Total Executive Compensation *

Gian Luca Clementi† Thomas Cooley‡ Sonia Di Giannatale §

First Version: 2/15/2008

Very Preliminary and Incomplete – Comments Welcome
Do Not Circulate or Quote Without the Authors’ Consent

Abstract

Moral hazard due to non–observability of managerial effort and the degree of technological complementarity between managerial inputs and firm’s assets may play an important role in shaping executive compensation. In this paper we model both factors in a dynamic principal–agent model with capital accumulation and we characterize their implications the constrained Pareto–optimal contract. We focus our attention on testable implications: (i) the relationship between compensation and firm size, (ii) the relative importance of current and deferred compensation, (iii) the sensitivity of compensation to innovations in shareholder wealth, and (iv) the relationship between such sensitivity and size. Very preliminary results show that when the marginal product of managerial effort is increasing in capital, our model is consistent with facts (i), (iii), and (iv).

Key words. CEO, Repeated Moral Hazard, Hidden Action, Pay–Performance Sensitivity, Capital Accumulation.

JEL Codes: D82, D86, D92, G30.

---

*We are very grateful to Alberto Bisin, Heski Bar–Isaac, Andre De Souza, and Laura Veldkamp, as well seminar attendants at the NYU’s Financial Economics Workshop, for their comments and suggestions. All remaining errors are our own responsibility.

†Department of Economics, Stern School of Business, New York University and RCEA. Email: clem@nyu.edu. Web: http://pages.stern.nyu.edu/~gclement

‡Department of Economics, Stern School of Business, New York University and NBER. Email: tcooley@stern.nyu.edu. Web: http://pages.stern.nyu.edu/~tcooley/

§Centro de Investigación y Docencia Económicas, México D.F., México. Email: sonia.digiannatale@cide.edu. Web: http://www.cide.edu/investigador/profile.php?IdInvestigador=179
1 Introduction

In this paper we investigate the role played by two factors in shaping executive compensation. The first factor is the technological relationship between assets in place and the executive’s actions. In other words, how the marginal product of managerial input varies with firm size. The second is the moral hazard that arises from the non-observability of such input. Rather than focus on one feature of compensation plans as is common in the literature, we assess our model by considering its implications for a large array of statistics.

We begin by documenting a series of stylized facts for two measures of executive compensation. One is the expected discounted value of the claims to the stochastic stream of cash payments that is promised to the executive at a certain date. The other is the sum of the variation in such measure and the cash compensation awarded during the year. Versions of the former, which we label expected life-time compensation, have been widely documented in studies of the “level” of compensation. The latter, which we will refer to as yearly total compensation (or, more simply, just as total compensation), has been employed in studies of the sensitivity of compensation to changes in shareholder wealth. Our model will provide implications for both measures.

Both definitions of compensation are shown to be positively correlated with measures of firm size. However, while expected life-time compensation has increased rather steadily during our sample period, yearly total compensation has not.

We then investigate the split of total yearly compensation between its current and deferred components. Current compensation includes all claims that can be instantaneously traded for consumption goods. Deferred compensation, the residual part, consists of the change in the current expected value of all claims over future consumption. It turns out that deferred component accounts for the largest part of total yearly compensation.

Finally, we document pay–performance sensitivity. Our estimates are largely consistent with those of Aggarwal and Samwick (1999), among others. When shareholder wealth rises by $1,000 dollars, CEO’s total compensation increase by $27 dollars for firms with the least volatile firms and by $2 dollars and 20 cents for the most volatile. We also find that most of this sensitivity is in deferred compensation and that sensitivity is essentially invariant to size.

Starting with Roberts (1956) if not earlier, several scholars have modeled the
technological relation between assets in place and managerial actions.\footnote{Among the most recent studies in this class are Gabaix and Landier (2008) and Tervio (2003).} Prompted by the evidence of a strong positive correlation between firm size and certain measures of compensation, most have focused on the role played by the marginal product of managerial input, however defined, and on how it varies with size. Unfortunately, this class of model does not generate testable implications for pay–performance sensitivity. Are the implications of this literature for the relationship between size and the “level” of compensation robust, once we generalize the framework in such a way that it is consistent with the data on pay–performance sensitivity? Gabaix and Landier (2008) argue that their model “explains the level of total compensation without appealing to effort considerations. Incentives would determine, in a second and subordinate step, the relative mix of total pay between salaries and incentives.” We will show that in our framework, their claim does not hold true.

At the same time, following the seminal contribution of Holmstrom (1979), a large number of papers has modeled the interaction between shareholders and executive as a Principal–Agent relationship. The shareholders’ payoff depends on the executive’s action, which is private information. Therefore, the constrained Pareto–optimal contract dictates that compensation will covary with the shareholders’ payoff. Because of this feature, these models have been employed to study pay–performance sensitivity. However, most models in this class do not deliver implications for the relationship between expected life–time compensation and size.

Our environment is a generalization of the dynamic moral hazard model of Spear and Srivastava (1987) and Wang (1997), augmented to allow for capital accumulation and for a non–trivial relation between utility cost of managerial effort and capital stock. First, we use it in order to characterize the interaction between moral hazard and the technological relation capital/managerial input. Second, we assess its descriptive power by comparing its implications for compensation with their empirical counterparts.

Jointly modeling moral hazard and the interaction between assets and managerial inputs produces novel insights and impacts all aspects of the constrained–optimal compensation contract. For example, consider the case in which the marginal utility cost of effort increases with size. Without moral hazard, a higher capital stock unequivocally calls for lower effort. With moral hazard, however, higher capital will also increase the benefit from eliciting higher. Therefore the net effect is not known a priori. Since in general higher effort also calls for higher expected compensation, the impact of increasing size on the latter can also be of either sign.
Very preliminary results show that our model produces implications that are qualitatively consistent with the stylized facts. In particular, in our framework (i) both measures of compensation we focus on are positively correlated with size and (ii) pay–performance sensitivity is higher, the lower firm–level volatility, but (iii) is essentially invariant to firm size. Finally, (iv) deferred compensation is substantially more sensitive than current compensation to changes in shareholder wealth.

Among the many questions that are still left unanswered, we feel that the most important one is whether the model can also account quantitatively for the features of the data that we document.

The reminder of the paper is organized as follows. In Section 2 we describe our dataset and we characterize a series of stylized facts for executive compensation. In Section 3 we study a simple, static model. The purpose is to provide intuition for the full–fledged dynamic model, which is introduced in Section 4. In Section 5 we conduct comparative statics exercises to highlight the model’s properties. Finally, in Section 6 we propose a calibration procedure and quantitatively assess the model. Section 7 concludes.

2 Data

We draw our data from the EXECUCOMP database, maintained by Standard & Poor’s. EXECUCOMP gathers data on the compensation of up to nine executive of each US company whose stock is traded on an organized exchange, from 1992 to date. Source for the database are the documents filed by the companies with the Security and Exchange Commission.

In this paper we confine our attention to CEOs, for the years 1992 to 2005. After cleaning, we are left with 4,643 executives, employed by 2,549 companies, for a total of 22,515 CEO–year observations. Details about the data and the construction of compensation measures can be found in Appendix A.

2.1 Measuring Compensation

Executives’ compensation packages are notoriously complicated objects. Murphy (1999) provides a detailed account of their many components. The four most important ones are salary, bonus, stock, and options. Among the others are severance payments, payments for unused vacation, 401K contributions, life–insurance premia, etc.
**Salary** It is the sum of cash and non-cash non-contingent compensation earned by the executive during the year.

**Bonus** It consists of the cash and non-cash compensation that is earned during the year, awarded upon the achievement of certain results – mostly set in terms of accounting figures.

**Stock** Companies award their CEOs shares in a variety of ways and with a variety of clauses. Besides granting shares at zero price, companies also allow their executives to purchase stock at discounted price during the year. Typically, the stock awarded is restricted, meaning that the CEO cannot dispose of it, and is forfeited upon leaving the company.

**Options** Stock options are generally awarded at the money, are restricted, and have exercise dates scattered over a number of years.

The summary measures of compensation used in the literature consist of different aggregates of such components. As mentioned in the Introduction, in this paper we will employ two measures. This choice is dictated by the model that will be introduced in Section 4. Expected lifetime compensation at a given time is simply the expected discounted value of all streams of cash that are promised to the executive. In principle, it includes the value of stock- and option-holdings and the expected discounted value of all future handouts, in form of cash and securities. Total yearly compensation is essentially the yearly variation in expected lifetime compensation, plus the cash payments received during the year. According to this approach, the stock-related yearly compensation is equal to the change in the value of stock holdings plus the proceeds from stock sales, minus the cost of purchase (which is zero in the case of stock grants). The option-related component consists of the change in the value of option holdings plus the gains from exercise.

For the sake of comparison, we will also consider a further measure, adopted by several studies such as Murphy (1999), Bebchuk and Grinstein (2005), Frydman and Saks (2006), and Gabaix and Landier (2008). It consists of the sum of the executive’s salary, bonuses, long-term incentive plans, the grant-date value of restricted stock awards, and the grant-date Black-Scholes valued of granted options. The reason why this measure is not at the center of our analysis is that our model does not generate implications for it.
2.2 Compensation and Firm Size

Beginning with Kostiuk (1990), a multitude of scholars have investigated the relationship between firm size and executive compensation. Kostiuk (1990) himself, Murphy (1999), Bebchuk and Grinstein (2005), and Gabaix and Landier (2008) among others, have found evidence of a positive correlation between the two variables. Consistently, we find that also in our sample their definition is unconditionally positively correlated with proxies for size such as sales and book value of assets.

The question is whether a similar pattern holds for our two measures of compensation. We find that it does. The two panels in Figure 1 depict median compensation for each decile in the distribution of book value of assets in our sample. The pattern is the same when we consider the distribution of sales instead.

Somewhat echoing the popular and business press, a host of recent papers have argued that the level of executive compensation has increased dramatically over time. For example, state that “The compensation of top executives increased by 6.8% per year from 1980 to 2003.” Bebchuk and Grinstein (2005) find that “Among S&P 500 firms, average CEO compensation climbed from $3.7 million in 1993 to $9.1 million in 2003 (an increase of 146%).” The main conclusion of Gabaix and Landier (2008) is that such increase can be explained by the rise in size of US corporations.

When using their definition, we also find that compensation has increased over time. Over the 14 sample–years, both mean and median compensation have increased at a much faster pace than compensation in other occupations. Interestingly, they have also increased in years in which most companies had rather poor financial results. This is the evidence used by those pundits that argue that CEOs never lose, even when

![Figure 1: Compensation and Firm Size](image-url)
their companies’ results are negative.

When we turn to our definitions, the message is not nearly as clear. Neither

definition increases monotonically over time. Indeed, the right panel shows that CEOs do lose when their companies perform poorly. Notice that both patterns are clearly influenced by stock market fluctuations. (Recall that the Dow Jones peaked in January 2000 and hit its minimum in September 2002.)

We find it convenient to partition compensation in current and deferred. Current compensation includes all claims that can be instantaneously traded for consumption goods. Deferred compensation, the residual part, consists of the current expected value of all claims over future consumption. Operationally, we define current compensation as the sum of salary, bonus, dividends, and net revenues from trade in stock. Deferred compensation is the sum of the yearly changes in the value of stock and options in portfolio, as well as the change in expected salary and in retirement benefit and other deferred payments. Interestingly, current compensation represents a small fraction of total compensation.

The reader will have noticed that we focused our attention on medians, rather than means. The reason is that the cross-sectional distribution of compensation is very skewed. Figure 4 depicts the distribution of total compensation in 1999 and 2003, respectively. In 1999, the average compensation of the top decile was close to a staggering one billion dollars. This is not that surprising, once one notices that (i) CEOs in the top deciles held substantial equity stakes in their companies and (ii) the stock market sky–rocketed in that year. In 1999, the most–highly compensated CEOs was Larry Ellison of Oracle, with 50 billion dollars. In that year, Oracle’s market capitalization rose from about 35.5 billions to close to 202. His equity stake was
24%. Ranked second and third were Bill Gates of Microsoft (16 billions) and Samuel Bronfman of Seagram (3.6). Their equity stakes were 15% and 14%, respectively. Skewness is not a characteristic of deferred compensation alone. A similar pattern emerges from the inspection of the distribution of current compensation. In 2003, for example, the average current compensation of the top decile was about 11 million dollars. For the next decile, this figure was only about 4 millions.

2.3 Sensitivity of Compensation to Changes in Shareholder Wealth

A further aspect of executive compensation that has attracted much attention from scholars is its sensitivity to changes in shareholder wealth.
Since Jensen and Murphy (1990), a large number of economists have tried to estimate the impact of innovations in shareholder wealth on the compensation of their CEOs. As pointed out by Hall and Liebman (1998), scholars have used a large number of different regression specifications. Here we follow Aggarwal and Samwick (1999) in estimating the following equation:

$$w_{ijt} = \gamma_0 + \gamma_1 \Delta MKT_{CAP}jt + \gamma_2 \Delta MKT_{CAP}jt \times F(\sigma_j) +$$

$$+ \gamma_3 F(\sigma_j) + \lambda_t + \varepsilon_{it},$$

where $i, j, t$ index the executive, the firm, and time, respectively. The letter $w$ denotes compensation, $MKT_{CAP}jt$ is total market capitalization, $\Delta$ is the one-period lag operator, and $\lambda_t$ is a dummy variable whose purpose is to control for aggregate shocks. Finally, $\sigma_{jt}$ denotes the dollar standard deviation in market capitalization. $F(\cdot)$ is the cumulative distribution of the distribution of standard deviations. As clearly explained in Aggarwal and Samwick (1999), the interaction term $\Delta MKT_{CAP}jt \times F(\sigma_j)$ was introduced because the impact on compensation of a 1,000 dollar change in shareholder wealth is expected to be larger, the smaller the average change in capitalization. The measure of sensitivity for company $j$ will be $\gamma_1 + \gamma_2 \times F(\sigma_j)$. The estimation results are reported in Table 1.

Table 1: Median Regression Estimates of Pay–Performance Sensitivities

<table>
<thead>
<tr>
<th>Dependent Variables:</th>
<th>Total</th>
<th>Current</th>
<th>Deferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shareholder_Gain</td>
<td>27.11</td>
<td>0.30</td>
<td>25.58</td>
</tr>
<tr>
<td></td>
<td>(0.1128)</td>
<td>(0.0125)</td>
<td>(0.0606)</td>
</tr>
<tr>
<td>Sh_Gain × Variance_distrib</td>
<td>-24.92</td>
<td>-261</td>
<td>-23.65</td>
</tr>
<tr>
<td></td>
<td>(0.07628)</td>
<td>(0.013)</td>
<td>(0.0633)</td>
</tr>
<tr>
<td>Variance_distrib</td>
<td>465.52</td>
<td>1585.33</td>
<td>-231.03</td>
</tr>
<tr>
<td></td>
<td>(246.48)</td>
<td>(26.155)</td>
<td>(129.27)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>14,233</td>
<td>14,940</td>
<td>14,233</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.076</td>
<td>0.084</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Note: Standard errors in parenthesis.

Table 2 reports the estimated effects on total, current, and deferred compensation, induced by a 1,000 dollar increase in shareholder wealth. Given the wide heterogeneity in dollar volatility of market capitalization, the effects are very different across firms. A 1,000 dollar increase in capitalization induces a 27 dollar increase in compensation for the firm with the lowest volatility and a 2 dollar increase for a firm with the highest. The table also highlights the fact that essentially all sensitivity is generated
by deferred compensation. Even for the firms with the lower variance, a 1,000 dollar increase in compensation has the effect of increasing compensation by only 30 cents.

Table 2: Sensitivity at different levels of volatility

<table>
<thead>
<tr>
<th>Dependent Variables:</th>
<th>Total</th>
<th>Current</th>
<th>Deferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest Variance</td>
<td>27.11</td>
<td>0.30</td>
<td>25.58</td>
</tr>
<tr>
<td>Median Variance</td>
<td>14.64</td>
<td>0.17</td>
<td>13.75</td>
</tr>
<tr>
<td>Largest Variance</td>
<td>2.18</td>
<td>0.04</td>
<td>1.93</td>
</tr>
</tbody>
</table>

Finally, we ask whether firm size has any effect on pay–performance sensitivity. To this end, we augment equation (1) by interacting the gain in shareholder wealth with assets lagged one period. It turns out that the effect of size is statistically significant, but economically insignificant.

Table 3: Median Regression Estimates of Pay–Performance Sensitivities

<table>
<thead>
<tr>
<th>Shareholder Gain</th>
<th>27.11</th>
<th>27.04</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.1128)</td>
<td>(0.1524)</td>
</tr>
<tr>
<td>Sh,( \text{Gain} \times \text{Variance,distrib} )</td>
<td>-24.92</td>
<td>-24.75</td>
</tr>
<tr>
<td></td>
<td>(0.07628)</td>
<td>(0.161)</td>
</tr>
<tr>
<td>Variance,distrib</td>
<td>465.52</td>
<td>17,787.14</td>
</tr>
<tr>
<td></td>
<td>(246.48)</td>
<td>(344.52)</td>
</tr>
<tr>
<td>Sh,( \text{Gain} \times \text{Asset(t-1)} )</td>
<td>-</td>
<td>-0.000382</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(0.000066)</td>
</tr>
<tr>
<td>Asset(t-1)</td>
<td>-</td>
<td>-0.0106</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>14,233</td>
<td>13,379</td>
</tr>
<tr>
<td>Pseudo ( R^2 )</td>
<td>0.076</td>
<td>0.08</td>
</tr>
</tbody>
</table>

3 The Static Model

The purpose of this Section is to build intuition for the analysis of the dynamic model to be introduced in Section 4. To this end, we consider a simple extension of the standard static hidden action problem of Holmstrom (1979). Our formulation is quite close to Baker and Hall (2004).

The principal is endowed with a stock of capital \( k \geq 0 \) and a production technology,
which is operated by the agent. The principal is risk neutral, while the agent is risk averse, with preferences represented by the utility function \( H(c, a) = u(c) - ak^\varphi \), where \( c \) and \( a \) denote consumption and investment, respectively, and \( u'(\cdot) > 0, u''(\cdot) < 0 \).

The marginal utility of effort is always negative, but varies with capital according to sign of \( \varphi \). When \( \varphi > 0 \), the marginal cost of effort is larger, the larger the capital. The opposite occurs when \( \varphi < 0 \). Notice that an increasing marginal cost of effort is observationally equivalent to a decreasing marginal product of managerial effort.

While the output of the production process is public information, the effort exerted by the agent constitutes his private information. This obviously implies that the agent’s consumption will depend on the output realization rather than on the actual managerial effort. Output is given by \( \theta f(k) \), where \( f : \mathbb{R}^+ \to \mathbb{R}^+ \), \( f'(\cdot) > 0 \), and \( \theta \in \{ \theta_h, \theta_l \} \), \( \theta_h > \theta_l > 0 \), with \( pr(\theta = \theta_h|a) = 1 - e^{-a} \).

The Constrained–Pareto efficient contract that offers an expected utility \( \omega \) to the agent is given by contingent consumption \( c_h, c_l \) and an effort recommendation \( a^* \) that solve the following optimization problem:

\[
\begin{align*}
\max_{c_h, c_l} & \quad [\theta_h f(k) - c_h][1 - e^{-a^*}] + [\theta_l f(k) - c_l]e^{-a^*}, \\
\text{subject to} & \quad u(c_h)[1 - e^{-a^*}] + u(c_l)e^{-a^*} - a^*k^\varphi = \omega, \\
& \quad a^* = \arg \max_a \ u(c_h)[1 - e^{-a}] + u(c_l)e^{-a} - ak^\varphi \\
& \quad a^* \geq 0.
\end{align*}
\]

For simplicity, now assume that \( u(c) = c^\sigma \) and define \( u_i \equiv u(c_i), i = h, l \). This implies \( c_i = u_i^{1/\sigma} \). Then rewrite the optimization problem as

\[
\begin{align*}
\max_{u_h, u_l} & \quad [\theta_h f(k) - u_h^{1/\sigma}][1 - e^{-a^*}] + [\theta_l f(k) - u_l^{1/\sigma}]e^{-a^*}, \\
\text{subject to} & \quad u_h[1 - e^{-a^*}] + u_l e^{-a^*} - a^*k^\varphi = \omega, \\
& \quad a^* = \arg \max_a \ u_h[1 - e^{-a}] + u_l e^{-a} - ak^\varphi \\
& \quad a^* \geq 0.
\end{align*}
\]

The incentive compatibility constraint yields \( a^* = -\log \left( \frac{k^\varphi}{u_h - u_l} \right) \). A greater effort implies a greater probability of a good outcome, but may also lead to a larger average agent’s consumption. In fact, a larger effort must be accommodated by a larger spread between contingent consumption levels. By risk aversion, this means a lower level of utility. If the increase in the probability of success is not large enough, the average level of consumption will have to grow in order for the principal to keep the promise of delivering \( \omega \).
Assuming $\sigma = 1/2$ allows for an even sharper characterization of the problem. Letting $x \equiv u_h - u_l$, we can rewrite the optimization problem as

$$\max_x \left( \frac{\theta_l - \theta_h}{x} f(k) k^{\varphi} + \theta_h f(k) + k^{\varphi} (2u_h - x) - u_h^2 \right)$$

subject to $u_h = \omega + k^{\varphi} \left[ 1 - \log \left( \frac{k^{\varphi}}{x} \right) \right]$

$$u_l = u_h - x$$

This formulation highlights the trade-off faced by the principal. On the one hand, increasing the spread between promised utilities has the effect of encouraging a greater effort provision. On the other hand, (i) by concavity of the utility function, a mean preserving spread of utilities also implies a higher expected cost in terms of consumption and (ii) the fulfillment of the promise of delivering $\omega$ requires the provision of a higher expected utility from consumption. Unless the increase in the success probability is substantial, this will also decrease the principal’s payoff.

The novelties of this environment with respect to the existing literature are that (i) we can do comparative statics exercises with respect to the size $k$ and (ii) the outcomes of these exercises depend crucially on the sign of $\varphi$.

Define average compensation as $c_h[1 - e^{-a^*}] + c_l e^{-a^*}$. Also, define the sensitivity of compensation to innovations in shareholder wealth as $\frac{(c_h - c_l)}{[\theta_h - \theta_l] f(k)}$.

Figure 5 depicts the standard case in which $\varphi = 0$, so that the marginal product of effort is invariant with respect to $k$. The recommended effort level increases with $k$. The reason is that the gain from obtaining a good outcome, $(\theta_h - \theta_l) f(k)$, also increases with $k$. We call this the volatility effect. Consistent with the data, average compensation increases (unconditionally) with size. The sensitivity of agent’s consumption with respect to changes in the principal’s payoff (shareholders’ wealth) also decreases with size. This unconditional relation holds also in the data, as larger firms have higher dollar standard deviation. How do the predictions change if we condition on volatility? We can do that simply by shutting down the volatility effect, i.e. by adjusting $\theta_h - \theta_l$ in such a way to keep the volatility of cash flows $(\theta_h - \theta_l) k$ constant as we increase $k$. In that scenario, as expected, both expected compensation and pay-performance sensitivity become invariant to size. We know from Section 2 that the first of these two conclusions is counterfactual. Figure 6 depicts a case in which the marginal product of the executive decreases with $k$, i.e. $\varphi > 0$. Because of this assumption, raising effort as $k$ increases will be less appealing. We call this the size effect. In this example, the volatility effect ends up being the dominating one. As a result, the constrained-efficient level of effort still increases with size.
Now refer to Figure 7. It depicts the same case considered in Figure 6, the only difference being that we shut down the volatility effect. As expected, recommended effort decreases with $k$. Conditionally on volatility, expected compensation decreases with size, while sensitivity increases with size. Both of these results conflict with the evidence presented in Section 2.

The simple exercises we have conducted hint that necessary condition for our framework to make sense of the stylized facts documented in Section 2 is that $\varphi < 0$: the marginal product of managerial input must be increasing in size. Rather than
continue with the analysis of this simple model, we now switch gears and study its dynamics version.

4 A Dynamic Model

The simple static model that we have outlined is definitely useful to outline the interaction between moral hazard and capital/manager complementarity. Its main shortcoming is that firm size is exogenous. By allowing the firm to optimally choose the level of capital, the dynamic model will pose restrictions on the joint dynamics of size and compensation.

For the sake of brevity we will make explicit only those assumptions that differ from the static model introduced above. Time is discrete and is indexed by \( t = 1, 2, \ldots \). Principal and agent discount future consumption streams at the common rate \( \beta \in (0, 1) \).

We assume that \( k_t \in [\underline{k}, \bar{k}] \in \mathbb{R}_+ \) and that \( a_t \in A \), where \( A = [a, \bar{a}] \in \mathbb{R}_+ \). Output \( (y_t) \) is given by

\[
y_t = \theta_t f(k_t),
\]

where \( \theta_t \in \Theta \) is a random variable distributed according to the time–invariant distribution function \( G(\theta_t|a_t) \). We assume that \( G \) has a density denoted by \( g \), and that \( g \) is twice continuously differentiable with respect to \( a \). We also assume that, for given \( a \), the distribution is i.i.d. from one period to the next and its support is compact.
We assume that the principal does not have access to any type of credit and that investment in the capital stock is the only means of transferring consumption to the future. Therefore, the following resource constraint needs to be satisfied at all $t$:

$$c_t + i_t \leq \theta_t f(k_t),$$

where, $c_t \geq 0$ is the agent’s consumption and $i_t$ represents investment. The law of motion for capital is the usual one:

$$k_{t+1} = (1 - \delta)k_t + i_t,$$

where $\delta \in (0, 1)$ denotes the depreciation rate. The last two expressions imply the following constraint:

$$c_t \leq \theta_t f(k_t) + (1 - \delta)k_t - k_{t+1}.$$

In the tradition of principal agent theory, we assume that the principal is risk-neutral, while the agent is risk-averse. The latter’s static preferences are represented by the utility function $u(c_t, m(a_t)l(k_t))$, which is assumed to be bounded, strictly increasing and strictly concave in $c_t$, and strictly decreasing in $m(a_t)l(k_t)$. This specification allows for the agent’s disutility from effort to depend on the size of the capital stock she manages. We assume that $m(a_t)$ is an increasing and convex mapping with respect to $a_t$, as is commonly assumed in standard agency models. The only restriction we impose on the function $l(k_t)$ is that it can only take non-negative values, that is $l : \mathbb{R}_+ \to \mathbb{R}_+$. We also assume that $u(c_t, m(a_t)l(k_t))$ is additively separable in the arguments $c_t$ and $m(a_t)l(k_t)$.

We allow the principal and the agent to employ history-dependent pure strategies, as in Spear and Srivastava (1987) and Wang (1997). History at time $t$ is given by the sequence $h^t = \{(y_s, k_s)\}_{s=0}^t$. The principal’s task is to design a compensation contract $\sigma = \{a_t(h^{t-1}), c_t(h^t)\}_{t=1}^\infty$ and a sequence of contingent capital levels $\{k_t(h^{t-1})\}_{t=1}^\infty$ that maximize her lifetime discounted expected utility. This notation reflects the assumption that the principal chooses investment at the beginning of every period, before the realization of $\theta_t$. The agent’s strategy consists of the sequence $\{a_t(h^{t-1})\}_{t=1}^\infty$.

The continuation profile of a contract $\sigma$ from date $t + 1$ on, given $h^t$, is denoted as $\sigma|h^t$. Conditional on the agent following the action recommendation given by $\sigma|h^t$, the agent’s continuation value is denoted by $\omega(\sigma|h^t)$, and that of the principal is denoted by $v(\sigma|h^t)$. 
A contract $\sigma$ is said to be feasible if the effort choice of the agent belongs to the set $A$, and the resource constraint is satisfied in every period, given the history of outputs. More formally,

**Definition 1** A contract $\sigma$ is feasible if

$$a_t(h^{t-1}) \in A, \forall t \geq 1, \forall h^{t-1},$$

and

$$0 \leq c_t(h^t) \leq \theta_t f(k_t) + (1 - \delta)k_t - k_{t+1}(h^{t-1}), \forall t \geq 1, \forall h^t.$$ (3)

**Definition 2** A contract $\sigma$ is incentive compatible if

$$a_t(h^{t-1}) \in \arg \max_a \int_\theta \{u(c_t(h^t), m(a)l(k_t)) + \beta w(\sigma|h^t)\} g(\theta_t|a_t(h^{t-1}))d\theta, \forall t \geq 1, \forall h^t.$$ (4)

This constraint ensures that the agent will not deviate from the principal’s effort recommendation plan in any future date, from period $t + 1$ on.

Since in this environment, $a_t(h^{t-1})$ is a continuous variable, we use the first-order approach to incentive compatibility, which is not universally valid. To ensure the validity of this approach, we assume that the Monotone Likelihood Ratio Property and the Convexity of the Conditional Distribution Condition hold, following Rogerson (1985) and Spear and Srivastava (1987).

Let $\Omega$ be the set of capital levels and agent’s expected discounted utilities $(k, \omega)$ such that there exists a feasible and incentive compatible contract that generates $\omega$, given $k$. That is, such that

$$\Omega \equiv \{(k, \omega) \in \Delta \mid \exists \sigma \quad \text{s.t.} \quad (2), (3), (4), \text{and,} \quad w(\sigma|h^0) = \omega\}.$$ 

Assume that $\Delta \in \mathbb{R}^2$ is non-empty and compact, and that it is endowed with a structure such that $\Omega$ is non–empty as well. Then, for every $(k, \omega) \in \Omega$, we define the following set:

$$\Phi(k, w) = \{v(\sigma|h^0) \mid (2), (3),(4), \text{and,} \quad w(\sigma|h^0) = \omega\},$$

where, $\Phi$ is the set of expected discounted utilities of the principal that can be generated by a contract $\sigma$ that is feasible and incentive compatible.

For given $(k, \omega)$, the principal’s problem is to choose a feasible and incentive compatible contract $\sigma$ that attains the supremum of $\Phi(k, \omega)$. 

15
In Appendix B we show that such contract exists, is unique. Furthermore, the optimization problem that we have described admits a recursive representation. The Bellman equation is

$$v(k, \omega) = \max_{a,k',c,\omega'(\theta)} \int_{\theta} \left\{ \theta f(k) - c - k' + (1 - \delta)k + \beta v(k', \omega'(\theta)) \right\} g(\theta|a) d\theta$$

s.t. \( \int_{\theta} \left\{ u(c(\theta), m(a)l(k)) + \beta \omega'(\theta) \right\} g(\theta|a) d\theta = \omega \) (5)

\( a \in \arg \max_a \int_{\theta} \left\{ u(c(\theta), m(a)l(k)) + \beta \omega'(\theta) \right\} g(\theta|a) d\theta \) (6)

\( 0 \leq c(\theta) \leq \theta f(k) - k' + (1 - \delta)k \ \forall \theta \) (7)

\( a \in A \) (8)

\( (k', \omega'(\theta)) \in \Omega \ \forall \theta \) (9)

Unfortunately an analytical characterization of the constrained Pareto–optimal contract is not possible. Therefore, we recur to numerical methods. First, we compute an approximation to the set \( \Phi \), using the algorithm developed by Abreu, Pierce, and Stacchetti (1990). Then, we approximate the solution to the Bellman equation by means of a variant of the value function iteration method. The details on the algorithm can be found in Appendix C.

4.1 An Example

The purpose of this section is to showcase the potential of our model, by running a simple numerical example. We begin by making assumptions about functional forms and parameters. The utility function takes the form \( u(c, m(a)l(k)) = \frac{c^{1-\sigma}}{1-\sigma} - ak^{\varphi} \). The production function is \( f(k) = k^\alpha \). Furthermore, we assume that \( A = [0, \bar{a}] \) and \( \Theta = \{ \theta_l, \theta_h \} \), with \( G(\theta_l|a) = e^{-a} \).
The parameter values are reported in Table 4. We set $\beta, \sigma, \alpha, \text{and } \delta$ to values that are standard in the macroeconomics literature. Given the indications emerged from the analysis of the static model, we set $\varphi$ to a negative value: the marginal cost of exerting effort decreases with firm size. The value of the outside option $\omega$ is chosen to ensure that the contract is renegotiation–proof. Finally, $\theta_h$ and $\theta_l$ are completely arbitrary.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$\delta$</th>
<th>$\varphi$</th>
<th>$\theta_h$</th>
<th>$\theta_l$</th>
<th>$\bar{\omega}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0.5</td>
<td>0.3</td>
<td>0.1</td>
<td>-0.4</td>
<td>1.5</td>
<td>0.4</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 4: Parameter Values

The value function is pictured in Figure 8. Notice that the set $\Phi$ is described by a correspondence that maps each value of the capital stock $k$ into a set $0, \bar{\omega}(k)$, where $\bar{\omega}(k)$ is monotone increasing in $k$.

Using the policy functions of the above problem, it is rather easy to compute the expected cost of delivering a promised utility $\omega$, given a capital stock $k$. For each pair $(k, \omega) \in \Phi$, the expected discounted cost of delivering $\omega$ is denoted by $C(k, \omega)$, where $C(k, \omega)$ solves the following functional equation:

$$C(k, \omega) = (1 - e^{-a^*}(k, \omega))\left[c^*_h(k, \omega) + \beta C(k', \omega_h) + e^{-a^*}(k, \omega)\right]$$

Figure 10 depicts the policy function for effort. As expected, recommended effort is increasing in the capital stock. Finally, Figure 11 depicts the policy functions for current and promised utility. In the left panel, we have plotted $u_i(k, \omega), i = h, l$. In
the right panel, we have pictured the contingent increment in deferred utility $\omega_i - \omega$, $i = h, l$.

Recall that our main objective is to compare the model’s predictions for executive compensation with the facts characterized in Section 2. The first task in order is to define theoretical counterparts for the empirical measures of compensation that we defined in that section. Our measure of firm size will be the capital stock $k$. Shareholder value will be $v(k, \omega)$. Profits are $\theta_i k^{\alpha_i} - c_i + k(1 - \delta) - k'$. Changes in shareholder value will be $v(k', \omega_i) - v(k, \omega)$. Here are our definitions of compensation:
$C(k, \omega)$ will be referred to as life–time compensation, $c_i(k, \omega)$ as current compensation, $C(k', \omega_1) - C(k, \omega)$ as deferred compensation, and $c_i(k, \omega) + C(k', \omega_1) - C(k, \omega)$ as yearly total compensation.

The exercise consists of generating a panel of firm–CEO pairs by simulating 6,000 paths of 20 periods each and computing the same statistics of Section 2. Refer to Figure 12: both life–time and total yearly compensation are unconditionally positively correlated with firm size. This holds true even when we condition on firm–level volatility.

We assess the model’s predictions for pay–performance sensitivity by running regression (1) on the same artificial panel. The results are shown in Table 5. Pay–
Figure 12: Compensation and Firm Size – Model

performance sensitivity is a lot larger for low-volatility firms, but varies very little
with firm size as measured by the capital stock. Furthermore, the sensitivity of de-
ferred compensation is sensibly higher than the that of current compensation. All of
these features are qualitatively consistent with the data.

Table 5: Median Regression Estimates of Pay–Performance Sensitivities – Model

<table>
<thead>
<tr>
<th>Dependent Variables:</th>
<th>Total</th>
<th>Current</th>
<th>Deferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shareholder Gain</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sh. Gain × Variance.distrib</td>
<td>−2.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance.distrib</td>
<td>1.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sh. Gain × Asset(t-1)</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asset(t-1)</td>
<td>−0.07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of observations 120,000
Pseudo $R^2$ 0.46

5 Comparative Statics

TBA
6  Quantitative Assessment
TBA

7  Conclusion
TBA
A Data

B Proofs and Lemmas

In order to prove that a contract $\sigma$ exists, we start by showing that the set of the principal’s expected discounted utilities $\Phi(k, \omega)$ is compact.\footnote{The proofs provided in this section follow the strategy of Wang (1997).}

**Proposition 1** $\Phi(k, w)$ is compact, $\forall (k, \omega) \in \Omega$.

**Proof.** Fix the pair $(k, \omega)$. We already know that $\Phi(k, w)$ is bounded. It is left to prove that it is also closed. Let $\{V_n\} \subseteq \Phi(k, w)$, where $V_n \rightarrow V_\infty$ when $n \rightarrow \infty$. We need to show that $V_\infty \in \Phi(k, w)$. In words, we need to demonstrate that there exists a contract $\sigma_\infty$ that satisfies $\Phi(2, 3, 4)$, $w(\sigma_\infty|h^0) = w$, and $v(\sigma_\infty|h^0) = V_\infty$. Now we will construct such an optimal contract $\sigma_\infty$. By the definition of $\Phi(k, w)$, there exists a sequence of contracts $\{\sigma_n\} = \{a^n_t(h^{t-1}), c^n_t(h^t)\}$ and capital $\{k^n_{t+1}(h^{t-1})\}$, where the constraints $\Phi(2, 3, 4)$, and $\omega(\sigma_n|h^0) = w$ are satisfied for every $n$. Therefore

$$V_\infty = \lim_{n \rightarrow \infty} \sum_{t=1}^{\infty} \beta^{t-1} \int [\theta_t f(k_t) - c^n_t(h^t) - k^n_{t+1}(h^{t-1}) + (1 - \delta)k_t\gamma(\theta_t)a^n_t(h^{t-1})] dh^t$$

For $t = 1$, notice that $\{a^n_1(h^0), c^n_1(h^1)\}$ and $\{k^n_2(h^0)\}$ are finite collections of bounded sequences. Therefore, there exist collections of subsequences $\{a^{n_1}_1(h^0), c^{n_1}_1(h^1)\}$ and $\{k^{n_2}_2(h^0)\}$ such that

$$\lim_{n_1 \rightarrow \infty} a^{n_1}_1(h^0) = a^{\infty}_1(h^0), \quad \lim_{n_1 \rightarrow \infty} c^{n_1}_1(h^1) = c^{\infty}_1(h^1), \text{ and } \lim_{n_2 \rightarrow \infty} k^{n_2}_2(h^0) = k^{\infty}_2(h^0).$$

We now consider $t = 2$. Notice that $\{a^n_2(h^1), c^n_2(h^2)\}$ and $\{k^n_3(h^1)\}$ are finite collections of bounded sequences, and we can define $\{a^{\infty}_2(h^1), c^{\infty}_2(h^2)\}$ and $\{k^{\infty}_3(h^1)\}$ similarly as we did for $t = 1$. If we iterate this procedure for $t = 3, 4, \ldots$, and let $\sigma_\infty = \{a^{\infty}_t(h^{t-1}), c^{\infty}_t(h^t)\}$ along with $k = \{k^{\infty}_{t+1}(h^{t-1})\}$, then it is easy to verify that the constructed contract $\sigma_\infty$ is what we desired for. \hfill \blacksquare

We will now formulate the Bellman equation to solve this infinite–horizon optimization problem. For all $(k, \omega) \in \Omega$, the highest level of expected discounted utility that can be achieved by the principal $\sigma|h^0$ is given by

$$v^*(k, \omega) \equiv \max \{v(\sigma|h^0) \in \Phi(k, \omega)\}.$$
Define the operator $T$ that maps from the space of bounded and continuous functions $v : \Omega \rightarrow \mathbb{R}$ into itself, with the sup norm, as:

$$T(v)(k, \omega) = \max \left\{ \int_\theta \{ \theta f(k) - c - k' + (1 - \delta)k + \beta v(k', \omega') \} g(\theta | a) d\theta \right\}$$

s.t. \left\{ \begin{align*}
    & a \in \arg \max \int_\theta \{ u(c, m(a)l(k)) + \beta \omega' \} g(\theta | a) d\theta \\
    & 0 \leq c \leq \theta f(k) - k' + (1 - \delta)k \\
    & a \in A \\
    & (k', \omega') \in \Omega
\end{align*} \right. \quad (17)

where the decision variables in the optimization process are the following: $a = a(k, \omega), c = c(\theta, k, \omega), k' = k'(k, \omega)$, and $\omega' = \omega'(\theta, k, \omega)$. The solution of this problem is Markovian stationary and perfect, in the sense that no deviation from the agent is expected in any period. Given that the just mentioned decision variables are expressed in stationary terms, then the history of the realizations of the output distribution is summarized by the state variables $(k, \omega)$. At this point, we need to demonstrate that $v^*(k, \omega)$ is a fixed point of $T$.

**Proposition 2** $v^*(k, \omega) = T(v^*)(k, \omega)$.

**Proof.** Fix $\omega$, the lifetime discounted utility ensured by the optimal contract to the agent, and $k$, the optimal capital level of the firm. First, we show that $T(v^*)(k, \omega) \leq v^*(k, \omega)$. This inequality is true if there exists a feasible and incentive compatible contract $\sigma$ such that $\omega(\sigma | h^0) = \omega$ and $v(\sigma | h^0) = T(v^*)(k, \omega)$. The desired contract $\sigma$ can be constructed in the following way. Let $a(k, \omega), c(\theta, k, \omega), k'(k, \omega)$, and $\omega'(\theta, k, \omega)$ denote the solution of the maximization problem associated with the definition of $T(v^*)(k, \omega)$. Now, let $a_1(h^0) = a(k, \omega), c_1(h^1) = c(\theta_1, k, \omega)$, and $k_2(h^0) = k'(k, \omega), \forall h^1$ along with $k_1 = k_0$. For the realization of $\theta$ in $t = 1$, denoted $\theta_1$ for the purpose of this proof, there exists a feasible and incentive compatible contract $\sigma_{\theta_1}$ that ensures a level of expected discounted utility $\omega'(\theta_1, k, \omega)$ to the agent, and $v^*(k'(k, \omega), \omega'(\theta_1, k, \omega))$ to the principal. Thus, we can say that $\sigma | h^1 = \sigma_{\theta_1}, \forall h^1$. It is obvious that the constructed contract $\sigma$ is what is desired.

We now need to show that $v^*(k, \omega) \leq T(v^*)(k, \omega)$. Let $\sigma^*$ be an optimal contract that ensures a level of expected discounted utility of $\omega$ to the agent, given $k$. In consequence, we can say that

$$v^*(k, \omega) = v(\sigma^* | h^0)$$
\[
v^*(k, \omega) = \int_{\theta} \left\{ \theta_1 f(k_1) - c_1^*(\theta_1) - k_2^*(h^0) + (1 - \delta)k_1 + \beta v^*(k_2^*(h^0), \sigma^*|h^1) \right\} g(\theta|a_1^*(h^0))d\theta,
\]
or, finally,
\[
v^*(k, \omega) \leq T(v^*)(k, \omega),
\]
where the last inequality is obtained by letting \( a(k, \omega) = a^*(h^0) \), \( c(\theta, k, \omega) = c_1^*(\theta_1) \), \( \omega'(\theta, k, \omega) = \omega'(|\sigma^*|h^0) \) along with \( k'(k, \omega) = k_2^*(h^0) \) and \( k_1 = k_0 \). This solution satisfies the constraints (17), (18), (19), (20), and (21).

Provided that we have a Bellman equation to solve for an optimal contract, we can operate recursively, as in Spear and Srivastava (1987). In this case the state variables are \( k \), the capital level of the firm, and \( \omega \), the expected discounted utility of the agent. For any \( t \), the state variables are \( k_t \), determined at the beginning of period \( t - 1 \) by \( k_t = k'(k_{t-1}, \omega_{t-1}) \), and \( \omega_t \), determined at the end of period \( t - 1 \) by \( \omega_t = \omega'(\theta_{t-1}, k_{t-1}, \omega_{t-1}) \). At the beginning of period \( t \), the agent decides his optimal effort given by \( a_t = a(k_t, \omega_t) \), and the principal decides on the optimal level of capital to be accumulated given by \( k_{t+1} = k'(k_t, \omega_t) \). Then, the value of the productivity shock is realized \( \theta_t \), and is observed both by the principal an the agent. Given this value, the principal pays the agent a current compensation level of \( c_t = c(\theta_t, k_t, \omega_t) \), and promises the agent a level of expected discounted utility of \( \omega_{t+1} = \omega'(\theta_t, k_t, \omega_t) \). Finally, the consumption of the principal in period \( t \) is given by \( (\theta_t f(k_t) - c_t - k_{t+1} + (1 - \delta)k_t) \), and a level of expected discounted utility of \( v_t = v^*(k_{t+1}, \omega_{t+1}) \). Now, the state variables are given by the pair \( (k_{t+1}, \omega_{t+1}) \). And the story is repeated for every period.

Given that \( \Omega \) is a convex subset of \( \mathbb{R}^2 \), that \( \Phi \) is non-empty, compact-valued and continuous, that the return function is bounded and continuous, and that \( \beta \in (0, 1) \), then we have that the operator \( T \) has a fixed point with the standard properties. This means that the principal’s problem defined in the last section has a solution, that can be obtained by a value function iteration process. The operator \( T \) satisfies the sufficient conditions of Blackwell for a contraction and, thus, the contraction mapping theorem ensures that the Bellman equation of Proposition 2 has a unique solution.

To perform the value function iteration process, we need first to find the set \( \Omega \). With this purpose in mind, we use the approach proposed by Abreu, Pierce, and Stacchetti (1990), called APS from now on. We have to demonstrate that \( \Omega \) is self-generating, and the APS approach will allow us to device an algorithm to compute \( \Omega \).
First, we have to define the construction of the set of initial values taken by the state variables. We assume that \( k \) is restricted to take values on a closed and bounded subset of \( \mathbb{R}_+ \), denoted \( X_k \). Let
\[
\Omega_0^k = \{ k : k \in X_k = [k_{\min}, k_{\max}] \}.
\]
The state variable \( \omega \) is allowed to take values on a closed and bounded subset of \( \mathbb{R}_+ \), denoted \( X_\omega(k) \). Let
\[
\Omega_0^\omega(k) = \{ \omega(k) : k \in \Omega_0^k \text{ and } \omega \in X_\omega(k) = [\omega_{\min}(k), \omega_{\max}(k)] \}.
\]
We set \( \omega_{\min} \) arbitrarily to a very small positive number.

To obtain \( \omega_{\max}(k) \), we solve the following dynamic optimization problem:
\[
W(k) = \max_{(a,c,k')} \int_\theta \{ u(c, m(a)l(k)) + \beta W(k') \} g(\theta | a) d\theta
\]
subject to \( 0 \leq c \leq \theta f(k) - k' + (1 - \delta)k \).

The solution of this problem exists given that \( u \) is bounded, strictly increasing and strictly concave in \( u \), and strictly decreasing in \( h(a)l(k) \). Also, the constraint space is convex with respect to the control variables of this problem. By solving this problem, we obtain the set of maximal and feasible values that the agent’s future discounted expected utility can take for each level of capital that belongs to \( \Omega_0^k \).

Let
\[
\Omega_0 = \{(k, \omega) : k \in \Omega_0^k \text{ and } \omega \in \Omega_0^\omega(k) \}.
\]

We will now use the concept of self-generation of APS. Let us define an operator \( B \) such that for any arbitrary \( \Sigma \in \mathbb{R}^2 \):
\[
B(\Sigma) = \{(k, \omega) \mid \exists \{(a,c,k'),\omega'\} \text{ s.t. } (17), (18), (19), (20), \text{ and } (k'(k,\omega),\omega'(|\theta,k,\omega)) \in \Sigma, \forall \theta \}
\]
The operator \( B \) is monotone in the following sense: \( \Sigma_1 \subseteq \Sigma_2 \) implies that \( B(\Sigma_1) \subseteq B(\Sigma_2) \). We say that \( \Sigma \) is self-generating if \( \Sigma \subseteq B(\Sigma) \).

**Proposition 3** (a) \( \Omega \) is self-generating. (b) If \( \Sigma \) is self-generating, then \( B(\Sigma) \subseteq \Omega \).

**Proof.** To prove (a), let \( (k, \omega) \in \Omega \). There exists a contract \( \sigma = \{a_t(h^{t-1}), c_t(h^t)\} \) and a sequence \( \{k_{t+1}(h^{t-1})\} \) which satisfy the constraints (2), (3), (4), and \( \omega(\sigma|h^0) = \omega \).

We now say that
\[
a(k, \omega) = a_1(h^0); \ k'(k, \omega) = k_2(h^0); \ c(\theta, k, \omega) = c_1(\{\theta\}), \ \forall \theta; \ \omega'(\theta, k, \omega) = \omega_2(\sigma|\{\theta\}), \ \forall \theta.
\]
It is obvious that \{a(k, \omega), c(\theta, k, \omega), k'(k, \omega), \omega' (\theta, k, \omega)\}, defined above, satisfies the constraints (17), (18), (19), (20), and (21). Therefore, \((k, \omega) \in B(\Omega)\), which demonstrates that (a).

To prove (b), let \(\Sigma\) be self-generating, and let \((k, \omega))_{k, \omega} \in B(\Sigma).\) We have to construct a contract \(\sigma = \{a_t(h^{t-1}), c_t(h^t)\}\) and a sequence \(k_{t+1}(h^{t-1}) = k_{h^0}\) that satisfy the constraints (2), (3), (4), and \(\omega(\sigma|h^0) = \omega_{h^0}.\) We construct such a contract recursively. First, there exist \(\{a(k_{h^0}, \omega_{h^0}), c(\theta, k_{h^0}, \omega_{h^0}), k'(k_{h^0}, \omega_{h^0}), \omega'(\theta, k_{h^0}, \omega_{h^0})\}\) that satisfies (18), (20), (21), and

\[
\int_{\theta} \{ u(c(\theta, k_{h^0}, \omega_{h^0}), m(a(k_{h^0}, \omega_{h^0}))l(k)) + \beta \omega'(\theta, k_{h^0}, \omega_{h^0}) \} g(\theta|a(k_{h^0}, w \omega_{h^0}))d\theta = \omega_{h^0},
\]

\(0 \leq c(\theta, k_{h^0}, \omega_{h^0}) \leq \theta f(k) - k_{h^0} + (1 - \delta)k.\)

For \(t = 1,\) let \(a_1(h^0) = a(k_{h^0}, \omega_{h^0})\) and \(c_1(h^1) = c(\theta_1, k_{h^0}, \omega_{h^0}), \forall h^1.\) Also, let \(k'_{h^0} = k_{h^1} = k'(k_{h^0}, \omega_{h^0})\) and \(\omega_{h^1} = \omega'(\theta, k_{h^0}, \omega_{h^0}), \forall h^1.\) Notice that \((k_{h^1}, \omega_{h^1}) \in \Sigma \in B(\Sigma)\) implies the existence of \(\{a(k_{h^1}, \omega_{h^1}), c(\theta, k_{h^1}, \omega_{h^1}), k'(k_{h^1}, \omega_{h^1}), \omega'(\theta, k_{h^1}, \omega_{h^1})\}\) that satisfies (18), (20), (21), and

\[
\int_{\theta} \{ u(c(\theta, k_{h^1}, \omega_{h^1}), m(a(k_{h^1}, \omega_{h^1}))l(k)) + \beta \omega'(\theta, k_{h^1}, \omega_{h^1}) \} g(\theta|a(k_{h^1}, w \omega_{h^1}))d\theta = \omega_{h^1},
\]

\(0 \leq c(\theta, k_{h^1}, \omega_{h^1}) \leq \theta f(k) - k_{h^1} + (1 - \delta)k.\)

We can iterate for \(t = 2, 3, 4, \ldots\) to construct the complete profile \(\sigma.\) We can then observe that, by construction, for any arbitrary \(t \geq 0\) and \(h^t,\)

\[
\omega(\sigma|h^t) - \omega_{h^t} = \int_{\theta} \beta[\omega(\sigma|h^{t+1}) - \omega_{h^{t+1}}]g(\theta|a(k_{h^t}, \omega_{h^t}))d\theta_{t+1}
\]

Since \(0 < \beta < 1\) and the utilities are bounded, the above equation implies that

\[
\omega(\sigma|h^t) = \omega_{h^t} \quad \forall \ t \geq 0 \quad \text{and} \quad \forall h^t.
\]

Hence, the contract that we have constructed is what is desired. ■

**Proposition 4**  
(a) \(\Omega = B(\Omega).\) (b) Let \(X_0 = \Delta,\) and let \(X_{n+1} = B(X_n),\) for \(n = 0, 1, 2, \ldots\) Then, \(\lim_{n \to \infty} X_n = \Omega.\)

**Proof.** Part (a) is obvious. To demonstrate part (b), we will first show that the sequence \(\{X_n\}\) is convergent. Obviously, \(B(X_0) \subseteq X_0.\) Next, we operate \(B\) on both sides of this expression and obtain \(X_{n+1} = B(X_n) \subseteq X_n, \forall n,\) because \(B\) is monotone increasing. Hence, \(\{X_n\}\) is a bounded and monotone decreasing set sequence with \(X_\infty = \lim_{n \to \infty} X_n = \bigcap_{n=0}^{\infty} X_n.\) Now, we show that \(\Omega \subseteq X_\infty.\) Given that \(\Omega \subseteq X_0,\) the
monotonicity property of $B$ ensures that $B(\Omega) \subseteq B(X_0)$. However, it must be true that $\Omega = B(\Omega)$, by part (a), and $B(X_0) = X_1$, by construction. Then, $\Omega \subseteq X_1$. By iteration we obtain $\Omega \subseteq X_n$, $\forall n \geq 0$, and consequently, $\Omega \subseteq X_\infty$. Now, we demonstrate that $X_\infty \subseteq \Omega$. Given the properties of the sequence $\{X_n\}$, we have that $B(X_\infty) = X_\infty$. Hence, $X_\infty$ is self-generating, and $X_\infty = B(X_\infty) \subseteq \Omega$. □

Proposition 4 ensures that if we start with a set $\Delta$, and iterate on it using the operator $B$, we will converge to the set $\Omega$. Moreover, this set $\Omega$, by part (a) of Proposition 3, is a fixed point of the operator $B$.

In this section we have proved that our contracting problem admits a Bellman equation representation which can be solved by using the contraction mapping theorem. We also established the validity of the self-generation concept in our environment which provides an algorithm to compute the space of the feasible and incentive compatible values of the agent’s expected utility.

C Algorithm

References


