Median	7.12	7.09	6.89	6.86	6.62	1.81	2.00

Actual totals	
CEO pay	7.12
CEO rent*	4.39
Market value	12,584.6

* Rent = Pay - Baseline pay ($2.7m)

Table 2. Total CEO rent under counterfactual firm size (2004).*
Pay over baseline pay; all values in $Billions.

						Counterfactual Market Value	
Reference firm	(A)	(a2)	(c)	(b2)	(B)	(A)	(B)
1000th largest	1.70	1.71	1.78	1.79	1.92	1,221.9	6,673.6
500th	5.77	5.74	5.59	5.55	5.30	4,140.5	8,945.4
1st	516.4	510.3	469.3	462.5	413.3	370,663.3	108,255.2

Actual totals	
CEO rent	4.39
Baseline pay × 1000	2.73
Market value	12,584.6

* See Table 1 for the details of parameters assumptions (A)-(B).