

# Empirical Content of a Continuous-Time Principal-Agent Model: The Case of the Retail Apparel Industry\*

Denis Nekipelov<sup>†</sup>

June 2010

## Abstract

There is a strong interest among corporate finance and industrial organization economists in estimating the effects of moral hazard and adverse selection in a variety of contractual settings. In this paper, I develop a methodology for the identification and non-parametric estimation of a continuous-time principal-agent model. My framework extends the existing literature on optimal dynamic contracts by allowing for the presence of unobserved state variables. To accommodate such heterogeneity, I develop an estimation method based on numerically solving for the optimal non-linear manager's response to the restrictions of the contract. Since my estimation methodology relies on observing optimal responses, as opposed to optimal contracts, it can be used to evaluate the optimality of particular executive contracts. To demonstrate this feature, I apply my methodology to executive contracts from the retail apparel industry. I then use the structural estimates of the contract parameters and the utility of the executive managers to evaluate the revenue and welfare effects of the introduction of the Sarbanes and Oxley act.

---

\*Corresponding author: Denis Nekipelov, Department of Economics, Duke University, e-mail: [denis.nekipelov@duke.edu](mailto:denis.nekipelov@duke.edu)

<sup>†</sup>I am deeply indebted to Han Hong, Paul Ellickson, Pat Bajari, Chris Timmins, and Curt Taylor for their help and continuous encouragement. I would like to thank all participants of the Microeconometrics and the Microeconomic Theory lunches at Duke, and the Econometrics Lunch at Stanford. The usual disclaimer applies.

# 1 Introduction

Empirical analysis of incentives provided by executive contracts is a practically important issue. In the empirical literature studying incentive contracts, it is a common practice to consider simple dependencies between components of contracts and characteristics of firm performance. In dynamic environments, this approach will yield unbiased predictions if both the firm owner and the manager observe current state variables, while the actions of the manager are unobserved by the owner. In the case where the manager has privately observed dynamic state variables, the optimal contract structure will be non-linear. In this paper, I argue that this result has the following implications. First, even if the owner of the firm uses a sub-optimal linear contract, the relationship between the contract parameters and the observable characteristics of the firm will be non-linear, while these characteristics become effectively endogenous. Second, endogeneity bias in this case cannot be corrected by conventional instrumental variable methods. This bias can lead to substantial errors in predicting the effects of analyzed characteristics on performance of the manager and, if these estimates are used for contract design, to sub-optimal contract structure.

To convey these arguments, I develop a continuous-time principal agent model with a dynamic state variable which is privately observed by the manager. In my model, motivated by managerial behavior in the retail apparel industry, I consider a single principal (a representative shareholder) and a single agent. State variables of the model include consumer demand and consumer tastes. Consumer demand is observed both by the principal and the agent. However, its dynamics are driven by consumer tastes, which are assumed to be observed only to the agent. This induces correlation between demand and consumer tastes even though random shocks in tastes and demand are uncorrelated. Consumer demand is also assumed to depend on the brand characteristic of the firm. The manager optimally sets this characteristic, exerting effort and decreasing her utility. The principal uses a linear contract which aims at setting incentives for the manager to set the brand characteristic optimally from the point of view of the principal. The optimal strategy of the manager will be a function of tastes and demand. Therefore, given that characteristics of firm's performance depend on consumers' tastes, they cannot be used as exogenous explanatory variables.

In this paper, I develop an estimation methodology which can alleviate this endogeneity problem. The method is based on explicit consideration of the manager's optimization problem and

using computed manager's best response to consistently estimate the incentive effects. I provide an empirical strategy to identify the model from the data and estimate its parameters using a non-parametric procedure, and adapt this methodology to a particular industry. The structural estimates can be used for the welfare analysis of both long and short-term effects of changes in legislation in the industries and for long-term predictions of the effect of these changes on the productivity of managers and firms' output. To illustrate my methodology, I use data from the ExecuComp dataset on executive contracts in the retail apparel industry. My structural estimates allow me to evaluate the welfare effect of the introduction of Sarbanes and Oxley act on the apparel retail industry.

The analytical procedure suggested in this paper proceeds in three steps. In the first step, I construct a continuous-time model to describe the manager's response to the contract incentives. I describe the structure of the state variables, information available to the principal and the agent, and the way the agent can influence the dynamics of the state variables. In practice, the analysis of contractual incentives should be industry-specific because the structure of observed and unobserved state variables can vary greatly by the industry, which can result in significant structural differences between contracts in different industries.

In the second step, I use the constructed structural model to design an estimation procedure to recover structural parameters for a particular industry. This step depends on the structure of the ExecuComp dataset which I use for empirical illustration. The estimation method used in the second step combines simulation from the model conditional on actually observed characteristics of the manager and the firm. It is based on comparing the simulated distribution of state variables with the actual distribution. I show the model which I study is non-parametrically identified. Simulating the model in the second step requires computation of the manager's best response. To fulfill this task, I develop a fast and reliable algorithm for solving for the best response of the manager and verify that the noise associated with simulations is infinitesimal as compared to the noise in the data.

In the third step, I illustrate the model by estimating its structural parameters and use the obtained structural estimates to analyze the effect of exogenous industry changes on profits and revenues of the firm under the presence of an agency problem.

For an empirical illustration, I use data for a particular industry: apparel retail. The sales

in this industry are driven by volatile consumer's tastes and display a significant amount of seasonality. The volatile structure of market tastes is frequently observable only by the manager (but not by the shareholders). This makes tastes an unobserved state variable. In the subsequent analysis I develop a model which is more general than Holmstrom and Milgrom (1987) to describe the executive compensation in the apparel retail industry. I estimate the model for ExecuComp data, and use the estimates to predict the effect on the industry due to the introduction of the Sarbanes and Oxley act (which imposed strict institutional restrictions on the actions of executive managers). Although I observe a behavioral response to the incentives provided in managerial contracts, a significant proportion of the compensation in the industry is in the form of fixed base salary. My analysis demonstrates a significant response to the introduction of the Sarbanes and Oxley act which increased legal manager's liability for the financial results of the firm. In particular, I find that the introduction of the act resulted in an increase both in the base salary and bonus parts of managers' compensation and has led to shifting the welfare loss towards the firm owners.

My paper fits broadly into the empirical literature on the economics of asymmetric information and, specifically, structural estimation of models of adverse selection and moral hazard. These issues have been studied, for instance in Abbring, Chiappori, Heckman, and Pinquet (Forthcoming), which deals with disentangling the effects of moral hazard and adverse selection in an equilibrium market setting. Lafontaine (1992) and Lafontaine and Shaw (1999) analyze the franchise decisions of the firms dealing with the effects of adverse selection in the market. Margiotta and Miller (2000) focuses on the cross-industry analysis of moral hazard in managerial compensation. The studies in Lemmon, Schallheim, and Zender (2000) and Callaghan, Saly, and Submaniam (2004) use linear models to estimate effect of contracts' incentives in the financial industry. My study differs substantially from the existing literature. First, from a theoretical perspective, I emphasize the importance of using a flexible dynamic model based on computed best responses to produce unbiased structural estimates. Second, I use a computational algorithm for the manager's best response and do not rely on its reduced-form specification. Third, my procedure can be used to produce counterfactual simulations directly. These features allow me to study dynamic incentives within the context of structural parameter estimation in which I nest the numerical solution for the best response of the manager into the estimation procedure.

My paper analyzes the role of dynamic incentives in contract design when the compensation of the manager's actions depends on her continuous - time performance by filling two gaps in the literature. First, although performance-based compensation for top management has been common for decades, it is unclear how close the structure of payments has been to the optimal contract structures. Second, as the structure of informational rent in the dynamic case is more complicated than that of the classical principal - agent problem, the structure of distribution of welfare associated with the distribution of informational rents has not been investigated in sufficient detail. In this paper, I address both issues and show for an example of contracts in the apparel retail industry, how one can produce the estimates of welfare effects from dynamic risk sharing and analyze how this effects change as the structure of incentive compensations changes.

The structure of this paper is as follows. In the first two sections, I develop a methodology for construction and estimation of a continuous-time model. Section 2 discusses the model and Section 3 discusses identification of the model and estimation technique. Section 4 then applies the developed methodology to estimation of the model for executives in the apparel retail industry. I also demonstrate how the obtained structural estimates can be used to produce the evaluations for welfare effects and changes in the production dynamics as a result of the introduction of the Sarbanes and Oxley act. Section 5 concludes.

## 2 A Continuous-Time Principal-Agent Model

In this section I develop a continuous-time principal-agent model. Using a particular structure for the state variables I specialize the model to the apparel retail industry.

### 2.1 Setup

In this section I present a continuous-time model of a managers's response to the contract incentives in the apparel retail industry. Similarly to the setup in Holmstrom and Milgrom (1987), I assume that the contract has fixed duration  $T$  and has to be renewed each period. The terms of the contract are designed by the owner of the firm (a representative shareholder). I simplify derivations in the analysis by assuming that the manager does not own a significant portion of the firm and thus relies only on compensation rather than on dividends.

The structure of the state variables in my model is significantly different from that in the classical setup of Holmstrom and Milgrom (1987). The state variable has two components. One component, which is denoted  $\theta_t$ , is idiosyncratic and represents consumers' tastes. The other component is the sales of the firm  $D_t$ , which the manager can control optimally. The dynamics of market sales is determined by the following equation:

$$dD_t = \psi(t, D_t, x_t - \theta_t) dt + \zeta(t, D_t, x_t - \theta_t) dB_t^D. \quad (1)$$

In this equation,  $D_t$  is the demand for the production of the firm at time  $t$ , where I assume that the firm does not experience stock-outs. The parameter characterizing the tastes of consumers in the market is denoted by  $\theta_t$ . The brand characteristic is denoted  $x_t$  which reflects the correspondence of the production variety offered by the firm to the variety of products demanded by consumers, and  $B_t^D$  is a Brownian motion. This form of demand dynamics reflects the cyclical structure of the ready-to-wear clothing market. The first component describes the deterministic trend driven by consumer tastes. The second component describes random deviations from the deterministic trend. Demand volatility increases if production variety offered by the firm does not reflect market demands.

There are two features which differentiate my model from the existing literature. First, I consider a state variable which is not observed by the principal and represents consumer's tastes. Second, I allow the agent to control both the drift and the diffusion of the state variable. In the presence of both features, the contract structure developed in Holmstrom and Milgrom (1987) is no longer optimal even in the class of linear contracts. While the first-best optimal contract in the presence of unobserved state variables is typically non-linear, I study optimality within a class of linear contracts which are computationally feasible and do not involve a numerically intractable problem with non-linear contracts. Some arguments in favor of the linear contract structure are given in Ou-Yang (2003), where the agent is also allowed to control the variance of the state variable.

The drift component of the demand  $\psi(t, D_t, x_t - \theta_t)$  is a function of time, as well as current demand volume  $D_t$  and a brand parameter  $x_t \in [0, 1]$ , which will be referred to as a "brand characteristic". It is assumed that  $\psi(t, D, z)$  is maximized at  $z = 0$  for any  $t \in [0, T]$  and  $D \in \mathbb{R}_+$ , that is, the deterministic component of demand is the highest if production fashionability completely reflects the consumer taste parameter. The function  $\psi(\cdot)$  reflects seasonal changes in

demand.

The diffusion component  $\zeta(t, D_t, x_t - \theta_t)$  is a function of the same set of variables that enter the drift component. The dependence of the diffusion term on the demand level reflects possibly higher fluctuations of demand in the firms with high volume of production. The dependence of the diffusion term on the difference between the market tastes for variety and the variety of production offered by the firm captures a probable increase in demand volatility if the brand characteristic of the firm is far from the market tastes.

Market demand in this paper is analyzed as a stochastic process. Such analysis implies a set of assumptions that I am imposing on both the demand dynamics and the industry structure. In particular, I assume that firms in the industry operate on a differentiated product market where each firm faces the demand for its particular market segment. Moreover, I assume that the demand on other market segments does not provide information relevant for predictions of consumers' behavior in a chosen segment. This assumption likely holds for the retail apparel industry, which exhibits a high degree of horizontal differentiation. Each firm in the apparel retail industry offers a mix of consumer products targeted at a particular niche of the consumers' market. I assume that the output of a firm can be expressed in terms of volume of a composite commodity. Moreover, the prices on the market correctly reflect the utility weights of particular products of the composite commodity. Therefore, the dollar value of firm's sales can be interpreted as the product of a price index and a quantity index. The price index reflects the unit value of the composite commodity. The quantity index reflects the quantity of the composite commodity where individual purchases are weighted by the utility indices. The brand characteristic of the firm then will represent the position of the firm in the differentiated product space. Therefore, the shares of purchases by a representative consumer from different firms will be driven by the relative locations of the preferences of the representative consumer and brand characteristics of firms. The demand of consumers for a particular brand is summarized by equation (1).

I consider the total sales of the firm rather than the physical volume of sales. While volume data on the apparel retail are available, they are problematic because, apart from the brand policy, the firms also have the price policy which can reflect the quality of the product. In fact the same model of clothes can be made of fabrics of different quality. The production cost and the price can differ significantly even across very similar pieces of clothing. Another important aspect is

in-season pricing. In particular, if after the introduction of a new product the firm discovers an unexpectedly high demand for the item, it can temporarily raise the price for this item to avoid stocking out before additional quantity is produced. The data for sales represents the value-weighted demand for the production of the firms and solves this problem. The tastes of consumers then determine the consumer's choice between both the quality and the price dimension of the production of the firm. Specified structure of sales of the firm reflects an additional assumption that consumers are generally price-takers and choose among a menu of objects on the market with different prices and qualities.

The consumers' taste parameter  $\theta_t$  is assumed to follow a driftless diffusion process:

$$d\theta_t = \gamma(t, \theta_t) dB_t^\theta. \quad (2)$$

The drift term represents a long-term dynamics in the tastes evolution which should not play a significant role in the decisions over a year although it might be important in the long-run. Moreover, the diffusion structure of equation (2) significantly simplifies the empirical analysis. The firms are assumed not to influence the dynamics of tastes. This equation suggests that the firm cannot influence consumers' tastes. This assumption is in line with modern research in apparel marketing and consumer psychology (such as, for instance Richardson (1996), Sproles (1981), Sproles (1985)), which suggest that modern apparel production is mostly consumer-driven.

The dynamics of tastes over time are driven by the change of fashion and fluctuations in the quality of output of a specific producer over time. For an individual consumer the movement of tastes changes her marginal utility of a specific brand and affects the demand. Managers of firms are different in their abilities of predicting the future values of  $\theta_t$ . The market share of the firm at a specific instant will be determined by the closeness of the firm's characteristic  $x_t$  to the consumer tastes characteristic  $\theta_t$ , given the price of the production of the firm relative to the prices of other firms.

There are two sources of asymmetric information in the agency problem of my model. First, I consider publicly traded firms with independent managers, in which there is a separation of ownership and control over the firm. Second, managers act in a volatile market environment and the exact financial outcomes of the firm at end of the contract period are unpredictable. To a large extent, the volatility of the market is caused by the heterogeneity of consumers across the market, and heterogeneity of aggregate consumer's tastes across time. With the pool of potentially

heterogeneous consumers, the market share of the firm at a specific time is determined by the price of the firm's output, the profile of tastes of the consumers in the market, and the profile of the stocks of the commodity accumulated by consumers. This consumer taste heterogeneity is associated with the randomness the financial outcome of the firm. A poor financial outcome in a specific period may be either the result of poor brand structuring policy by the manager or low seasonal demand for the product of the firm. The owner of the firm (who characterizes the representative shareholder) wants the manager to choose the best branding strategy to match the consumer tastes and to improve the market sales of the firm. Matching the market tastes is, however, costly to the manager because it requires a significant amount of market research each season to determine the optimal structure of the production mix. The lump-sum initial compensation for the manager will not create sufficient incentives for her to match the market tastes. The owner of the firm needs to design a contract for the manager to increase her effort level. However, the firm owner can only observe the continuous-time sales  $D_t$  and use it for contract design. The brand characteristic  $x_t$  as well as consumer's tastes  $\theta_t$  are not observed. The interpretation for this assumption is that: although the shareholders can observe the market sales of the firms from the market reports, to analyze the consumer tastes at a specific part of differentiated product market one needs to undertake costly market research.

I assume that the firm's instantaneous profit is a linear function of total sales  $\delta D_T$  where  $D_T$  is the cumulative sales at the terminal time  $T$ . The contract offered to the manager by the owner takes a linear form:

$$S(D) = \sigma(D_T) + \int_0^T \alpha(t, D_t) dt + \int_0^T \beta(t, D_t) dD_t, \quad (3)$$

so that the contract is defined by unknown functions  $\sigma(\cdot)$ ,  $\alpha(\cdot)$  and  $\beta(\cdot)$ , while  $S(D)$  is paid to the manager at the moment  $T$ . Most of the actual contracts used in practice have a linear structure because, first, it is tractable and, second, it is feasible for implementation. Moreover, a non-linear optimal contract schedule might not exist in general<sup>1</sup>.

Analyzing linear contracts both simplifies the computations and can approximate well the contract structure used in practice. The lump-sum part of the contract  $\sigma(D_T)$  characterizes the

---

<sup>1</sup>In some cases that the optimal incentive contract might suggest imposing infinite penalties on undesirable outcomes.

salary of the manager. The last two components characterize the incentive payouts (bonuses). The first part of bonus  $\int_0^T \alpha(t, D_t) dt$  sets incentives for the manager to achieve a high level of sales by the end of the contract period. The second part of bonus  $\int_0^T \beta(t, D_t) dD_t$  rewards the growth rate of sales during the contract period. This reward is related to risk aversion and can be justified by the loss aversion of the shareholders, who are concerned that higher growth numbers are better perceived by the stock market. Alternatively, a habit formation argument suggests that the shareholders have preferences for faster growing sales due to positive correlation between current share purchases and future share purchases.

In structure of the contract implied by equation (3) that the shareholders do not use the past values of the demand to evaluate the behavior of the manager retroactively. The manager therefore is not "rewarded" or "punished" for the behavior in the previous periods. Because of the diffusion assumption on the consumer tastes dynamics in my model past observations do not provide any basis recovering the future values of tastes. A more complex structure of the consumer tastes could potentially create problems with model identification because the tastes are not observed and I do not have a valid proxy for this variable. For these reasons I do not consider more complicated consumer tastes dynamics.

The manager is assumed to be risk-averse and derives utility from the monetary wealth of the manager minus the cost of effort on the job expressed in monetary terms. The effort of the manager will be called branding in this paper. Branding involves a decision about the production mix during the season and changes in the firm production between seasons. The instantaneous cost of branding is defined by  $c(x_t - \theta_t)$  so that the total cost for the manager is  $C(x) = \int_0^T c(x_t - \theta_t) dt$ . The function  $c(\cdot)$  is assumed to have the following properties:

- $c(\cdot)$  is symmetric so that  $c(-x) = c(x)$ ,
- $c(x) > 0$  for all  $x \in \mathbb{R}$ ,
- $\frac{\partial c(x)}{\partial x} < 0$  for all  $x > 0$ .

The branding cost  $c(x)$  derives from manager's effort to match the consumer tastes in the market which includes conducting the market research, organizing the work of a design team, making choices about the types and quality of fabrics, and pricing of products during the season to best target the demands of a specific category of consumers. These actions are costly to the manager

both in terms of the personal effort and in terms of the expenditures of the firm. In addition, effort can be costly because high production costs not offset by high sales can deteriorate the reputation of the manager. Given the contract structure and cost, the wealth of manager is defined as:

$$W_m = S(D) - C(x). \quad (4)$$

The manager is assumed to be risk averse with a utility function

$$U_m(W_m) = -\frac{1}{R_m} \exp\{-R_m W_m\}. \quad (5)$$

The wealth of the owner is the profit of the firm less the value of the contract offered to the manager:

$$W_o = \delta D_T - S(D). \quad (6)$$

The owner, similarly to the manager, is assumed to be risk - averse with the utility defined by

$$U_o(W_o) = -\frac{1}{R_o} \exp\{-R_o W_o\}. \quad (7)$$

The owner designs the contract to maximize the owner's utility over the contract space subject to the optimization problem of the manager given that the utility of the manager is not less than the reservation utility.

## 2.2 Manager's problem

The manager is assumed to be interested only in the final payout. Her utility is given by (5). The manager solves the problem of utility maximization by finding the optimal branding strategy for the firm given the contract and the dynamics of the demand and the consumer tastes.

The problem of the manager is to maximize her utility during the given period of time such that her wealth is a function of the sales of the firms and cost reflects effort of branding policy. The value function of the manager at time  $t$  is defined as the expected cumulative utility up to the end of the period given the information at time  $t$ . The value function of the manager at time  $t$  can be written as:

$$V(t, \theta_t, D_t) = -\frac{1}{R_m} \sup_{x_t} E_t \left\{ e^{-R_m \left( \sigma(D_T) + \int_t^T [\alpha(t, D_t) - c(x_t - \theta_t)] dt + \int_t^T \beta(t, D_t) dD_t \right)} \right\}. \quad (8)$$

The idea behind the optimization problem of the rational manager is that she will try to build a brand design strategy which will alleviate the non-systematic risk in the dynamics of her utility over time. The problem of the manager has the structure studied in the stochastic dynamic optimal control theory as in Gihman and Skorohod (1979). I derive the Bellman equation, which can be written in the form of a partial differential equation:

$$V_t + \sup_{x(\cdot)} \left\{ V_D (\psi - R_m \beta \zeta^2) + \frac{1}{2} V_{DD} \zeta^2 + \frac{1}{2} V_{\theta\theta} \gamma^2 - V R_m [\alpha + \beta \psi - c(x - \theta) - \frac{R_m}{2} \beta^2 \zeta^2] \right\} = 0. \quad (9)$$

In this equation, subscripts denote the corresponding derivatives of the value function. Equation (9) expresses the law of motion for the value function of the manager in terms of its derivatives over the state variable for each instant. The last term in the supremum shows that if the value function of the manager is constant in the state space, then it would be growing exponentially over time. The growth rate will increase if the "instantaneous" compensation of the manager increases and diminish if the costs or the variance of demand shock increases. This growth rate depends the risk aversion of the manager: if her risk aversion is sufficiently small, then an increase in the risk aversion increases the growth rate of value function, if it is very large, then an additional increase leads to the drop in the rate of growth of value function. The first term in the supremum reflects the influence of the marginal value of the manager with respect to the demand. If the demand is increasing over time rapidly enough, then a higher sensitivity of value function implies a lower increase in the value function over time. The second and the third term reflect the effect of the curvature of the value function and variances of shocks in the state variables on the dynamics. In general, a larger variance of shocks in the state variables leads to the slower growth in the value function. This means that the manager in this model "dislikes" uncertainty in the dynamics of state variable.

Intuitively, equation (9) can be obtained by equating to zero the systematic drift component in the stochastic Bellman equation for the manager. Informally, one can derive this condition by writing the Bellman equation for expected utility of the manager at time  $t$  in terms of utility at time  $t + dt$  and optimal actions. Taking the first-order condition, one can derive expression (9) using the rules of Itô calculus.

The structure of the optimal control problem suggests that at each instant of time the manager finds an optimal level of effort given the value of the taste parameter and realization of demand.

Then she recalculates the optimal policy at the next instant and use it to derive the continuation payoff given by the time derivative of value function. In this way the manager solves for the optimal effort at each value of the taste parameter, demand and time.

In the above derivation of the value function, the manager only maximizes the utility from the monetary payoffs. An implicit assumption here is that the managers do not have other objectives such as the size of the staff of the firm or the long-term value of the firm. This assumption is likely to hold in the industries with a significant control of firm owners over the top management.

### 2.3 Owner's problem

In the linear manager's contract, total compensation is formed by a fixed component and a stochastic component which depends on the dynamics of the firm's sales. In particular the contract structure might not only reward a high level of sales, but it also targets a specific growth rate of sales of the firm. The owner maximizes the expected utility derived from the share of the firm's profit minus the managerial compensation. Similarly to the manager, the owner has a CARA utility defined in (7). The owner's problem is to find the optimal parameters of the contract, including the salary and the bonus of the manager's contract, which maximize her expected utility at the end of the contract period<sup>2</sup>. Given the optimal response of the manager I rewrite the components of the contract to simplify the statement of the optimal contract design problem, that needs to be solved by the shareholders. In Ou-Yang (2003) it has been shown that the optimal linear contract structure is defined by the following expression:

$$S(D) = \overline{W}_m + \int_0^T c(x_t - \theta_t) dt + \frac{R_m}{2} \int_0^T \left( \beta - \frac{V_D}{R_m V} \right)^2 \zeta^2 dt + \int_0^T \left( \beta - \frac{V_D}{R_m V} \right) \zeta dB_t^D, \quad (10)$$

where  $\overline{W}_m$  is the manager's opportunity income. This contract is written in terms of the "effective" reward for demand growth  $\beta - \frac{V_D}{R_m V}$ . This effective reward will be lower if the value function of the manager is more sensitive to the changes in the market demand. Thus, the compensation will be more skewed towards the fixed salary payout. Such contract structure has a meaningful

---

<sup>2</sup>This is based on the assumption that I consider a competitive labor market for executive managers and thus, contracts can be signed again each period. In equilibrium with repeated contract signing under certain conditions the value function of the owner will depend only on the parameters of the current period.

interpretation. A non-idiosyncratic part of the contract is determined by the cost of the branding effort of the manager and the risk premium for the possibility of having low demand in spite of the high branding effort. This premium is defined by the conditional heteroscedasticity term in the consumer demand, contract compensation for the steady demand growth and the risk aversion of the manager. This formula suggests that more risk averse managers should have a higher compensation for the markets with high demand variance. The idiosyncratic compensation of the manager is determined by the stochastic demand shock and the compensation of the manager for the steady growth of the consumer demand. The optimal contracts depend only on the current but not past values of the consumer tastes parameter. This is due to the lack of path-dependence in the consumer tastes process which is driven by a diffusion process without a drift<sup>3</sup>. Therefore, the contract of the manager is determined only by the demand shocks and not by the taste shocks.

The value function of the firm owner can be written similarly to the value function of the manager in the form:

$$V^*(t, D_t) = -\frac{1}{R_o} \sup_{x, \beta} E_t \left\{ e^{-R_o \left( \delta D_T - \bar{W}_m - \int_t^T c(x_t - \theta_t) dt - \frac{R_m}{2} \int_t^T \tilde{\beta}^2 \zeta^2 dt - \int_t^T \tilde{\beta} \zeta dB_t^P \right)} \right\},$$

where  $\tilde{\beta} = \beta - \frac{V_D}{R_m V}$ . The equation for the owner's continuation payoff can be derived by applying Itô calculus to  $V^*(t, D_t)$ :

$$V_t^* + \sup_{x(\cdot), \tilde{\beta}(\cdot)} \left\{ V_D^* (\psi + R_o \zeta^2) + \frac{1}{2} V_{DD}^* \zeta^2 + V^* \left[ R_o c + \frac{R_o}{2} (R_o + R_m) \tilde{\beta}^2 \zeta^2 \right] \right\} = 0. \quad (11)$$

The problem of the owner resembles the problem of the manager. The optimal strategy of the owner is similarly computed in two stages. In the first stage, the owner calculates her preferred value of effort of the manager and a contract structure that can make the manager exert effort equal to the desired level. In the second stage, given the time derivative of her value function, the owner recomputes the value function at the next moment of time. This continuous-time backward induction allows me to compute the optimal contact structure along and the target level of effort.

One of the main assumptions here is that the firm owner knows the functional form of the demand process and the manager's cost so there is no asymmetric information about the manager's

---

<sup>3</sup>In the presence of fashion cycles one would expect that "tastes" will have a drift. However, if firms in the industry are aware of the existence of these cycles and take them into considerations, I can argue that I only model the responses to deviations from the long-term trend.

type. In the case where the manager’s type is unknown and her cost of effort can vary across types, then the contract structure can be more complicated. Then the contract will take into account additional ”truthful reporting” incentive constraints forcing the manager to behave optimally given her type. Comparing the optimal problem of the manager and the firm’s owner suggests that a higher cost of branding effort will decrease the growth rate of the value of the manager over time. It will also decrease the growth rate of the value of the firm owner over time, because the effort level of the manager will decrease. A similar effect is induced by risk aversion: a higher risk aversion reduces the growth rate of the value function over time for both the manager and the owner.

Together equations (9) and (11) describe the structure of the optimal dynamic contract between the principal and the agent. Equation (11) defines the optimal contract structure and the target level of effort from the point of view of the firm owner. Equation (9) defines manager’s effort given the contract structure. As was discussed above, the compensation and the bonus parts of the contract can be described in terms of the manager’s cost of the branding effort, the risk aversion parameter and functions describing the drift and the variance of the dynamics of consumers’ demand.

### 3 Identification and Estimation

In this section I develop an estimation strategy to estimate a continuous-time principal-agent model developed in the previous section. Before that, I investigate whether the model is non-parametrically identifiable given a sample of these variables of a sufficient size. The identification strategy is specialized for the data that I am using for structural estimation.

#### 3.1 Identification

In this section, I demonstrate that the theoretical model of dynamic optimal contract between the manager and the owner is non-parametrically identifiable given the structure of the data. Identification of the theoretical model of the manager’s behavior is a complex problem because the strategy of the manager is driven by the latent preference parameter  $\theta_t$ , which is unobserved both by the shareholders and by the econometrician. In the following I describe the general

structure of the data, outline the model assumptions, and discuss their role in the identification of the model given the data structure.

I assume that the information set of the econometrician is similar to that of the firm owner, who observes the demand process but not the consumer's tastes. The econometrician observes the demand realizations at the end of the period  $D_T$  and corresponding values of salary  $\sigma(D_T)$  and bonus  $S(D) - \sigma(D_T)$ . The entire path of demand and tastes during the contract period are not observed. The available data consist of the compensations (salaries  $W_{ik}$  and bonuses  $P_{ik}$ ) for a panel of managers  $i = 1, \dots, N$  observed over a collection of time periods  $k = 1, \dots, K$ . The available data include the sales of the firms for each period and a set of covariates consisting of the indicators of the firm performance for a specific period of time, and a set of individual-specific characteristics of the managers.

In the model the tastes parameter  $\theta_{ti}$  plays the role of the unobserved error term in the manager's decision problem and satisfies the following assumption:

**Assumption 1** *The taste parameter  $\theta_{ti}$  does not have a drift and is independent across managers  $i = 1, \dots, N$ . The compensation of the manager is only a function of the firm's sales and does not depend directly on the unobservable tastes.*

This assumption states that each firm is faced with the firm-specific tastes for a product on a differentiated market.

The unobserved dynamics of the tastes translates into the demand on the market for the production of a specific firm and subsequently affects the behavior of the manager. Given the dynamics of the tastes described by the measure of variance  $\gamma(\theta, t)^4$ , the manager solves the utility maximization problem given her risk aversion  $R_m$ , cost of effort  $c(x_t - \theta_t)$ , the parameters of demand  $\psi(\theta_t - x_t, t, D_t)$  and  $\zeta(\theta_t - x_t, t, D_t)$ , and the parameters of the contract  $\alpha(t, D_t)$  and  $\beta(t, D_t)$ . The optimal branding strategy of the manager  $x$  can be defined as a functional  $x(t, D_t, \theta, c(\cdot), \zeta(\cdot), \psi(\cdot), \alpha(\cdot), \beta(\cdot))$ . Note that the continuous-time structure of the model and Markov structure of state variables allow us to avoid considering their dependence from lagged variables. The dynamics of the sales will be a functional of the manager's actions. Therefore, the sales of the firm can be written as  $D(t, \theta, c(\cdot), \zeta(\cdot), \psi(\cdot), \alpha(\cdot), \beta(\cdot))$ . The following identification assumption is important and concerns the parameters of the contract.

---

<sup>4</sup>The variance of tastes at instant  $t$  will be determined by the integral of  $\gamma(\cdot)$  over time

**Assumption 2** For the set of covariates  $Z$ , functions  $\alpha(\cdot | Z)$  and  $\beta(\cdot | Z)$  given  $Z$ , which determine the wage and the bonus of the manager, are the same for all  $i = 1, \dots, N$

This suggests that the differences in the specific contracts can be attributed to the observable differences in the structure of the firm and features of managers. Conditional on this observed heterogeneity, this suggests that the contract structure is the same across the firms in the industry.

There are two approaches providing identification of the contract parameters. In the first approach I assume that the contract characteristics are homogeneous of degree 1 in time, and I assume that a large panel of data is available. Then, variation in the compensation across periods allows me to prove identification of contract parameters. In the second approach, I assume that the contracts are optimal from the point of view of the principal. As a result, the parameters of the contract will be functionally dependent on the basic features of the model: demand and cost function. The problem of identification of the contract parameters will reduce to identification of the demand and cost functions.

Consider identification of contract parameters from panel data. Assume that both  $\alpha(\cdot)$  and  $\beta(\cdot)$  are homogeneous of degree 1 in time. Using observations on the sales and covariates, functions  $\alpha(\cdot)$  and  $\beta(\cdot)$  can be estimated non-parametrically given covariates. Intuitively, for each data point  $i$  and  $k$ , I treat the values  $\alpha(t_k, D_{ki})$  and  $\beta(t_k, D_{ki})$  as unknowns. Moreover, I find the grid points for time as  $t_k = kT$ . Denote  $\tau_k = \frac{k}{K}$ . Then, writing the increase in the compensations as:

$$\Delta P_{ik} \approx \alpha(t_k, D_{ki})\Delta t + \beta(t_k, D_{ki})\Delta D_{ki} = \frac{T}{K} (\alpha(\tau_k, D_{ki})\Delta t + \beta(\tau_k, D_{ki})\Delta D_{ki})$$

I interpret the result as a system of  $NK$  equations with  $NK$  unknowns. In these equations  $\tau_k \in [0, 1]$  and represents the grid for time in the unit interval. Thus, one can recover the structure of the contract parameters from the data. Note that this result holds for any degree of homogeneity of contract parameters with respect to time.

It is not possible to simultaneously identify the scale of tastes separately from the functions  $\psi(\cdot)$  and  $\zeta(\cdot)$ . Consider the scale transformation  $\gamma(t, \theta) \mapsto Q\gamma(t, \theta)$  for  $Q > 0$ . Then I define the functions  $\bar{\psi}(t, \theta - x, D) = \psi(t, (\theta - x)/Q, D)$  and  $\bar{\zeta}(t, \theta - x, D) = \zeta(t, (\theta - x)/Q, D)$ . The demand equation corresponding to  $\bar{\psi}(\cdot)$  and  $\bar{\psi}(\cdot)$  remains exactly the same as before the transformation. A similar argument applies to a constant shift in the taste parameter. An additional scale assumption should be imposed on the time argument. Note that as the demand is observed only

as a cumulative quantity at the end of the auction, both scale and shift transformation change the shape of these functions without changing cumulative demand. For this reason I will impose either a scale or shift normalization. This leads us to the following location and scale identification assumption.

**Assumption 3** (i) Assume that contract parameters and demand parameters are either additively or multiplicatively separable in time in the form  $\psi(\cdot, t) = \varphi_1(\cdot) + \varphi_2(t)$  or  $\varphi_1(\cdot)\varphi_2(t)$ , where  $\varphi_1(\cdot)$  does not depend on time, while  $\varphi_2(0) = 0$ . In case of multiplicative separability an additional scale assumption is  $\varphi_2(1) = 1$ .

(ii) Assume that  $\gamma(t, 0) = 0$  and  $\gamma(t, 1) = 1$  so that only the shape of the term defining the heteroscedasticity of the tastes is defined but not its scale.

I consider here two possible kinds of data. The first kind are the panel data where manager and firm can be observed for several periods. The second kind of data contains i.i.d. observations of firms and managers over time. Identification arguments will be different for these two kinds of data. While the first kind of data allows us to identify the demand and parameters of the contract without imposing the condition of the contract optimality (11), the second kind of data requires this condition. The identification results can be summarized in the following theorem.

**Theorem 1** Suppose that assumptions 1-3 are satisfied. Then the following conditions identify parameters of the model for two possible kinds of data:

- If the data have panel structure where both dimensions asymptotically approach to infinity, then drift and diffusion terms in demand  $\psi(\cdot)$ ,  $\zeta(\cdot)$ , diffusion term of tastes  $\gamma(\cdot)$ , contact parameters  $\alpha(\cdot)$  and  $\beta(\cdot)$ , cost function  $c(\cdot)$  as well as risk aversion of the manager  $R_m$  are identified.
- If the data have panel structure then risk aversion of the firm owner  $R_o$  as well as fraction of profit in sales  $\delta$  can only be identified if the contract is optimal, i.e. its components solve (11).
- If the data have i.i.d. structure, then the components of the model  $\psi(\cdot)$ ,  $\zeta(\cdot)$ ,  $\gamma(\cdot)$ ,  $\alpha(\cdot)$ ,  $\beta(\cdot)$ ,  $c(\cdot)$ ,  $R_m$ ,  $R_o$ ,  $\delta$  are identified only if the contract is optimal, i.e. its components solve (11).

Note that this result basically states that for sufficiently rich data one can identify the parameters of the model without using the conditions of contract optimality. In these circumstances,

the problem of the owner adds extra constraints to the model. It is possible, therefore, both to test the validity of these constraints (and, thus, check the optimality of existing managerial contracts), and to build an overidentified model if these constraints appear to be valid.

The Brownian motion assumption restricts the dynamics of tastes and demand process to standard functionals of Brownian motion. This is an important assumption for model identification which simplifies the structure of optimal manager's response. A non-standard process for the demand and tastes dynamics is more complex. This is an area of future research.

The optimal contract mode described by assumptions 1-3 is similar to the class of non-linear models with non-separable errors that have been studied in recent literature. The role of the error term in my model is played by the taste parameter. Identification and estimation of general nonlinear non-separable models were described (among others) in Matzkin (2003) and Chesher (2003). Identification of my model is facilitated by the panel data structure of my dataset and the assumption that shocks in tastes and demand are uncorrelated. As a result, I can apply an appropriate technique applicable for non-separable models and estimate the parameters of the model.

### 3.2 Estimation

I use a conditional simulation strategy to estimate my model. The idea of estimation is to compare the joint distribution of the observed state variables with the joint distribution of the simulated state variables: salaries and bonuses of managers and firms' sales. Let us first elaborate the mechanism for recovering these joint distributions and then discuss the appropriate methods for their comparison.

The estimation proceeds in three steps. In the first step I flexibly estimate the joint distribution of salaries and bonuses of manager and sales of firms. This estimation is performed conditional on observable covariates. In the second step I simulate from random shocks and compute optimal behavior of the manager conditional on the observable covariates. The optimal strategy of the manager can be computed from the equation (9) describing the law of motion of manager's value function. Given the evolution of the taste shocks and demand shocks, I can generate the path of the firm's sales driven by the optimal strategy of the manager and, subsequently, compute bonuses and salaries of managers. This will produce the distribution of sales at each moment for

the considered set of firms.

In the third step, I match the distribution of the simulated state variables given the optimal behavior of the manager with the distribution of the observed state variables conditional on the covariates. To match these two distributions, I minimize a distance criterion with respect to parameters describing the structural functions of the model.

**Step 1.** For manager  $i$   $D_T^i$  denotes the total sales in the interval of length  $T$  (fiscal year) while  $W^i$  and  $P^i$  denote the total salary and bonus of the manager. For that period the joint distribution of these variables for the sample of independent observations can be evaluated, for example, using the kernel density estimator:

$$\widehat{f}_{\theta_0}(D, W, P) = \frac{1}{n h_D h_w h_p} \sum_{i=1}^n \kappa\left(\frac{D - D_T^i}{h_D}\right) \kappa\left(\frac{W - W^i}{h_w}\right) \kappa\left(\frac{P - P^i}{h_p}\right),$$

where  $\kappa(\cdot)$  is a kernel function and  $h_D$ ,  $h_w$  and  $h_p$  is a set of bandwidth parameters. The following set of assumptions are needed to assure consistency and asymptotic normality of the non-parametric estimates.

**A1** The four-dimensional realizations of the stochastic process  $\{D_\tau^i, W_\tau^i, P_\tau^i, \theta_\tau^i\}_{\tau \in [0, T]}$  are jointly independent across all  $i = 1, \dots, n$  for each  $\tau \in [0, T]$  and have density functions  $\varphi^\tau(D_\tau, W_\tau, P_\tau, \theta)$  for each  $\tau \in [0, T]$ .

**A2** The kernel function  $\kappa(\cdot)$  is continuous and bounded and satisfies

- $\int \kappa(z) dz = 1$
- $\int |\kappa(z)| dz < \infty$
- $\int z^2 \kappa(z) < \infty$

**A3** The three-dimensional random variable  $\{D^i, W^i, P^i\}_{i=1}^n$ , corresponding to the realization of the stochastic process  $\{D_\tau^i, W_\tau^i, P_\tau^i, \theta_\tau^i\}_{\tau \in [0, T]}$  at  $\tau = T$ , has a marginal density which is twice continuously differentiable in its arguments. The derivatives are continuous for  $\tau \in [T - \delta, T]$  for some  $\delta > 0$ .

**A4** As  $n \rightarrow \infty$ :  $h_w, h_D$ , and  $h_p \rightarrow 0$ ;  $n h_D h_w h_p \rightarrow \infty$ ;  $\sqrt{n h_D h_w h_p} (h_D h_w h_p)^2 \rightarrow 0$ .

Given these assumptions the estimate of the density function is converging pointwise to the

true density at the non-parametric rate and is asymptotically normal:

$$\sqrt{n h_D h_w h_p} \left( \hat{f}(D, W, P) - f_0(D, W, P) \right) \xrightarrow{d} N \left( 0, f_0(D, W, P) \left\{ \int_{-\infty}^{+\infty} \kappa^2(x) dx \right\}^3 \right)$$

The bias is negligible in the asymptotic distribution because of the undersmoothing assumption in **A.4**. As in standard kernel estimation theory, the consistency result  $\hat{f}(D, W, P) \xrightarrow{p} \int \varphi^T(D, W, P, \theta) d\theta$  relies on the twice-differentiability of the marginal joint density. Under the assumption that the bandwidth approaches zero as a function of  $n$ , the variance can be found by taking the expectation of the square of the individual term in the kernel density estimate.

**Step 2.** Conditional on the contract characteristics and the for each value of the structural parameter the distribution of observable state variables is obtained using the following sampling algorithm. Given a partition of time  $\{t_j\}_{j=1}^{N_\tau}$  where  $t_1 = 0$  and  $t_{N_\tau} = T$ , simulation increments of the Brownian motion as  $\Delta_\tau B_t^D \sim \sqrt{\tau} \epsilon_1$  and  $\Delta_\tau B_t^\theta \sim \sqrt{\tau} \epsilon_2$  where  $\epsilon_1$  and  $\epsilon_2$ , are independent standard normal random variables in each elements of partition. These draws are the used to recover the tastes and demand dynamics. While the Brownian innovations in the demand and tastes are independent, they are correlated due to function dependence. The paths of state variables are then computed conditioning on the observable covariates.

A naive simulation for the increments of Brownian motion-driven variable leads to the simulation error of order  $\sqrt{\tau}$ . However, in order to reduce simulation error I use a second-order Taylor-Itô approximation to improve simulation precision. Higher-order Taylor-Itô expansions are useful to obtain higher-order precision for solutions of partial differential equations, see for instance Kloeden and Platen (1992). For my purposes it is sufficient to consider a second-order approximation, leading to an error of order  $\tau$ . The dynamics of tastes can be then represented by recursion:

$$\theta_{t_j} = \theta_{t_{j-1}} + \gamma(t_j, \theta_{t_{j-1}}) \Delta_\tau B_{t_j}^\theta + \frac{1}{2} \gamma^\eta(t_j, \theta_{t_{j-1}}) \frac{\partial \gamma^\eta}{\partial \theta} \left[ \left( \Delta_\tau B_{t_j}^\theta \right)^2 - \tau \right]. \quad (12)$$

Using the tastes dynamics from equation (12), observed level of consumers' demand  $D_{t_j}$  at instant  $t_j$  given the tastes  $\theta_{T_j}$  optimal brand characteristic  $x_{t_j}$  is obtained by numerically solving equation

(9). Given the numerically computed brand characteristic, I simulate the sales at time  $t_{j+1}$  as:

$$D_{t_{j+1}} = D_{t_j} + \psi^\eta(t, D_{t_j}, x_{t_j} - \theta_{t_j}) \tau + \zeta^\eta(t, D_{t_j}, x_{t_j} - \theta_{t_j}) \Delta_\tau B_{t_j}^D + \frac{1}{2} \zeta^\eta \frac{\partial \zeta^\eta}{\partial D} \left[ \left( \Delta_\tau B_{t_j}^D \right)^2 - \tau \right]. \quad (13)$$

In equations (12) and (13)  $\eta$  specifies the parametrization of the functions of the model. The estimation procedure runs recursively up to instant  $T$  to obtain the simulated demand evolution and the total sales during the fiscal year  $D_T^{(s)}$ . Given the contract characteristics and the demand process, the simulated components of manager's compensation conditional on the observed covariates can be obtained as:

$$W_T^{(s)} = \sum_j \alpha(t_j, D_{t_j}) \tau,$$

and

$$P_T^{(s)} = \sum_j \beta(t_j, D_{t_j}) (D_{t_j} - D_{t_{j-1}}).$$

As the structural functions determining the evolution of the tastes and the demand process depend on the covariates, each simulated observation will correspond to one actual observation. The density of the simulated total sales of the firm as well as the components of the manager's contract can be again estimated using the kernel estimator:

$$\widehat{f}^{(s)}(D, W, P) = \frac{1}{n h_d h_w h_p} \sum_{j=1}^n \kappa \left( \frac{D - D_{jT}^{(s)}}{h_s} \right) \kappa \left( \frac{W - W_j^{(s)}}{h_w} \right) \kappa \left( \frac{P - P_j^{(s)}}{h_p} \right),$$

where  $\widehat{f}^{(s)}(D, W, P)$  reflects that it is computed from the simulated observations given the structural parameters  $\eta$ . In this expression I omit for brevity the component of the density reflecting conditioning on the observable covariates. The density for the structural parameter  $\eta$  is obtained as:

$$\widehat{f}_\eta(D, W, P) = \frac{1}{S} \sum_{s=1}^S \widehat{f}^{(s)}(D, W, P),$$

where  $S$  is the number of simulations per observation. The problem is now to compare the simulated distribution function  $\widehat{f}_\eta(\cdot)$  and the observed distribution function  $\widehat{f}_{\eta_0}(\cdot)$ . I then estimate

the model by finding the parameter  $\eta$  that minimizes the Hellinger distance between the observed and simulated distributions. Nekipelov (2007) uses the Kullback-Leibler divergence for estimation in similar circumstances. In this paper, I use those results and extend them to the estimator described here.

The proof of asymptotic normality of the structural estimates is provided in Appendix A.2. The proof starts with the analysis of simulation error connected with the approximation of the stochastic integrals by their discretized counterparts. In Appendix A.1 I show that if the step of discretization for simulation of the Brownian motion is decreasing fast enough, then the simulation error will be negligible. The methodology for simulations in equations (12) and (13) is essential and a similar order of approximation cannot be achieved using a naive sampling scheme. The results in Appendix A.1 show that under certain conditions the discretization error due to simulation is negligible and the only remaining source of error in the density estimation is the sampling noise. This means that the only source of error associated with the evaluation of the density of the joint distribution of sales, wage and bonus of the manager is a standard estimation error encountered in nonparametric density estimation. The estimated distribution will converge to the true distribution at a non-parametric rate. The asymptotic distribution for the simulated data then provides an accurate estimate for the actual distribution of state variables.

## 4 Analysis of incentives in the apparel retail industry

In this section I exemplify the modelling strategy that I develop in the previous sections to estimate structural parameters of executive contracts in the apparel retail industry. I then use the obtained estimates to analyze the effects of the introduction of the Sarbanes and Oxley act.

### 4.1 Data

I analyze the composition of compensation for the managers in the apparel retail industry. The data that I am going to use for analysis are collected from ExecuComp, a part of Standard & Poors' COMPUSTAT dataset, which contains detailed information about the structure of managerial compensations as well as the annual data on firms and compensation of executives. I only use the set of publicly traded companies in the ExecuComp dataset for three reasons.

First, it is likely that in the publicly traded companies there is more control of investors over the managerial performance. Second, the ExecuComp data are particularly interesting because it contains information that allows me to analyze the change in the structure of managerial contracts in response to the introduction of the Sarbanes-Oxley act. Lastly, the publicly traded apparel retail companies are more likely to adopt innovative production and distribution technologies. In addition, vertically vertically integrated production structure of these companies makes it easier to swiftly change the production structure in response to demand shocks.

The ExecuComp data for the apparel retail industry covers 26 apparel retail industry firms. For each firm it includes the compensation data for 7 executive managers per firm on average<sup>5</sup>. My sample includes the top 10 firms in the industry with the largest revenues. In addition to compensations, the dataset includes rich personal data for the managers as well as for the company. The subset of data used in this paper covers the time period from the year 1992 to the year 2005, which spans 12 years before the introduction of the Sarbanes-Oxley act and 2 years after the law became effective.

The company information includes a detail information for company sales, assets and other characteristics. The individual characteristics for manager contain standard demographic variables. The exact manager data available are: date when became CEO, date when joined the company, most recent title, rank by salary and bonus, gender, age, name. These characteristics are displayed in Table 1<sup>6</sup>.

Table 1 indicates that the average bonus is a large fraction of the base salary, indicating that the role of performance - based payment is significant in the total compensation of executive managers. Most demographic variables are available only for subset of the entire sample<sup>7</sup>. An exception is gender for which I control in the estimation procedure. The gender variable suggests that most executives are males. Females constitute only 21% of the total management. Moreover, the youngest observed age for the manager in the dataset is 42 and the oldest observed age is 78 with an average observed age of 54 years.

---

<sup>5</sup>I assume that there no unobserved heterogeneity in the sample, given the observed covariates, that creates no *ex ante* asymmetry in the data. Conditional on the executive's rank in the firm, it is reasonable to assume that managers solve similar tasks in firms with similar observable characteristics.

<sup>6</sup>The tables in this section are presented in Appendix B

<sup>7</sup>Some managers' gender can be inferred from their names and publicly available information source about the industry

Compensation data contain a wide range of components of the incentive contracts. Specifically, the data include the following items: salary, bonus, other annual compensations, option grants, amount of exercised options, restricted stock holdings (number of shares and value), options granted (number of shares and Black-Scholes value), long term incentive payout, and salary change per year. These variables provide the information about the available options and their prices, payments for participation at the executive boards, total cash paid to the managers and a list of other incentive payments such as available stock shares of the company. Summary statistics of the data are displayed in Table 2.

Table 2 shows that a significant amount of wealth is concentrated in different forms of options. In addition, there is a large proportion of unexercisable options as compared to exercisable options. The financial performance - based incentives for the managers provided by the option compensation are designed to maximize the performance of the market value of the firm.

## 4.2 Structural estimation

In this section I report the structural estimation results obtained from applying the estimation algorithm developed in the previous section to the data from the apparel retail industry. Observed heterogeneity in the apparel retail industry is controlled for by data about capital and asset structure of firms which includes the data on assets, market value of shares, profit, revenue from sales. The estimation procedure requires non-parametric estimation of the distributions of observed and simulated sales, bonuses and wages of managers, conditional on the covariates capturing cross-firm heterogeneity. The size of my dataset is not sufficiently large to make precise inference about joint densities of multiple variables. Therefore, I use a linear index  $\alpha'Z$  of conditioning variables to capture the dependence of the distribution of state variables from the observed heterogeneity:  $f(D, W, P | \alpha'Z)$ . The coefficient in the single index is estimated simultaneously with the structural parameters of the model.

The steps of the estimation procedure described in Section 3 are implemented as follows. For each of the data points, I simulated a pair of Brownian motions with the discretization step  $\tau = \frac{T}{100}$ . This pair of Brownian motions was fixed for each data point across the parameter search steps. For an initial "guess" of the parameters I evaluated the structural functions of the model: the linear index  $\alpha'Z$  in conditional distributions of observable state variables, the drift and the

diffusion components of the managerial contract, the drift and the diffusion part of the demand, the drift of the consumers' tastes, and the risk aversion parameter of the managers. Heterogeneous demand and tastes shocks across firms both capture the unobserved heterogeneity in the model. Additional heterogeneity in the demand structure will require additional assumptions about the distribution of random shocks. Introducing an additional unobserved heterogeneity parameter will further increase the computation cost of the model. In the current stage, the model is characterized by 37 parameters.

For each guess of parameter values and for each data point, the optimal manager's strategy is computed by numerically solving the partial differential equation on a square grid. The numerical partial differential equation solver uses the implicit finite difference algorithm, where the parabolic differential equation is transformed into a system of difference equations solvable in a linear number of steps. Equating the expected payoff of the manager at the end of the contract period to the final realized payoff at that time provides a boundary condition for the manager's problem.

Given the solution of the manager's problem, I obtained the optimal branding strategy for each data point. Using the pairs of Brownian motions simulated for each data point together with the optimal branding strategy I recovered the market demand, consumer's tastes, and compensations of the managers at each value of the structural parameters. The joint distribution of the simulated data is then evaluated using a multiplicative kernel density smoother. I obtain the structural estimates by minimizing with respect to parameter values the Hellinger distance between the distribution of the simulated data and the distribution of the actual data. Optimization is carried out using the MCMC, where I use the mean of 100,000 Monte Carlo draws for estimation and drop 10,000 of the initial draws.

I demonstrate in Section 3 that the optimal contract model is non-parametrically identified given a large enough dataset. In finite samples a flexible estimation strategy faces the bias-variance tradeoff. In particular, for this model it is computationally intensive to estimate because one needs to compute the optimal behavior of the manager. Thus, the model needs to be flexible but parsimonious enough to produce estimates in a reasonable time. In equations (14) - (19) I present specifications for the structural functions of the model which I use in the estimation procedure.

The drift of demand is specified as a linear function of time, demand and the distance between

the tastes and the brand parameter of the firm:

$$\psi(t, D, \theta - x) = a_0^\psi + a_1^\psi t + a_2^\psi D + a_3^\psi |\theta - x|. \quad (14)$$

The diffusion term is defined similarly to the drift term. This specification implies that the variance of demand shocks can both vary over time and can depend on the actions of the manager:

$$\zeta(t, D, \theta - x) = a_0^\zeta + a_1^\zeta t + a_2^\zeta D + a_3^\zeta |\theta - x|. \quad (15)$$

The dynamics of consumer's tastes is determined by the diffusion term which is specified as a function of time and a current value of the tastes parameter. To capture possible non-linear dependence of variance in tastes from the tastes parameter over time, I include the interaction term:

$$\gamma(t, \theta) = a_0^\gamma + a_1^\gamma t + a_2^\gamma \theta + a_3^\gamma t\theta. \quad (16)$$

The shape of the cost function reflects a particular feature of the model considered in this paper: the cost of matching the brand parameter of the firm with the consumers' tastes increases as the brand parameter approaches the tastes. In addition, the cost function is assumed to be symmetric in the distance between the brand parameter and consumers' tastes. Therefore, I specified the cost function as an inverse quadratic polynomial of the distance between the tastes and the brand parameter:

$$c(\theta - x) = \left( a_0^c + a_1^c |\theta - x| + a_2^c [\theta - x]^2 \right)^{-1}. \quad (17)$$

Finally, parameters of the contract are  $\sigma$  for fixed salary, and functions of demand and time for bonus. Both functions determining the bonus of the manager contain linear terms with time and price and an interaction term, reflecting that the dependence of bonus from the level of sales can change non-linearly over time. This produces:

$$\alpha(t, D) = a_0^\alpha + a_1^\alpha t + a_2^\alpha D + a_3^\alpha tD, \quad (18)$$

for compensation rewarding high level of sales and

$$\beta(t, D) = a_0^\beta + a_1^\beta t + a_2^\beta D + a_3^\beta tD, \quad (19)$$

for compensation rewarding sales growth. Since it is conventional to assume risk-neutrality of the representative shareholder, I chose to fix her risk-aversion parameter to be equal to zero. The risk-aversion parameter of the manager was estimated. Parameter estimates obtained from an MCMC estimation procedure are presented in Tables 3 and 4.

In estimation I recovered parameters of structural functions (14) - (19). The estimates of demand components show that demand is increasing within the period as indicated by the parameter  $a_0^\psi$ . If the brand parameter is further from the tastes of consumers, the growth rate is slower, as indicated by  $a_3^\psi$ . Coefficient  $a_3^\psi$  indicates that in order to achieve a high expected level of consumer demand at the end of the period the manager needs to match the consumer's tastes during the entire period. From parameter  $a_1^\zeta$  one can see that the variance of demand is decreasing during the period while the further the brand parameter is from consumers' tastes, the higher is the volatility of demand over time (indicated by positive value of parameter  $a_3^\zeta$ ). In total, demand parameters indicate that matching consumer tastes by the brand parameter of the firm the manager not only maintains high total sales of the firm but also reduces the volatility of sales within the period.

Parameters for the dynamics of tastes show a significant degree of heteroskedastisity in its behavior. In particular, parameter  $a_1^\gamma$  indicates decreasing variance of taste innovations over time, but positive  $a_2^\gamma$  leads to an increasing variance of taste innovations when the taste parameter is far from zero. The latter coefficient indicates that the stationary distribution of tastes has a substantial probability mass at the tails. As a result, prediction of future tastes becomes a complex statistical problem explaining high cost of tastes monitoring over time.

The cost function shows that it is increasingly costly to narrow the gap between the brand parameter and consumer tastes. Due to symmetric specification of the cost function, the maximum cost is achieved when tastes of consumers are perfectly matched by the brand parameter of the firm. The elasticity of cost with respect to the distance between tastes and the brand parameter linearly decreases as the distance increases.

Lastly, parameters for the characteristics of managerial contracts show that a significant response to the changes in the state variables is demonstrated in long-term compensation, which makes the manager put more emphasis on increasing the growth rate of sales at the end of the period (as indicated by  $a_1^\beta$ ).

Additional estimated parameters include the wage and the risk aversion parameters. For estimation purposes I normalized observed variables by equating the highest observed values of state variables to 100. Therefore, using the ratio of average total compensation to the maximal total compensation, the average relative risk aversion of the manager is equal to 7.4 which is close to existing estimates of risk aversions of stock managers.

The next set of results represent the parameters in the linear index as well as the normalizing constant for the contract and the risk aversion parameter of the manager. Estimates of the parameters in the single index show that the distributions of observable characteristics of the model are most affected by firms' market values, amounts of option grants awarded to the managers, shares owned by the manager and gender of the manager. These results suggest that structure of manager's compensation is most significantly differ by the size of the firm, the degree of firm ownership by the manager, and potential future ownership in terms of option grants.

### **4.3 Analysis of consequences of the Sarbanes-Oxley act**

#### **4.3.1 Overview of changes**

The Sarbanes-Oxley act (SOX) was passed by the US senate in July of 2002 as a response to a wave of corporate scandals involving large companies such as Enron and WorldCom. Before the SOX, some disclosure requirements on executive managers were set by the federal regulations while the rest were set by the state legislation. The provisions in the SOX set rigid restrictions on the actions of executive managers and the relation between corporations and auditors, and audit firms. Specifically, the SOX requires that the corporate audit must be conducted by an independent committee. The corporation is prohibited from buying non-audit services from the auditors. Corporate loans to the executive managers are prohibited. It is also required that the executive managers certify the corporate financial statement. In addition, the SOX requires that executive managers return the incentive-based part of the compensation from stock sales if the corporation reinstates earnings. In general, provisions of the Act impose more formal legal restrictions on the employment contracts of the executive managers.

The SOX has received a significant amount of criticism in the theoretical literature. Romano (2005) argues that the SOX was implemented during the corporate crisis and thus was not a result of multilateral objective consideration. For this reason, Romano (2005) suggests that the

SOX has set several unnecessary restrictions which do not prove to be effective in controlling malpractice by executive managers. Rather, it seems mostly to raise the compliance cost for the corporation. More importantly, Romano (2005) argues that the SOX has placed a relatively higher compliance cost burden on smaller corporations. Another concern reflected in Cohen, Dey, and Lys (2004) is that, due to the excessive restrictions in the SOX, the legislation might divert the decision of the managers from value-maximization. They can choose less risky projects in order to reduce personal risk exposure (deadweight loss from sub-optimal decisions). It may even lead corporations to switch to less public forms of ownership. The authors argue that, in response to increased managerial liability, the contracts of the managers should have a proportionally larger fixed component of the manager's compensation. Ribstein (2002) also criticizes the SOX, suggesting that the Act adds restrictions which do not set the correct incentives for the managers and cannot be considered an efficient instrument of prevention for corporate fraud. The general opinion in the literature is that the deadweight loss connected with the act's implementation and enforcement is very large (the estimates of losses by stock market investors in Zhang (2005) amount to \$1.4 trillion) while the benefits in terms of the risk reduction are limited.

The penalty imposed by the SOX leads managers to behave more cautiously in implementing policies. In the data for the years after the SOX was implemented I find evidence consistent with increased their effective risk aversions. In this section using the ExecuComp dataset for the years 2004 and 2005 when the SOX became effective I find the change in the risk aversion consistent with the observed change in the industry. In addition, I will use the dynamic model to predict the long-term effects of the SOX on the industry.

#### **4.3.2 Analysis of the SOX on the compensation structure**

The Sarbanes-Oxley Act imposes rigid restrictions on managerial performance resulting in substantial loss of compensation in case of undesirable economic outcome. Institutional changes associated with the introduction of the SOX can be modelled by considering an increase in the "effective" risk aversion amongst both managers and the investors. From the analysis of comparative dynamics of my model, I conclude that an increase in the manager's risk aversion should lead to an improved performance and, subsequently, to an increase in her compensation. For the Compustat data the effect of the SOX on managerial compensations can be illustrated by a

simple linear regression. In Table 5 I report the results of 2SLS regressions in which the dependent variables are the normalized sales of the firm (sales minus the average sales on the market normalized by the standard deviation of sales on the market in a specific year). Table 5 contains three sets of results. The first two sets represent the relationship between salaries and the bonuses of managers from the firm sales. The last set of results contains the estimates for the relation between the sales of the firm and firm's characteristics. In this last model I am concerned with possible endogeneity of such characteristics as the firm's value and I use relative variables such as capital structure of the firm as instruments. The data sample includes the firms doing retail sales business. As covariates I use the firm's market value, total assets, Tobin's  $q$ , and change in the assets within the year., with dummy variables for the firms specializing in the apparel retail. I also use a dummy variable for the years 2004 and 2005 when the SOX became effective. Table 5 demonstrates my estimation results for the system describing the salaries and bonuses of managers across firms in conjunction with sales of the firm. It shows that bonuses and salaries tend to grow with the normalized sales of the firm. More importantly, the introduction of the SOX led to an increase of \$27,624 in the base salaries and a \$43,494 raise in average bonuses of executive managers. An important economic question is whether compensation has been increased solely as a behavioral response to changing institutional environment or as a response of the structure of the industry to the introduction of the SOX. This question can be explicitly answered by my model. Specifically, the model can be used to predict the changes in the managerial compensation in response to changes in the manager's risk aversion. It is possible to find the change in the risk aversion that will generate the observed change in compensation. The new value of risk aversion will indicate the extent of managerial response which can explain the observed increase in the compensation.

### **4.3.3 Predictions for industry dynamics**

In the previous discussion I recover the effect of changes in the institutional environment on the "effective" risk aversion of executive managers. My continuous-time model of a manager in the apparel retail industry allows me to predict the effect of such an increase on the dynamic path of aggregate sales of the firm as well as the compensation package of the manager. Based on the simulated paths of the sales and the tastes parameter I find a change in the risk aversion which

matches the observed change in the compensation to the change in the compensation predicted by my structural model. I use mean-squared deviation of the simulated compensation from the average observed compensation as a criterion for search of the risk aversion and minimize this objective using a standard technique.

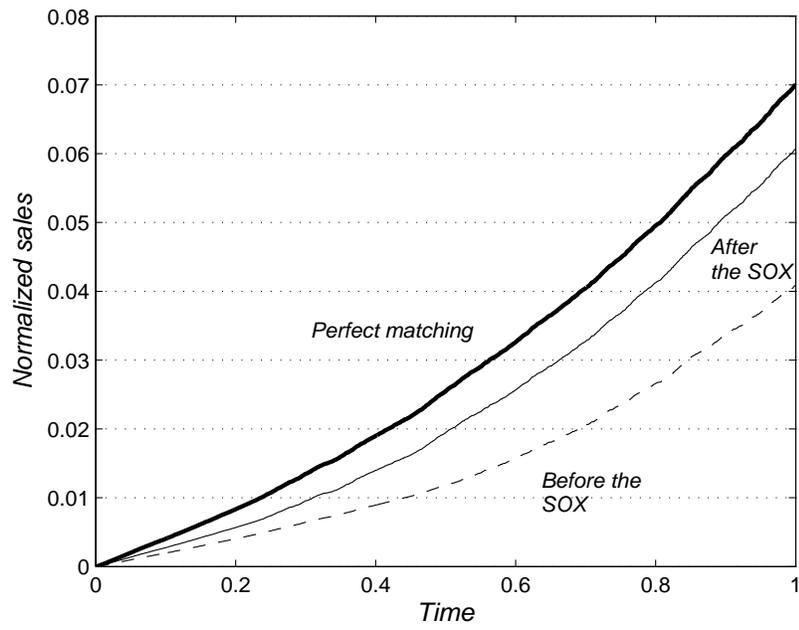
In simulations I assume that the risk aversion of the representative shareholder remains the same. The value of the absolute risk aversion of the manager that generates the observed change in the average managerial compensation is 4.02 which is 4 times bigger than the absolute risk aversion estimated on the data before the SOX. This is a very significant change in the risk aversion which was obtained holding the contract structure constant and it might have been a result of structural changes in the industry. However, observations from my data indicate that the profit of firms in the industry increases after the introduction of the SOX. Given that the period that I am considering is short, there was little incentive for the owners of the firm to reconsider the contract structure. My experiment in this case provides a lower bound for the effect of the introduction of the SOX on the long-term sales.

Next I simulate the effect of the estimated change in the absolute risk aversion on the dynamics of the sales of the firm. Simulations for the firm's sales dynamics are obtained directly from simulating (12) and (13) forward and then averaging over simulated paths. I used three cases to quantify the effect of Sarbanes and Oxley act on the sales of the firm. The first case, which was used as a reference point, is the case when the manager perfectly matches consumer tastes. On average, this case should yield the highest sales. The second case is the case of a risk-averse manager with the risk-aversion parameter obtained from the structural estimation. The third case reflects an increase in the effective risk aversion of the manager due to the introduction of the SOX. Figure 1 demonstrates the dynamics of average cumulative sales corresponding to the three cases. One can see that the case of perfect matching of consumer's tastes leads to the highest sales.

The case with the estimated risk aversion of the manager before the SOX leads to the smallest sales. One can also see that the effective increase in the risk aversion of the manager leads to an increase in sales, such that cumulative sales at the end of the period approach to the sales with perfect matching of tastes.

To summarize the simulation exercise, it is clear that the change in the effective risk aversion

Figure 1: Predicted cumulative sales of the firm



predicted from the structural model in response to the introduction of the Sarbanes and Oxley act is extremely large. In the model where contract structure remains the same after the legislative change the only variable affecting the amount of compensation is the effort of the manager. It is clear that the changes in the effort that would generate large increases in the compensation have to be induced by a large increase in the risk aversion. An alternative effect which can explain the observed compensation increase is the change in the contract structure induced by response in corporate structure and the structure of ownership of the firm. In general one could expect a combined effect as a result of changing both the effort of the manager and the structure of contracts. The period of observation of the post-SOX effect is relatively short, suggesting that the structure of the contract might not have completely reflected the institutional change. My predictions provide an explanation for one component of the overall response.

## 5 Conclusion

Modeling the dynamic principal-agent problem allows one to analyze the effect of incentives in the agent's contract on her long-run and short-run performance. In this paper I develop a methodology for implementing and structurally estimating a general continuous-time principal-agent model in the presence of the unobserved state variables. The non-linear response of the manager that I observe in my model extends the existing literature and it is more suitable for empirical implementation.

In this paper I provide conditions for identification of the continuous-time model from the data and I develop a methodology for empirical estimation of this model. Estimation methodology is based on matching the observed conditional distributions of state variables with the simulated distribution, for a particular set of structural parameters. Based on these distributions I use the Hellinger distance as a criterion for parameter estimation. I minimize the distance using a Markov Chain Monte Carlo technique.

I estimate the model for managerial contracts in the apparel retail industry. During the estimation procedure I recover the parameter of consumers' demand and characteristic of consumers' tastes. I also estimate the parameters of managerial contracts. I use the obtained estimates to analyze the effect of introduction of the Sarbanes and Oxley act. I find that the post-SOX compensations demonstrate a significant increase after the act. I assume that additional legislative

restrictions set by the SOX are increasing the effective risk aversion of the manager. Then I search for an increase in the risk aversion which could generate the observed increase in managerial compensation. Based on the estimated increase in the effective risk aversion I evaluate the effect on expected sales in the industry and I find that after the introduction of the SOX sales of a typical firm should increase. This analysis can be applied to other industries to perform welfare analysis and analysis of the output dynamics.

## References

- ABBRING, J. H., P.-A. CHIAPPORI, J. J. HECKMAN, AND J. PINQUET (Forthcoming): “Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish?,” *Journal of the European Economic Association, Papers and Proceedings*.
- ANDREWS, D. (1994): “Empirical Process Methods in Econometrics,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4. Elsevier Science.
- CALLAGHAN, S. R., P. J. SALY, AND C. SUBMANIAM (2004): “The Timing of Option Repricing,” *The Journal of Finance*, LIX(4), 1651–1676.
- CHESHER, A. (2003): “Identification of Nonseparable Models,” *Econometrica*, 71(5), 1405–1441.
- COHEN, D., A. DEY, AND T. LYS (2004): “The Sarbanes Oxley Act of 2002: Implications for Compensation Structure and Risk-Taking Incentives of CEOs,” *Working paper, Kellogg School of Management, Northwestern University*.
- GIHMAN, I. I., AND A. V. SKOROHOD (1979): *Controlled Stochastic Processes*. Springer-Verlag.
- HOLMSTROM, B., AND P. MILGROM (1987): “Aggregation and Linearity in the Provision of Intertemporal Incentives,” *Econometrica*, 55(2), 303–328.
- KLOEDEN, P., AND E. PLATEN (1992): *Numerical Solution of Stochastic Differential Equations*. Springer.
- LAFONTAINE, F. (1992): “Agency Theory and Franchising: Some Empirical Results,” *The RAND Journal of Economics*, 23(2), 263–283.
- LAFONTAINE, F., AND K. L. SHAW (1999): “The Dynamics of Franchise Contracting: Evidence from Panel Data,” *The Journal of Political Economy*, 107(5), 1041–1080.
- LEMMON, M. L., J. S. SCHALLHEIM, AND J. F. ZENDER (2000): “Do Incentives Matter? Managerial Contracts for Dual-Purpose Funds,” *Journal of Political Economy*, 108(21), 273–299.

- MARGIOTTA, M. M., AND R. MILLER (2000): “Managerial Compensation and the Cost of Moral Hazard,” *International Economic Review*, 41(3), 669–719.
- MATZKIN, R. (2003): “Nonseparable Estimation of Nonadditive Random Functions,” *Econometrica*, 71(5), 1339–1375.
- NEKIPELOV, D. (2007): “Entry Deterrence and Learning Prevention on eBay,” *Working paper*.
- OU-YANG, H. (2003): “Optimal Contracts in a Continuous-Time Delegated Portfolio Management Problem,” *Review of Financial Studies*, 16(1), 173–208.
- RIBSTEIN, L. (2002): “Market vs. Regulatory Responses to Corporate Fraud: A Critique of the Sarbanes-Oxley Act of 2002,” *Journal of Corporation Law*, 28(1), 1–67.
- RICHARDSON, J. (1996): “Vertical Integration and Rapid Response in Fashion Apparel,” *Organization Science*, 7(4), 400–412.
- ROMANO, R. (2005): “The Sarbanes-Oxley Act and the Making of Quack Corporate Governance.,” *Yale Law Journal*, 114(7), 1521–1613.
- SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- SPROLES, G. (1981): “Analyzing Fashion Life Cycles: Principles and Perspectives,” *Journal of Marketing*, 45(4), 116–124.
- (1985): “Behavioral science theories of fashion,” in *The Psychology of Fashion*, ed. by M. Solomon. Lexington Books.
- ZHANG, I. (2005): “Economic consequences of the Sarbanes-Oxley Act of 2002,” *Working paper*, *William E. Simon Graduate School of Business Administration, University of Rochester*.

# Appendix

## A Proofs

### A.1 Approximation properties of the simulation algorithm

Let us first analyze the precision of approximation of the variables  $D_T$ ,  $W_T$  and  $Z_T$  if the structural parameters in the simulations correspond to the structural parameters of the data generating process. Let

$x^*(t, D, \theta)$  be the optimal strategy of the manager given the time, market demand and consumers; tastes. Consider then the expectation of the difference between the simulated and the actual stochastic processes:

$$E \left\| D_T - D_T^{(s)} \right\|^2 = E \left\| \int_0^T \psi(t, D_t, x^* - \theta_t) dt + \int_0^T \zeta(t, D_t, x^* - \theta_t) dB_t^1 - \sum_{i=1}^{N_\tau} \left\{ \psi(t_i, D_{t_i}^{(s)}, x^* - \theta_{t_i})\tau + \zeta(t_i, D_{t_i}^{(s)}, x^* - \theta_{t_i})\Delta \widehat{B}_{t_i}^1 \right\} \right\|^2.$$

If the functions  $\psi(\cdot)$ ,  $\zeta(\cdot)$  and  $\gamma(\cdot)$  are continuously differentiable and bounded from above by some constant  $C$  in their domains then there exist constants  $L_i > 0$  such that:

$$\begin{aligned} & E \left\| \int_0^T \psi(t, D_t, x^* - \theta_t) dt + \int_0^T \zeta(t, D_t, x^* - \theta_t) dB_t^1 - \sum_{i=1}^{N_\tau} \left\{ \psi(t_i, D_{t_i}^{(s)}, x^* - \theta_{t_i})\tau + \zeta(t_i, D_{t_i}^{(s)}, x^* - \theta_{t_i})\Delta \widehat{B}_{t_i}^1 \right\} \right\|^2 \leq \\ & \leq L_1 \sum_{i=1}^{N_\tau} E \left\{ \sup_{t \in [t_i, t_{i-1}]} \|D_t - D_{t_i}^s\|^2 \right\} + L_2 \sum_{i=1}^{N_\tau} E \left\{ \sup_{t \in [t_i, t_{i-1}]} \|\theta_t - \theta_{t_i}^s\|^2 \right\} \leq \\ & \leq 2(L_1 + L_2) C^2 N_\tau \tau^2. \end{aligned}$$

For the uniform choice of the time grid  $N_\tau = T/\tau$  therefore,  $E \left\| D_T - D_T^{(s)} \right\|^2 \rightarrow 0$  as  $\tau \rightarrow 0$ . We can provide the similar conditions to assure the convergence of the simulated variables  $W_T^{(s)}$  and  $Z_T^{(s)}$ .

I obtained that the simulated paths of the considered stochastic process will have the moments close enough to the moments of the actual stochastic process. We should now establish the same result for the distribution of the simulated response and the actual process. For this reason consider the distribution function of the actual random variable  $\mathbf{P} \{D_T < \delta, W_T < \omega, P_T < \pi\}$  and the distribution of the simulated random variable  $\mathbf{P} \{D_T^{(s)} < \delta, W_T^{(s)} < \omega, P_T^{(s)} < \pi\}$  and consider the difference between these distribution in the fixed point  $(\delta, \omega, \pi)$ . For the distance between two distributions we can write the following chain of expressions:

$$\begin{aligned} & \left| \mathbf{P} \{D_T < \delta, W_T < \omega, P_T < \pi\} - \mathbf{P} \{D_T^{(s)} < \delta, W_T^{(s)} < \omega, P_T^{(s)} < \pi\} \right| = \\ & \left| E [\mathbf{1} \{D_T < \delta\} \mathbf{1} \{W_T < \omega\} \mathbf{1} \{P_T < \pi\}] - E [\mathbf{1} \{D_T^{(s)} < \delta\} \mathbf{1} \{W_T^{(s)} < \omega\} \mathbf{1} \{P_T^{(s)} < \pi\}] \right| \leq \\ & \leq \left| E [\varphi^\gamma(D_T, W_T, P_T) - \varphi^\gamma(D_T^{(s)}, W_T^{(s)}, P_T^{(s)})] \right|, \end{aligned}$$

where the function  $\varphi^\gamma(\cdot)$  approaches the indicator function as  $\gamma \rightarrow 0$  and is absolutely integrable and infinitely differentiable with respect to its arguments. Assuming that the support of  $\varphi^\gamma(\cdot)$  corresponds to that of the distribution of the random vector  $(D_T, W_T, P_T)$ , then the expectation of the derivative  $\varphi^\gamma(\cdot)$  has a finite limit. As a result we can evaluate the distance from above:

$$\begin{aligned} & \left| \mathbf{P} \{D_T < \delta, W_T < \omega, P_T < \pi\} - \mathbf{P} \{D_T^{(s)} < \delta, W_T^{(s)} < \omega, P_T^{(s)} < \pi\} \right| \leq \\ & \leq C_1^\gamma E \|D_T - D_T^s\| + C_2^\gamma E \|W_T - W_T^s\| + C_3^\gamma E \|P_T - P_T^s\| \leq \Lambda^\gamma \tau \sqrt{N_\tau}. \end{aligned}$$

The last inequality follows from the relationships established for each of the components of the considered stochastic process. The previous inequality follows from differentiability of  $\phi^\gamma(\cdot)$  and the fact that  $\int |\phi^\gamma(x, y, z)| dx dy dz < \infty$ . Thus, we have established the result that the distribution function of the simulated stochastic process approaches the distribution function of the observed stochastic process. If the distribution functions are assumed to be smooth, then the density  $\widehat{f}_\theta^{(s)}(D, W, P)$  is converging to the density of  $(D_T^{(s)}, W_T^{(s)}, P_T^{(s)})$ .

It has been also shown that the rate of the mean square convergence is at least  $\sqrt{\tau} \sim N_\tau^{-1/2-\psi}$ , for  $\psi > 0$ . As a result, the simulation error will be negligible as compared with the density estimation error if the size of the simulation grid  $N_\tau$  significantly exceeds the value  $nh_D h_w h_p$  determining the rate of non-parametric convergence of the density estimate.

**A7** There exists some  $\psi > 0$  such that  $N_\tau^{-\psi} nh_D h_w h_p \rightarrow 0$  as  $n \rightarrow \infty$ .

In fact we will need to provide the asymptotic distribution of the non-parametric estimate of the density of the simulated sample. We can provide the following chain of expressions:

$$\begin{aligned} & \sqrt{nh_D h_w h_p} \left[ \widehat{f}_\theta^{(s)}(D, W, P) - f_\theta(D, W, P) \right] = \\ & \sqrt{nh_D h_w h_p} \left[ \widehat{f}_\theta^{(s)}(D, W, P) - f_\theta^{(s)}(D, W, P) \right] + \sqrt{nh_D h_w h_p} \left[ f_\theta^{(s)}(D, W, P) - f_\theta(D, W, P) \right] = \\ & = \sqrt{nh_D h_w h_p} \left[ \widehat{f}_\theta^{(s)}(D, W, P) - f_\theta^{(s)}(D, W, P) \right] + \Lambda N_\tau^{-\psi} \sqrt{nh_D h_w h_p}. \end{aligned}$$

The last component of the presented expression vanishes due to the assumption **A7**. For the first component we can apply a standard result from the kernel density estimation. It suggests that under assumptions **A1** - **A6** the density estimate is asymptotically normal. As a result for the whole expression we can write:

$$\sqrt{nh_D h_w h_p} \left( \widehat{f}_\theta^{(s)}(D, W, P) - f_\theta(D, W, P) \right) \xrightarrow{d} N \left( 0, f_\theta(D, W, P) \left\{ \int_{-\infty}^{+\infty} \kappa^2(x) dx \right\}^3 \right),$$

given the assumption imposed on the bandwidth parameters and kernel function and an additional assumption that  $nh_D h_w h_p / N_\tau \rightarrow 0$  as  $n \rightarrow \infty$ .

One of the specific problems of the used approach is that the empirical density of the model response for different parameter values will be smooth with respect to the structural parameter only in the limit. In general, if we rely on the likelihood inference there is a need in computing the derivative of the density with respect to the structural parameters. To simplify the expressions denote  $x = (D, W, P)$  and let  $\varphi_h(\cdot)$  be the kernel function of argument  $x$ . Let  $x_i^{(s)}(\theta)$  be the elements of the simulated sample of model response. The density at point  $x$  has the standard expression  $\widehat{f}_\theta^{(s)}(D, W, P) = \frac{1}{n} \sum_{i=1}^n \varphi_h(x - x_i^{(s)}(\theta))$ . It

is appropriate to define the derivative of the density with respect to the structural parameter as<sup>8</sup>:

$$\frac{\partial \widehat{f}_\theta}{\partial \theta} = \lim_{\delta \rightarrow 0} \frac{\widehat{f}_{\theta+\delta}^{(s)}(D, W, P) - \widehat{f}_\theta^{(s)}(D, W, P)}{\delta}$$

Considering the limiting behavior of this sum. First, take the expectation of the individual term assuming that the density function is three times differentiable:

$$E \left\{ \frac{\varphi_h(x - x_i^{(s)}(\theta + \delta)) - \varphi_h(x - x_i^{(s)}(\theta))}{\delta} \right\} = \frac{\partial f_\theta(x)}{\partial \theta} + \frac{1}{2} \frac{\partial^2 f_\theta(x)}{\partial \theta^2} \delta + \frac{1}{2} \frac{\partial^3 f}{\partial x^2 \partial \theta} h^2 \delta + o(\|\delta h^2\|).$$

This expression suggests that the bias due to discretization of the derivative will be of order  $\delta$ . Note again that discretization error is irreducible because the derivative needs to be computed for the samples of the finite size. Therefore, such error will be present even if the sample size approaches infinity. The variance of the individual term can be obtained as:

$$V \left( \frac{\varphi_h(x - x_i^{(s)}(\theta + \delta)) - \varphi_h(x - x_i^{(s)}(\theta))}{\delta} \right) = \frac{2}{\delta^2} f_\theta(x) \int h \varphi_h^2(u) du - \frac{2}{\delta^2} [f_\theta(x) \int \varphi_h(u) du]^2 + o_h(\|\delta^{-2}\|).$$

As the standard CLT applies under the independence assumption, given that  $n h_D h_w h_p \rightarrow \infty$  while  $h_D h_w h_p \rightarrow 0$ :

$$\sqrt{n h_D h_w h_p} \left( \frac{\partial \widehat{f}_\theta}{\partial \theta} - \frac{\partial f_\theta(x)}{\partial \theta} - \frac{1}{2} \frac{\partial^2 f_\theta(x)}{\partial \theta^2} \delta \right) \xrightarrow{d} N \left( 0, \frac{2}{\delta^2} f_\theta(x) \left[ \int_{-\infty}^{+\infty} \kappa^2(u) du \right]^3 \right).$$

This expression illustrates that the attempts to reduce the bias due to discretization lead to the significant increase in the variance of the derivative estimate. A similar result holds when instead of kernel estimation one uses sieves to compute the density function. In that case the bias and the variance of density derivative with respect to the structural parameters will be inversely related.

The main conclusion from these derivations is that the brute-force gradient method is likely to fail in my case if one uses it to find the structural parameters. This also means that the expansions for the sample loss functions similar to those used to derive the asymptotic properties of M-estimators cannot be obtained for the simulation-based estimation proposed in this paper. A possible solution to this problem is to consider differentiability in  $\mathbf{L}^2$  sense which will allow one to obtain correct asymptotic expansions under much lighter conditions.

---

<sup>8</sup>The major problem here is that the size of the response sample will always coincide with the size of the dataset because the simulations are performed using the conditioning on the observation-specific covariates. This suggests that the specified limit does not necessarily exist for finite sample sizes or sequence of sample sizes approaching infinity. For the practical purposes then it is reasonable to suggest some finite  $\delta$  and consider the finite difference as an approximation of the function derivative.

## A.2 Hellinger distance and asymptotic estimates of structural parameters

Assuming that the structural parameters of the model are contained in some compact set  $\Theta \subset \mathbb{R}^k$  we can assure the convergence rate  $\delta(n)$  for the density estimate of  $(D, W, Z)$  pointwise for each  $\theta \in \Theta$ . The reason for using the concept of Hellinger distance is that it allows one to provide the structure of the asymptotic behavior of the econometric model even in the cases where the density function approaches zero and so log-likelihood of the model approaches infinity. To extend the class of the analyzed functions even further the authors consider the concept of Hellinger differentiability which requires the square root of the density function to be differentiable in  $\mathbf{L}^2$  norm. Note that under this definition "problematic" density functions such as those similar to the Epanechnikov density are differentiable even though they do not have a pointwise derivative on the entire support. Such concept can be useful in the semiparametric procedures because it allows one to avoid the problems with non-differential density functions when, for instance, B-splines of low order are used for estimation.

For probability measures  $\mathbb{Q}$  and  $\mathbb{P}$  the squared Hellinger distance has the expression:

$$H^2(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \int \left( \sqrt{d\mathbb{P}} - \sqrt{d\mathbb{Q}} \right)^2$$

If the likelihood ratio is defined  $\mathcal{L} = \frac{d\mathbb{Q}}{d\mathbb{P}}$  then the Hellinger distance can be rewritten as:

$$H^2(\mathbb{P}, \mathbb{Q}) = 1 - E_{\mathbb{P}} \left\{ \sqrt{\mathcal{L}} \right\}.$$

Let us adopt this definition for my case so that let  $\hat{f}_{\theta_0}(\cdot)$  be the density of the data and  $\hat{f}_{\theta}^{(s)}(\cdot)$  be the density of the simulated response. For notational simplicity denote  $x = (D, W, Z)$ . Then for the empirical density measures we can write the expression for the Hellinger distance as:

$$\hat{H}_n(\theta, \theta_0) = 1 - \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{\hat{f}_{\theta}^{(s)}(x_i)}{\hat{f}_{\theta_0}(x_i)}}$$

Consider the Hellinger expansion of the distance for  $\xi_i(\theta, \theta') = \sqrt{\hat{f}_{\theta}^{(s)}(x_i)/\hat{f}_{\theta'}(x_i)}$ . In this case the density expansion in  $\mathbf{L}^2$  sense can be determined as:

$$\sqrt{\hat{f}_{\theta}(x)} = \sqrt{\hat{f}_{\theta_0}(x)} + \frac{1}{2}(\theta - \theta_0)' \Delta(x) \sqrt{\hat{f}_{\theta_0}(x)} + r_{\theta}(x).$$

Note that  $\hat{f}_{\theta}(x)$  is differentiable in  $\mathbf{L}^2$  sense and the same expansion can be written for it. Therefore, the Hellinger distance can be represented as:

$$\begin{aligned} \hat{H}_n(\theta, \theta_0) = 1 & - \\ & - \frac{1}{n} \sum_{i=1}^n \frac{\sqrt{\hat{f}_{\theta_0}^{(s)}(x_i) + \frac{1}{2}(\theta - \theta_0)' \Delta_1^{(s)}(x_i) \sqrt{\hat{f}_{\theta_0}^{(s)}(x)} + \frac{1}{4}(\theta - \theta_0)' \Delta_2^{(s)}(x_i)(\theta - \theta_0) \sqrt{\hat{f}_{\theta_0}^{(s)}(x)} + r_{\theta}(x)}}{\sqrt{\hat{f}_{\theta_0}(x_i)}}. \end{aligned}$$

As  $\widehat{f}_\theta(\cdot)$  is asymptotically normal, we can find a random variable  $\xi(x)$  and a parameter  $\eta(n) = (n h_d h_w h_z)^{-1}$  so that  $E[\xi^2(x)] < M$  for some positive constant  $M$  and  $\widehat{f}_{\theta_0}^{(s)}(x) = f_{\theta_0}(x) + \xi(x)\eta(n)$ , and  $\widehat{f}_{\theta_0}(x) = f_{\theta_0}(x) + \widetilde{\xi}(x)\eta(n)$ . Substituting these expressions into the expression for the Hellinger distance:

$$\widehat{H}_n(\theta, \theta_0) = \frac{1}{2}(\theta - \theta_0)' \frac{1}{n} \sum_{i=1}^n \Delta_1^{(s)}(x_i) + \frac{1}{4}(\theta - \theta_0)' \frac{1}{n} \sum_{i=1}^n \Delta_2^{(s)}(x_i)(\theta - \theta_0) + \frac{\Lambda_n \eta(n)}{2},$$

where  $E[\Lambda_n^2] < \infty$ . Now we can analyze this transformed distance function to obtain the asymptotic behavior of the minimum distance estimates. For this purpose we can use the results from the theory of empirical processes, for instance, from Andrews (1994).

Consider a single realization of the stochastic process  $\{x_t^{(k)}\}_{t \in [0, T]} = \{D_t^{(k)}, W_t^{(k)}, Z_t^{(k)}\}_{t \in [0, T]}$ . For this realization one can write:

$$\sum_{i=1}^{N_\tau} \sqrt{\left[ \frac{\widehat{f}(x_i^{(k)})}{\widehat{f}_\theta(x_i^{(k)})} \right]} = \int_0^T \sqrt{\left[ \frac{\widehat{f}(x_t^{(k)})}{\widehat{f}_\theta(x_t^{(k)})} \right]} dP_{N_\tau},$$

where  $P_{N_\tau}(\cdot)$  is the empirical measure, generating the discretized stochastic process  $\{x_t^{(k)}\}_{t \in [0, T]}$ . Let  $\xi_1$  and  $\xi_2$  are nuisance parameters in non-parametric estimates of  $f_{\theta_0}(\cdot)$  and  $f_\theta(\cdot)$ . Let us denote the latter stochastic integral by  $J^{(k)}(\theta, \xi_1, \xi_2)$ .

Note now that the Hellinger distance can be written as

$$\widehat{H}(\theta_0, \theta, \widehat{\xi}_1, \widehat{\xi}_2) = 1 - \frac{1}{n} \sum_{k=1}^n J^{(k)}(\theta, \widehat{\xi}_1, \widehat{\xi}_2) = 1 - J_n(\widehat{\theta}, \widehat{\xi}_1, \widehat{\xi}_2).$$

Note that the structure of the estimator is similar to that of M-estimators, but the deterministic moment condition in this case is substituted by a stochastic moment condition. Under correct specification the estimator for  $\theta_0$  satisfies:

$$J_n(\widehat{\theta}, \widehat{\xi}_1, \widehat{\xi}_2) - 1 = 0$$

with probability 1 as  $n \rightarrow \infty$ .

To prove the consistency and derive asymptotic properties of the obtained estimator we follow first the reasoning in Andrews (1994). In fact note that one can write down the mean value expansion as:

$$o_p(1) = J_n(\widehat{\theta}, \widehat{\xi}_1, \widehat{\xi}_2) - 1 = \sqrt{n} J_n(\theta_0, \widehat{\xi}_1, \widehat{\xi}_2) + \frac{\partial}{\partial \theta'} J_n(\theta^*, \widehat{\xi}_1, \widehat{\xi}_2) \sqrt{n} (\widehat{\theta} - \theta_0) - 1$$

As the integral over empirical measure converges to the integral over the Wiener measure:

$$\frac{\partial}{\partial \theta} J_n(\theta^*, \widehat{\xi}_1, \widehat{\xi}_2) \xrightarrow{p} -Q = -E \left\{ \int_0^T \frac{\partial}{\partial \theta} \sqrt{f(x_t \theta_0)} dP_t \right\}$$

The latter integral is finite as long as the distribution has a mean square expansion of its density.

The expression for the parameter under consideration can be written as:

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= Q^{-1}\sqrt{n}J_n(\theta_0, \hat{\xi}_1, \hat{\xi}_2) + o_p(1) = \\ &= Q^{-1}\left[\sqrt{n}\left(J_n(\theta_0, \hat{\xi}_1, \hat{\xi}_2) - J_n^*(\theta_0, \hat{\xi}_1, \hat{\xi}_2)\right) + \sqrt{n}J_n^*(\theta_0, \hat{\xi}_1, \hat{\xi}_2)\right],\end{aligned}\tag{20}$$

where  $J_n^*(\theta, \xi_1, \xi_2) = \frac{1}{n} \sum_{k=1}^n E\{J^{(k)}(\theta, \xi_1, \xi_2)\}$ .

As  $\xi_1$  and  $\xi_2$  are nuisance parameters in the non-parametric density estimation, it has been shown in above that, similarly to standard kernel estimators as in Silverman (1986) such estimates are pointwise asymptotically normal. Using Fubini theorem one can argue that

$$\text{var}_\xi \left[ J^{(k)}(\theta_0, \hat{\xi}_1, \hat{\xi}_2) \right] = E \left\{ \int_0^T \text{var}_\xi \left[ \sqrt{\left[ \frac{\hat{f}(y_t, t)}{\hat{f}_{\theta_0}(y_t, t)} \right]} dP_t \right] \right\}.$$

The existence of this variance is justified by the existence of finite variances of  $\sqrt{\hat{f}(\cdot)}$  and  $\sqrt{\hat{f}_\theta(\cdot)}$  which is guaranteed by the existence of the mean square expansion for the square root of the density. In particular, if the simulated sample is independent from the actual sample of trajectories we can write the asymptotic expression for the variance given that the density estimates are obtained from the kernel smoother with a kernel function  $\kappa(\cdot)$  given assumptions A4 and A5 as:

$$\text{var}_\xi \left[ \sqrt{nh_t h_y} J_n^*(\theta_0, \hat{\xi}_1, \hat{\xi}_2) \right] \xrightarrow{n \rightarrow \infty} \left( \int_0^\infty \kappa^2(\psi) d\psi \right)^2 E \left\{ \int_0^T \frac{dP_t}{\sqrt{f(y_\tau, \tau, \theta_0)}} \right\} = \Omega_\xi$$

Note that this variance is only driven by the variance in the estimation of joint density but not by the stochastic process *per se*. The reason for this is that if we could have a perfect estimate of the distribution of the process  $x$  and the timing of its jumps, then the Hellinger distance under true parameter values will be equal to zero and thus the variance of the corresponding stochastic integral would be equal to zero as well. The only source of variance in  $J_n^*$  is therefore the error in non-parametric density estimate.

This integral exists as the considered Wiener measure has a Radon-Nykodim density.

We can further write then that:

$$\sqrt{nh_t h_y} J_n^*(\theta_0, \hat{\xi}_1, \hat{\xi}_2) \xrightarrow{d} N(0, \Omega_\xi).$$

Denote now  $v_n(\xi) = \sqrt{n}(J_n(\theta_0, \xi_1, \xi_2) - J_n^*(\theta_0, \xi_1, \xi_2))$ . Under true  $\xi = \xi^0$  as  $J^{(k)}(\cdot)$  are independent while  $v_n(\xi^0) \equiv 0$ .

Assuming that the mean square expansion of the density can be evaluated with the Lipschitz constant  $\Lambda$ , then:

$$E \left\{ \sup_{\|\xi - \xi'\| \leq \delta} \left| J^{(k)}(\theta_0, \xi') - J^{(k)}(\theta_0, \xi) \right|^2 \right\} \leq 4\Lambda\delta^2,$$

by Doob's inequality. Then given mean - square convergence of non-parametric estimates of density under  $\mathbf{L}^2$  norm, according to Andrews (1994),  $v_n(\cdot)$  is stochastically equicontinuous in  $\xi$  which implies that  $v_n(\hat{\xi}) - v_n(\xi^0) \xrightarrow{p} 0$ .

This assures that the estimate  $\hat{\theta}$  minimizing the Hellinger distance is asymptotically normal and:

$$\frac{1}{\sqrt{\eta(n)}} (\hat{\theta} - \theta_0) \xrightarrow{d} N(0, Q^{-1} \Omega_\xi Q^{-1}),$$

where  $Q = \frac{\partial}{\partial \theta} E \left\{ \sqrt{\hat{f}_{\theta_0}(x_i)} \right\}$  and  $\Omega_\xi = \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta'} E \left\{ \sqrt{\hat{f}_{\theta_0}(x_i)} \right\}$ .

Note that to make this expansion one does not need differentiability of the estimated density with respect to the structural parameters. The imposed requirement concerns the differentiability of the expected estimate of the density with respect to the structural parameter. The existence of this derivative is ensured by the existence of the derivative of the true density function  $f_\theta(\cdot)$  with respect to parameter  $\theta$  which seems to be a reasonable requirement. Moreover, even this assumption can be weakened by only requiring that the integral  $\int \{f_\theta(x)\}^{3/2} dx$  is differentiable with respect to  $\theta$ . This assumption is likely to hold even in the non-regular cases when the support of the distribution  $f_\theta(\cdot)$  depends on the parameter or when the density has first-order discontinuities.

## B Tables

Table 1: General Data for Executive Managers

variable	N. obs.	mean	stdev	min	max
Salary (thous. \$)	1851	430.7159	305.2276	.001	2279.138
Bonus (thous. \$)	1851	268.1957	500.8492	0	5672.5
Executive age	651	54.4424	7.5858	42	78
Gender (Female=1)	1851	.2106969	.4079137	0	1

Table 2: Some Characteristics of Compensations (thousands \$)

variable (values in thous. \$)	N. obs.	mean	stdev	min	max
Total compensation	1851	698.8374	733.2091	.001	7826.584
Restricted stock grants	1851	235.2837	1504.514	0	40250
Restricted stock holdings (value)	1630	44.65941	258.5092	0	4500
Options granted (B-S value)	1851	122.6672	422.6474	0	5000
Long term incentive payouts	1851	17.35361	146.881	0	3484.613
Options exercised	1630	654.7266	3015.383	0	88491.23
Unexercised exercisable op.	1630	233.4002	900.3827	0	14523.19
Unexercised unexercisable op.	1630	373.8253	1322.936	0	19629.75
Tot. compens. change (% per year)	1501	454.6537	15378.86	-92.954	595551.3

Table 3: Structural estimates of demand and contract parameters

Parameter	Estimate	Std. deviation
<i>Drift in demand</i>		
$a_0^\psi$	2.929683	1.610359 **
$a_1^\psi$	-0.1407972	0.7875199
$a_2^\psi$	-0.031096	1.410139
$a_3^\psi$	-2.976491	0.6215653 ***
<i>Diffusion in demand</i>		
$a_0^\zeta$	0.5304351	0.3612863
$a_1^\zeta$	-1.890973	0.6208012 ***
$a_2^\zeta$	-0.2784005	0.9215241
$a_3^\zeta$	3.534504	1.61069 **
<i>Diffusion in tastes</i>		
$a_0^\gamma$	-1.508633	0.6848587 **
$a_1^\gamma$	-2.400086	1.275214 **
$a_2^\gamma$	1.553593	0.905026 **
$a_3^\gamma$	1.547807	1.160901 *
<i>Cost function</i>		
$a_0^c$	1.301216	0.583414 **
$a_1^c$	0.1607157	0.083076 **
$a_2^c$	-1.23243	1.154534
<i>Reward for level of sales</i>		
$a_0^\alpha$	0.3655538	0.7095321
$a_1^\alpha$	-0.0092213	0.588669
$a_2^\alpha$	-0.6352032	0.9920091
$a_3^\alpha$	-0.1107006	0.4291889
<i>Reward for growth of sales</i>		
$a_0^\beta$	-0.3333785	0.6646826
$a_1^\beta$	0.7126854	0.3604874 **
$a_2^\beta$	-0.338622	0.5106897
$a_3^\beta$	-1.50507	0.4819816 ***
<i>Wage and risk aversion</i>		
$\sigma$	0.9577366	0.4618519 **
$R_m$	0.8246784	0.4039156 **

(\*\*\*) reflects significance on 1% level, (\*\*) - 5% significance, (\*) - 10% significance

Table 4: Structural estimates of parameters of the single index

Variable	Estimate	Std. deviation
Sales % change	-0.1380467	0.0977934
Market value	-1.197663	0.5734218 **
Pre-tax/average income	0.7703222	0.5608631
Book-to-market ratio	0.1993998	0.4151506
Stock grant options/options	1.603688	0.9515046
Option grants	2.09016	0.5417078 ***
Value of shares/bonus	-3.376088	1.56608 **
Gender (female=1)	-2.501264	1.533577 *

Table 5: Estimation results for the system describing firm's sales and manager's compensation

<i>Variable</i>	<i>Coefficient</i>	<i>std. error</i>	<i>t-statistic</i>
Dependent variable: Salary ( $\mathbf{R}^2=0.212$ )			
Sales	51.51886	2.50402	20.57
Firm's value	0.0021807	0.0004241	5.14
Tobin's q	2.452688	0.5870775	4.18
Change in assets	-0.0034975	0.0035935	-0.97
Apparel	67.90431	7.229979	9.39
SOX	27.62433	11.58859	2.38
Time	14.29167	0.9717439	14.71
Constant	292.5628	6.395592	45.74
Dependent variable: Bonus ( $\mathbf{R}^2=0.189$ )			
Sales	70.91936	4.438576	15.98
Firm's value	0.0090904	0.0007518	12.09
Tobin's q	11.67322	1.041005	11.21
Change in assets	-0.0048182	0.0063721	-0.76
Apparel	51.95567	12.8201	4.05
SOX	43.49443	20.54893	2.12
Time	19.77868	1.723088	11.48
Constant	74.17933	11.34068	6.54
Dependent variable: Sales ( $\mathbf{R}^2=0.414$ )			
Firm's value	0.0001162	1.76E-06	66.01
Tobin's q	-0.0233784	0.0033326	-7.02
Change in assets	-0.0000402	0.0000205	-1.96
Apparel	-0.4840632	0.0406628	-11.9
SOX	-0.034501	0.0661104	-0.52
Time	-0.0556195	0.0054865	-10.14
Constant	0.1261532	0.036442	3.46
N. obs =6756			