Estimating a Tournament Model of Intra-firm Wage Differentials*

Jiawei Chen
Department of Economics
University of California, Irvine

Matthew Shum†
Division of Humanities and Social Sciences
California Institute of Technology

August 24, 2009

Abstract

We consider the estimation of a tournament model with moral hazard (based on Rosen (1986, *AER*)) when only aggregate data on intra-firm employment levels and salaries are available. Equilibrium restrictions of the model allow us to recover parameters of interest, including equilibrium effort levels in each hierarchical stage of the firm. We illustrate our estimation procedures using data from major retail chains in the US. We find that only a fraction of the wage differential directly compensates workers for higher effort levels, implying that a large portion of the differentials arises to maintain incentives at lower rungs of the retailers.

JEL Classification: C13, J31, L81
Keywords: Tournament Model, Estimation, Wage Differentials, Incentive Effect, Retail Chains

1 Introduction

Wage differentials within retail chains, even at the lowest levels, can be quite large: for example, full-time sales staff at a major clothing retail chain are paid roughly $16,000

*We are grateful to Joe Harrington for providing us access to the data used in this analysis, and for useful comments. We also thank Mike Baye, Rick Harbaugh, Nick Hill, Ben Klemens, Philippe Marcoul, Joanne Roberts, Chris Ruebeck, Nadia Soboleva, Ayako Suzuki, Quang Vuong, and seminar participants at Chicago, Georgetown, Indiana, Northwestern, Yale, the 2004 IIOC conference, and the 2005 SITE conference for useful suggestions.

†Corresponding author. Tel.: +1 626 395 4022; fax: +1 626 432 1726; e-mail: mshum@caltech.edu.

per annum on average, but store managers earn around \$33,600, over twice that amount. Further up the hierarchy, district managers make on average over \$60,800 (in 1986 dollars).[1] Why do these differentials arise? One explanation is that employees at higher levels of a firm are paid more, because they work harder, or are more productive. Alternatively, the tournament literature proposes that wages at the top of a hierarchy must be kept high in order to provide incentives for workers, even in low levels of the hierarchy, to exert effort.

Despite the sizable theoretical literature on tournament models (see McLaughlin (1989) and Bolton and Dewatripont (2005, chap. 8) for surveys), empirical work related to these models is limited. Previous empirical work on tournament models have mainly focused on testing the predictions of these models. This includes papers on executive compensation (cf. Main, O'Reilly, and Wade (1993), Eriksson (1999)), professional sports (eg. Ehrenberg and Bognanno (1990a), (1990b), Bronars and Oettinger (2001)), and agricultural (poultry) production (Knoeber and Thurman (1994)). One common feature of these papers is their focus on industries for which productivity measures at the individual worker level are observable.

However, for many industries, these worker productivity measures are difficult to obtain (or unavailable). In this paper, we consider the structural estimation of tournament models when we observe only aggregate firm-level information on employment and wages at each hierarchical level within the firm. Importantly, our empirical strategies do not require observation of workers' productivity levels. By exploiting the equilibrium restrictions of the elimination tournament model, we derive estimates of model unobservables — including workers' equilibrium effort levels — which are consistent with the observed wage data. Hence, the aim of this paper is not to test the tournament theory (as in the previous empirical work), but to use it as a guide to obtain values for the structural elements of the tournament model.

Recently, several papers have estimated structural models of tournament models, as is done in this paper. Ferrall (1996), (1997), estimates structural models of internal labor markets within, respectively, law firms and engineering firms.[2] Zheng and Vukina (2007) estimate a rank-order tournament model using the Knoeber and Thurman (1994) data, and use

[1]These figures are for 1997. These data are drawn from the National Retail Federation Specialty Store Wage and Benefit Survey, which will be described below. For confidentiality reasons, the names of the stores used in this study cannot be mentioned in the paper. By definition, full-time sales staff must work at least 30 hours a week.

[2]Relatedly, Ferrall and Smith (1999) develop and estimate a structural tournament-like model of sports championship series.

the estimated structural parameters to simulate outcomes under an alternative cardinal compensation scheme. The empirical approaches taken in those papers and the present paper are quite different, due to both differences in the underlying theoretical tournament model being considered, as well as differences in the datasets used for estimation.

We illustrate our methodologies by estimating tournament models using a dataset of wages and employment levels within a number of large American retail chains (including many retailers found in typical shopping malls). Our empirical analysis focuses on the lower hierarchical levels (i.e., the sales staff, assistant store manager, and store manager positions) of these firms. We are able to estimate the equilibrium effort levels in these positions, consistent with the observed data and the theoretical tournament model.

We find that effort is generally increasing at higher levels in the firm. Moreover, our results suggest that only a small fraction – typically less than 50% – of the observed wage differentials directly compensates workers for higher effort at higher levels of the hierarchy, implying that over half of the differentials arise purely to maintain incentives at lower rungs of the retailers. Using our estimates, we also simulate counterfactual compensation levels that employees would be paid if effort were observable and contractible, and compare them with actual compensation levels (which we interpret as information-constrained "second-best" wages).

In the next section, we present a store-level tournament model, based on the model in Rosen (1986), that we will employ in this paper. In Section 3, we discuss the identification of this model, and develop two estimation methodologies. In Sections 4 and 5 we present the empirical illustration. We conclude in Section 6.

2 Economic model

The tournament view of a firm's internal labor market differs in important respects from the efficiency wage literature on labor contracts, which likewise focuses on contracts as a means to provide incentives to workers to provide effort. The non-tournament efficiency wage literature has focused on "absolute" compensation schemes, in which each worker is paid according to how her observed performance measures against some objective, absolute benchmark. As long as a worker's observed performance depends (even stochastically) on her effort or productivity, these non-tournament compensation schemes ultimately need not generate intra-firm wage differentials if effort or productivity is unchanging across different hierarchical levels within the firm.

This is not the case with tournaments, which are a form of "relative" compensation schemes, whereby a given worker is paid (or promoted) depending on her performance relative to her co-workers. In an elimination tournament, workers must exert effort and be productive even at low levels of the hierarchy in order to remain in contention for the larger prizes which are available at higher levels of the firm.[3] As Rosen (1986) points out, intra-firm wage differentials can arise even when workers exert identical effort levels at each stratum of the firm, because the pay differentials at higher levels of the hierarchy motivate effort exertion at lower levels of the hierarchy.

The goal of this paper is to develop empirical strategies to estimate a tournament model when only aggregate firm-level information on employment and wages, at each hierarchical level within the firm, are available. Next, we introduce our model, which is based on the elimination tournament model of Rosen (1986). In this model, each worker's career within a firm transpires as a progression through a tournament, in which workers compete against each other at each hierarchical level of the firm, with the winners at each level advancing to compete at higher levels.

A given firm has $S+1$ hierarchical levels, indexed by s, with $s=0$ corresponding to the highest level, and $s=S$ corresponding to the entry level. $W_1,\ldots,W_{S+1}$ denote the payoffs (wages) at each level of the firm. (Note our indexing convention, whereby W_{s+1} is the wage that the "losers" at stage s make.) Hence, W_{S+1} is the salary earned by the employees at the lowest level in the firm hierarchy, and it can be interpreted as a "reservation wage" for all the workers in our model.

Let $n_0,\ldots,n_S$ denote the number of workers at each level of the firm: hence, the total number of workers from level s who are advanced up to level $s-1$ are $\sum_{s'=0}^{s-1} n_{s'} \equiv m_{s-1}$. Note that, by this definition, $m_{s-1} < m_s$, for $s=0,\ldots,S$.

[3] Another strand of the tournament literature has focused on rank-order tournaments, in which workers are paid according to their relative performance in identical tasks (cf. Lazear and Rosen (1981), Green and Stokey (1983), and Holmstrom (1982)). The goal of rank-order tournaments is to encourage high effort in a homogeneous task in the presence of common (across all workers) unobserved productivity shocks, which differs from the goal of an elimination tournament, which is generally to provide incentives for continued (and perhaps higher) effort levels at higher hierarchical strata of the company. This paper focuses solely on the elimination-type of tournaments.

2.1 Competition

Within each level of the firm, we specify a model of competition which seems especially relevant for the retail environment. Within level s, we assume that the firm divides the m_s contenders into L_s equal-sized subgroups, each consisting of m_s/L_s workers (abstracting away from integer issues). A tournament is played among the members of each subgroup, with the $m_{s-1}/L_s \geq 1$ best performers selected to advance to the next level $s-1$. Let

$$f_s \equiv m_s/L_s, \quad g_s \equiv m_{s-1}/L_s$$

denote, respectively, the number of contenders and winners per subgroup. In the retail application below, a subgroup is interpreted as either a store or a region, depending on the stage of the tournament.[4]

We assume that all workers are homogeneous, and focus on a symmetric pure strategy Nash equilibrium in which each worker exerts the identical effort level x_s^* (where $x > 0$) at level s.[5] An agent who exerts an effort level $\bar{x}$ during level s while all of the rivals in her subgroup exert the equilibrium level of effort x_s^* advances with probability

$$P_s(\bar{x}; x_s^*) = \frac{h(\bar{x}) + (g_s - 1) * h(x_s^*)}{h(\bar{x}) + (f_s - 1) * h(x_s^*)}. \tag{1}$$

In the above, $h(\cdot)$ is a function translating individual effort levels into the advancement probability.[6] The $P_s(\cdots)$ function captures, in reduced-form, the procedure whereby the firm selects winners at each stage of the tournament, and is induced ultimately by the information structure of the game (i.e., what signals of effort the firm observes). Since our dataset includes no measures of productivity for any employee, we avoid more detailed modeling of the information structure, and adopt the reduced-form advancement probability given in Eq. (1).[7]

[4] In Rosen's (1986) paper, and in most sports tournaments (eg. tennis tournaments), $f_s = 2$ for $s = 1, \ldots, S$, so that the number of subgroups at each level $L_s = \frac{1}{2} m_s$.

[5] While it would be interesting to consider a model with worker heterogeneity, it would be difficult to identify and estimate such a model without individual work-level employment histories, which our dataset does not have.

[6] This parameterization of the advancement probability conditional on effort follows Rosen. For the $f_s = 2$ case considered by Rosen, if $h(x) = \exp(x)$, then the advancement probability (1) is a binary logit probability. The logit probability function can, in turn, be justified by a model where the worker i with the higher productivity y_{is} advance out of stage s, and the productivity measure y_{is} is equal to the effort x_{is} plus an additive random noise term which follows the type I extreme value distribution. This structural interpretation of the parameterization (1) no longer applies when $f_s \neq 2$.

[7] See Ferrall (1996), pp. 814–815, for an example where the form of the advancement function $P_s(\cdots)$ is explicitly derived given additional assumptions on the information structure of the game.

In the symmetric equilibrium, all employees exert identical levels of effort, so that $\bar{x} = x_s^*$ and the probability of advancing beyond level s does not depend on the equilibrium effort level x_s^*:

$$p_s^* \equiv P_s(x_s^*; x_s^*) = \frac{g_s}{f_s} = \frac{m_{s-1}}{m_s}. \tag{2}$$

While the functional form for the advancement probability in Eq. (1) is not arbitrary, as remarked above, the form of the equilibrium probability (2) obtains very generally, requiring only that, in equilibrium, any g_s-subset of the f_s contestants in stage s of the tournament are chosen to advance with equal probability. This is a reasonable requirement because, in any symmetric equilibrium, every contestant should expend an identical level of effort.

Note that *any* symmetric equilibrium in which all m_s contestants exert identical levels of effort (including zero effort) will yield the same winning probability m_{s-1}/m_s in equilibrium. Therefore, the specific values of f_s and g_s matter only insofar as it affects the players' incentives, and therefore the effort levels that they choose. Indeed, as we will see below, the toughness of competition (as parameterized by f_s and g_s) has a crucial effect on the amount of effort exerted in equilibrium.[8]

2.2 Equilibrium

The equilibrium sequence of effort levels $\{x_s^* : s = 1, \dots, S\}$ is determined by a dynamic optimization problem.[9] Let V_s denote the (equilibrium) value of progressing to (and potentially beyond) level s. If a given worker chooses effort level $\bar{x}$ at level s, her value V_s is implicitly defined via the Bellman equation

$$V_s = \max_{\bar{x}} \{P_s(\bar{x}; x_s^*) V_{s-1} + (1 - P_s(\bar{x}; x_s^*)) W_{s+1} - c(\bar{x})\} \tag{3}$$

where $c(\cdot)$ is the cost of effort function.[10] In the above display, the first term within the curly brackets denotes the worker's expected payoff from advancing to the next ($s-1$-th)

[8]The symmetric equilibrium has a prisoner's dilemma quality: every worker would be better off if nobody exerted any effort (since the equilibrium winning probability P_s is the same no matter how much effort is exerted), but this is not an equilibrium.

[9]With $S+1$ hierarchical levels, there are only S stages to the tournament, because there is no more competition at the top ($s = 0$) stage.

[10]Implicitly, we are assuming a discount rate equal to 1, so that both competition and payoffs occur simultaneously. When the discount rate is less than one, the Bellman equation (3) would also be consistent with a model where costs are paid in the current period, but the rewards (either W_{s+1} or V_{s-1}) are not accrued until next period, in which case the costs should be divided by the discount rate, in order for the units of $c(\cdot)$ to be comparable with V_{s-1} and W_{s+1}. We thank a referee for suggesting this alternative interpretation.

round, while the second term is the payoff from "losing" in stage s and obtaining the wage W_{s+1}. At level s, a given worker chooses an effort level $\bar{x}$ to maximize the right-hand side of (3). Throughout, we assume workers are risk-neutral, so that in this model there is no insurance aspect to contracting.

In the symmetric equilibrium, all workers expend identical effort levels x_s^* in level s; this effort level must satisfy the following first-order condition:

$$P_{s,1}\left(x_s^*; x_s^*\right)\left(V_{s-1} - W_{s+1}\right) - c'(x_s^*) = 0 \tag{4}$$

where $P_{s,1}(\cdots)$ denotes the derivative of $P_s(\cdots)$ with respect to the first argument. Given the odds-ratio parameterization (1) of the $P_s(\cdots)$ function, in equilibrium

$$p_{s,1}^* \equiv \left.\frac{\partial P_s\left(x; x_s^*\right)}{\partial x}\right|_{x_s^*} = \frac{h'(x_s^*)}{h(x_s^*)}\frac{1 - p_s^*}{f_s}. \tag{5}$$

By substituting Eq. (5) into the first-order condition (4), we obtain

$$\frac{h'(x_s^*)}{h(x_s^*)}\frac{1 - p_s^*}{f_s}\left(V_{s-1} - W_{s+1}\right) = c'(x_s^*). \tag{6}$$

The above equation is the main equation which characterizes equilibrium effort levels in our tournament game, and our estimation strategies will exploit this optimality equation.

2.3 Remarks

Before proceeding to discuss estimation, however, we raise several remarks about features of the model. First, we assume that workers are homogeneous, so that the tournament has no selection aspect (ie., it does not screen for the more productive workers), but rather only a moral hazard effect, arising from the (assumed) non-contractibility of effort. We make this assumption because we only observe aggregated data. Without observations of individual worker-level data, it is difficult to identify and estimate models with worker-specific heterogeneity. (However, in all our empirical results, we are completely flexible in allowing for chain-level heterogeneity, by estimating the model separately for each retail chain.)

Second, the model places restrictions on workers' employment paths within a firm. The model does not accommodate voluntary attrition (quits) out of the firm, and also only allows workers to enter the firm at the lowest ($s = S$) level of the hierarchy. As above, these assumptions are mainly motivated by data considerations; without observing individual worker-level employment paths, it is difficult to identify and estimate a model with these features. We will discuss these assumptions in more detail below, when we discuss our dataset.

Third, because of data constraints, we use cross-sectional data to estimate an inherently dynamic model. That is, we use annual wage to represent the full value of losing a promotion competition, which implicitly assumes that annual wages are proportional to expected lifetime utilities, which in turn makes sense only in an infinite-horizon model. If the horizon is finite, then annual wages may not be proportional to expected lifetime utility, due to life-cycle concerns. For example, many of the mid-level workers are in their middle age; if they fail to get promoted, their wage is "spread out" over roughly twenty years. On the other hand, low-level workers may have "lost" the competition earlier in their careers, implying that their wage could be spread out over forty years.

Relatedly, in our empirical work we use stocks of workers to estimate flows (promotion rates). This correspondence between the two may not hold if "losers" are given a chance to compete again in the future. If such repeated competition is taken into account, promotion rates may differ from employment ratios because the latter includes previous losers.

3 Empirical strategy

The goal of estimation is to recover values of the equilibrium effort levels $x_1^*, \ldots, x_S^*$ as well as (to the extent possible, as we will be precise about later) the functions $h(\cdot)$ and $c(\cdot)$. The data at hand consists of observations of wages $W_1, \ldots, W_{S+1}$ and number of employees $n_1, \ldots, n_S$ at each hierarchical level of a firm. Moreover, observing $n_1, \ldots, n_S$ implies that we observe the advancement probabilities $p_1^*, \ldots, p_S^*$ because these are, in equilibrium, ratios of the number of employees at different levels of a firm (cf. Eq. (2)).

Our empirical strategies are based on the first-order condition (6), which summarizes the equilibrium relations between the observables (p_s^*, W_{s+1}, f_s) and the parameters of interest (x_s^* and the h and c functions).[11] Next, we point out an underidentification problem which arises in estimating this model. Subsequently, we develop two estimation strategies which overcome this underidentification problem in different ways.

[11]The reliance of our estimation strategies on first-order conditions bears some qualitative similarities to the approach taken in the empirical auction literature (eg., Guerre, Perrigne, and Vuong (2000)) and the empirical equilibrium search literature (eg., Ridder and van den Berg (1998), Hong and Shum (2006)).

3.1 Discussion of Identification

The first-order condition (6) can be rewritten as

$$\mu(x_s^*)\frac{1-p_s^*}{f_s}\left(V_{s-1}-W_{s+1}\right)=c(x_s^*), \tag{7}$$

where

$$\mu(x)\equiv\frac{h'(x)}{h(x)}/\frac{c'(x)}{c(x)}. \tag{8}$$

In matrix notation, this is

$$\begin{pmatrix} \frac{\mu(x_1^*)}{f_1}(1-p_1^*) & 0 & \cdots & 0 \\ 0 & \frac{\mu(x_s^*)}{f_2}(1-p_2^*) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\mu(x_S^*)}{f_S}(1-p_S^*) \end{pmatrix}\left[\begin{pmatrix} V_0 \\ V_1 \\ \vdots \\ V_{S-1} \end{pmatrix}-\begin{pmatrix} W_2 \\ W_3 \\ \vdots \\ W_{S+1} \end{pmatrix}\right]=\begin{pmatrix} c(x_1^*) \\ c(x_2^*) \\ \vdots \\ c(x_S^*) \end{pmatrix}. \tag{9}$$

This can be conveniently used in order to derive a recursive formulation for V_s: by plugging Eq. (7) into Eq. (3), we obtain

$$V_s=\beta_s V_{s-1}+\left(1-\beta_s\right)W_{s+1},\ \ s=1,\ldots,S \tag{10}$$

where

$$\beta_s\equiv p_s^*-\mu(x_s^*)\frac{1-p_s^*}{f_s}.$$

Using the initial condition $V_0=W_1$, we can solve Eq. (10) forward to derive

$$\begin{aligned} V_1 &= \beta_1 W_1+(1-\beta_1)W_2 \\ V_2 &= \beta_2\beta_1 W_1+\beta_2(1-\beta_1)W_2+(1-\beta_2)W_3 \\ &\vdots \\ V_S &= \beta_1\ldots\beta_S W_1+(1-\beta_1)\beta_2\ldots\beta_S W_2+(1-\beta_2)\beta_3\ldots\beta_S W_3+\cdots+(1-\beta_S)W_{S+1} \end{aligned}$$

or

$$\begin{pmatrix} V_0 \\ V_1 \\ V_2 \\ \vdots \\ V_S \end{pmatrix}=\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \beta_1 & (1-\beta_1) & \cdots & 0 & 0 \\ \beta_1\beta_2 & \beta_2(1-\beta_1) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_1\cdots\beta_S & (1-\beta_1)\beta_2\cdots\beta_S & \cdots & (1-\beta_{S-1})\beta_S & (1-\beta_S) \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_S \\ W_{S+1} \end{pmatrix}. \tag{11}$$

Substituting (11) into (9), we obtain

$$\begin{pmatrix} \frac{\mu(x_1^*)}{f_1}(1-p_1^*) & 0 & \cdots & 0 \\ 0 & \frac{\mu(x_s^*)}{f_2}(1-p_2^*) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\mu(x_S^*)}{f_S}(1-p_S^*) \end{pmatrix} \left[\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \beta_1 & (1-\beta_1) & \cdots & 0 \\ \beta_1\beta_2 & \beta_2(1-\beta_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1\cdots\beta_{S-1} & (1-\beta_1)\beta_2\cdots\beta_{S-1} & \cdots & (1-\beta_{S-1}) \end{pmatrix} \right.$$
$$\left. \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_S \end{pmatrix} - \begin{pmatrix} W_2 \\ W_3 \\ \vdots \\ W_{S+1} \end{pmatrix} \right] = \begin{pmatrix} c(x_1^*) \\ c(x_2^*) \\ \vdots \\ c(x_S^*) \end{pmatrix}. \tag{12}$$

The system of equations (12) gives us, for each firm, S equations but $2S$ unknowns $c(x_1^*), \ldots, c(x_S^*)$, $\mu(x_1^*), \ldots, \mu(x_S^*)$.

Clearly, with only S equations, we cannot identify the $\mu(\cdot)$ and $c(\cdot)$ functions nonparametrically. We proceed by restricting $h(\cdot)$ and $c(\cdot)$ to lie within a parametric family:

$$\begin{aligned} h(x) &= x^\gamma \\ c(x) &= x^\alpha. \end{aligned} \tag{13}$$

That is, we assume power (constant-elasticity) specifications for both the h function and the c function.

We next examine the system of equations (12). First recall that all the β's are functions of μ only. Second, once we assume the power specifications for the h and c functions, all the μ's are equal to γ/α. So γ and α enter into the left-hand side of (12) only in the form of γ/α, and x_s^* does not enter into the left-hand side at all. Furthermore, the right-hand side of each equation in the system is $x_s^{*\alpha}$. As a result, each FOC can be expressed as $\phi_s(\gamma/\alpha) = x_s^{*\alpha}$, where we use $\phi_s(\gamma/\alpha)$ to denote the left-hand side of an FOC in order to emphasize that it depends only on the ratio γ/α. So if $(\alpha, \gamma, x_1^*, \ldots, x_S^*)$ satisfy the FOC's, then $(1, \gamma/\alpha, x_1^{*\alpha}, \ldots, x_S^{*\alpha})$ also satisfy them. That means the models with these multiple values of parameters are observationally equivalent, and to proceed we need to normalize one way or another. To do this we set $\alpha = 1$. Consequently, $c(x) = x$.

With these assumptions, a total of $S+1$ parameters—$\theta \equiv (x_1^*, \ldots, x_S^*, \gamma)$—are to be estimated. For this specification of $h(\cdot)$, γ parameterizes the responsiveness of the advancement probability P_s to an individual's effort level. Hence, in a setting where effort is observable with noise, it is reasonable to interpret a larger value of γ as implying that the observations of effort are less noisy, which may imply, in turn, that a better monitoring technology is in place.

With these assumptions, we can simplify $\mu(x_s^*) = \gamma$ for $s = 1, \ldots, S$, and the system of first-order conditions (12) reduces to

$$
\begin{pmatrix}
\frac{\gamma}{f_1}(1-p_1^*) & 0 & \cdots & 0 \\
0 & \frac{\gamma}{f_2}(1-p_2^*) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{\gamma}{f_S}(1-p_S^*)
\end{pmatrix} *
$$
$$
\left[\begin{pmatrix}
1 & 0 & \cdots & 0 \\
\beta_1 & (1-\beta_1) & \cdots & 0 \\
\beta_1\beta_2 & \beta_2(1-\beta_1) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\beta_1 \cdots \beta_{S-1} & (1-\beta_1)\beta_2 \cdots \beta_{S-1} & \cdots & (1-\beta_{S-1})
\end{pmatrix}
\begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_S \end{pmatrix}
-
\begin{pmatrix} W_2 \\ W_3 \\ \vdots \\ W_{S+1} \end{pmatrix}
\right]
=
\begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_S^* \end{pmatrix}
\tag{14}
$$

and $\beta_s = p_s - \gamma \frac{1-p_s}{f_s}$ for $s = 1, \ldots, S$. We are still underidentified, as the system of equations (14) contains S equation but $S+1$ unknowns $(x_1^*, \ldots, x_S^*, \gamma)$. Below, we describe two approaches which circumvent this fundamental underidentification issue.

Before describing our estimation procedures, however, we look at some comparative statics from the first order conditions (eqs. (6) and (14)), which yield some economic intuition for this underidentification result. Given our assumptions on $h(x)$ and $c(x)$ (in Eqs. (13)), if we increase the number of contestants f_s, while holding the advancement probability fixed (i.e., by always adjusting the number of winners $g_s = f_s \cdot p_s^*$, for some pre-specified level of p_s^*), then $\frac{dx_s^*}{df_s}|_{p_s^* \text{fixed}} < 0$: equilibrium effort levels are smaller when the number of contestants increases.[12] This suggests that an increase in the "toughness of competition" (as measured by f_s) actually dilutes equilibrium incentives to provide effort: when f_s is higher, the marginal effect of additional effort on the winning probability, which is equal to $\gamma * \frac{1-p_s}{f_s}$, is lowered.

[Figure 1 about here.]

[12]In making this calculation, we assume that any change in f_s leaves p_{s-1}^* unaffected.

[Figure 2 about here.]

In Fig. 1, we illustrate this comparative static. We have graphed pairs of best response curves corresponding to different values of f and g, values of (respectively) the contenders and winners from the lowest (sales staff) level of a tournament played at the San Francisco-area stores of retailer I.[13] In the solid lines, we graph the best response curves corresponding to the case where f (the number of contending sales staff) is 10.25, and the number of winners g (those who "win" the sales floor tournament) is 4.38, the actual observed values. The lines marked with circles are the best response curves for the case where both f and g are halved: the reduction in the number of competitors has raised workers' incentives to exert effort, and the equilibrium effort levels (measured in money units) double from about \$1500 to \$3000. In contrast, if we double the number of contenders, effort levels decrease by about half, from about \$1500 to \$800 units (as indicated by the intersection of the third set of best-response curves, marked in the crossed lines). Hence, configurations of data analogous to that presented in Figure 1 would be informative regarding the rankings of equilibrium effort levels in different stages: hypothetically, if there were two stages s, s' in the tournament which were completely identical, except for $f_s > f_{s'}$, then we could conclude that $x_s^* < x_{s'}^*$.

However, while variation in f_s across stages is informative as to the rankings in x_s^* across stages, it is not enough to pin down the magnitudes of the effort levels, which depend on γ. With the power function parameterization, we obtain that $\frac{dx_s^*}{d\gamma} > 0$, $\forall s$: the equilibrium effort levels are increasing in γ, as illustrated in Fig. 2. The solid lines are a pair of best-response curves corresponding to $\gamma = 1.5$. When γ is increased to 2.0, implying more responsiveness in the probability of winning to worker effort, we see that the circle-marked best response curves intersect at a higher point, implying that equilibrium effort levels rise, by around 15%. When γ is decreased to 1.0 (illustrated by the best-response curves marked with crosses), the equilibrium effort levels decrease by around 15%. These considerations suggest that the actual magnitude of the effort levels is not separately determined apart from γ. Essentially, for given values of the observables $(f_s, \ldots, f_S,\ p_1^*, \ldots, p_S^*,\ W_1, \ldots, W_{S+1})$, a whole continuum of values of $(\gamma, x_1, \ldots, x_S)$ satisfy Eqs. (14).[14] Hence, additional restrictions would be required to jointly pin down the effort levels with the curvature parameter

[13] In these simulations, we fixed γ at 2.2, and also the wages fixed at the values observed in the San Francisco-area stores for this retailer.

[14] On the other hand, given the monotonicity of equilibrium effort levels x_s^* in γ, if one is willing to restrict γ to a certain interval (e.g. $\gamma \in [\underline{\gamma}, \bar{\gamma}]$), then bounds on equilibrium effort levels could be obtained by solving the system (14) at $\underline{\gamma}$ and $\bar{\gamma}$. We leave this intriguing possibility for future work.

γ. Next, we consider two alternative sets of restrictions.

3.2 First approach

In the first approach, we address the underidentification issues by adding additional equations regarding the supply-side of the model. Specifically, we assume that the employment at the S levels of the firms are chosen to maximize profits at the (chain, year, and geographic location)-level.[15] In specifying the supply-side, we assume that firms set $n_1, \ldots, n_S$ optimally, but take the functional form of the advancement probability function $P_s(\cdots)$ as given.[16] Here we depart from Rosen's (1986) model, which does not have a supply-side.[17]

We assume that firms choose $n_1, \ldots, n_S$ to maximize the following profit objective:

$$\max_{n_1,\ldots,n_S} \eta Q\left(n_1, \ldots, n_S; x_1, \ldots, x_S\right) - \sum_{s=0}^{S} n_s W_{s+1} \tag{15}$$

where $\vec{n} = (n_1, \ldots, n_S)$ and $\vec{W} = (W_1, \ldots, W_{S+1})$.[18] The production function is assumed to take the CES form

$$Q\left(n_1, \ldots, n_S; x_1, \ldots, x_S\right) = \left(\sum_{s=1}^{S} \left(n_s x_s^*(\vec{n}, \vec{W})\right)^{\rho}\right)^{1/\rho} \tag{16}$$

in the aggregate effort levels exerted in the different levels of the company. η is a multiplicative profitability parameter (scaling from effort to revenue units), and the substitution parameter ρ should lie in $[-\infty, 1]$ for the production function to be concave.

[15] We thank Quang Vuong for this suggestion.

[16] We could have also assumed that firms set wages $W_1, \ldots, W_{S+1}$, and also used the first-order conditions arising from those choices. We did not do this for two reasons. First, the first-order conditions for $n_1, \ldots, n_s$ imply enough additional restrictions for recovering the effort levels $x_1, \ldots, x_S$ and curvature parameter γ. Second, and more importantly, it turns out that the first-order conditions for $W_1, \ldots, W_{S+1}$ are multiplicative in γ, and so do not allow us to identify that parameter (once we take the ratio of first-order conditions, as is done in Eq. (17) below).

[17] Note that in our model, wages are exogenously given but the ratio of workers at different levels are endogenously chosen by the firm to maximize profit, whereas in Rosen's model, competition is always one-to-one and the question is what wage structures would lead to constant efforts throughout the hierarchy. So our model can be considered as the flip side of Rosen's model. We thank a referee for pointing out this comparison.

[18] This profit specification resembles the specifications used in models of hierarchical firms (eg. Qian (1994)).

The first-order conditions for this problem imply:

$$\frac{W_{s+1}}{W_{S+1}} = \frac{(n_s x_s^*)^{\rho-1} x_s^* + \sum_{t=1}^{S} (n_t x_t^*)^{\rho-1} n_t \frac{\partial x_t^*}{\partial n_s}}{(n_S x_S^*)^{\rho-1} x_S^* + \sum_{t=1}^{S} (n_t x_t^*)^{\rho-1} n_t \frac{\partial x_t^*}{\partial n_S}},\ s = 1, \ldots, S-1 \tag{17}$$

which yield $S-1$ equations, with the extra parameter ρ. (By dividing two first-order conditions, we eliminate the scaling parameter η, which we are not interested in estimating.) The partial derivatives $\frac{\partial x_t^*}{\partial n_s}$ are obtained from the system of equations (14).

For our empirical work, $S = 3$, so that the supply-side adds two equations. These two equations, combined with the S equations given in (14), enable us to recover the equilibrium effort levels $x_1^*, \ldots, x_S^*$, the power parameter γ of the $h(\cdot)$ function, as well as the new ρ parameter of the CES production function. This can be done separately for each firm, year, and geographical location (so we can allow equilibrium effort levels to differ arbitrarily across geographical locations and over time, for a given retail chain).

3.3 Second approach

The main benefit of the first approach is that we are able to recover distinctive effort levels for each set of (firm-, geographic region-, and year-) observations. However, this is done at the cost of making potentially restrictive assumptions regarding the supply-side. In the second approach, we dispense with the supply-side assumptions, but substitute instead the assumption that for a given firm, the effort levels $x_1^*, \ldots, x_S^*$, as well as the functional form of the $h(\cdot)$ function, are constant over both time and geographic locations.

However, once we assume that $x_1, \ldots, x_S$ and γ remain unchanged over time and geographic locations for each retail chain, we must introduce additional randomness into the model to accommodate the variation in employment and wages across time and locations which we observe in the data. Therefore, we do this by allowing the wages $W_1, \ldots, W_{S+1}$ to be observed with error.[19]

In particular, we assume that W_{ismt}, the observed wage for firm i, strata s, location l, and year t, is equal to the actual (but unobserved) wage W_{is}^* perturbed with additive measurement error ϵ_{ismt}:[20]

$$W_{islt} = W_{is}^* + \epsilon_{islt},\ s = 1, \ldots, S \tag{18}$$

[19] One may be justified in assuming the presence of measurement error, because the observed wages are obtained by surveys, where respondents report the average salaries earned in each hierarchical level.

[20] Note that given our assumptions that the effort levels $x_1, \ldots, x_S$ and γ are identical across all markets m and period t for a given firm i, the actual wage W_{is}^* should be unchanged across markets m and period t.

where ϵ_{islt} is a mean zero measurement error assumed independent across s, m, and t for a given firm i, and also independent of W^*_{is}. (Note that we assume that W_{iS+1lt}, the wage at the lowest stratum of the company, is not observed with error. The reason for this will be noted below.)

We estimate the parameters $x^*_1, \ldots, x^*_S$, as well as the parameters of the $h(\cdot)$ function, separately for each firm i. (In what follows, we drop the firm i subscript for convenience.) For each firm, we will estimate these parameters by method of moments. In order to derive the estimating equations, we combine Eqs. (14), (5), and (18) to obtain

$$\begin{pmatrix} \frac{\gamma}{f_1}(1-p^*_1) & 0 & \cdots & 0 \\ 0 & \frac{\gamma}{f_2}(1-p^*_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\gamma}{f_S}(1-p^*_S) \end{pmatrix} *$$
$$\left[\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \beta_1 & (1-\beta_1) & \cdots & 0 \\ \beta_1\beta_2 & \beta_2(1-\beta_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1\cdots\beta_{S-1} & (1-\beta_1)\beta_2\cdots\beta_{S-1} & \cdots & (1-\beta_{S-1}) \end{pmatrix} \begin{pmatrix} W_1-\epsilon_1 \\ W_2-\epsilon_2 \\ \vdots \\ W_S-\epsilon_S \end{pmatrix} - \begin{pmatrix} W_2-\epsilon_2 \\ W_3-\epsilon_3 \\ \vdots \\ W_{S+1} \end{pmatrix} \right] = \begin{pmatrix} x^*_1 \\ x^*_2 \\ \vdots \\ x^*_S \end{pmatrix} \tag{19}$$

or, in shorthand,

$$\begin{aligned} \mathbf{A}\left[\mathbf{B}\left(\vec{W}_{1:S}-\vec{\epsilon}_{1:S}\right)-\left(\vec{W}_{2:S+1}-\vec{\epsilon}_{2:S}|0\right)\right] = \vec{x} \;\Rightarrow \\ \vec{\epsilon} = -\tilde{\mathbf{B}}^{-1}\left[\mathbf{A}^{-1}\vec{x}-\mathbf{B}\vec{W}_{1:S}+\vec{W}_{2:S+1}\right] \end{aligned} \tag{20}$$

where $\tilde{\mathbf{B}}$ is the matrix $\mathbf{B}$ minus a $S \times S$ matrix with ones in the $(i, i+1),\ i = 1, \ldots, S-1$ spots and zeros everywhere else. (In the first display, $\vec{\epsilon}_{2:S}|0$ denotes the S-vector where the first $S-1$ elements are $\epsilon_2, \ldots, \epsilon_S$ and the S-th element is a zero.) Because the system of equations (19) is only S-dimensional, we cannot accommodate an additional measurement error in the wage W_{S+1}.[21]

[21] Since this is ad-hoc, we also obtained results where we allowed W_{S+1} to be observed with error, but assumed instead that the observed W_1 (the district manager salary) contained no measurement error. Overall, the magnitude of the results remained quite stable.

For each firm i, the population moment conditions exploited in the estimation is

$$E[\epsilon Z] = 0 \tag{21}$$

where Z is an M-vector of instruments. The sample analog of the above is

$$\mathbf{m}_{LST}(\theta) \equiv \begin{bmatrix} \frac{1}{LST}\sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{s=1}^{S}\epsilon_{lst}(\theta)Z_{sl1t} = 0 \\ \vdots \\ \frac{1}{ST}\sum_{t=1}^{T}\sum_{s=1}^{S}\epsilon_{lst}(\theta)Z_{slMt} = 0 \end{bmatrix}$$

where the dependence of ϵ_{st} on the parameters θ emphasizes the fact that, at each value of θ, the ϵ's are obtained as residuals, via Eq. (20). Let $M \geq S+1$ be the total number of moments conditions employed in estimating θ. We seek the minimizer of the quadratic form

$$\theta_{LST} \equiv \text{argmin}_{\theta}\ \mathbf{m}_{LST}(\theta)'\Omega\mathbf{m}_{LST}(\theta)$$

and our estimator has the limiting distribution (as T goes to infinity)

$$\sqrt{ST}\,(\theta_{LST}-\theta_0) \xrightarrow{d} N\left(0,(J'\Omega J)^{-1}J'\Omega V\Omega J(J'\Omega J)^{-1}\right)$$

where

$$J = E_0\frac{\partial\mathbf{m}(\theta_0)}{\partial\theta_0}$$
$$V = Var_0\ \mathbf{m}(\theta_0) = E_0\mathbf{m}(\theta_0)\mathbf{m}(\theta_0)'$$

and $\mathbf{m}(\theta)$ denotes the M-vector of moment conditions, and Ω is a $M\times M$ weighting matrix. In the results below, we use a two-step GMM procedure in which an estimate of the optimal weighting matrix $\Omega = V^{-1}$ is used in the second step, so that the limiting variance of our estimator reduces to $(J'\Omega J)^{-1}$.

In estimation, we employ seven moment conditions to estimate the four parameters in the model. The instruments which we use for a given observation of firm i, stratum s, location l, and year t are (i) n_{islt}, the number of workers in the firm at this location, year and level; and (ii) $w_{isl't}$, the wages of workers in the same level and during the same year, but at six different locations $l' \neq l$. For firms for which we have a relatively small number of observations, we only use observations in years when at least 7 markets are observed for the same firm (in order to have enough instruments). That explains why, in column 2 of Table 5, the number of observations used in estimation in the second approach is smaller than the number of observations reported in Table 1. For example, for firm A, we have observations for all years 1997-1999, but we use only the eight observations for 1997, because only for this year do we have observations for at least 7 markets.

Second-order optimality conditions In both estimation approaches given above, we assume that the first-order condition (4) characterizes the optimal effort levels chosen by the workers. However, second-order conditions should also hold at the optimal effort levels, given the tournament parameters. In our empirical work, therefore, we check each estimated set of effort levels $x_1^*, \ldots, x_S^*$ to ensure that they satisfy the second-order condition which corresponds to the first-order condition in Eq. (4):

$$\hat{c}'(\hat{x}_s^*) * \left[\frac{\hat{h}''(\hat{x}_s^*)}{\hat{h}'(\hat{x}_s^*)} - 2 \frac{\hat{h}'(\hat{x}_s^*)}{f_s * \hat{h}(\hat{x}_s^*)} \right] - \hat{c}''(\hat{x}_s^*) < 0 \tag{22}$$

for $s = 1, \ldots, S$, with the hats ($\hat{}$) denoting estimated quantities. For the parametric assumptions on the $h(\cdot)$ and $c(\cdot)$ functions (Eqs. (13)), the second-order conditions reduce to

$$\frac{\hat{\gamma} - 1}{\hat{x}_s^*} - 2 \frac{\hat{\gamma}}{f_s \hat{x}_s^*} < 0.$$

In what follows, we report only the empirical results which satisfy these second-order conditions.

4 Empirical illustration

As an illustration of the methodologies developed above, we estimate the model using data on wage differentials and employment levels at a number of large US retail chains. Most of these retailers typically have locations in shopping malls and centers in the suburban US. After presenting the results from the first approach and the second approach, respectively, we proceed to consider two extensions of the model. In the first extension we accommodate effort at the highest level by modeling it as observed by firms, and in the second extension we allow workers to quit. The estimates from these extensions are presented in the appendix, and show that our main findings are robust to these alternative specifications.

4.1 Data

The dataset is drawn from the Specialty Store Wage and Benefit Survey performed by the National Retail Federation (NRF), for the years 1997-1999. This survey contains aggregated information on the number of employees and average annual salary for employees in each hierarchical level, for a number of large national retail chains. For confidentiality reasons, we cannot identify the chains by name, but refer to them by letters (see Table 1 for a list of the 14 chains considered in the empirical exercise).

The data are aggregated up to the (retail chain–geographic area) level, so we cannot distinguish between different stores within the same chain and geographic area. Therefore, in our analysis, we essentially treat all the stores within the geographic area as identical stores. We focus on four levels of employment within each chain: Sales Staff ($s = 3$), Assistant Store Manager ($s = 2$), and Store Manager ($s = 1$), with the District Manager ($s = 0$) position taken as the prize in the tournament. We focus on these levels because they are the most similar and, hence, comparable across chains. At higher levels of the chains, the hierarchies can differ substantially across chains. In stages 2 (asst store managers) and 3 (sales staff), the tournament is played at the (chain-region-store-year) level, whereas in stage 1 (store managers), it is played at the (chain-region-year) level. Hence, in stages 2 and 3, each subgroup is a store, whereas in stage 1, each subgroup is a region.

According to the tournament model, we interpret the observed number of workers at stages $s = 1, \ldots, S$ as the number of "losers" (n_s) at that stage. The number of district managers, n_0, is interpreted as the number of "winners" in stage $s = 1$ of the tournament game.[22] As we discussed in Section 2.1 above, we need to make assumptions on the level of competition at each stage of the tournaments. At the sales staff and assistant store manager levels, we take the number of subgroups to be $L_2 = L_3 = n_1$, the number of store managers in each geographic region. That is, we assume that competition takes place within stores, and that the number of stores that a retail chain operates in each geographic region is equal to the total number of store managers employed in each region. At the store manager level, we assume that $L_1 = n_0$, the total number of district managers in the geographic region. That is, we assume that the firm divides up the stores within each geographic region in a number of equal-sized districts, each headed by a single district manager. (Throughout, we ignore integer issues for convenience.)

In Table 1 we give summary statistics on the wage and tournament parameters (number of contenders and winners in each subgroup, at each of the three stages considered) for the retail chains which we will consider in our study, averaged across geographic locations and years. (For full-time Sales Staff, salary is calculated as hourly wage*40 hours*50 weeks.) There are large variations across stores both in wages as well as the intensity of competition, so that we perform estimation on a store-by-store basis. Across most retailers, the assistant

[22] We do not model the changes across years in employment at the (chain, geographic location)-level. This would require extra assumption and modeling of the flow of workers in and out of the firm, which is beyond the scope of this paper (especially given that we have no data on the job tenure of the workers at each hierarchical level). In essence, then, we treat observations across years for a given retailer as independent observations.

store manger/store manager salary gap is at least as large, and often larger, than the sales staff/assistant store manager gap. Note also that g_2, the number of "winners" in the second stage (the competition among assistant store managers) is always slightly more than one; this is because, even though competition takes place within each store in stages 2 and 3, each store must promote slightly more than one winner, in order to produce the required number of employees to cover both the store manager and district manager positions.[23]

In Table 2, we define and describe some retailer characteristics which will be used in our analysis. These include variables for the retailers' line of business ($DCLOTH$, $DFOOT$, $DHOUSE$), store location and size ($DMALL$, $DLARGE$). Finally, $DCOMMIS$ indicates whether more than 25% of the compensation for sales staff are commission-based. We will use this information on commissions as an informal check on our estimates below.

As we mentioned earlier, our theoretical model is stylized, and abstracts away from several potentially important features of the retailers that we consider. First, the model does not allow workers to quit. In practice, attrition could be very common, especially among the sales staff of a retail chain.[24] However, this possibility need not affect our analysis since we interpret the data at hand as representing the *desired* level of employment at each hierarchy. Therefore, we allow for sales staff to quit, as long as they are replaced by other workers who "take their place" in the tournament (and as long as the assistant store managers and store managers are chosen from among the sales staff, not from the outside).

Second, we assume that all employees enter the hierarchy by the lowest (sales staff) stage. In practice, however, it may be possible to be directly hired as an assistant store manager, or store manager. Because we do not have individual-level data, we are unable to assess how important this is among the retailers in our dataset. Hence, our empirical application should be considered primarily as an illustration of our estimation methodologies, rather than a full examination of tournament incentives within retail chains.

[23]Note that we do not consider part-time sales staff in this paper, and assume that these employees do not participate in the tournament. Typically, the number of part-time sales staff outnumber full-time sales staff by a ratio of 5:1 or 6:1 in most retail chains. We only report the number of full-time sales staff in Table 1, which is why the average number of sales staff per store seems so small.

[24]For a subset of the retailers, we observe aggregated (chain level) turnover rates for each hierarchical level. Among this subset of firms, the median turnover rate (defined as the number of terminations divided by the number of workers employed at a given level) across firms is 14%, 23%, 33%, and 38% at the district manager, store manager, assistant store manager, and (full-time) sales staff stages, respectively. In an earlier version of the paper, we estimated an extension of the model which allows employees to leave the firm if they receive higher outside wage offers. However, we omit this extension from this version of the paper, because there are only four retailers which report enough turnover data.

[Table 1 about here.]

[Table 2 about here.]

4.2 Results: First approach

Using the first approach, we are able to recover the values of γ, ρ, x_1^*, x_2^*, x_3^* for every (chain, location, year) observation of $(W_1, \ldots, W_{S+1}, n_0, \ldots, n_S)$. Table 3 presents the average and standard deviation of the recovered values for each retail chain, where the average is taken across all (year, geographic region) pairs observed in the dataset for that particular chain.[25]

[Table 3 about here.]

Two features are noticeable. First, higher effort levels are exerted at higher levels of the firm, implying that at least part of the observed wage differentials across the store manager, assistant store manager, and sales staff stages can be justified for effort reasons alone. However, for several retailers, the average effort level drops at higher hierarchical levels: for example, for retailers B and C the average effort level at the assistant store manager stage exceeds the average effort level at the store manager stage. Second, for a majority of the firms, there is a larger gap in absolute effort between the assistant store manager and store manager stages than between the sale staff and assistant store manager stages. This mimics the relative sizes of the wage differentials, as given in Table 1.

Next, we see that the average $\gamma > 1$ in all the retail chains, (excepting chain J, which is the sole eyewear retailer in our dataset). The implies that for most chains, the $h(\cdot)$ function, which measures the sensitivity of the advancement probability to effort levels, is convex on average. Finally, for a large number of the retailers, the average ρ is quite negative, implying a low elasticity of substitution between the various types of workers which we consider.

Because we are able to recover effort levels, as well as γ and ρ, at the (chain, region, year)-level, we next examine the variation in these quantities. We ran regressions to see whether store characteristics (listed in Table 2) could explain the variation in these quantities across

[25] As discussed above, for each set of estimates, we checked that the second-order conditions in Eqs. (22) hold. We also dropped observations for which the nonlinear equation solver did not converge, as well as observations where the recovered $\rho > 1$, implying that the production function is not concave. Among these criteria, the non-convergence of the nonlinear solver accounted for the most eliminations – 56% of the observations.

retailers. The regression results are reported in Table 4. γ, the parameter in the $h(\cdot)$ function linking effort to the probability of winning, is smaller in shopping mall stores (coefficient on $DMALL$ -0.778) and in larger stores (coefficient on $DLARGE$ is 0.341). The coefficients are negative across the line-of-business dummies $DCLOTH$, $DFOOT$, and $DHOUSE$ (relatively to the omitted category, which consists of the children's retailer G and the eyewear retailer J).

The $DCOMMIS$ variable, an indicator for the use of commissions for sales staff, allows us to perform some ad-hoc specification checks on the estimates. First, we see that γ is significantly higher in stores which pay sales staff using commissions (the coefficient on $DCOMMIS$ is 1.097 and statistically significant). Since a larger γ can be interpreted as less noisy observations of effort, this finding could imply that firms offer sharper incentives (such as sales commissions) when effort signals are more accurate. This is consistent with Holmstrom and Milgrom's (1994) model of incentive contracts within a supermodular framework.

Moreover, $DCOMMIS$ enters positively (and significantly) in the regression where x_3, the sales staff effort, is the dependent variable, confirming, as we would expect, a positive relationship between commissions and effort level.[26] Since we do not impose this relationship in recovering the effort levels, this finding offers support for our model, and for our interpretation of the x's as effort levels.

[Table 4 about here.]

4.3 Results: second approach

Table 5 presents estimates of effort levels and the γ parameter using the second, GMM-based approach.[27]

[Table 5 about here.]

Generally, the estimates are reasonably precisely estimated. As with the first-approach estimates reported before, we check that, at the estimated quantities, the second-order

[26] We do not include $DCOMMIS$ as a regressor in the x_1 and x_2 regressions, because the survey only asks retailers to report whether commissions were used for sales staff.

[27] We varied the starting values in obtaining our estimates, to ensure that the reported estimates are reasonably robust and stable.

conditions (in Eq. (22)) are satisfied at the estimated effort levels. As a more formal specification check, the GMM J-statistic (and the associated p-value under the null of correctly-specified moment conditions) is given in the last column. For all 14 retail chains, we could not reject the null hypothesis (at reasonable significance levels) that the moment conditions are indeed satisfied at the estimated values.

The estimates for the $h(\cdot)$ function parameter, γ, indicate that this function is convex across all retailers[28], ranging from a low value of 1.34 for retailer J, to 3.86 for retailer H. Furthermore, for a majority of the choices, the estimated γ using the second approach is higher than the average recovered using the first approach. Correspondingly, it is not surprising to find that, generally, the effort levels estimated using the second approach are higher than those reported earlier because, as we discussed above, equilibrium effort levels increase when γ increase.

In addition, effort levels are generally higher at higher levels of the company. The sole exception is retailer J, for which a pronounced drop in effort occurs between the assistant store manager stage (where effort costs of \$3341.8 are expended) and the store manager stage (where effort costs of \$1912.7 – about a 40% drop – are expended). Retailer J is the sole eyewear retailer, and arguably this retailer is the one where sales staff are required to have the most training, as their duties include fitting and ordering eyewear for customers.[29]

5 Decomposing wage differentials

Next, we use our recovered parameter values to decompose wage differentials into the part which compensates workers for higher effort, and the part which serves to incentivize employees. In a perfect-information, perfectly-competitive setting, intra-firm wage differentials between two positions should just compensate employees for their effort cost differentials across the two positions. In a tournament setting, this need not be true, since wage differentials between levels i and $i+1$ must also serve to give incentives for more effort at lower levels $s > i$ of the company.

[28]Given $h(x) = x^{\gamma}$ and $c(x) = x$, this finding can also be interpreted as implying that the $h(\cdot)$ function is more convex than the cost of effort function across all 14 retailers.

[29]We also ran regressions of the estimated γ's and effort levels on retailer characteristics, analogously to the regressions reported in Table 4 for the first approach results. However, none of the coefficients were precisely estimated in these regressions. This is not surprising, because in the second approach we pool observations across geographic markets and years, resulting in only 14 observations in each of these regressions. For this reason, we do not report these regression results here.

Using the results obtained above, we can directly measure how much of the observed intra-firm wage differentials between stages i and $i+1$ (which, given our indexing convention, is $\Delta w_i \equiv w_{i+1} - w_{i+2}$), directly rewards higher effort differentials, $\Delta x_i \equiv x_i - x_{i+1}$. We introduce the notation $\delta_s \equiv 100 * \left(\frac{\Delta x_s}{\Delta w_s}\right)$, which denotes the ratio of effort to wage differential between stages s and $s+1$. In Table 6 we present the values for δ_1 (between the assistant store manager and store manager stages) and δ_2 (between the sales staff and assistant store manager stages) implied by our model estimates. The results using the first and second approaches, which are both reported in Table 6, differ numerically because the underlying estimating approaches are quite different. In the following discussion, we focus on qualitative results which are robust across both empirical approaches.

[Table 6 about here.]

For the first approach, δ_1 and δ_2 was calculated for each firm, year, and geographic location, and we report the mean and standard deviation of these values for each retail chain. We see that, on average, both δ_1 and δ_2 are typically less than 50. Since wages at a given stage i can provide incentives for effort only at stages $s > i$ prior to i, this implies that, for most chains, wages at the assistant store manager stage are an important source of incentives for effort provision on the sales floor and, similarly, wages at the store manager level also compensate for effort exerted in earlier stages.

Indeed, for several stores (retailers B, C, and E between the assistant store manager and store manager stages, in the first approach results), the average δ is *negative.* This occurs only when the effort differential between stages is estimated to be negative (because the wage differentials are never negative), and implies that the wage differential exists completely to compensate effort at previous stages. However, the mean δ_2 for chain M, and the mean δ_1 for chain D both exceed 100 (implying that the increase in effort outstrips the increase in wage between two levels). This makes sense only in a tournament setting, where workers perceive an option value of "staying alive" in the tournament.

5.1 Counterfactual: effort-based compensation

The results from the previous section suggest that a large proportion — typically over 50% — of the observed wage differentials arise to provide workers incentives to exert effort. The need for these incentives would only arise when effort is non-contractible. In order to gauge how well the second-best tournaments are performing for the retail chains which we study, we compare each retail chain's observed wage bill under the tournament setting to its wage bill under an alternative scheme in which workers' effort levels are contractible, and therefore compensation based directly on a worker's effort.

At each hierarchical level s, we compute the alternative wage (assuming firms have complete bargaining power), for a given effort level x_s in stage s, as

$$W_s^{CF} = W^R + x_s$$

where W^R is some reservation wage (assumed constant across all stages s), and x_s is the effort level at stage s, in money units. In calculating the counterfactual wages for each store, we assume that the firm's desired effort levels at each stage correspond to the effort levels estimated before.[30] Furthermore, we set W^R for each store to be equal to $W_{S+1} - x_S$, the sales staff salary less the cost of effort for each salesperson.[31]

These wage bill results are given in Table 7. We see that, indeed, for all of the stores, and across both approaches, the counterfactual wage bill is lower than the observed wage bill (which we interpret throughout this paper as being generated from a second-best tournament). The percentage differences between the two wage bills (reported in columns 3 and 6 of Table 7 for, respectively, the first and second approach estimates) hover between 20-25% for most retailers, ranging from 19% for retailer N, up to 48% for retailer J (using the first approach results). Therefore, while tournaments are second-best, in some cases the firms are not doing much worse using the tournaments, compared with the scenario where workers could be compensated directly for their effort.[32]

[Table 7 about here.]

6 Conclusions

In this paper, we considered the estimation of a tournament model with moral hazard when only aggregate data on intra-firm employment levels and salaries are available. We show

[30] Strictly speaking, then, these are not first-best effort levels, which would be the levels which maximized productivity less effort costs at each hierarchical level.

[31] We note that we cannot compute the entire counterfactual wage bill for all stages of the tournament (from $s = 0$ to $s = 3$) because we cannot estimate the equilibrium effort level exerted by district managers (the "winners" in each tournament).

[32] Ideally, one would like to compare the tournament wage bill to the wage bill from another second-best incentive pay scheme. However, this would require additional assumptions regarding the performance measures that the firms observe, upon which the workers' wages would depend. This is difficult because we observe no information on these matters in the data.

how the equilibrium restrictions of the tournament model allow us to recover parameters of interests, including the equilibrium effort levels in each hierarchical stage of the firm. We illustrated our estimation procedures using data from major retail chains in the United States. The estimates suggest that effort levels are generally higher at higher strata of employment within a firm, but that only a small fraction of the wage differential directly compensates workers for higher effort levels: at the estimated effort levels, we find that typically less than 50% of the observed wage differentials are for rewarding higher effort levels at higher levels of the corporations, implying that over half of the differentials arise purely to maintain incentives at lower rungs of the retailers.

There are also general extensions of the current work. One important implicit assumption made in this paper is that the tournament framework is correct, and no attempt has been made to test the tournament framework versus alternative models of the data generating process for the observed wage data. Moreover, we have assumed here that workers are homogeneous within a firm, across all hierarchical levels. It will be interesting to consider whether a model with both adverse selection and moral hazard (such as MacLeod and Malcomson (1988)) can be identified and estimated with the type of data considered in this paper.

A Extensions

In this appendix, we discuss two extensions of the model. In the first extension we accommodate effort at the highest level by modeling it as observed by firms, and in the second extension we allow workers to quit.

A.1 Effort at the highest level

So far we have assumed that the equilibrium effort for workers at the highest level is zero because there is no competition at that level. While this assumption is reasonable in the context of a sports tournament such as the soccer World Cup (the winner in the final gets the trophy and there is no more competition), in the retail chains industry workers at the highest level have to exert effort to keep their jobs. To accommodate effort at the highest level, we assume that firms observe such effort and offer wage contracts for those workers as follows:

$$\bar{W}_1 = \begin{cases} W_1 & \text{if } x_0 \geq x_0^* \\ 0 & \text{otherwise,} \end{cases}$$

where x_0^* is chosen by firms.

With the normalization that $c(x) = x$, payoff for workers at the highest level is

$$V_0 = \begin{cases} W_1 - x_0 & \text{if } x_0 \geq x_0^* \\ 0 - x_0 & \text{otherwise.} \end{cases}$$

As long as firms set x_0^* below W_1, in equilibrium workers at the highest level always choose to exert x_0^* and obtain $V_0 = W_1 - x_0^*$. We therefore obtain the following system of equations by replacing W_1 with $W_1 - x_0^*$ in the system of equations (12):

$$\begin{pmatrix} \frac{\gamma}{f_1}(1-p_1^*) & 0 & \cdots & 0 \\ 0 & \frac{\gamma}{f_2}(1-p_2^*) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\gamma}{f_S}(1-p_S^*) \end{pmatrix} *$$
$$\left[\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \beta_1 & (1-\beta_1) & \cdots & 0 \\ \beta_1\beta_2 & \beta_2(1-\beta_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1 \cdots \beta_{S-1} & (1-\beta_1)\beta_2 \cdots \beta_{S-1} & \cdots & (1-\beta_{S-1}) \end{pmatrix} \begin{pmatrix} W_1 - x_0^* \\ W_2 \\ \vdots \\ W_S \end{pmatrix} - \begin{pmatrix} W_2 \\ W_3 \\ \vdots \\ W_{S+1} \end{pmatrix} \right] = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_S^* \end{pmatrix}$$

We take the first approach to estimation. Firms choose $n_1, \ldots, n_S$, and x_0^* to maximize the following profit objective:

$$\max_{n_1,\ldots,n_S,x_0^*} \eta Q\left(n_0, n_1, \ldots, n_S; x_0^*, x_1, \ldots, x_S\right) - \sum_{s=0}^{S} n_s W_{s+1},$$

where

$$Q\left(n_0, n_1, \ldots, n_S; x_0^*, x_1, \ldots, x_S\right) = \left((n_0 x_0^*)^\rho + \sum_{s=1}^{S} \left(n_s x_s^*(\vec{n}, \vec{W}) \right)^\rho \right)^{1/\rho}.$$

The first-order conditions for this problem imply:

$$\begin{aligned} \frac{W_{s+1}}{W_{S+1}} &= \frac{(n_s x_s^*)^{\rho-1} x_s^* + \sum_{t=1}^{S} (n_t x_t^*)^{\rho-1} n_t \frac{\partial x_t^*}{\partial n_s}}{\left(n_S x_S^*\right)^{\rho-1} x_S^* + \sum_{t=1}^{S} (n_t x_t^*)^{\rho-1} n_t \frac{\partial x_t^*}{\partial n_S}}, \; s = 1, \ldots, S-1, \\ 0 &= (n_0 x_0^*)^{\rho-1} n_0 + \sum_{s=1}^{S} (n_s x_s^*)^{\rho-1} n_s \frac{\partial x_s^*}{\partial x_0^*}. \end{aligned}$$

Both $\frac{\partial x_t^*}{\partial n_s}$ and $\frac{\partial x_s^*}{\partial x_0^*}$ are obtained from the system of equations above. Compared to the specification in which effort at the highest level is assumed to be zero, here we add one more equation, the FOC w.r.t. x_0^*, and we estimate one more parameter, x_0^*.

The results for this extension are presented in Table 8. There are three features. First, the efforts (measured in money units) exerted by workers at the highest level (district managers) range from \$4,886 to \$13,066 and average at \$8,230. These efforts are the highest among the four levels of workers. Second, the efforts exerted by worker at the second highest level (store managers) are substantially lower than the estimates obtained from the original first approach (the average drops from \$3,419 to \$1,689). This is as expected, as the efforts that are now required of district managers reduce their payoffs, thus lowering the incentive for store managers to exert effort. Finally, compared to the original first approach results, the estimated efforts exerted by the lowest two levels (assistant store managers and sales staff) are little affected.

Table 9 presents the average percentage of wage differentials accounted for by effort differentials. These results are qualitatively similar to those reported in Table 6 for the original first approach. In particular, we continue to see that the δ's are typically below 50 and that some δ's are even negative, suggesting that a large portion of the wage differentials arises to maintain incentives for workers at lower levels.

[Table 8 about here.]

[Table 9 about here.]

A.2 Accommodating turnover rates

In this subsection, we describe how the model estimated in the paper could be extended to allow workers to quit the firm if they receive a wage offer which exceeds their current salary. In our modification of the model, we allow employees to leave the firm once they find their place in the firm (i.e., after they "lose" and remain in some stage s). We do not allow workers to quit while they are still "active" in the tournament. In addition, we maintain the assumption that the firm can only hire workers from the outside at the lowest level of the firm (level S), and that positions at levels $s < S$ can only be filled by advancing workers from lower stages.

In the dataset, we observe turnover rates $\lambda_0, \ldots, \lambda_S$, where λ_s is defined as the ratio of the total workers terminated in stage s divided by the total number of workers employed at stage s (which includes both the terminated and non-terminated workers).[33] These turnover rates are observed at the firm and year level, but only at the national level (i.e., not broken down by geographic locations).

Let $F_s,\ s = 0, \ldots, S$ denote the CDF of outside wages for employees in stage s. We can interpret the observed turnover rates as

$$\lambda_s = 1 - F_s(W_{s+1}),\ s = 0, \ldots, S. \tag{23}$$

That is, the observed turnover rate at stage s is interpreted as the probability of obtaining an outside wage offer exceeding the stage s salary, which is W_{s+1}. The workers' Bellman equation, for stage s, is

$$V_s = \max_x \left\{ p_s(x; x_s^*) V_{s-1} + (1 - p_s(x; x_s^*))\, E_{R_s} \max\left(W_{s+1}, R_s\right) \right\} \tag{24}$$

where R_s denotes the outside wage offer for a stage s worker, and $R_s \sim F_s$. Obviously,

$$E \max\left(W_{s+1}, R_s\right) = F_s(W_{s+1}) W_{s+1} + (1 - F_s(W_{s+1}))\, E\left[R_s | R_s > W_{s+1}\right].$$

Let $\tilde{W}_{s+1} \equiv E \max\left(W_{s+1}, R_s\right)$.

In order to estimate this amended model, we need to make additional assumption on the outside wage distributions $F_0, \ldots, F_S$. In the following, we assume that each of the wage distributions is uniform:

$$R_s \sim \mathcal{U}[0, \kappa_s]$$

where $\kappa_s,\ s = 0, \ldots, S$ are unknown parameters. With this distributional assumption:

$$\begin{aligned} F_s(W_{s+1}) &= \frac{W_{s+1}}{\kappa_s} = 1 - \lambda_s \\ &\Rightarrow\ \kappa_s = \frac{W_{s+1}}{1 - \lambda_s},\ s = 0, \ldots, S \end{aligned} \tag{25}$$

$$\begin{aligned} E\left[R_s | R_s > W_{s+1}\right] &= \frac{1}{2}\left(W_{s+1} + \kappa_s\right) \\ &= \frac{1}{2} W_{s+1} \frac{2 - \lambda_s}{1 - \lambda_s} \end{aligned}$$

$$\tilde{W}_{s+1} = W_{s+1} * \left[(1 - \lambda_s) + \frac{1}{2} \frac{\lambda_s (2 - \lambda_s)}{(1 - \lambda_s)} \right].$$

[33] Obviously, firms do not fire workers in our model. All terminations arise because the worker receive a higher outside wage offer and quit.

Hence, after plugging in these items into Eq. (24), we can estimate as before.

For the first approach, given observations of $\lambda_0, \ldots, \lambda_S$, we can back out $\kappa_0, \ldots, \kappa_S$ using Eq. (25). Then we can construct $\tilde{W}_1, \ldots, \tilde{W}_{S+1}$ and then estimate $x_1^*, \ldots, x_S^*$, as well as γ and ρ, using the system of equations (14), substituting in $\tilde{W}_s$ in place of W_s, for $s = 1, \ldots, S+1$.

For the second approach, we again assume that the observed wages are contaminated by additive measurement error. With this assumption:[34]

$$\begin{aligned}\tilde{W}_s &= (W_s - \epsilon_s) * \left[(1-\lambda_s) + \frac{1}{2}\frac{\lambda_s(2-\lambda_s)}{(1-\lambda_s)}\right], \ s = 1, \ldots, S \\ \tilde{W}_{S+1} &= W_{S+1} * \left[(1-\lambda_{S+1}) + \frac{1}{2}\frac{\lambda_{S+1}(2-\lambda_{S+1})}{(1-\lambda_{S+1})}\right].\end{aligned}$$

Let $\Psi_{1:S}$ denote the $S \times S$ diagonal matrix with $(1-\lambda_s) + \frac{1}{2}\frac{\lambda_s(2-\lambda_s)}{(1-\lambda_s)}$ in the s-th diagonal position. Then, for this case, the estimating equation corresponding to Eq. (20) in the main text is

$$\begin{aligned}\mathbf{A}\left[\mathbf{B}\Psi_{0:S-1}\left(\vec{W}_{1:S} - \vec{\epsilon}_{1:S}\right) - \Psi_{1:S}\left(\vec{W}_{2:S+1} - \vec{\epsilon}_{2:S}|0\right)\right] = \vec{c}' \ \Rightarrow \\ \vec{\epsilon} = -\tilde{\tilde{\mathbf{B}}}^{-1}\left[\mathbf{A}^{-1}\vec{c}' - \mathbf{B}\Psi_{0:S-1}\vec{W}_{1:S} + \Psi_{1:S}\vec{W}_{2:S+1}\right]\end{aligned}$$

where

$$\tilde{\tilde{\mathbf{B}}} \equiv \mathbf{B}\Psi_{0:S-1} - \tilde{\Psi}_{2:S-1}$$

and $\tilde{\Psi}_{2:S-1}$ denotes the $\Psi_{1:S-1}$ matrix bordered at the bottom and the left with, respectively, a row and column of zeros.

We continue to assume that the observed employment levels for each firm, year, and geographic location are the desired employment levels for the firm. For both the first and second approaches, the presence of turnover implies that firms must hire, and promote, more workers in order to achieve the desired employment levels at each stage of the firm. Hence, we need to reconstruct our measures of the number of competitors at each stage. The L_s's (number of subgroups) stay the same. We must redefine the number of "losers" and "contenders" in each stage as (for $s = 0, \ldots, S$):

$$\begin{aligned}\tilde{n}_s &= \frac{n_s}{1-\lambda_s} \\ \tilde{m}_s &= \sum_{s=0}^{s} \frac{n_s}{1-\lambda_s}.\end{aligned}$$

[34] As above, we need to assume that one of the wages – in this case, W_{S+1} – is not contaminated by measurement error.

Results using the second approach for the four retailers for which we were able to obtain turnover rates at the sales staff, assistant store manager, store manager, and district manager levels, for at least a single year, are reported in Table 10. Noticeably, the J-test specification checks have substantially lower p-values for this model, relative to the results in the main text, which are obtained from a model which does not accommodate turnover.[35]

However, the qualitative implications of the results are similar to those in the main text. For the model with turnover, Table 11 reports the average percentage of wage differentials accounted for by effort differentials, and Table 12 reports the comparison between the observed wage bill and the counterfactual wage bill implied by the estimates. The patterns in these tables are consistent with their counterparts in the no-turnover model, Tables 6 and 7, respectively. In particular, we continue to find that an important purpose of the wage differentials is to provide workers incentives to exert efforts (Table 11), and that the counterfactual wage bill (when workers' effort levels are assumed to be contractible) is always lower than the observed wage bill, with the percentage differences ranging from 19% to 33% (Table 12).

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

[35] However, the second-order conditions in Eq. (22) continue to hold for these estimates.

References

Bolton, P., and M. Dewatripont (2005): *Contract Theory.* MIT Press.

Bronars, S., and G. Oettinger (2001): "Effort, Risk-Taking, and Participation in Tournaments: Evidence from Professional Golf," Manuscript, University of Texas at Austin.

Ehrenberg, R., and M. Bognanno (1990a): "Do tournaments have incentive effects?," *Journal of Political Economy*, 98, 1307–1324.

——— (1990b): "The incentive effects of tournaments revisited: Evidence from the European PGA Tour," *Industrial and Labor Relations Review*, 43, S74–S88.

Eriksson, T. (1999): "Executive Compensation and Tournament Theory: Empirical Tests on Danish Data," *Journal of Labor Economics*, 17, 262–280.

Ferrall, C. (1996): "Promotions and Incentives in Partnerships: Evidence from Major U.S. Law Firms," *Canadian Journal of Economics*, 29, 811–827.

——— (1997): "Empirical Analysis of Occupational Hierarchies," *Journal of Human Resources*, 34, 1–34.

Ferrall, C., and A. Smith (1999): "A Sequential Game Model of Sports Championship Series: Theory and Estimation," *Review of Economics and Statistics*, 81, 704–719.

Green, J., and N. Stokey (1983): "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91, 349–364.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–74.

Holmstrom, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340.

Holmstrom, B., and P. Milgrom (1994): "The Firm as an Incentive System," *American Economic Review*, 84, 972–991.

Hong, H., and M. Shum (2006): "Using Price Distributions to Estimate Search Costs," *RAND Journal of Economics*, 37, 257–275.

Knoeber, C., and W. Thurman (1994): "Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production," *Journal of Labor Economics*, 12, 155–179.

Lazear, E., and S. Rosen (1981): "Rank-order Tournaments as Optimal Labor Contracts," *Journal of Political Economy*, 89, 841–864.

MacLeod, W., and J. Malcomson (1988): "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy*, 96, 832–854.

Main, B., C. O'Reilly, and J. Wade (1993): "Top executive pay: Tournament or teamwork?," *Journal of Labor Economics*, 12, 606–628.

McLaughlin, K. (1989): "Aspects of Tournament Models: A Survey," in *Research in Labor Economics*, ed. by R. Ehrenberg. JAI Press, Inc.

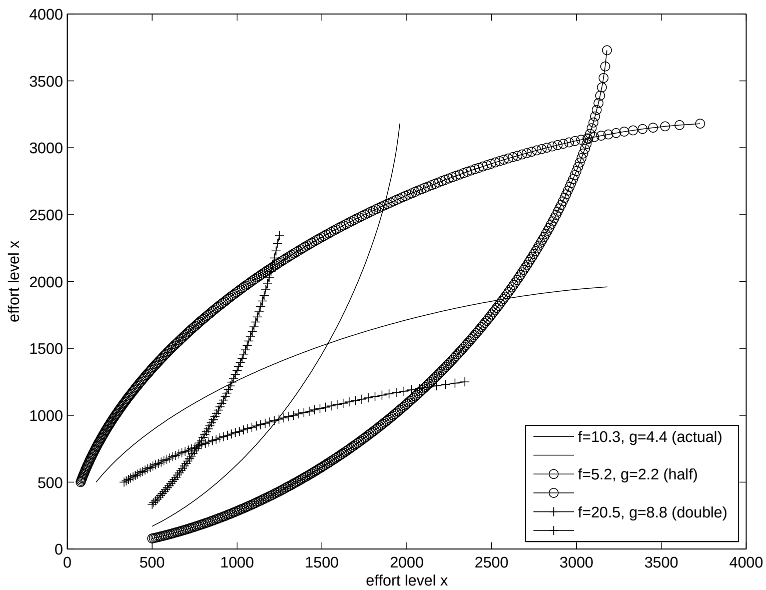
Qian, Y. (1994): "Incentives and Loss of Control in an Optimal Hierarchy," *Review of Economic Studies*, 61, 527–544.

Ridder, G., and G. van den Berg (1998): "An Empirical Equilibrium Search Model of the Labor Market," *Econometrica*, 66, 1183–1221.

Rosen, S. (1986): "Prizes and Incentives in Elimination Tournaments," *American Economic Review*, 76, 701–715.

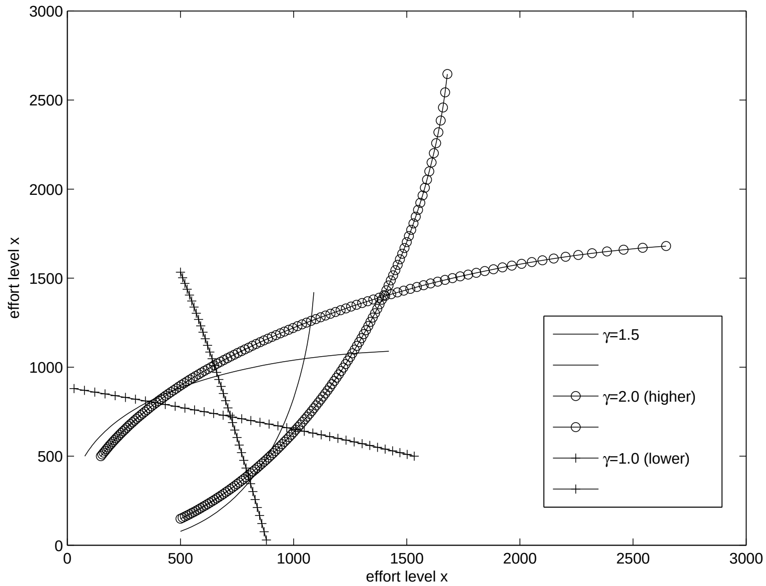
Zheng, X., and T. Vukina (2007): "Efficiency Gains from Organizational Innovation: Comparing Ordinal and Cardinal Tournament Games in Broiler Contracts," Forthcoming, *International Journal of Industrial Organization*.

Figure 1: Comparative statics of tournament model



Effects of changes in toughness of competition

Figure 2: Comparative statics of tournament model



Effects of changes in γ (curvature parameter for $h(\cdot)$ function)

Table 1: Average wage and tournament parameters in major retail chains

Store identifier	# mkt/yr obs.	Dist mgr ($s = 0$)	Store mgr ($s = 1$)	Asst store mgr ($s = 2$)	Sales staff ($s = 3$)
A (Clothing)	13	31978.39	19171.11	13875.25	8645.20
		(3968.13)	(1550.75)	(1835.10)	(894.61)
		1.23	4.69	7.77	2.08
			1/4.92	1.32/3.00	3.00/3.48
B (Athletic footwear)	8	37327.43	18051.92	11334.86	8016.21
		(5421.04)	(1970.93)	(553.25)	(495.38)
		1.00	15.50	17.00	9.00
			1/16.5	1.07/2.19	2.19/2.78
C (Non-athletic Footwear)	9	25302.87	14628.8	8048.46	6069.92
		(2654.97)	(916.31)	(348.82)	(1755.60)
		1.22	13.33	16.22	9.11
			1/11.56	1.11/2.39	2.39/3.18
D (High-end specialty)	12	45563.5	24468.35	16019.71	9665.78
		(4838.46)	(4261.09)	(1415.18)	(999.65)
		1.00	2.50	2.17	1.83
			1/3.5	1.5/2.46	2.46/3.40
E (Clothing)	38	35767.75	20855.25	14304	11025.61
		(6502.10)	(2998.42)	(2310.14)	(1196.92)
		1.05	14.58	18.32	21.87
			1/14.71	1.12/2.47	2.47/3.90
F (High-end specialty)	12	41695.05	25456.24	18018.76	9943.56
		(7138.76)	(3971.62)	(2238.46)	(1358.79)
		1.00	6.08	5.33	18.50
			1/7.08	1.24/2.10	2.10/6.04
G (Children)	13	40410.57	25120.34	18464.37	11328.93
		(7636.94)	(4064.84)	(1916.36)	(1006.76)
		1.23	4.54	12.38	7.92
			1/4.77	1.32/4.12	4.12/5.89
H (Athletic footwear)	21	29811.7	17989.24	10203.93	8403.99
		(3389.65)	(1799.55)	(1013.59)	(748.64)
		2.48	19.76	49.24	9.33
			1/8.75	1.18/3.71	3.71/4.53
I (Clothing)	42	38017.61	21691.99	16480.2	10361.4
		(3214.85)	(1720.96)	(3088.39)	(714.72)
		8.36	50.81	111.07	126.62
			1/7.33	1.19/3.53	3.53/6.46
J (Eyewear)	14	45495.18	25910.98	20066.61	9382.12
		(4900.57)	(2287.53)	(2478.05)	(600.92)
		1.36	11.71	10.79	15.57
			1/9.55	1.15/2.11	2.11/3.46
K (Clothing)	21	43205.93	24275.2	18040.99	12208.09
		(4135.74)	(2255.50)	(1466.40)	(1175.39)
		1.52	7.81	19.10	14.95
			1/6.21	1.27/4.35	4.35/7.07
L (Household items)	17	47210.02	32884.84	23046.81	10885.75
		(5267.12)	(3002.86)	(2175.07)	(2252.73)
		1.00	1.94	7.06	16.41
			1/2.94	1.79/5.66	5.66/13.07
M (Non-athletic Footwear)	45	42655.28	20606.52	10298.17	8651.43
		(4090.03)	(2132.92)	(974.77)	(703.60)
		3.76	32.73	27.31	43.73
			1/9.15	1.32/3.15	3.15/5.20
N (Books)	20	32946.63	19329.51	12181.77	8871.47
		(3571.39)	(2123.77)	(1400.18)	(971.85)
		1.15	4.75	5.50	6.40
			1/5.5	1.35/2.69	2.69/4.35

Top entry in each cell is annual salary (in 1986 dollars). Second entry is standard deviation of salary. Third entry in each cell gives n_s, and fourth entry gives g_s/f_s, the ratio of "winners" from each subgroup to the size of each subgroup. For Sales Staff, salary is calculated as hourly wage*40 hours*50 weeks.

Table 2: Retailer characteristics

Definitions:
$DCLOTH$: =1 if clothing retailer
$DFOOT$: =1 if footwear retailer
$DHOUSE$: =1 if housewares retailer
$DMALL$: =0 if stores mostly located in shopping centers; =1 if in shopping malls
$DLARGE$: =0 if stores mostly < 20,000 sq. ft.; =1 if $\geq$ 20,000
$DCOMMIS$: =1 if over 25% of sales-staff pay is based on commissions

Retail Chain	$DCLOTH$	$DFOOT$	$DHOUSE$	$DMALL$	$DLARGE$	$DCOMMIS$
A	1	0	0	1	1	0
B	0	1	0	1	1	1[a]
C	0	1	0	1	1	1[b]
D	0	0	1	1	1	0
E	1	0	0	1	1	0
F	0	0	1	1	1	1[c]
G	0	0	0	1	0	0
H	0	1	0	0	0	0
I	1	0	0	1	0	0
J	0	0	0	1	1	0
K	1	0	0	1	0	0
L	0	0	1	0	0	0
M	0	1	0	0	1	0
N	0	0	0	1	1	0

[a]Only for 1999
[b]Only for 1998
[c]Only for 1999

Table 3: Average Parameters for Retail Chains: First Approach

RETAIL CHAIN	# (Yr-Location) Obs.[a]	Avg x_1 (stdev)	Avg x_2 (stdev)	Avg x_3 (stdev)	Avg γ (stdev)	Avg ρ (stdev)
A	6	4556.08 (2322.71)	1639.03 (870.62)	398.82 (172.03)	2.03 (1.07)	-1.56 (0.46)
B	8	2309.14 (991.30)	2521.79 (689.53)	527.70 (203.40)	2.07 (0.65)	-1.71 (0.82)
C	5	1388.00 (788.33)	1651.83 (468.95)	209.73 (55.91)	1.15 (0.32)	-0.83 (0.29)
D	4	9063.02 (6284.87)	1208.78 (865.31)	491.68 (381.74)	1.50 (1.00)	-0.11 (1.19)
E	25	1315.08 (905.33)	1967.67 (1443.30)	542.39 (433.93)	1.41 (1.04)	-8.33 (28.35)
F	0[b]	—	—	—	—	—
G	3	1989.53 (1800.92)	1024.25 (1065.05)	259.74 (298.83)	1.04 (1.25)	-2.43 (1.48)
H	6	2020.78 (2084.50)	1454.38 (489.05)	215.72 (91.43)	1.05 (0.40)	-1.09 (0.53)
I	20	2373.51 (836.12)	1516.83 (397.51)	299.66 (134.12)	1.36 (0.47)	-2.45 (0.77)
J	2	820.72 (1147.47)	768.11 (1071.53)	852.58 (1194.34)	0.55 (0.75)	-10.74 (2.57)
K	13	3523.32 (1994.66)	1314.03 (896.51)	372.74 (408.95)	1.32 (1.04)	-3.83 (4.76)
L	2	5533.15 (1646.92)	2993.55 (971.07)	1787.60 (327.81)	2.69 (0.90)	-3.34 (0.74)
M	23	5488.56 (4237.27)	2656.54 (1105.44)	500.64 (230.57)	1.82 (0.75)	-2.18 (1.56)
N	14	5409.00 (2794.65)	2058.66 (720.71)	651.07 (190.33)	2.31 (0.64)	-3.49 (3.82)

[a] We only included those (Yr-Location) observations for which the calculated effort levels satisfy the two conditions given at the end of Section 3.3 of the main text.

[b] There were no (year-location) observations which both satisfied the two model optimality conditions, as well as yielded convergent estimates for the nonlinear solver used to solve the supply-side first-order conditions.

Table 4: Regressions of parameters on retailer characteristics: first approach

	Dependent Var:				
Regressors:	γ	$x_1/1000$	$x_2/1000$	$x_3/1000$	ρ
DMALL	-0.778**	-2.085**	-0.639*	-0.513***	1.831
	(0.391)	(0.956)	(0.321)	(0.151)	(5.921)
DLARGE	0.341**	0.698	0.564***	0.182***	-3.591
	(0.177)	(0.612)	(0.206)	(0.069)	(2.686)
DCLOTH	-0.338	-2.216**	0.255	-0.103	-0.055
	(0.256)	(0.885)	(0.297)	(0.099)	(3.877)
DFOOT	-0.873**	-2.444**	0.354	-0.574***	3.460
	(0.405)	(1.113)	(0.374)	(0.157)	(6.139)
DHOUSE	-0.488	2.134	-0.271	0.069	12.294*
	(0.439)	(1.517)	(0.510)	(0.170)	(6.658)
DCOMMIS	1.097**	—	—	0.520***	-3.754
	(0.511)	(—)	(—)	(0.198)	(7.744)
year dummies	yes	yes	yes	yes	yes
city dummies	yes	yes	yes	yes	yes
N	131	131	131	131	131
R^2	0.305	0.316	0.341	0.388	0.224

Table 5: Parameter estimates: Second approach

Retail Chain	#obs[a]	x_1: SM effort (std er)	x_2: ASM effort (std er)	x_3: Sales effort (std er)	γ[b] (std er)	J-statistic[c] (p-value)
A	8	4362.4	832.9	487.9	1.94	0.1544
		(249.3)	(181.2)	(78.1)	(0.17)	(0.9846)
B	8	2673.1	2262.7	583.59	2.23	0.4129
		(680.4)	(200.3)	(54.8)	(0.19)	(0.9376)
C	9	2255.5	2183.7	362.74	2.49	0.2756
		(120.6)	(178.1)	(36.3)	(0.15)	(0.9646)
D	12	3758	2083.7	1493.0	1.57	0.8159
		(735.9)	(870.0)	(88.5)	(0.28)	(0.8457)
E	38	2557.5	1025.1	778.5	2.46	0.0445
		(1222.1)	(506.9)	(67.8)	(0.43)	(0.9975)
F	12	2977.7	2812.9	1600.3	1.85	0.0609
		(177.49)	(416.65)	(217.26)	(0.32)	(0.9961)
G	13	3321.4	2465.8	570.9	1.98	0.4706
		(176.4)	(618.5)	(81.4)	(0.07)	(0.9253)
H	21	5613.0	1302.0	96.5	3.86	0.1691
		(494.9)	(89.5)	(33.5)	(0.44)	(0.9824)
I	42	4253.0	1546.6	601.7	2.21	0.1439
		(384.8)	(377.7)	(89.8)	(0.21)	(0.9861)
J	14	1912.7	3341.8	1551.0	1.34	0.1731
		(776.8)	(2135.6)	(1174.5)	(0.22)	(0.9818)
K	21	3779.0	1930.5	435.4	1.63	0.0151
		(944.3)	(355.4)	(33.5)	(0.19)	(0.9995)
L	17	4140.3	3507.3	1030.9	2.47	0.1744
		(1491.5)	(1830.9)	(326.2)	(0.76)	(0.9816)
M	45	3316.1	1412.9	1661.4	2.34	0.2271
		(2045.7)	(3086.2)	(1250.4)	(1.59)	(0.9731)
N	16	3709.8	2122.0	510.8	1.87	0.4765
		(3699.1)	(964.2)	(116.4)	(0.93)	(0.9240)

[a]As in Table 1, each observation denotes that wages w_1, w_2, w_3, w_4 and n_0, n_1, n_2, n_3 were observed for a particular (location,year) combination. That is, the number of observations is equal to LST, using the notation in Section 3.3.

[b]Exponent on $h(\cdot)$.

[c]asymptotically distributed $\chi^2(3)$ under null that the moment conditions in Eq. (21) hold.

Table 6: Average percentage of wage differentials accounted for by effort differentials

$$\delta_s \equiv 100 * \left(\frac{x_s - x_{s+1}}{w_{s+1} - w_{s+2}}\right), \; s = 1, 2$$

	First approach[a]		Second approach[b]	
RETAIL CHAIN	mean δ_1 (stdev)[c]	mean δ_2 (stdev)	δ_1 (stder)[d]	δ_2 (stder)
A	51.25	32.25	65.03	8.25
	(36.64)	(12.21)	(6.39)	(5.14)
B	-2.12	60.86	6.38	52.47
	(17.05)	(20.66)	(10.08)	(7.20)
C	-3.90	93.43	1.11	114.53
	(8.26)	(18.24)	(3.76)	(12.93)
D	100.51	14.45	21.44	9.69
	(75.81)	(12.50)	(17.99)	(14.29)
E	-4.79	23.99	28.19	2.65
	(19.99)	(156.10)	(19.02)	(5.76)
F			2.36	15.31
			(7.02)	(7.85)
G	14.42	15.43	13.63	28.35
	(12.53)	(13.53)	(11.38)	(10.16)
H	5.54	64.17	57.00	78.05
	(18.81)	(36.30)	(7.36)	(7.03)
I	13.83	29.65	78.05	18.82
	(11.10)	(6.61)	(20.81)	(9.16)
J	0.83	-0.53	-26.79	17.53
	(1.20)	(0.78)	(35.65)	(16.40)
K	33.43	18.85	35.74	27.17
	(23.52)	(10.01)	(20.39)	(6.60)
L	25.07	10.00	6.81	21.70
	(4.71)	(6.41)	(35.14)	(18.79)
M	27.80	121.28	19.06	-17.17
	(40.50)	(57.60)	(26.85)	(157.22)
N	48.41	43.90	23.33	55.88
	(46.51)	(20.38)	(45.91)	(32.10)

[a] Corresponding to recovered parameter values summarized in Table 3.

[b] Corresponding to GMM estimates from Table 5.

[c] δ_1 was calculated separately for each retail chain, year, and geographic location. We report the standard deviations of δ_1 across all years and geographic locations, for a given retail chain.

[d] Standard error computed using delta method, for the estimates in Table 5.

Table 7: Observed vs. counterfactual total wage bill implied by estimates
Counterfactual: if effort were contractible

	First approach			Second approach[a]		
Retail Chain	Observed wage bill ($mills)	Counterfactual wage bill ($mills)	% Diff (Obs-CF)/Obs	Observed[b] wage bill ($mills)	Counterfactual wage bill ($mills) (stder)	% Diff (Obs-CF)/Obs
A	1.49	1.13	0.24	1.808	1.301 (0.018)	0.28
B	4.35	3.16	0.27	4.356	3.160 (0.104)	0.27
C	1.33	1.02	0.23	3.347	2.498 (0.337)	0.25
D	0.39	0.31	0.21	1.349	0.836 (0.239)	0.38
E	20.62	15.96	0.23	30.344	23.793 (0.934)	0.22
F	—	—	—	5.286	3.872 (0.055)	0.27
G	0.91	0.68	0.25	5.893	4.239 (0.109)	0.28
H	2.91	2.27	0.22	19.318	17.222 (0.197)	0.11
I	58.95	44.84	0.24	178.431	137.952 (1.785)	0.23
J	1.60	0.84	0.48	9.296	5.344 (0.254)	0.43
K	9.29	7.46	0.20	14.977	11.815 (0.192)	0.21
L	0.73	0.50	0.32	6.939	5.044 (0.224)	0.27
M	32.75	26.32	0.20	58.289	41.084 (3.123)	0.30
N	3.41	2.77	0.19	3.343	2.714 (0.324)	0.19

[a]Corresponding to GMM estimates from Table 5.
[b]Note that figures in first and third columns may not coincide due to (i) rounding errors; and (ii) for some (firm/geographic locations/year) observations, we were not able to obtain convergent estimates for the first approach.

Wages bills only for store managers, assistant store managers, and sales staff.

Table 8: Average Parameters for Retail Chains: First Approach with Effort at the Highest Level

Retail Chain	# (Yr-Location) Obs.[a]	Avg x_0^* (stdev)	Avg x_1 (stdev)	Avg x_2 (stdev)	Avg x_3 (stdev)	Avg γ (stdev)	Avg ρ (stdev)
A	2	5045.89	2109.35	2009.95	376.05	2.09	-1.29
		(2014.57)	(1726.94)	(1740.64)	(145.03)	(1.34)	(1.43)
B	6	10269.3	876.89	2824.9	495.08	1.88	-1.19
		(2127.49)	(248)	(869.85)	(143.46)	(0.45)	(0.43)
C	5	5542.02	670.3	1587.56	205.1	1.1	-0.67
		(1050.44)	(369.94)	(507.44)	(56.01)	(0.35)	(0.32)
D	2	13065.91	6905.79	5799.9	1303.1	4.29	-0.37
		(8732.43)	(2780.73)	(4784.27)	(1192.94)	(2.99)	(0.51)
E	22	7719.66	593.43	2056.91	493.49	1.29	-1.79
		(3780.57)	(310.4)	(1517.48)	(367.85)	(0.87)	(1.02)
G	3	6316.65	2229.35	1780.6	738.01	2.03	-1.74
		(3328.55)	(1460.68)	(955.63)	(740.4)	(1.56)	(1.45)
H	5	4885.94	1081.26	1404.94	222.06	0.99	-0.67
		(824.6)	(948.41)	(262.85)	(63.51)	(0.22)	(0.33)
I	20	7456.69	1170.03	1594.39	295.71	1.35	-2.18
		(1294.54)	(292.57)	(528.6)	(129.87)	(0.47)	(0.62)
J	1	12193.99	2880.49	6981.91	3971	9.22	-2.34
		(-)	(-)	(-)	(-)	(-)	(-)
K	13	6963.95	2016.8	1650.33	522.44	1.58	-2.32
		(2639.77)	(983.94)	(1372.91)	(699.38)	(1.55)	(2.5)
L	2	5555.04	2993.45	3799.26	1946.44	2.97	-0.78
		(2160.41)	(623.53)	(1070.49)	(274.27)	(0.12)	(0.35)
M	20	11811.45	2592.14	3194.08	532.32	1.99	-1.26
		(2624.33)	(1932.64)	(1156.24)	(233.23)	(0.8)	(0.75)
N	13	7711.23	2344.37	2432.69	653.43	2.22	-1.23
		(3369.86)	(1191.82)	(698.52)	(207.2)	(0.73)	(0.78)

[a] We only included those (Yr-Location) observations for which the calculated effort levels satisfy the two conditions given at the end of Section 3.3 of the main text.

Table 9: Average percentage of wage differentials accounted for by effort differentials: first approach with effort at the highest level

$$\delta_s \equiv 100 * \left(\frac{x_s - x_{s+1}}{w_{s+1} - w_{s+2}} \right), \ s = 0, 1, 2$$

	First approach[a]		
RETAIL CHAIN	mean δ_0 (stdev)[b]	mean δ_1 (stdev)	mean δ_2 (stdev)
A	29.91 (3.36)	2.18 (1.52)	41.34 (41.69)
B	51.33 (3.46)	-28.19 (8.17)	70.36 (21.78)
C	42.13 (3.44)	-13.52 (4.33)	89.53 (22.97)
D	16.32 (45.61)	14.5 (23.01)	62.42 (23.26)
E	43.23 (11.94)	-18.7 (15.63)	17.2 (160.08)
G	31.15 (12.78)	7.32 (15.42)	18.35 (6.07)
H	35.6 (7.62)	-5.35 (9.25)	59.75 (26.98)
I	40.58 (6.51)	-6.36 (6.98)	31.27 (8.3)
J	60.65 (-)	-68.99 (-)	30.83 (-)
K	29.38 (12.38)	7.14 (19.88)	20.72 (11.88)
L	21.45 (20.49)	-7.36 (16.26)	15.55 (12.7)
M	44.6 (11.41)	-6.78 (20.07)	151.79 (71.26)
N	37.69 (14.35)	-0.56 (17.85)	54.4 (16.19)

[a] Corresponding to recovered parameter values summarized in Table 8.

[b] δ_s was calculated separately for each retail chain, year, and geographic location. We report the standard deviations of δ_s across all years and geographic locations, for a given retail chain.

Table 10: Parameter estimates: incorporating turnover rates

Retail Chain	#obs	x_1 (std er)	x_2 (std er)	x_3 (std er)	γ^a (std er)	J-statistic^b (p-value)
A	32	3322.3	3588.8	207.81	1.107	1.853
		(506.02)	(2106.6)	(504.02)	(0.238)	(0.603)
B	32	1822.8	3733.1	1861.4	1.822	1.887
		(649.07)	(5201.8)	(6434.5)	(0.276)	(0.596)
G	24[c]	3614.6	2448.2	691.47	0.918	2.018
		(772.51)	(10047)	(1387.5)	(0.056)	(0.569)
M	60	3057.6	5681.9	2549.3	1.128	1.730
		(1497.4)	(4339.5)	(23681)	(0.187)	(0.630)

[a]Exponent on $h(\cdot)$.

[b]asymptotically distributed $\chi^2(2)$ under null that the moment conditions in Eq. (21) hold.

[c]Fewer observations are available for this retail chain (as compared to the number of observations used for the results in Table 5) because turnover rates were not reported for some years. Same applies for Retail Chain M.

Table 11: Average percentage of wage differentials accounted for by effort differentials: incorporating turnover rates

	Second approach[a]	
Retail Chain	$\% \left(\frac{x_1 - x_2}{w_1 - w_2}\right)$ (stder)	$\% \left(\frac{x_2 - x_3}{w_2 - w_3}\right)$ (stder)
A	-4.9105	80.879
	(36.219)	(61.988)
B	-29.701	58.486
	(82.244)	(143.26)
G	20.257	33.765
	(183.45)	(177.58)
M	-25.35	230.93
	(53.648)	(2057.6)

[a] Corresponding to GMM estimates from Table 10.

Table 12: Observed vs. counterfactual total wage bill implied by estimates: incorporating turnover rates

	Second approach[a]		
Retail Chain	Observed[b] wage bill ($mills)	Counterfactual wage bill ($mills) (stder)	% Diff (Obs-CF)/Obs
A	1.808	1.455 (0.189)	0.19
B	4.356	2.923 (1.292)	0.33
G	1.478	1.190 (0.301)	0.19
M	37.245	27.203 (43.21)	0.27

[a]Corresponding to GMM estimates from Table 10.

[b]Note that figures in first and third columns may not coincide due to (i) rounding errors; and (ii) for some (firm/geographic locations/year) observations, we were not able to obtain convergent estimates for the second approach.

Wages bills only for store managers, assistant store managers, and sales staff.