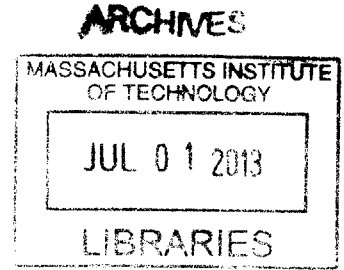Sustained Rents in Imperfect Labor Markets: Essays on Recruitment, Training, and Incentives

by

Alan Benson

B.S. Industrial and Labor Relations
Cornell University, 2007

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

© Alan Benson 2013. All rights reserved.

Author: _____
Alan Benson
Ph. D. Candidate
Thesis Author

Certified by: _____
Paul Osterman
NTU Professor of Management
Thesis Co-Supervisor

Certified by: _____
David Autor
Professor of Economics
Thesis Co-Supervisor

Accepted by: _____
Ezra Zuckerman Sivan
Professor of Strategic Management
Chair of the Sloan PhD Committee

Sustained Rents in Imperfect Labor Markets: Essays on Recruitment, Training, and Incentives

by

Alan Benson

Submitted to the MIT Sloan School of Management
on April 18, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Management

Abstract

This thesis is composed of three papers, each relating to labor market imperfections and their implications for firms' staffing practices. In the first paper, I examine why hospitals provide direct financial support to nursing schools and faculty. This support is striking because nursing education is clearly general, clearly paid by the firm, and information asymmetries appear minimal. Using AHA and survey data, I find hospitals employing a greater share of their MSA's registered nurses are more likely to provide such support, net of size and other institutional controls. I interpret this result as evidence that technologically-general skills training may be made *de facto*-specific by mobility frictions.

In the second paper, I present a theory of couples' job search whereby women sort into lower-paying geographically-dispersed occupations due to expectations of future spouses' geographically-clustered occupations and (thereby) inability to relocate for work. Results confirm men segregate into geographically-clustered occupations, and that these occupations involve more-frequent early career relocations for both sexes. I also find that the minority of the men and women who depart from this equilibrium experience delayed marriage, higher divorce, and lower earnings. Results are consistent with the theory's implication that marriage and mobility expectations foment a self-fulfilling pattern of occupational segregation, with individual departures deterred by earnings and marriage penalties.

In the third paper, I examine the use and misuse of authority and incentives in organizational hierarchies. Through a principal-supervisor-agent model inspired by sales settings, I propose organizations delegate authority over salespeople to front-line sales mangers because they can decompose performance measures into ability and luck. The model yields the result that managers on the cusp of a quota have a unique personal incentive to retain and adjust quotas for poor performing subordinates, permitting me to distinguish managers' interests from those of the firm. I parametrically estimate the model using detailed person-transaction-level microdata from 244 firms that subscribe to a "cloud"-based service for automating transaction processing and compensation. I estimate 13-15% of quota adjustments and retentions among poor performers are explained by the managers' unique personal interest in meeting a quota. I use agency theory to evaluate firms' mitigation practices.

Thesis Co-Supervisor: Paul Osterman, NTU Professor of Human Resources and Management
Thesis Co-Supervisor: David Autor, Professor of Economics

Acknowledgements

Contents

Chapter 1

Firm-Sponsored General Education and Mobility Frictions: Evidence from Hospital Sponsorship of Nursing Schools and Faculty[1]

Why do hospitals provide direct financial support to local nursing schools and faculty? To hospital nursing officers, the answer is simple and intuitive: it is a long-term investment to fill nursing vacancies and address local nursing shortages. However, human capital theory implies these resources would be better-spent raising wages, given the literature's general consensus that nursing markets are competitive (Adamache and Sloan 1982; Sullivan 1989; Hirsch and Schumacher 1995, 2005).

Using unique data on staffing practices collected from surveys matched with American Hospital Association (AHA) data, this paper presents evidence that weak local competition prompts hospitals to provide direct financial support to nursing schools and faculty. This finding makes three distinct contributions. First, because this form of firm-sponsored general education supports students for whom information asymmetries appear to be minimal and for whom support cannot be deducted from wages, it corroborates Acemgolu and Pischke's (1999a) hypothesis that technologically-general skills may be made *de facto* specific by mobility frictions. Second, it infers monopsony rents in a market whose evidently highly-elastic labor supply calls to question monopsony's relevance to labor markets. Third, it poses an explanation for long stretches of perceived nursing shortages in the U.S. and abroad. In short, this study contributes a case of search-induced monopsony to the human capital literature, a case of firm-sponsored general education induced by search frictions to the monopsony literature, and by combining the human capital and monopsony literatures, this study contributes a set of policy recommendations.

In the classic human capital model, employers do not pay for general training because doing so requires raising wages commensurately to prevent competitors from opportunistically recruiting trained employees (Becker 1964). For incumbent employers to recoup training costs, productivity must rise faster than wages with respect to training, implying imperfect competition and "wage compression" (Acemoglu and Pischke 1999a). Recent research emphasizes how private information may prompt employers to bear the costs of general skills training. In

particular, incumbent firms may place more-informed bids than competitors following an investment in general skills because they possess private information of the quality of education (Katz and Ziderman 1990, Chang and Wang 1996) or workers' innate abilities (Acemoglu and Pischke 1997, 1998), or acquire private information through training (Autor 2001).

However, from the perspective of human capital theory, hospitals' direct financial support for nursing schools and faculty is remarkable; it overcomes the typical dilemmas of empirical human capital research and allows weak local competition to be largely-isolated from the firm-sponsored general education's usual explanations. Because nursing students are pursuing degree programs and a nationally-recognized nursing license, training is clearly general. Because grades, licensure, and the quality of the nursing school are largely publicly-observed, information asymmetries appear to be minimal. Because support for nursing schools benefits students who are not employees, training costs are clearly borne by employers and cannot be discounted from wages (although alternative support could be, such as tuition reimbursement for current employees pursuing nursing licensure). Because AHA data allow firms' local employment shares to be calculated, it is relatively-feasible to distinguish firms facing stark local competition from those facing weak local competition.

Using the matched surveys, I estimate that 45% of hospitals provide financial support to nursing schools (either independently or jointly), and about one-third pay for nursing faculty. Net of size and other institutional controls, I estimate that hospitals in the top quartile of nurse employment share are about three times as likely to support nursing schools and pay for nursing faculty as hospitals in the bottom quartile of nurse employment share. I interpret these results as evidence that weak local competition motivates hospitals to pay for nurses' basic education. Lastly, I discuss how the applicability of human capital and monopsony models implied by these results can inform manpower policies designed to boost the supply of nurses.

I.   Hospitals' Justifications for Nursing School Support

Unfortunately, much of the literature on hospitals' staffing practices is composed of descriptive studies of specific cases, with less attention to explaining the wide differences in nursing vacancies and staffing policies across hospitals and regions. Recent major studies of US hospitals' educational strategies include the Community Tracking Study (see May, Bazzoli, and Gerland 2006) and the Future of Nursing Study of the Institute of Medicine and Robert Wood Johnson Foundation (IOM 2010).

In 2008 (when this study's first survey was taken), the nursing administration literature and administrators interviewed for this study referred to support for local nursing education as an "investment" and a method of addressing a perceived nursing shortage.[2] While labor shortages are not a well-defined equilibrium concept, the belief in persistent local nursing shortages was

---

[2]   Perceptions of a nursing shortage have declined since the recession. From 2000 to 2008, registered nursing unemployment was between 1% and 2%, while about 10% of hospitals' budgeted positions for RNs were vacant. Despite the recession, health care and nursing has added jobs, and unemployment among RNs remained about 2-3% in 2009 and 2010 (Benson 2011; Buerhaus Auerbach, and Staiger 2003).

widespread among the hospital administrators who pushed providing support to local schools. A majority of RNs (82%), physicians (81%), hospital CEOs (68%), and hospital Chief Nursing Officers (74%) surveyed by Buerhaus et al. (2007) believed their hospital is facing a nursing shortage. Similarly, 77% of Chief Nursing Officers surveyed in 2008 for this study believed their hospital "is currently experiencing a shortage of registered nurses."

For hospital nursing officers, the shortage manifests in difficulty recruiting nurses, implicitly at prevailing wages. In July 2007, the AHA estimated that American hospitals posted 116,000 budgeted but vacant positions for FTE registered nurses, or 8.1% of all hospital registered nursing positions. In 2008, nursing officers surveyed for this study reported a mean of 10.5% of RN positions were "budgeted but vacant." To contrast, estimated unemployment among RNs was 1.5% in 2008. In a series of articles, Buerhaus, Staiger, and Auerbach (2000, 2003, 2007a, 2007b) projected the vacancy gap will be widened by accelerating retirement of RNs and the increased demand for health care among baby-boomers. The 2008-2009 edition of the BLS Occupational Outlook Handbook (2008) plainly declared a "shortage" for both clinical and faculty nurses. Several state and federal initiatives also sought to support nursing schools and boost the supply of nurses and nurse educators.[3] A study conducted by the Institute of Medicine (2010) concluded that employment numbers may understate a nursing skills deficiency, since technological changes in the practice of nursing are raising the demand for nurses familiar with electronic medical records, team nursing, and evidence-based practice.

Traditionally, nursing managers used loan forgiveness to attract new nurses. However, some managers view loan forgiveness with skepticism, noting that their local nursing schools are often capacity constrained and that attractive wages and benefits may simply reward those who won a spot in a nursing program while doing little to increase the supply. Corroborating this claim, the American Association of Colleges of Nursing (2008, 2009) estimates that U.S. nursing schools denied admission to 40,285 qualified applicants in 2006 and 49,900 qualified applicants in 2009. Berlin, Wilsey, and Bednash (2005) estimate that 33,000 qualified applicants to nursing programs were declined admission in 2004-2005. As a result, hospitals became increasingly involved in local nursing education, including the provision of direct financial support for nursing schools and faculty.

Interviewees believed that direct financial support is more likely to be given by larger regional hospitals less prone to local competition—a claim consistent with survey results. In areas with several major hospitals, nursing officers may ask their counterparts at other area hospitals to

---

[3]   Examples include the Nursing Reinvestment Act of 2003, the President's 2003 High Growth Job Training Initiative, earmarks in the Graduate Assistance in Areas of National Need program of the Department of Education, and the health professions training grants in the American Recovery and Reinvestment Act. Firms hiring foreign nurses through the H1-B program are also given a special exemption to the requirement they demonstrate an inability to recruit a domestic nurse for the same position. Since 2008, the recession simultaneously reduced demand for preventative and elective procedures and increased the nursing supply by attracting nurses to re-enter the labor force. Unemployment among registered nurses rose to about 2% in 2009 and 2010, continuing its trend at one-fourth the national unemployment rate (Benson 2011; Buerhaus, Auerbach,and Staiger 2009). Surveyed nursing managers reporting a shortage of registered nurses dropped from 77% to 20% from 2008 to 2010, and the mean vacancy rate dropped from 10.5% to 4.1%.

"share the burden" by making commitments of their own. This concern is consistent with the contractual solution of the collective action problem under oligopsony, and is also explored in the empirical section. Evidence from this study's interviews and survey results, as well as existing literature, suggest that joint programs are not uncommon. One interviewed hospital administrator described a "nursing collaborative" between major regional hospitals and nursing schools that also included endowing a school of nursing education. Such partnerships, it was noted, may be organized by hospital systems or regional hospital associations.

II.   Firm-Sponsored General Education, Monopsony, and Nursing Education

In the classic human capital model, firms do not finance general education because *ex post* wage competition between the incumbent firm and potential poachers eliminates potential rents (Becker 1964). Models of firm-sponsored general education therefore invoke market imperfections that allow firms to retain trained workers while paying wages below their productivity, thereby compressing the wage profile relative to the productive returns to education (Acemoglu and Pischke 1999a).

In a series of papers, Acemoglu and Pischke (1997, 1999a, 1999b) hypothesize mobility frictions, oligopsony, or credit constraints may allow firms to pay trained workers less than their marginal product and recoup training investments. However, testing these models is challenging. One dilemma is that firm-sponsored general education may be discounted against workers' earnings—for example, firm that pay for employees' tuition in MBA programs typically require a "work commitment," as well as reimbursement if the worker reneges on the commitment to return. Tests for wage-discounting have been performed using the U.S. Equal Opportunity Pilot Project (Barron, Berger, and Black 1999; Grossberg and Sicilian 1999), National Longitudinal Survey of Youth (Loewenstein and Spletzer 1998), and British Household Panel Survey (Booth and Bryan 2005). These studies have generally found little or no evidence of wage discounting in the presence of firm-sponsored training. However, testing for wage discounting in these data remains problematic because it is not clear whether training is general, whether the provision of training is attributable to private information, or what the earnings would be in the absence of firm-sponsored training.

Rather than mobility frictions, recent research emphasizes the role of asymmetric information in inducing firms to sponsor general education. Theoretical models by Katz and Ziderman (1990) and Chang and Wang (1996) propose that firms may provide general training if its quality is unobserved by potential poachers. Acemoglu and Pischke (1998) hypothesize that employers enjoy *ex post* informational monopsony power if workers who quit after receiving education are expected by potential employers to be adversely-selected. Citing evidence from a temporary help agency, Autor (2001) hypothesizes firms providing general skills training enjoy private information regarding trainees' aptitude, allowing them to pay less than workers' productivity until their tenure signals their aptitude to other employers.

However, asymmetric information does not appear to be driving hospitals' financial support to nursing schools and faculty. Direct financial support to nursing schools cannot be discounted against beneficiaries' wages and cannot be conditioned upon beneficiaries' abilities. Moreover, it seems unlikely that hospitals receive private information regarding graduates' aptitudes or

quality of their training as a condition of their financial support, and also unlikely that non-sponsoring hospitals believe graduates who do not work for their sponsors are adversely-selected.[4] Indeed, in this case of firm-sponsored education, the beneficiaries of training subsidies include non-employees who may not even be aware that their education is being supported by a local employer.

Rather, hospital administration literature refers to such support as an "investment," implying their belief that local labor markets are partly-shielded from external competition (and thereby not perfectly elastic). As such, hospitals' provision of financial support for schools and faculty appears to be most-consistent with Acemoglu and Pischke's (1997) hypothesis that mobility frictions may allow firms to extract rents on newly-trained local workers and induce them to pay for general skills training.

Nursing's institutional and demographic features make it a classic setting for testing monopsonistic search frictions. Institutionally, hospitals are faced with the competing pressures of increasing scale to offer a wider variety of treatments and dispersing to be near the population, resulting in hospitals' geographic sparsity and a perception that single hospitals or hospital systems can be dominant employers within small metropolitan areas. This sparsity may inhibit nurses' inter-firm mobility since switching employers may involve longer commutes or the pecuniary and psychic costs of home relocation (Bhaskar and To 1999; Bhaskar, Manning, and To 2002; Manning 2003).

Recent research on "new" monopsony models emphasizes how firms may enjoy equilibrium monopsony rents due to workers' search costs, forfeited accumulated skills, and foregone earnings (Burdett and Mortensen 1998, Manning 2003). Furthermore, a majority of RNs are married (71%), are women (94%), and are supporting dependents (52%); these traits have been found to reduce the likelihood of relocation and potentially affect incumbent employers' decisions as to whether to match non-local wage offers (McKinnish 2008, Mincer 1978, Pixley and Moen 2003, Sandell 1977). Institutional and demographic features have thereby led researchers to hypothesize that regionally-powerful hospitals are able to "lock-in" nurses at lower wages than competitors are willing to bid.

However, research has generally concluded that nurse employment is too sensitive to wage changes in the long run for hospitals to enjoy significant monopsony power in wage-setting. Using non-federal hospital data in 1979, Adamache and Sloan (1982) find that, after controlling for population density and costs of living, entry-level real wages among RNs employed at non-federal hospitals are uncorrelated with hospital concentration. Using wages in MSAs and in rural state areas from the Current Population Survey, Hirsch and Schumacher (1995, 2005) find no evidence that nurse/non-nurse relative wage rates are correlated with hospital density or market size. Using a cross-section of nurse wage data from an AHA personnel survey, Sullivan (1989) estimates that the nurse labor supply is quite inelastic; however, Sullivan's estimates do not differ from large and small markets, which on-face is contrary to "new" monopsony theory's

---

[4] Although hospitals that offer clinical rotation sites for nursing programs may extract private information regarding nursing students' abilities, direct financial support for the school is typically not a precondition; in the data, two-thirds of hospitals that serve as clinical rotations sites do not provide direct financial support for nursing schools.

predictions. Using an exogenous change in minimum nurse wages at Veterans Affairs hospitals, Staiger, Spetz and Phibbs (2010) find that non-VA hospitals rose short-run wages by about 12.8% the increase in VA wages, providing perhaps the strongest evidence for monopsonistic wage setting in the nursing market. Although evidence is mixed and evidently sensitive to the empirical strategy, a review by Boal and Ransom (2003) concludes that the rate of monopsonistic exploitation of nurses is "close to zero."

As Boal and Ransom (1997) note, skepticism toward the "classic" monopsony hypothesis in nursing poses a paradox in light of high sustained vacancy rates and perceived shortage for registered nurses. An upward-sloping supply for registered nurses would prompt the profit-maximizing hospital to post excess vacancies at the monopsony wage rate. Likewise, because this wage rate is below the "market" wage rate, Mincerian rents on workers' investments in a nursing education are inefficiently low, discouraging potential entrants—a result analogous to a worker's unwillingness to pay for firm-specific skills.

Hospital administrators' stated reasons for providing financial support to nursing schools rather than raising wages vary widely. Particularly in major regional hospitals, administrators express a deep awareness of the well-being and productivity of local nursing schools and are often involved in efforts to expand their programs. The implementation of support also varies widely. Support may involve lump sum financial transfers to create new programs or expand existing ones, paying for new nursing faculties' salaries (or providing new faculty with a supplementary "fellowship" or stipend), or providing schools with lecture or office space.

III. Testing for Firm-Sponsored General Education and Mobility Frictions

For the reasons described above, the provision of direct financial support from hospitals to nursing schools presents as a fruitful setting for testing the intersection of human capital and monopsony theory. On one hand, human capital theory implies some form of monopsony power must exist to induce wage compression. On the other, evidence for monopsony in nurse wage-setting is mixed and generally treated with skepticism.

While previous literature focused on estimating wage elasticities (i.e. Adamache and Sloan 1982; Hirsch and Schumacher 1995, 2005; Sullivan 1992; Staiger, Spetz and Phibbs 2010), the empirical strategy presented here provides evidence for monopsony rents in the form of hospitals' willingness to pay for nurses' education in a fashion consistent with monopsonistic mobility frictions and not with fluid cross-metropolitan competition for nurses. While inferring rents in this fashion may appear indirect, it also has key advantages. In particular, cost-of-living, non-wage compensation, and an absence of data on productivity (with respect to tenure, market size, or training) complicate estimation of wage inelasticity; this study abstracts from these concerns, using the willingness of hospitals with large employment share to sponsor nursing schools as evidence of monopsonistic rents.

In particular, I hypothesize hospitals pay for general skills training due to oligopsony rents and allocative efficiencies in the provision of training. Although both oligopsony rents and allocative efficiencies are strong assumptions, the former represents the wage compression shown to be a necessary condition for firm-sponsored general education by Acemoglu and Pischke (1999a),

and the latter is perhaps the simplest explanation for why hospitals attract workers into nursing by paying for nursing schools and faculty rather than raising wages.[5] These assumptions are motivated both by earlier studies and by fieldwork investigating how financial support for nursing education is implemented.

Early studies hypothesized hospitals collude to fix nurse wages, either in the form of explicit "wage standardization" agreements or by using wages at major local hospitals as reference points when setting their own (Sullivan 1989, Hirsch and Schumacher 1995, Yett 1975). Wage-fixing has also been alleged in a series of lawsuits around the time of the first survey.[6]

Interviewed hospital administrators report that, even if hospitals do not explicitly exchange wage data, nursing managers are generally aware of salaries in their area's major hospitals and refer to them when setting their own hospitals' salaries. Federal law requires that the Centers for Medicare and Medicaid Services collect nurse wage data for hospitals participating in Medicare and publish geographically-adjusted wage indices.[7] Wage rates at local hospitals may also be available through collective bargaining agreements or explicitly reproduced through pattern bargaining. These factors result in a sentiment that major local hospitals act as "wage leaders." Wage-fixing would also be consistent with that the extent of shortages are particularly poignant and certain metropolitan areas (Bureau of Labor Statistics 2008; May, Bazzoli, and Gertrand 2006), and a class of temporary agency nurses (typically referred to by administrators as "travelers") work in geographic regions facing shortages (Bureau of Labor Statistics 2010).

Even in perfect monopsony, firms may not subsidize training because doing so is only as-appealing to workers as commensurately raising wages. However, nursing students may benefit more from a dollar spent by hospitals on their education than a dollar spent on their future wages. This may be the case in the event that external sources (such as the government or private foundations) match contributions by hospitals. Hospitals' support may also utilize slack

---

[5]   Rents might also be explained in a "new monopsony" framework. In the absence of perfect wage discrimination, raising wages to attract new nurses may require raising wages for incumbent nurses for whom hospitals can extract rents (for example, due to firm-specific learning, psychic attachment to the firm, switching costs). As evidence of this, nursing compensation (particularly for new recruits) has risen rapidly since the early 1990s, leading some hospitals to "frontload" compensation to new recruits through immediate or deferred signing bonuses, tuition reimbursement, or other means. Bloom, Alexander, and Nuchols (1997) find that nurses' productivity grows more quickly with experience than do their earnings, consistent with the hypothesis hospitals may extract rents on experienced nurses. Using shocks from contract nurses, who are typically more costly to hospitals than staff nurses but with less (if any) within-unit experience, Bartel et al. (2011) show that nurses' within-unit tenure reduces patient length of stay, providing further evidence of wage compression.

[6]   From June 2006 to January 2008, class-action lawsuits were filed against seventy-three hospitals and sixteen health systems in Detroit, Albany, Chicago, Memphis, and San Antonio, alleging that they had violated Section 1 of the Sherman Act "by agreeing to fix or depress the wages they pay RNs and agreeing to exchange wage information about the amounts they were paying, and planned to pay, RNs working in area hospitals" (Miles 2007: 305). St. John's Health System's $13.6 million settlement with Detroit nurses has since led to increasing concern that hospitals pay practices will face increasing scrutiny.

[7]   These reports are available online and without fee at http://www4a.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/SNFPPS/WageIndex.html, accessed 04/27/12.

resources, such as unused lecture hall space. Hospitals may also provide paid leave for nurses who teach during the off-season. As a result, subsidies may represent "allocative efficiencies," and some amount of hospital-sponsored training may also be efficient.

Second, potential nurses may discount the future more than hospitals, face credit constraints, possess imperfect information regarding the cost or benefits of a nursing education, or may be more averse to the risk of dropping out of nursing school. Each of these would result in an underinvestment by workers in general skills training (see Bennett, Glennerster, and Nevinson 1993 for a theoretical treatment).

To see how oligopsony and allocative efficiencies may induce firms to sponsor general education, consider the following model. In period 2, workers in a given metropolitan area produce y for their employer if they are trained and nothing otherwise. This seems reasonable because all U.S. states require registered nurses to be licensed.[8] Suppose oligopsonists match competitors' wage offers in a "tit-for-tat" strategy, where they pay the greater of their competitors wages or the monopsonist's wage w_m, and assume workers facing identical wages sort into firms i in proportion to their existing labor share s_i. Let w denote the oligopsony wage rate, such that a firm's rents on trained nurses are $y - w$. Workers in period 1 may discount wages paid in period 2 at a different rate than hospitals, such that the present value of wage w in period 2 is δw in period 1.

In period 1, firms choose training subsidies. Let $t_i \geq 0$ denote per-worker training paid for by firm i, and let T denote the total per-worker training paid for by all firms. Training subsidies provided by firms may reduce trainees' costs by some multiplier (e.g. due to matching grants, use of slack physical or human resources, etc.), such that students benefit at a rate αT. Suppose workers pursue training if the net present value of the training subsidies and earnings exceed the worker's idiosyncratic outside option, such that the quantity of workers pursuing a nursing education is $q(\alpha T + \delta w)$.

In this environment, the monopsonist's wage w_m solves $\max_w (y - w)q(\alpha T + \delta w)$, and rents accruing to incumbent firms at this wage are $(y - w_m)s_i q(\alpha T + \delta w_m) - t_i$. Consider the marginal returns to raising wages and paying for training. Using subscripts to denote derivatives, when competitors match wage offers, paying for training is superior to raising wages if

$$s_i * (y - w)q_t(\alpha) - 1 > -s_i q(\alpha T + \delta w_m) + s_i(y - w)q_w(\delta) \tag{1}$$

and paying for training is superior to the status quo if

$$s_i * (y - w)q_t(\alpha) - 1 > 0 \tag{2}$$

Solving for s_i, firm i pays for general skills training if $s_i > [q + (y - w)(q_t(\alpha) - q_w(\delta))]^{-1}$ and $s_i > [(y - w)q_t(\alpha)]^{-1}$. If allocative efficiencies are positive and workers discount the future more than firms, $\alpha > 1$ and $\delta < 1$, implying $q_t(\alpha) - q_q(\delta) > 0$. Firms contribute a nonzero amount to general

---

[8]   All state boards of nursing require registered nurses to graduate from an approved nursing program and pass the NCLEX-RN examination, which tests for basic competence.

skills training if their local employment share is sufficiently great, and this critical threshold declines with oligopsony rents (ie. $y - w$), and nursing supply is sufficiently sensitive to training subsidies (ie. $q_t(\alpha)$). Firms prefer paying for general skills training rather than raising wages if the local employment share is sufficiently great, and this critical threshold decreases with an increase in oligopsony rents (ie. $y - w$) the gap between the allocative efficiencies and workers' excess discounting (ie. $q_t(\alpha) - q_q(\delta)$). Note that if either rents or the allocative efficiency-discount gap are nonpositive, no employment share is sufficiently great to induce a firm to pay for training. The empirical strategy is to infer rents and efficiencies by showing firms with high local employment shares often pay for general skills training.

In the research setting, firm-sponsored general skills training takes the form of independent and direct financial support for nursing schools and faculty. The model thereby yields the two following hypotheses:

H1:   Metropolitan hospitals that enjoy a greater share of a metropolitan area's registered nurses (net of absolute size and other controls) are more likely to finance general skills training by independently providing direct financial support to local nursing schools.

H2:   Metropolitan hospitals that enjoy a greater share of a metropolitan area's registered nurses (net of absolute size and other controls) are more likely to finance general skills training by independently paying or sponsoring local nursing faculty.

In addition, I examine the hypothesis using joint, rather than independent, financial support as the dependent variable. This hypothesis emerged due to interviewees' suggestion that major hospitals in semi-competitive markets may want to foster local nursing education, but not wish to do so alone. Joint sponsorship of nursing education also represents a contractual solution to free-ride on other hospitals' educational investments under oligopsony. Interviews and theory thereby inform the third hypothesis:

H3:   Metropolitan hospitals that enjoy a greater share of a metropolitan area's registered nurses (net of absolute size and other controls) are more likely to pay for general skills training by jointly providing direct financial support to nursing schools.

Alternative results could be that, contrary to existing (non-survey) studies, instances of hospitals financing nursing schools are exceedingly rare, regardless of hospital characteristics. This finding would imply that rents and allocative efficiencies, if they exist, are generally insufficient to induce firm-sponsored general skills training. Another alternative result would be that *not* rare, but that hospitals' metropolitan nurse employment share is independently a poor predictor of training, in which case mechanisms orthogonal to the hospital share (such as information access or visibility) may be more important.

IV.   Methods and Results

Data include institutional, metropolitan, and local labor market characteristics of U.S. hospitals. Institutional characteristics come from the AHA Annual Survey of Hospitals. Nurse recruitment practices come from mail surveys conducted in the summers of 2008 and 2010.

For each survey, hospitals were randomly sampled from metropolitan hospitals included in the AHA hospital survey. For the 2008 survey, of 1,286 first-round surveys deliverable to Chief Nursing Officers at hospitals with at least twenty-five beds, 140 were returned. A follow-up survey sent to non-respondents yielded 133 second-round respondents, for a cumulative response rate of 21.2%. The response rate was higher in 2010. Of 1,401 deliverable surveys in a single round, 304 were returned, for a response rate of 21.7% and a total of 577 hospital-year observations. Interviewees suggested the survey would more likely be filled by an administrative assistant rather than the Chief Nursing Officer personally.

Table 1 presents the response rates by round and hospital size. A chi-squared test suggests response rate depends on governance, but otherwise there is insufficient evidence to conclude, at a 10% significance level for either survey, that other variables predict response. Logistic regressions find that log-population and log-population density are not statistically significant predictors of response.

[TABLE 1]

The surveys were developed through an iterative process of referring to nursing management literature and consultation with chief nursing officers, labor representatives, and deans at geographically and institutionally-diverse hospitals and nursing schools. Rather than focusing only on practices relating to educational partnerships, the accompanying cover letter described the study as "a survey of hospital registered nurse recruitment strategies," and asked chief nursing officers to identify whether they used a wide variety of nurse staffing practices. The surveys collected fifty-four items on registered nurse employment. In addition to questions related to training, the surveys collected data on: turnover, vacancies, collective bargaining representation, agency nurse utilization, scheduling, signing bonuses, relocation bonuses, international recruitment, magnet status, recruitment costs, and the numbers of local and non-local recruits. To mitigate response bias with respect to the dependent variable of interest, the survey's cover letter did not explicitly express an interest in nursing education. In addition, education questions were posed near the end of the survey instrument, followed only by questions regarding vacancies, turnovers, and contact information. The cover letter offered summary statistics to respondents who left contact information. Table 2 presents dependent variables of interest.[9]

[TABLE 2]

Although nearly the same number of hospitals sponsor schools as sponsor faculty (ninety-seven versus ninety in 2008, and ninety-one versus eighty-five in 2010), these do not necessarily represent the same hospitals. Among respondents to both items, fifty-seven hospitals in 2008

---

[9] As a rudimentary check against alternative hypotheses (e.g. information asymmetries), the survey also asked nursing officers to identify whether their hospital provided direct financial support a non-local nursing school, of which only four reportedly did.

and fifty hospitals in 2010 provided direct financial support to both schools and faculty. Three "yes-or-no" questions regarding hospital's independent and joint sponsorship of nursing schools and faculty were favored over the collection of continuous measures of financial support.

The staffing survey is matched with the AHA Survey, which provides the independent variables. The AHA survey is conducted at the establishment-level, and hospitals within the same system are treated as distinct hospitals for the purposes of data collection and the calculation of employment share, although findings are not sensitive to this choice.

The hospital's share of full-time equivalent registered nurses (FTE RNs) in the hospital's own MSA[10] is measured from the population of hospitals in the AHA survey.[11] Hospital bed size is an important control since larger hospitals may have more policies (such as support for nursing schools) net of employment share. Although larger hospitals may intuitively also have larger employment shares, the great variation in the population of US cities makes the correlation between log-beds and log-employment share positive but not particularly strong (r = 0.47). As a result, there are hospitals with more than 600 beds that employ less than 2% of their MSA's hospital nurses and hospitals with fewer than fifty beds that employ the majority of their MSA's hospital nurses. A hospital system is a group of hospitals that are owned and/or contract-managed by a central organization. A hospital network is a group of hospitals that formally coordinate the delivery of a broad spectrum of services, and may not be members of the same system. Lastly, I control for hospital governance with three classifications: government, for-profit, and non-profit (including church-affiliated).

[TABLE 3]

Table 3 disaggregates the share of hospitals sponsoring nursing schools and faculty for each independent variable. To construct the tabular presentation, hospital employment shares are made discrete by reporting above-median and below-median shares. A simple chi-squared test concludes that the three sponsorship rates are different (p<0.01) between hospitals with below-median and above-median employment shares in both 2008 and 2010.

[TABLE 4]

Table 4 presents results from four multinomial logistic regressions predicting the likelihood of independent and joint direct financial support to local nursing schools, corresponding to hypotheses H1 and H3. Because hospital employment shares are highly skewed, I use its

---

[10] The Federal Office of Management and Budget uses the Metropolitan Statistical Area (MSA) to designate "a core area with a population nucleus [of at least 50,000 inhabitants], plus adjacent communities having a high degree of economic and social integration with that core," making it an appealing demarcation of a local labor market.

[11] RNs also work in physicians' offices, residences, nursing care facilities, and other settings. CPS MORG data suggests about 40% of RNs work outside of hospitals, independently of metropolitan characteristics (such as population). Although the CPS MORGs feature enough nurses to show the relation between non-hospital employment and metropolitan characteristics are weak, they do not feature enough nurses to calculate non-hospital employment share reliably for any individual MSA, potentially introducing measurement error. Including non-hospital employment in employment shares has little substantively effect on results.

logarithm to avoid excessive statistical leverage. The log-share of FTE RNs has a mean of -3.294 (corresponding to a log-mean share of 3.71%), a standard deviation of 1.745, a minimum of -8.046, and a maximum of 0.

Consistent with hypothesis H1, hospitals with a greater share of metropolitan FTE RNs are more likely to provide independent direct financial support to nursing schools. In both the 2008 and 2010 regressions, the magnitude of the effect of hospital share is mildly reduced after controlling for the size of the hospital (as measured by the number of hospital beds) and other hospital characteristics.

Support for hypothesis H3 is mixed. Among hospitals surveyed in 2008, those with a greater employment share are more likely to sponsor nursing schools jointly (using a 5% significance test). Among hospitals surveyed in 2010, hospitals with a greater employment share are not significantly more likely to sponsor nursing schools jointly. It is also notable that control variables, including bed-size, have relatively mild effects on the likelihood of school sponsorship, net of employment share.

[TABLE 5]

Table 5 presents results from six binomial logistic regressions assessing hospitals' sponsorship of nursing faculty. Regression results using the 2008 survey show a strong relationship between hospital employment share with the propensity to pay or sponsor nursing faculty. This result is significant at 1% in all three models. However, regression results from 2010 suggest that the effect of hospital-share is captured by hospital size. Indeed, further analysis reveals that large hospitals steeply increased their propensity to sponsor nursing faculty from 2008 to 2010, while smaller hospitals, including those that enjoy high local employment shares, became less-likely to sponsor nursing faculty. This shift poses a mystery, particularly given the robustness of the effect of hospital shares to controlling for hospital size for faculty sponsorship in 2008, and direct financial support to nursing schools in 2008 and 2010.

[TABLE 6]

Table 6 presents the estimated likelihood hospitals pay for nursing schools and nursing faculty by nurse employment. I calculate point estimates and standard errors using the delta method from the full regression models (i.e., Table 4, Regressions 2 and 4; and Table 5 Regressions 3 and 6). To isolate the independent effect of hospitals' employment share, I present point estimates at the mean value for control variables. Including infra-marginal hospitals, those in top quartile in local nurse employment share are about three times as likely to provide financial support than a hospital with the bottom quartile share, after hospital size and other institutional controls. The model predicts a hospital has a 50% probability of paying for nursing schools if it employs at least 30% of a metropolitan area's registered nurses.

[TABLE 7]

Table 7 presents results for three common alternative forms of firm-sponsored general education: tuition support for non-employees in exchange for work commitment, tuition support for

Licensed Practical Nurses (LPNs) enrolled in Bachelor's of Science in Nursing (BSN) programs,[12] and tuition support for employees enrolled in Masters of Science in Nursing (MSN) Programs. Although these forms of firm-sponsored general education could, in principal, be discounted commensurately against workers' wages, recent empirical work suggests that employers that pay for workers' general education do not fully-discount wages by employers' contributions, contrary to the predictions of human capital theory under perfect competition (Barron, Berger, and Black 1999; Leuven 2005). If employers are partial-claimants on the returns to general skills training, and bear part of the training costs (in contrast to direct financial support for schools and faculty), then we may expect hospital share to be a positive, albeit weaker, predictor of these forms of support. Consistent with cost-sharing and rent-sharing, results for 2008 suggest that a hospital's local employment share predicts tuition support for a non-employee work-commitment, employees in BSN programs, and employees in MSN programs, net of the full controls. However, results for 2010 are not statistically significant, and both the shares of hospitals offering these forms of tuition support and its relationship to hospital share declined. One hypothesis is that the recession relaxed the market for nurses, and because wages are downwardly-rigid, hospitals reduced benefits such as tuition reimbursement instead.

As robustness checks, I test the hypotheses under alternative specifications and using instrumental variables for hospital share. Because the Herfindahl-Hirschman Index[13] is classically used as the independent variable in monopsony studies, all regressions were also run replacing the log-share with the employment log-HHI as the primary independent variable of interest. Consistent with hypothesis H1, hospitals in high-HHI metropolitan areas are significantly more likely (at 1%) to provide independent financial support to nursing schools in both the 2008 and 2010 samples. Estimates for joint sponsorship are not statistically significant. Selecting an instrument that specifically affects hospital share is challenging, and three were used. First, using Census data, I used the MSA's log-population density as an instrument for both log-employment share and for the log-HHI. High population density is negatively correlated with low mean employment shares and HHIs; intuitively, low density MSAs tend to include very large rural areas with small metropolitan centroids where hospitals are sparse. I also used per capita income and Catholic hospital penetration (which may prompt struggling hospitals to merge) as instruments for the employment share and the HHI. Each instrument has no substantive effect on results, and Durbin-Wu tests performed for each of the full regressions and for each instrument failed to find evidence that either hospital share or HHI is endogenous.[14]

Some caveats deserve special mention.

---

12 Licensed Practical Nurses (LPNs) are also known as Licensed Vocation Nurses (LVNs) in the U.S., and are roughly equivalent to "enrolled nurses" (ENs or SENs) in other Anglo countries. LPNs have less formal skills than Registered Nurses, and LPN-to-BSN programs typically train LPNs to become RNs. In 2008, the survey finds these three programs are offered by 31%, 85%, and 79% of sampled hospitals, respectively; in 2009, the survey finds these three programs are offered by 20%, 75%, and 79% of sampled hospitals, respectively.

13 In this case, the HHI for metropolitan area j is the summation of hospitals' squared shares of full-time equivalent registered nurses, taken from the universe of hospitals in the AHA database (rather than the sample frame). Intuitively, the HHI for MSA j gives the probability that two registered nurses chosen at random from MSA j's hospitals are employed by the same hospital.

14 Technical details and results are available upon request.

First, while log-share intends to capture probabilistic monopsony rents, it may also be capturing the role of omitted metropolitan and institutional features positively correlated with financial support to schools and faculty. For example, hospitals with greater employment shares (net of the number of hospital beds) may simply be more bureaucratic, and therefore more likely to establish formal financial arrangements with nursing schools and faculty. While instrumental variables are the standard method for addressing this concern, it is difficult to find one that would narrowly affect hospital employment shares, particularly given that interviews and existing research provide little guidance as to what omitted variables may be a concern.

Second, although the challenges inherent to conducting an employer survey were anticipated and considered throughout the data-collection process, response bias remains a concern. Hospital administrators may respond to the survey for a variety of reasons—due to an interest in academic research in nursing recruitment, for access to summary statistics, because they perceive a shortage at their own hospital to be a problem, or because they believe they are effectively avoiding one. Although response is not evidently conditioned on observable characteristics, these sources of response bias may lead to biased estimates of the levels at which hospitals are sponsoring nursing schools and faculty. More troubling is the possibility that these sources of response bias are positively correlated with both the propensity to sponsor education and the employment share, which would increase the likelihood of a Type I error.

Lastly, though financial support of local nursing programs is correlated with employment shares, it is still a common-enough practice in metropolitan areas to warrant consideration of the other reasons for sponsoring education. Financial support may expand the pool of nursing students whom hospitals screen through complementary training practices, such as clinical rotations or career ladders for LPNs. A hospital's interest in improving its visibility or its reputation is not captured in the model. Supporting nursing education might also promote loyalty and commitment among employed nurses, reducing turnover or improving productivity. If hospitals in competitive markets are compelled to pay for training by visibility or goodwill, the expected sign would be the opposite as that found in the results. If hospitals in competitive and non-competitive markets were equally-compelled by visibility or goodwill concerns, the effect would be neutral. If these motivations are positively correlated with hospital share and not captured by controls (such as size or non-profit status), they would increase the likelihood of committing a Type I error.

Although this study is subject to the usual challenges of conducting an employer survey, it confirms that hospital financial support for nursing schools and faculty is indeed fairly common, and suggests that such support is generally provided by hospitals with large shares of local nursing employment, net of size and other institutional controls.

V. Policy Implications

Results lend support for the application of frictional search models to the study of nursing manpower. These models present a set of policy implications and offer directions for further evaluation. First, evidence that employers extract rents on locally-trained nurses suggests that the potential for non-competitive wage setting should be taken seriously. Indeed, extended and

sustained shortages up to the most-recent recession are themselves theoretically surprising, as it appears that wage gains were not sufficient to entice workers into the nursing profession.

Second, this study suggests there may be efficiency gains to be made by improving collaboration and the allocation of resources between hospitals and nursing schools. As shown in the model, firm-sponsored general education requires not only frictional job search, but also some form of allocative efficiencies in the provision of nurse training, excessive discounting among students, or liquidity constraints among students. Likewise, interviews conducted for this study identified several methods by which hospitals used slack resources, such as hospital lecture space or non-seasonal work, to promote nurse training. Further study may identify other ways in which hospitals may efficiently coordinate nurse training, and other ways institutions may help ease students' credit constraints (such as through subsidized loans or work commitments).

Third, by providing evidence for monopsony in the nursing market, this study invites future research to consider whether monopsony may explain other supplementary staffing practices, such as the widespread extensive reliance of travelling contract nurses, mandatory overtime, and foreign recruitment.

Fourth, state or local institutions may help make nursing wages more competitive by improving the portability of their education. One way to do so is to adopt national training guidelines for transition-to-practice programs. The National Council of the State Boards of Nursing offers a model for doing so (for a discussion, see IOM 2010; NCSBN 2009).

Lastly, in areas with weak local competition for registered nurses, public subsidies designed to benefit nursing students (e.g. of the form taken the Nurse Reinvestment Act or most state nurse workforce initiatives) may be less effective than would be the case in perfect competition if part of the rents are captured by hospitals rather than potential students. While supporting local hospitals is not in itself an undesirable outcome, interventions intending to boost the supply of nurses in areas with weak local competition may require hospitals to increase their own contributions to nursing education.

VI. Concluding Remarks

Hospitals' provision of direct financial support to nursing schools and their faculty poses a striking exception both to the usual explanations for firm-sponsored general skills training and to evidence suggesting the market for hospital nurses is not subject to monopsony. Rather, the tendency of this support to be provided by hospitals enjoying weak local competition suggests that frictional search models at the intersection of human capital theory and the "new" monopsony literature are more applicable to nursing education. This finding makes two theoretical contributions.

First, this setting overcomes the typical empirical dilemmas in human capital theory and provides support for the hypothesis that mobility frictions may induce firms to pay for technologically-general skills. Hospitals' support for nursing education is an instance where the training is highly-general, clearly paid by the firm (rather than, for example, being discounted against earnings), and is a setting in which private information is minimal. By associating financial

support with hospital's employment share, the study suggests local employers expect to extract rents on locally-trained nursing graduates.

Second, hospital sponsorship of nursing schools implies the existence of monopsony rents by using a non-standard method (inference from patterns of firm-sponsored training). While recent studies generally reject that employers behave like monopsonists by estimating high labor supply elasticities, this conclusion has left open the puzzle of excessively high vacancy rates at prevailing wages. Results suggest that hospitals pattern compensation off of major employers, and use alternative practices to attract "marginal" nurses and to supplement staffing to fill nurse vacancies. These practices, such as aggressive relocation and signing bonuses, overtime, temporary nurses, and foreign nurses, are consistent with monopsonistic wage discrimination but are overlooked by traditional testing instruments. Hospital nursing may thereby be a rich setting to examine the human resource practices consistent with the market failure and the presence of rents.

By distinguishing nursing education as a case of firm-sponsored general education induced by mobility frictions, this study invites human capital and monopsony to be applied to the study of nursing markets. Specifically, this study implies that public support designed to increase the nursing supply should require that matching support hospitals when local labor markets appear to be weak; that hospitals, nursing schools, and the state should seek out methods of promoting nursing education that exploit allocative efficiencies and slack resources; and that interventions designed to promote the portability of a nursing education may promote wage competition and nursing employment.

Bibliography

Acemoglu, Daron, and Jörn-Steffen Pischke, 1997. Training and Innovation in an Imperfect Labor Market. Review of Economic Studies 64(3): 445-64.
_____, 1998. Why do firms train? Theory and Evidence. The Quarterly Journal of Economics 113(1): 79-119.
_____, 1999a. The Structure of Wages and Investment in General Training. Journal of Political Economy 107: 539-72.
_____, 1999b. Beyond Becker: Training in Imperfect Labor Markets. Economic Journal 109(453): F112-42.
Adamache, Killard, and Frank Sloan, 1982. Unions and hospitals: Some unresolved issues. Journal of Health Economics 1(1):81-108.
Autor, David, 2001. Why Do Temporary Help Firms Provide Free General Skills Training? The Quarterly Journal of Economics 116(4):1409-48.
American Association of Colleges of Nursing, 2006. State Legislative Initiatives to Address the Nursing Shortage. Issue bulletin.
American Association of Colleges of Nursing, 2008. 2007-2008 Enrollment and Graduations in Baccalaureate and Graduate Programs in Nursing.
_____, 2009. 2008-2009 Enrollment and Graduations in Baccalaureate and Graduate Programs in Nursing.
Barron, John, Mark Berger, and Dan Black, 1999. Do Workers Pay for On-The-Job Training? Journal of Human Resources 34(2): 235-52.

Bartel, Ann, Ciaran Phibbs, Nancy Beaulieu, and Patricia Stone. 2011. "Human Capital and Organizational Performance: Evidence from the Healthcare Sector." NBER Working Paper No. 17474.

Becker, Gary, 1964. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. Chicago: University of Chicago Press.

Bennett, R. Glennerster, H., and Nevinson, R, 1993. Learning Should Pay. LSE and BP monograph.

Benson, Alan, 2011. The (New) Economics of Staffing Registered Nurses. Working Paper.

Berlin, Linda, Stefanie Wilsey, and Geraldine Bednash, 2005. 2004-2005 enrollment and graduations in baccalaureate and graduate programs in nursing. Washington DC: American Association of Colleges of Nursing.

Bhaskar, Vira, and Ted To, 1999. Minimum Wages for Ronald McDonald Monopsonies. Economic Journal 109, 190-203.

Bhaskar, Vira, Alan Manning, and Ted To, 2002. Oligopsony and Monopsonistic Competition in Labour Markets. Journal of Economic Perspectives, 16: 155-74.

Bloom, Joan R., Jeffrey A. Alexander, and Beverly Nuchols, 1997. Staffing Patterns and Hospital Efficiency in the United States. Social Science and Medicine 44(2): 147-55.

Boal, William, and Michael Ransom, 1997. Monopsony in the Labor Market. Journal of Economic Literature, 35: 86-112.

_____, 2003. Monopsony in American Labor Markets. Robert Whaples ed. EH.Net Encyclopedia.

Booth, Alison, and Mark Bryan, 2005. Testing Some Predictions of Human Capital Theory: New Training Evidence from Britain. Review of Economics and Statistics 87(2): 391-4.

Buerhaus, Peter, Karen Donelan, Beth Ulrich, Linda Norman, Catherine DesRoches, and Robert Dittus, 2007. Impact of the nurse shortage on hospital patient care: Comparative perspectives. Health Affairs 26(3): 853-62.

Buerhaus, Peter I. Staiger, David, and Douglas Auerbach, 2000. Implications of an aging registered nurse workforce. The Journal of the American Medical Association 283(22): 2948-54.

_____, 2003. Trends: Is the Current Shortage of Hospital Nurses Ending? Health Affairs, 22(6): 191-8.

_____, 2007a. Better Late Than Never: Workforce Supply Implications of Later Entry into Nursing. Health Affairs, 26(1): 178-85.

_____, 2007b. Recent trends in the registered nurse labor market in the US: Short-run swings on top of long-term trends. Health Affairs, 25(2): 191-8.

_____, 2009. The Recent Surge in Nurse Employment: Causes and Implications. Health Affairs 28(4): w657-68.

Burdett, Kenneth, and Dale T. Mortensen, 1998. Wage differentials, employer size, and unemployment. International Economic Review 39(2): 257-73.

Bureau of Labor Statistics, U.S. Department of Labor, 2008. Occupational Outlook Handbook: Registered Nurses. 2008-2009 ed.

_____, 2010. Occupational Outlook Handbook: Registered Nurses. 2010-2011 ed.

Chang, Chun, and Yijiang Wang, 1996. Human capital investment under asymmetric information: The Pigouvian conjecture revisited. Journal of Labor Economics 14(3): 505-19.

Hicks, John, 1939. The Foundations of Welfare Economics. Economic Journal. 49(196): 696-712.

Hirsch, Barry T, and Edward Schumacher, 1995. Monopsony Power and Relative Wages in the Labor Market for Nurses. Journal of Health Economics, 14(4): 443-73.

———, 2005. Classic monopsony or new monopsony: Searching for Evidence in Nursing Labor Markets. Journal of Health Economics 24(5): 969-89.

IOM (Institute of Medicine). 2012. *The Future of Nursing: Leading Change, Advancing Health.* Washington, D.C.: The National Academies Press.

Katz, Eliakim, and Adrian Ziderman, 1990. Investment in General Training: The Role of Information and Labour Mobility. Economic Journal 100: 1147-58.

Leuven, Edwin. 2005. The Economics of Private Sector Training: A Survey of the Literature. Journal of Economic Surveys 19(1): 91-111.

Loewenstein, Mark, and James Spletzer, 1998. Dividing the Costs and Returns to General Training. Journal of Labor Economics 16(1): 142-71.

Manning, Alan, 2003. Monopsony in Motion: Imperfect Competition in Labor Markets. Princeton, NJ: Princeton University Press.

May, Jessica, Gloria Bazzoli, and Anneliese Gerland, 2006. Hospitals' responses to nurse staffing shortages. Health Affairs, 25: 316-23.

McKinnish, Terra, 2008. Spousal Mobility and Earnings. Demography 45(4): 829-49.

Miles, Jeff, 2007. The Nursing Shortage, Wage-Information Sharing Among Competing Hospitals, and the Antitrust Laws: The Nurse Wages Antitrust Litigation. Houston Journal of Health Law and Policy 7: 305-78.

Mincer, Jacob, 1978. Family Migration Decisions. The Journal of Political Economy, 86(5): 749-773.

NCSBN (National Council of State Boards of Nursing). 2009. "NCSBN Model Nursing Practice Act and Model Nursing Administrative Rules." Retrieved from https://www.ncsbn.org/1455.htm, April 26, 2012.

Pixley, Joy E. and Phillis Moen, 2003. Couples Career Priorities, in Phillis Moen (ed.) It's about Time: Couples and Careers. Ithaca, NY: ILR Press.

Sandell, Steven, 1977. Women and the Economics of Family Migration. The Review of Economics and Statistics 59(4): 406-14.

Grossberg, Adam, and Paul Sicilian, 1999. Minimum Wages, On-the-Job Training, and Wage Growth. Southern Economic Journal 65(3): 539-56.

Staiger, Douglas, Joanne E. Spetz, and Ciaran S. Phibbs. 2010. Is There Monopsony in the Labor Market? Evidence from a Natural Experiment. *Journal of Labor Economics* 28(2): 211-36.

Sullivan, Daniel, 1989. Monopsony Power in the Market for Murses. *Journal of Law and Economics* 32:S135-78.

Yett, Donald, 1975. *An Economic Analysis of the Nurse Shortage.* Lexington, Mass: Lexington Books.

Table and Figures

Table 1. Response Rates by Hospital Bed Size

2008	Respondent		Non-Resp.	Rate
	1st Rd	2nd Rd		
25-49	15	15	176	14.6%
50-99	19	32	215	19.2%
100-199	40	44	249	25.2%
200-299	28	15	204	17.4%
300-399	15	15	112	21.1%
400-499	10	5	57	20.8%
500+	13	7	102	16.4%
Total	140	133	1013	21.2%
2010	Respondent		Non-Resp.	Rate
	1st Rd.	(N/A)		
25-49	42		155	21.3%
50-99	53		229	18.8%
100-199	78		307	20.3%
200-299	55		169	24.6%
300-399	27		111	19.6%
400-499	22		49	31.0%
500+	27		77	26.0%
Total	304		1097	21.7%

Notes: Includes US Metropolitan Hospitals, Summer 2008 and 2010.

Table 2. Tabulations of Dependent Variables of Interest

Survey Wording	2008			2010		
	Yes	No	Pct.	Yes	No	Pct.
"Our hospital independently provides direct financial support for local nursing school(s)"	46	175	20.8%	81	214	27.5%
"Our hospital jointly (with other area hospitals) provides direct financial support to local nursing schools"	51	175	22.6%	45	243	15.6%
"Our hospital pays or sponsors nursing faculty"	90	180	33.3%	98	204	32.5%

Notes: Sampling errors < 6%.

Table 3. Summary Statistics

	Sample Shares	Summary of Dependent Variables, 2008			Summary of Dependent Variables, 2010		
		percent supporting nursing schools alone	percent supporting nursing schools alone or jointly	percent supporting nursing faculty	percent supporting nursing schools alone	percent supporting nursing schools alone or jointly	percent supporting nursing faculty
Metro RN FTE Share							
Above Median	50%	31.4%	47.1%	45.6%	39.2%	43.7%	39.8%
Below Median	50%	11.7%	25.2%	21.7%	12.6%	17.6%	23.3%
Bed Size							
25-49	11.0%	7.7%	20.0%	16.7%	7.7%	10.5%	7.5%
50-99	18.0%	18.2%	29.4%	28.6%	17.6%	22.0%	20.8%
100-199	30.8%	13.1%	35.4%	28.8%	20.8%	27.6%	23.7%
200-299	15.8%	32.3%	48.8%	44.2%	34.6%	38.8%	43.6%
300-399	11.0%	27.3%	42.9%	53.3%	37.0%	40.0%	48.1%
400-499	5.5%	35.7%	40.0%	35.7%	55.0%	55.6%	50.0%
500+	7.3%	38.9%	45.0%	42.1%	51.9%	59.3%	66.7%
Control							
Non-Profit/Church	59.7%	24.4%	39.6%	40.0%	30.0%	37.4%	39.1%
Government	21.6%	15.4%	24.1%	22.4%	31.0%	29.1%	30.5%
For-Profit	18.7%	18.9%	40.0%	27.7%	16.4%	18.5%	14.0%
System Membership							
Yes	56.4%	22.8%	39.1%	37.6%	31.4%	37.6%	37.0%
No	43.6%	20.0%	32.7%	29.3%	22.6%	24.6%	26.8%
Network Membership							
Yes	30.8%	23.4%	25.3%	34.1%	29.5%	30.1%	41.8%
No	69.2%	17.3%	41.3%	33.9%	26.8%	33.0%	28.7%

Table 4. Multinomial Logistic Regressions of Hospital Financial Support to Nursing Schools

	Direct Financial Support, 2008				Direct Financial Support, 2010			
	Regression 1		Regression 2		Regression 3		Regression 4	
	Jointly	Indep.	Jointly	Indep.	Jointly	Indep.	Jointly	Indep.
log-RN FTE Share	0.218*	0.488**	0.253*	0.461**	0.128	0.542**	0.0382	0.446**
	(0.098)	(0.111)	(0.113)	(0.122)	(0.149)	(0.0938)	(0.170)	(0.105)
log-Hospital Beds			0.114	0.287			0.328	0.425*
			(0.211)	(0.228)			(0.300)	(0.186)
Owner: Non-Profit			ref.	ref.			ref.	ref.
Owner: Government			-1.014*	-0.501			-0.968	0.0132
			(0.502)	(0.465)			(0.782)	(0.371)
Owner: For-Profit			-0.259	-0.115			-1.289	-0.485
			(0.456)	(0.550)			(0.818)	(0.461)
Member of System			0.853*	-0.143			0.630	0.560
			(0.392)	(0.392)			(0.523)	(0.314)
Member of Network			-1.528**	-0.786			-0.897	-0.455
			(0.454)	(0.414)			(0.613)	(0.330)
Constant	-0.484	0.126	-0.845	-0.946	-1.84**	0.829**	-3.56*	-1.735
	(0.349)	(0.338)	(1.240)	(1.330)	(0.580)	(0.307)	(1.858)	(1.120)
LR $\chi2$	23.2**		49.5**		39.78**		57.36**	
n	267		267		290		290	

Notes: Standard errors are in parentheses. Reference dependent variable includes hospitals not providing financial support to nursing schools. * significant at 5%; ** significant at 1%; two-tailed tests.

Table 5. Binomial Logistic Regressions of Hospital Financial Support of Nursing Faculty

	Supports Faculty, 2008			Supports Faculty, 2010		
	(1)	(2)	(3)	(4)	(5)	(6)
log-RN FTE Share	0.372**	0.320**	0.329**	0.292**	0.120	0.0800
	(0.084)	(0.089)	(0.092)	(0.077)	(0.085)	(0.088)
log-Hospital Beds		0.263	0.196		0.816**	0.734**
		(0.167)	(0.173)		(0.176)	(0.176)
Owner: Non-Profit			ref.			ref.
Owner: Government			-0.768*			-0.251
			(0.372)			(0.346)
Owner: For-Profit			-0.673			-1.044**
			(0.421)			(0.448)
Member of System			0.486			0.514*
			(0.306)			(0.285)
Member of Network			-0.400			0.193
			(0.317)			(0.292)
Constant	0.478	-0.996	-0.521	0.251	-4.461**	-4.334**
	(0.280)	(0.972)	(1.017)	(0.273)	(1.054)	(1.072)
LR $\chi2$	21.69**	24.23**	31.97**	15.54**	40.44**	49.70**
n	265	265	265	298	298	298

Notes: Standard errors in parentheses. * significant at 5%; ** significant at 1%; two-tailed tests.

Table 6. Estimated Likelihood of Supporting Nursing Schools and Faculty by Own-Metropolitan Employment Share Percentile, Controlling for Bed Size, Governance, System, and Network Status

YEAR	Outcome	Hospital Local Employment Share Percentile				
		10th	25th	50th	75th	90th
2008	Pays for Schools	0.165	0.224	0.303	0.409	0.524
		(0.086)	(0.08)	(0.067)	(0.07)	(0.106)
	Pays for Faculty	0.155	0.207	0.275	0.369	0.474
		(0.082)	(0.075)	(0.064)	(0.066)	(0.103)
2010	Pays for Schools	0.135	0.183	0.248	0.341	0.447
		(0.071)	(0.066)	(0.058)	(0.068)	(0.112)
	Pays for Faculty	0.258	0.271	0.287	0.305	0.324
		(0.1)	(0.076)	(0.059)	(0.065)	(0.098)

Notes: Standard errors in parentheses. Point estimates and standard errors calculated by the delta method from the full logistic regressions. By their respective row, models used to construct point estimates are: Table 4, Regression 2; Table 5, Regression 3; Table 4, Regression 4; and Table 5, Regression 6.

Table 7. Binomial Logistic Regressions Predicting Other Forms of Educational Support

	2008			2010		
	Commit	LPN-BSN	BSN-MSN	Commit	LPN-BSN	BSN-MSN
	(1)	(2)	(3)	(4)	(5)	(6)
log-RN FTE Share	0.195*	0.292**	0.214*	0.190	-0.173	-0.0565
	(0.089)	(0.107)	(0.099)	(0.097)	(0.092)	(0.103)
log-Hospital Beds	-0.015	-0.198	-0.149	0.085	0.249	0.627**
	(0.168)	(0.217)	(0.187)	(0.177)	(0.163)	(0.182)
Owner: Non-Profit	ref.	ref.	ref.	ref.	ref.	ref.
Owner: Government	-0.528	-0.991*	-0.718*	-0.210	-0.852*	-1.065**
	(0.359)	(0.418)	(0.361)	(0.383)	(0.333)	(0.367)
Owner: For-Profit	-1.021*	-0.251	0.282	0.171	-0.969*	-1.052*
	(0.436)	(0.555)	(0.511)	(0.435)	(0.376)	(0.413)
Member of System	0.511	0.170	0.223	0.149	0.407	0.915**
	(0.302)	(0.411)	(0.346)	(0.311)	(0.289)	(0.327)
Member of Network	-0.450	0.752	0.293	0.573	0.233	0.159
	(0.313)	(0.492)	(0.373)	(0.313)	(0.324)	(0.373)
Constant	0.019	3.900**	2.800*	-1.467	-0.599	-1.832
	(0.989)	(1.306)	(1.111)	(1.086)	(0.986)	(1.081)
LR $\chi2$	15.11	17.74	12.78	12.50	17.63	38.23
n	264	270	269	308	303	309

Notes: Standard errors in parentheses. "Commit" denotes tuition support for non-employees for work commitment. "LPN-BSN" denotes tuition support for employee Licensed Practitioner Nurses to pursue Bachelor's of Science in Nursing degrees. "BSN-MSN" denotes tuition support for employees to pursue Masters of Science in Nursing degrees. * significant at 5%; ** significant at 1%; two-tailed tests.

Chapter 2

A Theory of Dual Job Search and Sex-Based Occupational Clustering[1]

In 2010, 67% of families with at least one labor force participant featured dual-earner couples, up from 41% in 1980.[2] Given the frequency that young, career-oriented workers relocate for work or schooling and the impact these decisions may have on the career trajectories of the partner, it is not surprising that researchers' attention to the colocation problem has also grown. Mincer (1978) introduces a neoclassical approach to household relocation decisions, noting that decisions optimizing the family's net career outcomes may impair the careers of "tied stayers" or "tied movers." Costa and Kahn (2000) attribute the increasing concentration of college-educated "power-couples" in large metropolitan areas in part to their ability to sustain two simultaneous careers. Empirical studies have challenged the symmetry of these models due to evidence household relocation decisions favor husbands' careers. Compton and Pollak (2007) find only the education of the husband predicts migration to large metropolitan areas. Others find household migration usually advances husbands' careers to the detriment of their wives (Boyle et al. 2001, Clark and Huang 2006, Jacobsen and Levin 2000, McKinnish 2008, Long 1974).

However, by emphasizing why and for whom household relocation decisions are made, research on couples' job search neglects the broader challenges and opportunities presented by variation in the geographic flexibility of jobs and potential endogeneity in occupational choice. Specifically, newly-trained physicists, computer scientists, naval architects, and nuclear engineers are typically geographically-constrained and require calculated moves early in their careers. In contrast, administrative assistants, dentists, school teachers, nurses, and general managers typically enjoy relative geographic-flexibility. Given that formative training and career investments are typically incurred prior to marriage, and given the difficulty of reconciling two careers each demanding calculated moves, do young men and women self-segregate into geographically-constrained and geographically-flexible occupations in anticipation of the colocation problem?

Section I introduces a model whereby early specific career investments, couples' desire to colocate, and variation in the geographic flexibility of jobs result in a coordination problem that prompts men and women to sort into geographically-clustered and dispersed occupations in advance of marriage, a phenomenon I refer to as "sex-based occupational clustering."[3] In the model, occupations are clustered or dispersed for reasons exogenous to their sexual composition, and highly-clustered occupations compensate for the disamenity of geographic constraints but penalize couples with two such careers due to the colocation problem. The model is consistent with a number of well-known features of inequality, including the tendency for women to be "tied-movers" when a family relocates for work and the segregation of women into lower-paying jobs. It also yields specific testable predictions regarding the marriage and earnings penalties that deter desegregation, and implies these predictions should be greatest for those a high degree of formal education.

Section II constructs an occupational clustering index. Using the Decennial Census, for each occupation, I use the index to calculate the share of workers in that occupation that would need to relocate to equalize its workers per capita in every MSA. I find workers in clustered occupations tend to have higher rates of relocation-for-work among never-married and single-earning men and women.

Section III confirms sex-based occupational clustering in which occupations dominated by women tend to be geographically-dispersed while those dominated by men tend to be geographically-clustered, after controlling for occupations' hours worked, mean age, physical strength requirements, and the extent it involves assisting and caring for others. To provide an intuition of the magnitude, prime age dual-earner couples featuring a husband in an occupation that is more clustered than his wife's outnumber the reverse about two-to-one.

Section III also suggests men and women are deterred from departing from this pattern by earnings and marriage penalties. Women who enter geographically-clustered occupations experience lower earnings, later marriage, and higher divorce rates compared with those entering dispersed occupations. Men, in contrast, enjoy earnings premia and lower divorce rates in clustered occupations. Both sex-based occupational clustering and related effects on earnings, mobility, and marriage are most-pronounced among the college-educated, suggesting the highly-educated employ more occupationally-specific skills.

I.   A Model of Occupational Clustering and Sex Segregation

This section provides sufficient conditions for equilibrium sex-based occupational clustering. The model considers risk-neutral workers who initially differ only in sex and who are interested

---

[3]   The model and timing is similar to others in which the household division of labor and childrearing expectations affect men's and women's *ex ante* training investments (see, for example, Becker 1991, Echevarria and Merlo 1999, Engineer and Welling 1999, and Hadfield 1999). However, while these models' *ex ante* investments follow from specialization in household production (and the inability to efficiently contract domestic work), *ex ante* investments in the model proposed here follow from variation in the geographic constraint and flexibility across occupations.

in low marriage search costs and high family earnings. It is symmetric in that earnings and the costs of marriage search are not conditioned directly on sex.[4] The timing is as follows:

1. Men and women choose to receive training[5] in either a clustered or dispersed occupation. Training is costless and binary, but may only be made in one occupation. Workers differ only in sex and are otherwise homogenous (i.e. in earnings by occupation and in marriage search costs). Workers choose training to maximize expected utility, which is additively-separable in their expected Period 2 marriage search costs and their Period 3 expected cumulative family earnings. If a sex is indifferent, its workers choose training in dispersed occupations.

2. Men and women engage in a costly marital search. Marriage is assortive, but only imperfectly so due to probabilistic "unconditional love" between pairs with two clustered or two dispersed occupations.

3. Families choose between two locations to maximize joint earnings. The key assumption is that compensating differentials make expected personal wages in clustered occupations superior to those in dispersed occupations only if that worker can choose location to maximize personal wages. Therefore, expected family wages are greatest when one spouse enters a clustered occupation and the other enters a dispersed occupation.

The environment results in a coordination problem in *ex ante* human capital investments analogous to the classic "battle of the sexes" game, with two pure-strategy Nash equilibria conforming to sex-based occupational clustering and an unstable mixed-strategy equilibrium.

Conceptually, occupations are assumed to be clustered or dispersed for reasons exogenous to their sexual composition. This may be due to production technology (engineers or scientists with highly-specialized skills may benefit from local knowledge spillovers), input supply (mining is clustered around natural resources), or product markets (physicians and hairdressers are dispersed because they provide services that are universally demanded and must be done in-person). Ellison and Glaeser (1997) examine industrial agglomeration in greater detail. Household production might also be understood as a "dispersed" occupation, in that the productivity of a non-labor force participant is likely to be robust to exogenous relocations.[6]

---

[4] Relaxing symmetry by conditioning earnings on sex initially shifts the mixed strategy Nash equilibrium and makes one pure strategy Nash equilibrium superior to the other, and at the extreme, results in a unique equilibrium. Conditioning earnings directly on sex may be justified by pregnancy (Polachek 1981; Engineer and Welling 1999), comparative advantage in childrearing (Becker 1965, 1985; Mincer and Polachek 1974), discrimination (Becker 1971), or natural developmental factors advantaging women in educational attainment (Buchmann, DiPrete, and McDaniel 2006; Goldin, Katz, and Kuziemko 2006). Conditioning the disutility of marriage search by sex may be justified, for example, by differences and men's and women's fertility by age.

[5] "Training" in an occupation should be interpreted broadly, and may include formal schooling, on-the-job training, progression through career ladders, and accrued tacit knowledge of how to succeed in one's field, formation of professional networks, and so on.

[6] Engineer and Welling (1999) propose a similar model in which men and women marry out of love (analogous to $p_u = 1$ in period 2 of this model), resulting in a coordination problem that prompts men and women to specialize in domestic or external labor in advance of marriage. While their model's coordination problem arises from the complementarities of external and household labor (*i.e.*

32

The exposition proceeds by backward induction.

A. Period 3: Job Search Environment and Expected Household Earnings

The earnings subgame adapts a household relocation framework in the spirit of Mincer (1978). Unlike Mincer's model, employment opportunities in some occupations are location-invariant (dispersed occupations, such as secretaries) while employment opportunities in others differ by location (clustered occupations, such as nuclear engineers).

Let W_{ij} denote expected family earnings, which is the expected sum of the heads' wage rates, where i and j represent the occupations of the family heads. Occupations i and j may be clustered ("c") or dispersed ("d"). Because wages are not directly conditional on sex, W_{ij} can take three values: W_{cc}, W_{cd}, and W_{dd}.

To characterize W_{ij}, consider the following job search environment. Assume there are two locations, no relocation costs, and spouses must work in the same location. Families choose location to maximize expected joint earnings and locate randomly if indifferent between the two locations. Because both spouses gain utility from W_{ij}, consider the decision rule of the family as the decision rule consistent with the individual interests spouses. Wages in dispersed occupations are location-invariant and fixed at w_D, while wages in clustered occupations are w_H in one location and w_L in the other, with $w_H > w_L$. Wages w_L may represent the best outside-option for geographic mismatch, which may involve foregoing opportunities for career-advancing work relocations and job transfers, switching occupations, or exiting the labor force entirely.[7] The next section confirms occupational clustering is strongly and positively correlated with relocation for work for never-married and for married men and women.

The first key assumption is that wages in clustered occupations compensate workers for the disamenity of geographic constraint and the *ex ante* risk of the colocation problem (*i.e.* displacement resulting in w_L), such that $w_H > w_D > w_L$. The second key assumption is that, if the worker is employed in one of the two locations randomly (*i.e.* due to locating for the spouse), expected wages in dispersed occupations are higher than those in clustered occupations, ie. $w_D \geq 0.5w_H + 0.5w_L$.[8] Intuitively, if a household chooses location to maximize one worker's wages (*i.e.* because the other spouse has a dispersed occupation and earns w_D regardless of location), expected personal wages will be superior in the clustered occupation, and if the household does not choose location for the worker (*i.e.* because the spouse has a clustered occupation), expected

---

domestic production cannot be efficiently outsourced), this model coordination problem arises from the complementarities of clustered and dispersed occupations (*i.e.* individuals are bound by the desire to colocate).

[7] Relocation is treated here as a single event that occurs after the location of w_H is revealed in Period 3. However, since relocation is costless, there is no discounting, and workers are risk-neutral, there is no material difference between this model and one in which a worker with a clustered occupation must dynamically relocate to receive w_H.

[8] These assumptions could be treated as properties of a model where compensating differentials lead workers of one type (e.g. men) to segregate into occupations featuring a disamenity (e.g. geographic constraint), just as dangerous occupations pay a compensating differential and attract workers who are less risk-averse with regard to injury. Here, the disamenity of constraint for one sex is reduced (increased) by the segregation of the other sex into dispersed (clustered) occupations.

personal wages are superior in the dispersed occupation. I present evidence for compensating differentials for men and colocation penalties for women in the results section.

Note that, in the case of pure segregation by sex, workers of the sex segregating into clustered occupations always earn w_H, while workers of the other sex always earn w_D and are deterred from clustered occupations by the probability of w_L. While the first assumption prompts workers of one sex to enter clustered occupations, the second assumption captures the anti-complementarity of clustered occupations that exacerbates the colocation problem and deters workers of the other sex from also entering clustered occupations.[9]

For families with two clustered occupations in which high-wage opportunities are geographically-uncorrelated, there is a probability of 0.5 that wages w_H do not overlap and that one spouse earns w_L.[10] Expected wages are therefore $W_{cc} = 0.5(w_H + w_H) + 0.5(w_H + w_L)$. For families with exactly one clustered occupation, the family selects the location where the spouse with the clustered occupation earns w_H, and the family earns $W_{cd} = w_H + w_D$.[11] For families with two dispersed occupations, the location choice is immaterial and the family earns $W_{dd} = 2w_D$. The first key assumption implies $W_{cd} > W_{dd}$ and the second key assumption implies $W_{cd} \geq W_{cc}$.

B. Period 2: Marriage Search Environment and Expected Marriage Costs

In Period 2, workers engage in a marriage search. Suppose unmarried workers incur cost $c > 0$ to randomly-encounter an unmarried member of the opposite sex with whom, given that they have the opposite occupation type (they have "compatible careers"), they would marry. Furthermore, a share of matches $p_u \in [0, 1]$ would marry out of "unconditional love" (i.e. even if they have the same occupation type).[12] This quality is assumed to be match-specific and independent of past match quality. Therefore, with probability $(1 - p_u)$, a couple with the same occupational type will

---

9 Because clustering is treated as an exogenous feature of occupations and the empirical section treats clustering as relative, "pooling" into either clustered or dispersed occupations is neither meaningful or testable. Rather, we might instead interpret the latter assumption: the exogenously-determined distribution of occupations sufficiently wide such that some occupations are indeed sufficiently constrained, while others are indeed sufficiently flexible, to meaningfully impact the careers of spouses. Alternatively, rejecting the model's predictions may suggest the exogenous geographic distribution of occupations is insufficient to substantially impact spouses' careers.

10 In practice, opportunities are likely to be correlated, as about 4.5% of all dual-earner couples in the 2000 Census 5% PUMS work in the same occupation, more than would be expected by chance. The career compatibility of marriage partners who meet in the same city or in the same industry or occupation are also likely correlated. However, to the degree restricting the marital search to those within your same occupation, industry, or expected future location results in costly delayed marriage (ie. in period 2), the model's predictions are upheld.

11 In equilibrium, this is also consistent with research finding women are more-likely than men to be tied-movers.

12 Alternatively, in a continuous-time approach, potential marital matches may be thought of as a Markovian arrival process with a random non-monetary match-specific utility component, where workers adopt a higher reservation match quality when the spousal candidate has the same occupation (and thereby lower expected future family earnings in Period 3). Then p_u arises endogenously as the probability the match quality exceeds the reservation value, increasing with the variance of the non-monetary match-specific utility, and decreasing with the wage gaps. Parameters c and $p_u c$ would then respectively denote the average search costs associated with encountering the set and the subset of workers whom would marry.

34

not marry and will delay marriage by re-entering the "pool" of unmarried workers, re-incurring the search cost. Let ψ_m and ψ_f denote the share of unmarried men and women in the pool of unmarried workers who work in clustered occupations.

Noting that pairs only delay marriage if they are in the same occupation type and they are not members of the subset p_u, the probability a worker of sex i in a clustered occupation marries in a given period of search is $[1 - \psi_j(1 - p_u)]$, and the probability a worker of sex i in a dispersed occupation marries in a given period is $[1 - (1 - \psi_j)(1 - p_u)]$. Because workers who re-enter the pool are again randomly-matched to a spousal candidate of the opposite sex, the probability a worker of sex i marries a worker of sex j in the opposite occupation type is equal to the probability the worker marries a worker in the opposite occupation type given that the worker marries that period. For example, the probability a worker of sex i in a clustered occupation marries a worker of sex j in a dispersed occupation is $(1 - \psi_j)[(1 - \psi_j) + p_u\psi_j]^{-1}$. These probabilities are used to calculate expected future family earnings in Period 1.

Let p_i denote the share of sex i entering clustered occupations, where $i \in \{m, f\}$. Noting that the expected search duration is the inverse of probability of marriage in any search period, we may weight p_i by the search durations to compute share of the unmarried pool of sex i in a clustered occupation:

$$\psi_i = \frac{p_i[1 - \psi_j(1 - p_u)]^{-1}}{(1-p_i)[1 - (1 - \psi_j)(1 - p_u)]^{-1} + p_i[1 - \psi_j(1 - p_u)]^{-1}} \tag{1}$$

Lastly, note that the expected marriage search costs are given by the expected search duration multiplied by the marriage search cost c.

C. Period 1: Training Choice

In Period 1, risk-neutral men and women consider expected future wages and marriage search costs, and select their training. Because workers are risk-neutral and the utility of wages and disutility of marriage search costs are additively separable, a male trains in a clustered occupation if

$$\left(\frac{p_u\psi_f}{(1-\psi_f) + p_u\psi_f}\right)W_{cc} + \left(\frac{1-\psi_f}{(1-\psi_f) + p_u\psi_f}\right)W_{cd} - c\left(1 - \psi_f(1 - p_u)\right)^{-1} >$$

$$\left(\frac{\psi_f}{\psi_f + p_u(1-\psi_f)}\right)W_{cd} + \left(\frac{p_u(1-\psi_f)}{\psi_f + p_u(1-\psi_f)}\right)W_{dd} - c\left(p_u + \psi_f(1 - p_u)\right)^{-1} \tag{2}$$

And a female trains in a clustered occupation if

$$\left(\frac{p_u\psi_m}{(1-\psi_m) + p_u\psi_m}\right)W_{cc} + \left(\frac{1-\psi_m}{(1-\psi_m) + p_u\psi_m}\right)W_{cd} - c\left(1 - \psi_m(1 + p_u)\right)^{-1} >$$

$$\left(\frac{\psi_m}{\psi_m + p_u(1-\psi_m)}\right)W_{cd} + \left(\frac{p_u(1-\psi_m)}{\psi_m + p_u(1-\psi_m)}\right)W_{dd} - c\left(p_u + \psi_m(1 - p_u)\right)^{-1} \tag{3}$$

There are three Nash equilibria. First, consider the training decision of the atomistic male or female in the case that all men pursue training in clustered occupations and all women pursue training in dispersed occupations, such that $p_m = 1$ and $p_f = 0$. By equation (1), $p_m = 1$ and $p_f = 0$ implies $\psi_m = 1$ and $\psi_f = 0$. In this scenario, (2) implies it is incentive-compatible for the atomistic

male to enter clustered occupations if $W_{cd} - c > W_{dd} - cp_u^{-1}$, and (3) implies it is incentive-compatible for the atomistic female to enter dispersed occupations if $W_{cd} - c \geq W_{cc} - cp_u^{-1}$. These inequalities follow from the key assumptions regarding the wage environment, $p_u^{-1} \geq 1$, and $c > 0$. The second pure strategy Nash equilibrium is the symmetric case where all females pursue training in clustered occupations and all males pursue training in dispersed occupations. The third equilibrium is the unstable mixed strategy Nash equilibrium, whereby men and women sort into clustered occupations in an equal proportion that increases with W_{cc}, declines with W_{dd}, approaches 0.5 with W_{cd} for $p_u \neq 0$, and approaches 0.5 with c for $p_u \neq 1$. The pure strategy equilibria correspond to Hypothesis 1.

Comparative statics show that as sex j departs from the pure strategy Nash equilibrium, sex i's rate of substitution of expected costs in marriage search with costs in expected wages depends on p_u. When $p_u = 1$ (*i.e.* workers only marry out of love, regardless of occupational type), then all costs of sex i are borne by expected wages at a rate of $\psi_j W_{cc} + (1 - \psi_j) W_{cd}$ if sex i segregates into clustered occupations and at a rate $(1 - \psi_j) W_{cd} + \psi_j W_{dd}$ if sex i segregates into dispersed occupations. When $p_u = 0$ (*i.e.* workers would not marry out of love someone of the same occupational type), then all costs of sex i are borne by marriage search costs at a rate $-c\psi^{-1}$ for workers entering clustered occupations and $-c(1 - \psi)^{-1}$ for workers entering dispersed occupations. For $p_u \in (0, 1)$, workers substitute between the expected wage penalties of the colocation problem and the expected costs of delayed marriage, yielding both wage penalties and marriage search penalties for men and women who depart from the equilibrium. These correspond to Hypotheses 2 and 3.

The results section tests for and identifies the pure-strategy Nash equilibrium. It then tests the predictions corresponding to the Nash equilibrium—that men and women departing from this equilibrium incur wages and marriage search costs. Specifically, these predictions are[13]:

(i)   Sex-based occupational clustering: The degree to which a worker's occupation is geographically-clustered varies by sex

(ii)   Wage penalties for deviation: The sex segregating into clustered occupations has higher wages in clustered occupations than in dispersed occupations, and the sex segregating into dispersed occupations has higher wages in dispersed occupations than in clustered occupations

(iii)   Marriage penalties for deviation: The sex in which the majority sort into clustered occupations marries earlier in clustered occupations than in dispersed occupations, and the sex in which the majority sort into dispersed occupations marries earlier in dispersed occupations than in clustered occupations

---

[13]   Although this paper does not examine predictions regarding educational investments, existing empirical research has found that career pathways are formed early in one's life course and reflect the segregation in skill investments as predicted by the model. Daymont and Andrisani (1984) find that women are more likely than men to major in education, the humanities, and health or biology, and less likely to major in engineering, science or math, or professional studies, the latter of which will be shown to typify geographically-clustered occupations. Studies by McDonald and Thornton (2007) and Black et al. (2008) find that the majority of the sex pay gap among college-educated workers may be explained by choice of college majors, which is also consistent with segregation by women into majors leading to lower-paying, geographically-dispersed occupations.

In addition, results are expected to be strongest in occupations that require the greatest formal skills training, specifically those where a majority of workers have bachelor's degrees. In the United States, education through high school is highly-general, with specialization largely beginning with college coursework, post-secondary vocational training, and internships. It is this specialization that determines the geographic scope of the job search, the importance of calculated relocations for work, the robustness of the job to relocations for a spouse's career, and the magnitude of the sunk training costs involved in switching occupations in response to a colocation problem. In contrast, workers with less formal skills training may find job-switching more tenable. For example, although gaming cage workers (who account for and exchange chips at casino cages) are highly-clustered and dominated by less-educated women, relocation to a city without a gaming industry may not be very costly since foregone accumulated skills are minimal and potentially transferable to other less-skilled service-sector occupations (such as a cashiers).

D. Discussion on Potential Enrichments

The model makes a number of simplifying assumptions to focus on the coordination problem posed by occupationally-specific educational investments made in advance of knowing the spouse or from where job opportunities will come.

In the model, workers care only for joint family earnings and marriage. In reality, workers also choose their occupations according to their personal tastes. If some proportion of workers would select their occupation out of love of the occupation, the remaining will segregate into clustered and dispersed occupations, although the wage and marriage penalties for departing will fall as the proportion of workers selecting their occupation out of love for the occupation rises.

In the model, all workers receive training, all workers get married to partners of the opposite sex, and all people work. The results section considers exceptions conceptually. I interpret workers with less formal education as workers who should be able to change jobs cheaply (reducing the predicted magnitude of results. I present results for non-marriage. I also present separate results for married non-labor force participants, whose domestic work might be thought of as a "dispersed" occupation. I do not examine same-sex couples in this paper because predictions are theoretically ambiguous, and due to methodological issues.

In the model, workers are sequentially trained, are married, and then observe wage offers in different locations. In reality, foundational career investments, marriage search, and work relocation decisions are often long and overlapping processes rather than discrete sequential decisions.[14] Men and women may delay marriage search until after they have settled down in a long-term job, or they may intentionally search for a spouse within the same occupation or industry so that the geographic distribution of career advancement opportunities, while restricted, will be correlated. As a "Battle of the Sexes" game, the model requires some cost for avoiding the colocation problem after investments into a clustered or dispersed occupation have been made. As such, the model's predictions are robust to foregoing marriage until long-term location is known, restricting the marriage search to those whose location or occupation is known be

---

[14] However, the general timing of the model is consistent with Gautier, Svarer, and Teulings (2010), who interpret the tendency of young, single Danish workers to migrate to large cities, marry, and then move outside the city as evidence initial early-career location decisions are partly governed by the desire to reduce marriage search costs.

compatible, divorce due to differences in geographic preferences, or switching occupations after the spouse is known, as long as these alternatives are costly. Predictions are not robust to costlessly delaying the marriage search, costlessly restricting the marriage search, or costlessly switching among occupations or spouses.

Exceptions to these simplifying assumptions present a wide range of opportunities for testing other specific predictions of the model. The objective of this paper is to introduce a general life course framework for analyzing the colocation problem, and test its basic predictions for segregation, earnings, and marriage by sex and education.

II.   Measuring Occupational Clustering

The model's predictions are driven by anti-complementarities when both spouses in a dual-earner couple work in occupations that reward self-serving relocation decisions for career advancement, and the benefits of having one spouse in an occupation that is geographically-flexible and robust to work relocations for the spouse. Operationalizing the geographic constraint and flexibility of occupations is not straightforward, and directly examining work relocations by occupation is problematic for at least three reasons. First, job search behaviors and opportunities are likely to be endogenous by age, sex, and marital status. Second, the penalties for tied-spouses are difficult to estimate using data on relocations, particularly among "tied-stayers." Third, data sources with work relocation information such as the CPS have insufficient statistical power to estimate the likelihood of relocation for work for individual occupations. The key advantages of focusing instead on occupational clustering is that it avoids these dilemmas, as the geographic distribution of occupations is likely to be (relatively) exogenous to their sexual composition and demographics, occupational clustering captures geographic constraint while ignoring tied moves and stays, and occupational clustering may be estimated with the (much more powerful) Decennial Census files. An outstanding limitation is that using the occupation's geographic distribution may feature measurement error in the degree occupations truly reward calculated relocation decisions or the degree to which they are truly robust to spousal relocations.

To measure occupational clustering, I construct an occupation-level data set from labor force participants residing in metropolitan areas in the 5% public-use microdata sample of the 2000 U.S. Decennial Census. The Census features the large number of observations (7.6 million metropolitan workers) across the 283 metropolitan statistical areas and the 474 six-digit SOC occupations necessary to analyze occupational clustering. I control for two occupational characteristics quantified by the O*NET Database[15]: the degree to which aptitude for "explosive physical strength" and "assisting and caring for others" affects job performance. Measures for each vary from 0 to 100. These represent the typical human capital and taste-based explanations for occupational segregation (see Anker 1997 for a review). I then derive a Duncan-type index for polytomous dissimilar variables. Duncan's D was originally used by Duncan and Duncan (1955) to study occupational stratification by summing the absolute differences between the proportions of two groups (such as men and women). Because the variable of interest is

---

[15]   O*NET (short for the Occupational Information Network) is sponsored by the Department of Labor and the Employment and Training Administration, and has served as the standard reference for job definitions since the Dictionary of Occupational Titles was phased out in 1991. It provides continuously-updated and standardized measures of occupational characteristics and is constructed from surveys of job incumbents and occupation analysts.

polytomous rather than dichotomous dissimilarity (metropolitan areas rather than sex), the generalized version allows analysis of clustering. This "clustering index" of occupation j takes the form

$$C_j = \frac{1}{2} \sum_{i=1}^{I} \left| \frac{n_{ij}}{n_j} - \frac{n_i - n_{ij}}{n - n_j} \right|$$

where I represents the set of metropolitan areas, n represents the counts of workers aged 18-65 in the labor force, and subscripts denote counts within metropolitan areas i and occupations j. For illustration, if occupation j represents exactly 1% of the workforce in every metropolitan area, the absolute sum of the differences (and the clustering index for that occupation) is zero, regardless of the sex composition across metropolitan areas or other occupations. As workers are concentrated into fewer metropolitan areas, the clustering index converges to one. Like the original Duncan's D, the generalized "clustering index" also has an intuitive interpretation: it is the proportion of workers within an occupation that would need to relocate for the share of the occupation to be uniform across metropolitan areas. For example, in 2000 and within metropolitan areas, there were seventeen registered nurses per thousand labor force participants in the United States, and the (relatively low) clustering score of 0.084 for registered nurses implies 8.4% of registered nurses would need to relocate for there to be seventeen registered nurses per thousand workers in every metropolitan area in the United States. Following this interpretation, it is also intuitive to see that the index places greater weight on clustering in more populous cities. Functionally, occupational clustering in more-populous cities affects the national mean (the right hand term) more-so than clustering in small cities.

[TABLE 1]

Table 1 lists the three non-military occupations with the highest and lowest clustering indices by education and sex. Educational categories are assigned from the most-likely educational category for a randomly-sampled worker in that occupation (*i.e.* the modal educational attainment). Highly-clustered occupations that employ highly-educated men tend to be in specialized sciences and engineering, while those for highly-educated women tend to be diverse and smaller occupations, such as museum curators and archivists. Highly-educated and dispersed occupations that are dominated by men include physicians and managers, while those dominated by women are larger occupations, many of which are in teaching and education.

Among occupations employing less-educated workers, men dominate technical occupations, including both dispersed occupations (such as computer and auto repair) and clustered occupations (assemblers). Highly-dispersed, some-college, female-dominated occupations are some of the largest in terms of total employment, including registered nurses, secretaries and administrative aides, auditing clerks, customer service representatives, and nursing home aides.

Interestingly, clustered occupations dominated by men with less formal skills training (for example, machinists, technicians, and extraction workers) appear to involve a great degree of highly-specific vocational skills. Workers in these occupations are likely to involve foregone wages for exogenous relocations, but may not be recognized by measures of formal education. In contrast, less-educated but highly-clustered occupations dominated by females appear to involve highly-general and transferable skills, and are likely to be less-penalized by relocation. I find that

less-educated men earn a premium in clustered occupations, but not women (see the results section).

This index is then normalized by log-transformation. Among labor force participants aged 18-65, the log-transformed clustering index has a mean of -2.17, a standard deviation of 0.55, a minimum of -3.26, and a maximum of -0.23 across all occupations.

[TABLE 2]

Table 2 summarizes the clustering index by sex across educational attainment and marital status. To provide an intuition, the mean clustering index of the occupations held by men is 0.247 log-points (28%) greater than that for women. The relatively large difference in the columns than the rows implies sex is a much more powerful predictor of the degree to which a worker's occupation is clustered than other demographic characteristics, and two-sample t-tests conclude that the mean log-clustering index is lower for women across each educational and marital category. Clustered occupations have relatively high shares of workers who have graduate degrees, are married with absent spouses or separated, and who are working full-time, though effects for these are small (2%-6%). Clustered occupations differ little in whether workers have children, although the interpretation is complicated by selection bias caused by the lower labor force participation for women with children.

Before checking the model's predictions for occupational clustering and related wage and marriage penalties, I examine the model's implication that occupational clustering prompts workers to relocate for work, since wages are location-invariant only in dispersed occupations. I do so by estimating the probability workers and couples "relocate for new job or job transfer" using the March CPS Supplements (2003-2010) that use the 2000 Census SOC codes. Households attribute about 10% of all relocations primarily for work-related reasons, with other major reasons including family, change in marital status, and upgrading housing. I restrict the sample to prime age workers (25 to 40) because most workers are done with schooling, are highly at-risk of both relocation and marriage, and likely to be selecting job offers to accumulate skills. The clustering variable is appended to the March CPS data from calculations performed using the 2000 Census. Analysis is conducted for never-married workers, one-earner couples, and two-earner couples. For two-earner couples, propensity to relocate is examined using the maximum of the heads' log-clustering indices. The effect of clustering is calculated separately when this corresponds to the husband's or wife's occupation. Specifically, logistic regression's estimated log-odds a family relocates for work is given by

$$\ln\left(\frac{\widehat{p}_i}{1-\widehat{p}_i}\right) = \widehat{\beta}_0 + \widehat{\beta}_1 \ln C_{i,H} X_i + \widehat{\beta}_2 \ln C_{i,W} (1 - X_i)$$

where $\widehat{p}_i$ is the logit function's estimated probability that household i relocates for work, $C_{i,H}$ is the husband's clustering index, and X_i is an indicator for $C_{i,H} \geq C_{i,W}$. Results are shown in Table 3.

[TABLE 3]

Consistent with the model, never-married prime-age men and women are both more likely to relocate for work when they work in clustered occupations. This lends support for the model's

implication that geographically-clustered occupations reward and induce relocation for work among those who are most-able to relocate.

Among households where on the husband works, families in which the husband's occupation is more-clustered are more-likely to relocate for work. The CPS includes very few married couples where only the wife works that move for work-related reasons, making similar estimates for women unreliable. Regressions 7 and 8 show that the propensity to relocate for work among two-earner couples are about equally-sensitive to whether the husband or wife has the more-clustered occupation. Children have a negative effect on propensity to relocate in all regressions. Restricting the sample to bachelor's degree-holding workers and households (not shown) yields estimated effects of occupational clustering on relocation for work that are, for all models, higher than those from the full sample.

III.   Results for Segregation, Wages, and Marriage

First, I test for sex-based occupational clustering (Hypothesis 1). Upon establishing the clustering of men into clustered occupations, I test the corresponding predictions regarding the wage penalties (Hypothesis 2) and marital penalties (Hypothesis 3) predicted to affect the minority who depart from this equilibrium.

A.   Sex-Based Occupational Clustering in the Family and Labor Market

The model predicts men or women will segregate differentially into clustered and dispersed occupations, rather than occupational clustering being unconditioned on sex. These results will be used to identify the corresponding equilibrium predictions around marriage and wages.

Table 4 compares labor force participation and occupational clustering within families and breaks down tabulations by the mean age of the heads. Consistent with the model, men in both age groups are significantly and substantially more-likely to work in the more-clustered occupation.

[TABLE 4]

The table also provides intuitive measures of the magnitude of sex-based occupational clustering. In 2000, among both younger and older non-power couples and power-couples (those with two bachelor's degree holders), couples in which the husband's occupation had a clustering score greater than the wife's occupation outnumbered the reverse nearly two-to-one. Male-breadwinner families outnumber female-breadwinner families more than six-to-one. Household labor may be thought to be a "very-dispersed" occupation, and highly-amenable to relocation.[16]

Table 5 examines sex-based occupational clustering in the economy as a whole by presenting logistic regression results estimating the likelihood a worker is female given the log-clustering index of his or her occupation using the Decennial Census. First, I run bivariate regressions for

---

[16]   However, household labor may also be complementary to external labor market labor for other reasons other than geographic constraint and flexibility. Either complementarities in domestic and external production, or Ricardian specialization in domestic and external production along with an inability to efficiently outsource domestic production, are typically essential to household models (see Becker 1991 for a discussion).

41

all 474 occupations. Next, I perform additional analyses to isolate the effect of geographic clustering from alternative neoclassical explanations for occupational segregation. To address potential alternative explanations for occupational segregation, I control for the occupation's usual hours, its mean age, its requirements of physical strength, and its requirements of assisting and caring for others. Next, I restrict the sample to the subset of "highly-educated" occupations that further excludes 366 of those where the majority of workers have not obtained a bachelor's degree (or greater).[17] This subset focuses on occupations for which early-career training investments are likely to be greatest and most occupationally-specific, raising the cost of mid-career occupation switching and generating the theory's specific prediction that segregation will be most punctuated among high-skill occupations. In contrast, sex-based occupational clustering should break down if switching were costless and did not require leaving productive skills fallow, which may be the case among low-skill occupations. For example, switching from a gaming cage worker (clustered) to a cashier (dispersed) due to a spouse's relocation may involve less foregone income than switching from one highly-technical specialty to one employing broader skills at the destination city.

[TABLE 5]

Consistent with the theory, Table 5 finds a pattern of sex-based occupational clustering whereby men tend to enter more geographically-clustered than do women. Controls for the occupation's mean age and usual hours worked are shown to be weak predictors of occupational clustering. Not surprisingly, among all occupations, women segregate into jobs requiring less physical strength and greater care work. Among highly-educated occupations, the relationships between these occupational characteristics and the sex composition are weaker. Including further controls reduces the magnitude but retains the significance of occupational clustering in the full regressions. As predicted, occupational clustering is much more-pronounced in occupations in which the majority of workers have bachelor's degrees. Among these occupations, controlling for occupational characteristics calculated from the Census and O*NET has a small and statistically insignificant effect on the occupational clustering coefficient.

To provide an intuition for the magnitude, Regression 2 estimates that a worker in an occupation with a log-clustering score one standard deviation above the mean (more-clustered) has a 36% probability of being female, and one with a log-clustering score one standard deviation (more-dispersed) below the mean has a 55% probability of being female. For workers in high-skill occupations, Regression 6 estimates these to be 34% and 78%, respectively.[18]

[FIGURE 1]

Figure 1 provides a scatterplot of the clustering index against female composition for both the full set of non-military occupations and then for the subset of highly-educated occupations where

---

[17]   The only occupation omitted for being "physical" that would otherwise appear among the "highly-educated" occupations is "55-1010: Military Officer Special and Tactical Operations Leaders/Managers." This occupation is male-dominated, highly-clustered, and omitted in both subsamples. Incidentally, military officers have attracted special research attention because they are perceived to be very challenging for dual earner couples. For example, Gill and Haurin (2002) examine how the decision to pursue military officer training is affected by their wives' earnings and labor force attachment.

[18]   Estimates are calculated at the mean values for average age and usual hours.

the majority of workers have bachelor's degrees or greater. In both samples, there is a downward and statistically significant slope indicating that highly clustered occupations tend to have fewer women, particularly among highly-educated occupations. It is also clear that larger occupations tend to be less clustered and more female-dominated, while smaller occupations tend to be more clustered and more male-dominated. Generally, the largest occupations—secretaries, retail salespeople, school teachers, cashiers, customer service representatives, and nurses—all perform tasks that must be done in-person and serve a population's basic needs. On the other hand, the smaller occupations often perform specialized tasks specific to industries that benefit from clustering, and tend not to produce non-transportable in-person services, with very few exceptions ("septic tank servicers and sewer pipe cleaners" and "motion picture projectionists" are among the smallest twenty occupations and are relatively dispersed). It also may be the case that larger occupations are not only dispersed, but also offer well-functioning and competitive ("thick") labor markets in any given city, making them attractive to potential tied-movers.

Graphical analysis and Cooks D show that influential observations yielding the significant downward OLS regression line include teaching, health, and administrative support occupations (highly female, highly dispersed) and engineering occupations (highly male, highly clustered). Influential observations working against the downward trend include sewing machine operators and gaming cage workers (highly female, highly clustered) and truck drivers (highly male, highly dispersed). More broadly, exceptions tend to be among occupations requiring less formal education, and thereby potentially less of a sunk cost for switching less-skilled occupations. Indeed, the second panel of Figure 1 illustrates Table 5's finding that occupational clustering is more-pronounced in the more highly-educated occupations, with few exceptions. The most influential residuals among occupations where the majority of workers have bachelor's degrees are "Chief Executives" and "Clergy," both of which are large, male-dominated, highly-dispersed occupations, although these are problematic for other reasons.

Both family and economy-wide results are consistent with sex-based occupational clustering in which men segregate into clustered occupations. Following the model, this then yields predictions regarding how wage premia of clustered occupations vary by sex.

B. Wage-Age Profiles by Sex, Education, and Clustering

Having established the segregation of men into geographically-clustered occupations, this section tests the model's corresponding predictions regarding the wage premia in clustered occupations by sex. Formally, in perfect segregation men always enjoy the compensating differential $w_H > w_D$ in clustered occupations, while women are deterred from clustered occupations by the risk post-marital geographic displacement (w_L). In contrast, the model assumes that dispersed occupations are robust to spousal relocations, and fixed at w_D. The model thereby predicts higher wages in clustered occupations for men and higher wages in dispersed occupations for women. The model predicts these properties keep equilibrium segregation stable.

I also examine related properties of the model. First, the sex-pay gap should be driven by clustered occupations, not by dispersed occupations. This is because women are more-likely to

be penalized by geographic displacement in clustered occupations, while wages in dispersed occupations are fixed for both sexes (and all geographies) w_D.[19]

To analyze wage premia in clustered occupations at different ages by sex and education, I separate workers in the 2000 Decennial Census into terciles by occupational clustering score; workers in occupations with occupational clustering scores in the sixty-seventh percentile and above (greater than 0.164) are denoted as workers in "clustered occupations," and workers with scores in the thirty-third percentile and below (less than 0.092) are denoted as workers in "dispersed occupations." Table 6 reports estimated median wages in clustered and dispersed occupations by education, sex, and age. Estimates are made using quantile regression to predict real median wages as a function of age (linear and quadratic terms), estimated separately for each sex and at four education levels.

[TABLE 6]

Consistent with the model, median wages among men in clustered occupations are greater than median wages among men in dispersed occupations throughout the working life and at every educational category. Consistent with the model's prediction regarding the relative wage gap, the premium for working in clustered occupations among women is relatively small in the early-career and negative in the late-career. For both men and women, I expect the premia for clustered occupations to overestimate the premium for the family because it neglects the effect on the spouse's earnings. The overestimate is believed to be particularly great for women, since husbands tend to be in more-clustered occupations than their wives.

Table 7 also confirms that the sex wage gap is more pronounced in clustered occupations, suggesting wage penalties may have a deterrent effect on women. Within highly-educated occupations, the male wage premium is between -9% and 12% in dispersed occupations and between 5% and 40% in clustered occupations, depending on age and education. While the wage gap is similar throughout the life course in dispersed occupations, the gap in clustered occupations tends to grow throughout the life course. Similarly, among less-educated occupations, the wage gap grows more rapidly in clustered occupations.

[FIGURE 2]

Figure 2 illustrates the striking magnitude of the wage gap for highly-educated women in clustered occupations at later ages, which is absent for women in dispersed occupations. At early ages, women in clustered occupations enjoy greater median wages than those in dispersed occupations. However, median wages for highly-educated women in clustered occupations rapidly decelerate at later ages. For women with graduate degrees, median wages peak around age forty, and fall below highly-educated women in dispersed occupations in the mid-forties, while wages for highly-educated women in dispersed occupations continue to rise and (in

---

[19]   Selection bias is likely to complicate straightforward testing since labor force exit and occupational switching are not directly observed and would be expected to mitigate the earnings penalties of the colocation problem. For estimation purposes, this is particularly troublesome since workers may be prompted to exit the labor force or switch occupations due to a spouse's higher earnings (a labor supply-reducing income effect) or demanding career, both of which would be expected to be characteristics of clustered occupations. This bias may attenuate results.

contrast) closely track wages for highly-educated men in dispersed occupations. As a result, women in clustered occupations earn much lower wages than observably comparable men, while women in dispersed occupations have wages that are similar for comparable men.

One interpretation for the rapid deceleration is that highly-educated women in clustered occupations are more likely than men to be (or to have been) tied-stayers or tied-movers, and these career impediments manifest lower median wages, particularly when compared to men of equal age and education in clustered occupations. In contrast, wages for highly-educated women in dispersed occupations are very similar to those of men of equal age; the wage disparity is stark among highly-educated women in clustered occupations, and this disparity grows along the life course. This interpretation is consistent with Bertrand, Goldin, and Katz's (2009) finding that much of the sex pay gap among MBAs may be explained by the likelihood of job discontinuity (particularly from childrearing) within ten years after completing education. The distinction between highly-clustered and highly-dispersed occupations suggests that the fall from career ladders may be more common or pronounced among women who enter clustered occupations such as banking or consulting rather than dispersed occupations such as medicine or education; like childrearing, the tied-mover/tied-stayer problem affects women more than men.

Among women with some college and associate's degrees, those with clustered occupations do tend to earn more than those in dispersed occupations, but differences are very small. One interpretation is that retraining (ie. job-switching) costs for these occupations are low, and therefore it is feasible to retrain cheaply if a spouse relocates. Indeed, many of the female-dominated less-educated occupations that are highly-clustered (eg. textile occupations and gaming workers) appear to involve less-accumulation of on-the-job and tacit skills as those dominated by men (largely in the technical trades, manufacturing, and extraction).

Analysis of wage profiles by clustering, age, sex, and education supports the model's prediction that clustered occupations are associated with wage premiums for educated members of the sex generally entering clustered occupations (men), and high-but-decelerating premiums for educated members of the sex generally entering dispersed occupations (women).

C. Occupational Clustering and Marital Search

Because men are more-likely to enter geographically-clustered occupations, the model predicts that women who depart from occupational selection patterns by entering geographically-clustered occupations will, on average, experience later first-marriage. Following the conceptual framework, this is due to the relatively-slow arrival rate of marriage candidates with compatible (dispersed) occupations. Likewise, men who depart from occupational selection patterns by entering geographically-dispersed occupations will also experience a longer average marital search, as the arrival rate of marriage candidates with clustered occupations is more slow. Formally, in the case of pure segregation, men and women always marry in the first period at cost c, and are deterred from the other occupation by the expected cost cp_u^{-1}, where $p_u \in [0, 1]$ is the probability of marrying a random spousal candidate out of love and $p_u^{-1} \geq 1$.

[FIGURE 3]

Figure 3 examines this hypothesis by reporting the probability of being ever-married by age, education, and whether the individual's occupation is above or below median clustering.

Consistent with the hypothesis, women who enter clustered occupations experience later first-marriage. The quartile ages for first-marriage among highly-educated women are 24, 28, and 34 in clustered occupations and 23, 26, and 30 in dispersed occupations. By age thirty, 63% of highly-educated women in clustered occupations are ever-married, versus 74% of those in dispersed occupations. The effect among less-educated women is smaller, with an age-at-first-marriage quartiles at 21, 24, and 30 in clustered occupations and 21, 24, and 29 in dispersed occupations. Evidence is weaker for men. Quartiles among highly-educated men are the same in clustered and dispersed occupations: 24, 28, and 34. Among less-educated men, working in a clustered occupation has an opposite effect as women; quartiles in clustered occupations are 23, 25, and 32, while those in dispersed occupations are 23, 26, and 34.

One challenge of the Census data is that it does not distinguish cohabitating partners from married partners. This may bias results if women who enter clustered occupations are also more-likely to prefer cohabitation to marriage.

[FIGURE 4]

As noted by Mincer (1978), the colocation problem may cause marital tension. If the colocation problem is particularly-likely in families where women enter geographically-clustered occupations as is argued here, then Mincer's logic may be extended to use occupational clustering to predict divorce. Figure 4 shows that ever-married women in clustered occupations are more-likely to be currently divorced and not remarried.[20]

Consistent with the theory, women who enter clustered occupations are more likely to delay or forego marriage, particularly among the college-educated. While theory predicts the reverse for men (who, when entering dispersed occupations, delay marriage to find a female with a clustered occupation), evidence for this is weak among non-college-educated men, and there is no evidence of delayed marriage among college-educated men entering dispersed occupations. One interpretation is that college-educated men in dispersed occupations have distaste for delaying marriage to seek out a wife with a clustered occupation who would be the higher-earning "leading spouse."

IV.   Discussion and Future Work

This paper has three principal goals: to broaden the theoretical scope of the colocation problem to include its effects on men's and women's occupations, marriage timing, and wages; to introduce a clustering index capturing the variation in the constraint or flexibility of occupations as a tool for research on household relocation decisions; and to test a basic set of predictions regarding equilibrium segregation and the penalties predicted to keep it stable. A broader

---

[20]   Differences were also tested in a regression framework as follows. Eight regressions were run—one for each combination of education, sex, and clustering. The sample frame includes non-never-married individuals (*i.e.* those at-risk of being divorced). The dependent variable is a "divorced" marital status, and the independent variables are clustering, age, age-squared. All coefficients in all four regressions are statistically significant with $p<0.01$. Evaluated at age forty, women in clustered occupations had a higher probability of being divorced and men had a lower probability of being divorced, each with a $p<0.01$. The estimated difference for less-educated men, however, is very small (2%, standard error of 0.4%).

implication of the theory is that men's and women's expectations of future mobility may help perpetuate occupational segregation, the sex-pay gap, disproportionate marriage market penalties for women entering geographically-clustered jobs. A further implication is that these features of inequality are perpetuated by their aggregately self-fulfilling expectations of future mobility.

Results confirm the segregation of one sex (men) into geographically-clustered occupations.[21] Results suggest men are discouraged from entering dispersed occupations primarily by career costs, and women are discouraged from entering clustered occupations primarily by family costs. However, results suggest less-educated men suffer later marriage when entering dispersed occupations and more-educated women suffer lower mid- and late-career median wages in clustered occupations.

As an explanation of occupational segregation, the principal concern is that geographically-clustered occupations possess other qualities affecting segregation. Indeed, it is clear that dispersed and clustered occupations differ in several ways. Clustered occupations tend to be specific to clustered industries, which may be subject to the natural and technological agglomerative forces described by Ellison and Glaeser (1997), while dispersed occupations either tend to be specific to dispersed industries (e.g. education or healthcare) or be employed in supportive capacities across potentially-clustered industries (e.g. general managers or administrative assistants). While occupational clustering's relative-exogeneity with respect to sex is an appealing property, the underlying features that make occupations clustered or dispersed may also yield segregation through mechanisms unrelated to geography. However, I find little evidence that controls for other neoclassical explanations—such as average working hours, wage, physicality, or requirement of assisting and caring for others—reduce the explanatory power of geographic clustering on sex. An omitted variable of concern must also explain why the effect of clustering grows with formal education and has a differential effect on wages and marriage outcomes by sex.

In addition to explaining occupational segregation, wages, and marriage outcomes, the model may also help contextualize other labor market phenomena. First, because workers in dispersed occupations implicitly pay for the amenity of geographic flexibility, the model presents an explanation for why women segregate into lower-paying occupations and college majors with lower earnings (Blau and Kahn 2000, McDonald and Thornton 2007, Black et al. 2008). Second, the model provides an explanation for occupation switching and equilibrium skills mismatch, and may explain why women are more-likely to possess fallow skills (Johansson and Katz 2006, Ofek and Merrill 1997; see also McGoldrick and Robst 1996 for a critical view). Third, the model may explain why the expansion of women with bachelor's and graduate degrees are being absorbed by the most highly-dispersed occupations, including physicians, lawyers, dentists, and general managers. Fourth, the model helps explain "statistical discrimination" in training by proposing why women have higher turnover rates due to exogenous spousal relocation (for studies explaining discrimination in training by differences in turnover rates by sex, see Duncan

---

[21] It remains unclear why all societies appear to exist on the equilibrium where women segregate into dispersed occupations. One possibility is that men's rational expectations to be able to relocate for work, and women's expectation to be exogenously relocated, is a legacy of women's historic emphasis on domestic work. Alternatively, childrearing may disproportionately interrupt women's careers, and workers sort to avoid interrupting the dominant career.

and Hoffman 1979, Gronau 1988, Lillard and Tan 1986, and Royalty 1996). Fifth, the model explains why women are persistently more-likely than men to be tied-movers (Mincer 1978, Nivalainen 2004). In short, a wide variety of labor market phenomena might be better-understood by adopting a life-cycle approach and treating occupational choice as endogenous to geographic constraints.

Works Cited

Anker, Richard. 1997. "Theories of Occupational Segregation by Sex: An Overview." *International Labour Review* 136(3): 315-339.

Becker, Gary. 1965. "A Theory of the Allocation of Time." *Economic Journal*, 75(299): 493-517.

_____. 1971. *The Economics of Discrimination*. Chicago: University of Chicago Press.

_____. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics*, 3(1): S33-S58.

_____. 1991. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bertrand, Marianne, C. Goldin, and L. F Katz. 2009. Dynamics of the gender gap for young professionals in the corporate and financial sectors. Working Paper No. 14681, National Bureau of Economic Research, Cambridge, MA.

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor. 2008. "Gender Wage Disparities Among the Highly Educated." *Journal of Human Resources*, 43(3): 630-59.

Boyle, Paul et al. 2001. "A Cross-National Comparison of the Impact of Family Migration on Women's Employment Status." *Demography*, 38(2): 201-213.

Buchmann, Claudia, and Thomas A. DiPrete. 2006. "The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement." *American Sociological Review*, 71(4): 515-41.

Clark, William A. V., and Youqin Huang. 2006. "Balancing move and work: women's labour market exits and entries after family migration." *Population, Space and Place*, 12(1): 31-44.

Compton, Janice, and Robert A. Pollak. 2007. "Why Are Power Couples Increasingly Concentrated in Large Metropolitan Areas?" *Journal of Labor Economics* 25(3): 475-512.

Costa, Dora L., and Matthew E. Kahn. 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990." *Quarterly Journal of Economics*, 115(4): 1287-1315.

Daymont, Thomas N., and Paul J. Andrisani. 1984. "Job Preferences, College Major, and the Gender Gap in Earnings." *The Journal of Human Resources*, 19(3): 408-428.

Duncan, Greg J., and Saul Hoffman. 1979. "On-The-Job Training and Earnings Differences by Race and Sex." *The Review of Economics and Statistics*, 61(4): 594-603.

Duncan, Otis Dudley, and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review*, 20(2): 210-217.

Echevarria, Cristina, and Antonio Merlo. 1999. "Gender Differences in Education in a Dynamic Household Bargaining Model." *International Economic Review* 40(2): 265-86.

Ellison, Glenn, and Edward L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy*, 105(5): 889-927.

Engineer, Merwan, and Linda Welling. 1998. "Human Capital, True Love, and Gender Roles: Is Sex Destiny?" *Journal of Economic Behavior and Organization*, 40(2): 155-78.

Gautier, Pieter, Michael Svarer, and Coen Tuelings. 2010. "Marriage and the City: Search Frictions and Sorting of Singles." *Journal of Urban Economics* 67(2): 206-18.

Gill, Leroy, and Donald Haurin. 2002. "Wherever He May Go: How Wives Affect their Husband's Career Decisions." *Social Science Research*, 27(3): 264-79.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives*, 20(4): 133-4A.

Gronau, Reuben. 1988. "Sex-Related Wage Differentials and Women's Interrupted Labor Careers-the Chicken or the Egg." *Journal of Labor Economics*, 6(3): 277-301.

Hadfield, Gillian. 1999. "A Coordination Model of the Sexual Division of Labor." *Journal of Economic Behavior and Organization* 40(2): 125-53.

Jacobsen, Joyce P., and Laurence M. Levin. 2000. "The effects of internal migration on the relative economic status of women and men." *Journal of Socio-Economics*, 29(3): 291-304.

Johansson, M., and K. Katz. 2007. "Wage differences between women and men in Sweden: the impact of skill mismatch." An Institute for Labour Market Policy Evaluation working paper, No. 2007: 13.

King, Miriam, et al. 2010. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. Machine-Readable Database. Minneapolis: University of Minnesota.

Lillard, L. A, and H. W Tan. 1986. "Private Sector Training. Who Gets It and What Are Its Effects?" Rand Corporation.

Long, Larry H. 1974. "Women's Labor Force Participation and the Residential Mobility of Families." *Social Forces* 52: 342-48.

McGoldrick, KimMarie, and John Robst. 1996. "Gender Differences in Overeducation: A Test of the Theory of Differential Overqualification." *The American Economic Review*, 86(2): 280-284.

McDonald, Judith A., and Robert J. Thornton. 2008. "Do New Male and Female College Graduates Receive Equal Pay?" *Journal of Human Resources*, 42(1): 32-48.

McKinnish, Terra. 2008. "Spousal Mobility and Earnings." *Demography.* 45(4): 829-849.

Mincer, Jacob. 1978. "Family Migration Decisions." *The Journal of Political Economy*, 86(5): 749-773.

Nivalainen, Satu. 2004. "Determinants of Family Migration: Short Moves vs. Long Moves." *Journal of Population Economics*, 17(1): 157-75.

Ofek, Haim, and Yesook Merrill. 1997. "Labor Immobility and the Formation of Gender Wage Gaps in Local Labor Markets." *Economic Inquiry*, 35(1): 28-47.

Polachek, Solomon William. 1981. "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure." *The Review of Economics and Statistics*, 63(1): 60-69.

Royalty, Anne Beeson. 1996. "Effects of Job Turnover on the Training of Men and Women." *The Industrial and Labor Relation Review*, 49(3): 506.

Ruggles, Teven, J. et al. 2010. Integrated Public Use Microdata Series: Version 5.0. Machine-readable database. Minneapolis: University of Minnesota.

Tables and Figures

TABLE 1–Most and Least-Clustered Occupations by Education and Sex

Modal Education	Type	Occupations with Male Majority	Occupations with Female Majority
Grad. Degree (n=30)	Most-Clustered	Economists (0.46) Astron. & Phys. (0.45) Legislators (0.38)	Oth. Health Pract. (0.41) Audiologists (0.27) Archivists & Curators (0.27)
	Most-Dispersed	Physicians (0.12) Dentists (0.13) Clergy (0.14)	Educ. admins (0.08) Counselors (0.10) Spec. Ed. Teachers (0.15)
	No. Occs	*19*	*11*
College Degree (n=85)	Most-Clustered	Mine/Petrol Engineers (0.69) Marine Engineers (0.63) Nucl. Engineers (0.57)	Bio. Technicians (0.36) Rec. Therapists (0.34) Budget Analysts (0.3)
	Most-Dispersed	Wholesale Sales Reps (0.09) Other Managers (0.1) Pharmacists (0.12)	Elementary Teachers (0.06) Secondary Teachers (0.08) HR Managers (0.09)
	No. Occs	*52*	*33*
Some College/ Asso. (n=146)	Most-Clustered	Avionics Techs (0.54) Geo/Petrol Techs (0.49) Tool & Die Makers (0.47)	Occ. Therapist Aides (0.42) Brokerage Clerks (0.42) New Acct. Clerks (0.34)
	Most-Dispersed	Retail Sales Supers (0.05) Gen. Ops Managers (0.07) Non-Ret. Sales Supers (0.08)	Bookkeeping Clerks (0.05) Admin Support Supers (0.05) Secretaries (0.06)
	No. Occs	*69*	*77*
HS or Less (n=213)	Most-Clustered	Tire Builders (0.7) Shoe Machinists (0.71) Oth. Extraction Work. (0.65)	Textile Wind. Setters (0.79) Gaming Cage Work. (0.67) Textile Knit. Setters (0.64)
	Most-Dispersed	Food Serv. Managers (0.06) Janitors (0.07) Auto Service Techs (0.08)	Retail Salespersons (0.04) Receptionists (0.05) Hairdressers (0.06)
	No. Occs	*171*	*42*

Notes: Data Source: 2000 Decennial Census 5% PUMS. Clustering indices in parentheses. Four military occupations are excluded, each of which are highly-clustered and majority-male.

TABLE 2–Differences in Log-Clustering Indices to Unconditional Mean Among Females, by Sex, Education, Marital Status, Full-Time Status, and Children

	Males		Females	
	Mean	SD	Mean	SD
Total	0.247	(0.563)	$\equiv 0$	(0.523)
By Education				
High School Equiv. or Less	0.250	(0.568)	0.026	(0.556)
Associate's/ Some Coll.	0.194	(0.557)	-0.052	(0.489)
College Degree	0.220	(0.537)	0.002	(0.484)
Graduate Degree	0.301	(0.488)	0.088	(0.488)
By Marital Status				
Married, Spouse Present	0.246	(0.548)	-0.012	(0.514)
Married, Spouse Absent	0.275	(0.583)	0.039	(0.544)
Separated	0.242	(0.547)	0.056	(0.522)
Divorced	0.235	(0.548)	0.029	(0.513)
Widowed	0.237	(0.553)	0.029	(0.531)
Never married	0.201	(0.563)	-0.002	(0.523)
By Full-Time Status				
Works $\geq$ 35 hours per week	0.246	(0.550)	0.023	(0.512)
Works < 35 hours per week	0.211	(0.561)	-0.020	(0.522)
By Children				
Has Own-Children in Household	0.248	(0.547)	0.005	(0.513)
Does Not	0.223	(0.558)	-0.005	(0.522)
Observations	3,954,006		3,691,964	

Notes: Data Source: 2000 Decennial Census 5% PUMS. Log-clustering indices are normalized to the unconditional mean among females (-2.286) by subtraction. Standard errors by sex are both less than 0.0003. Standard errors by subcategory are all less than 0.002. Includes labor force participants aged 18-65.

TABLE 3–Logistic Regression Estimating Probability of Relocating for Work, All Education Levels

	Never-Married Workers		Married Couples with Spouses Present					
			One Spouse ILF				Both ILF	
	Men	Women	Husb-Only ILF		Wife-Only ILF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log-Clustering	0.207**	0.295**	0.309**	0.298**	0.113	0.127		
	(0.054)	(0.065)	(0.069)	(0.069)	(0.281)	(0.282)		
Husband's Log-C x (Husb More Clust)							0.19**	0.175*
							(0.069)	(0.070)
Wife's Log-C x (Wife More Clust)							0.163*	0.148*
							(0.069)	(0.070)
Family Has Child				-0.6**		-0.8*		-0.95**
				(0.112)		(0.31)		(0.07)
Constant	-3.0**	-3.0**	-3.0**	-2.5**	-3.8**	-3.2**	-3.9**	-3.2**
	(0.113)	(0.145)	(0.139)	(0.167)	(0.637)	(0.676)	(0.129)	(0.136)
LR Chi-Squared	14.7**	20.2**	19.8**	44.9**	0.2	5.8*	7.8*	172.4**
Observations	47354	40616	27201	27201	2810	2810	63233	63233

Notes: From CPS March Supplement 2003-2010. Standard errors in parentheses. Excludes workers reporting the primary reason for relocation was for reasons other than work or job transfer. Occupations are those reported within one year of migration.

** Significant at the 1 percent level.

 * Significant at the 5 percent level.

TABLE 4–Husbands and Wives Labor Force Participation and Relative Occupational Clustering by Mean Age and Whether Both Have Bachelor's Degrees

Family Type	All Couples		"Power" Couples	
	Age 25-45	Age 46-65	Age 25-45	Age 46-65
Dual-Earner Couples				
Husband's Occ More Clustered	61.6%	62.6%	57.2%	61.0%
Wife's Occ More Clustered	34.2%	32.8%	33.9%	30.5%
Same Occupation	4.2%	4.6%	8.9%	8.5%
Single-Earner Couples				
Only Husband ILF	86.5%	71.6%	92.6%	78.8%
Only Wife ILF	13.5%	28.4%	7.4%	21.2%
Observations	1,334,326	938,715	224,702	160,367

Notes: Data Source: 2000 Decennial Census 5% PUMS. Includes married couples with spouse present. "Power" couples are defined as families where both heads have bachelor's degrees. dual-earner couples, 31% of couples are single-earner couples, and 7% of couples have no labor Among all couples, 62% of couples are force participants. Among power couples, 71% are dual-earner couples, 26% are single-earner couples, and 3% have no labor force participants. "Ages" are mean age of husband and wife.

TABLE 5–Logistic Regression Predicting Workers' Sex (Female = 1, Male = 0) by Occupational Clustering

	All Occupations		Majority-Bachelor's Holding Occupations	
	(1)	(2)	(3)	(4)
Log-Clustering	-0.818**	-0.472**	-1.536**	-1.347**
	(0.183)	(0.170)	(0.252)	(0.196)
Occ. Mean Age		-0.006		-0.118**
		(0.019)		(0.0383)
Occ. Mean Hours		-0.024**		-0.026**
		(0.001)		(0.001)
Occ. Explosive Strength		-0.063**		0.042*
		(0.017)		(0.019)
Occ. Assisting and Caring		0.042**		0.008
		(0.008)		(0.005)
Constant	-1.868**	-1.784	-3.063**	2.925
	(0.348)	(0.986)	(0.436)	(1.494)
Clusters (Occupations)	474	429	97	88
Unique Observations	8,295,671	7,567,351	1,475,610	1,409,053
Pseudo R-Squared	0.034	0.060	0.086	0.124

Source. 2000 Decennial Census 5% PUMS.

Notes: Standard errors clustered by occupation in parentheses. The occupation's mean age and weekly hours are calculated from the Census. The required ability of "explosive strength" and work activity of "assisting and caring for others" is calculated from O*NET. Some miscellaneous occupations cannot be matched in O*NET.

** Significant at the 1 percent level.
* Significant at the 5 percent level.

TABLE 6–Quantile Regression-Estimated Median Wages and Clustered Occupation Wage Premiums by Education, Age, and Sex

	Age 25			Age 40			Age 55		
	Men	Women	Gap	Men	Women	Gap	Men	Women	Gap
(i) Graduate									
Clustered	$16.58	$15.80	5%*	$26.39	$20.99	26%*	$27.53	$20.74	33%*
	(0.16)	(0.13)		(0.06)	(0.05)		(0.07)	(0.07)	
Dispersed	$12.24	$13.51	-9%*	$22.92	$21.09	9%*	$25.56	$23.00	11%*
	(0.21)	(0.10)		(0.07)	(0.03)		(0.07)	(0.04)	
Clust. Premium	35.3%*	16.9%*		15.1%*	-0.4%*		7.73%*	-9.8%*	
(ii) Bachelor's									
Clustered	$15.25	$13.67	12%*	$23.87	$18.40	30%*	$24.02	$17.26	39%*
	(0.06)	(0.08)		(0.03)	(0.05)		(0.04)	(0.07)	
Dispersed	$12.06	$12.26	2%*	$18.91	$16.86	12%*	$19.41	$17.50	11%*
	(0.09)	(0.05)		(0.05)	(0.03)		(0.06)	(0.03)	
Clust. Premium	26.4%*	11.4%*		26.2%*	9.10%*		23.7%*	-1.3%*	
(iii) Some Coll./A.D.									
Clustered	$10.86	$9.34	16%*	$17.04	$13.02	31%*	$17.71	$13.03	36%*
	(0.02)	(0.02)		(0.01)	(0.01)		(0.01)	(0.02)	
Dispersed	$9.74	$8.66	12%*	$14.37	$12.22	18%*	$14.60	$12.76	14%*
	(0.02)	(0.01)		(0.02)	(0.01)		(0.02)	(0.01)	
Clust. Premium	11.4%*	7.81%*		18.6%*	0.06%*		21.3%*	2.10%*	
(iv) High School/Less									
Clustered	$9.15	$7.57	21%*	$12.87	$9.33	38%*	$13.19	$9.60	37%*
	(0.01)	(0.01)		(0.00)	(0.01)		(0.00)	(0.01)	
Dispersed	$8.61	$7.22	19%*	$11.63	$9.15	27%*	$11.90	$9.60	24%*
	(0.01)	(0.00)		(0.01)	(0.00)		(0.01)	(0.00)	
Clust. Premium	6.34%*	4.88%*		10.6%*	1.94%*		10.8%*	0.03%	

Notes: Data Source: 2000 Decennial Census 5% PUMS. Median wage estimates are calculated from sixteen quantile regressions (two sexes, two occupational categories, and four educational attainments) with linear and quadratic terms for age. Standard errors are calculated by bootstrapping.

* Significant at the 1 percent level.

FIGURE 1
Scatterplot and OLS Fit of Female Share versus Log-Clustering Index

(i) All Occupations

(ii) Highly-Educated Occupations
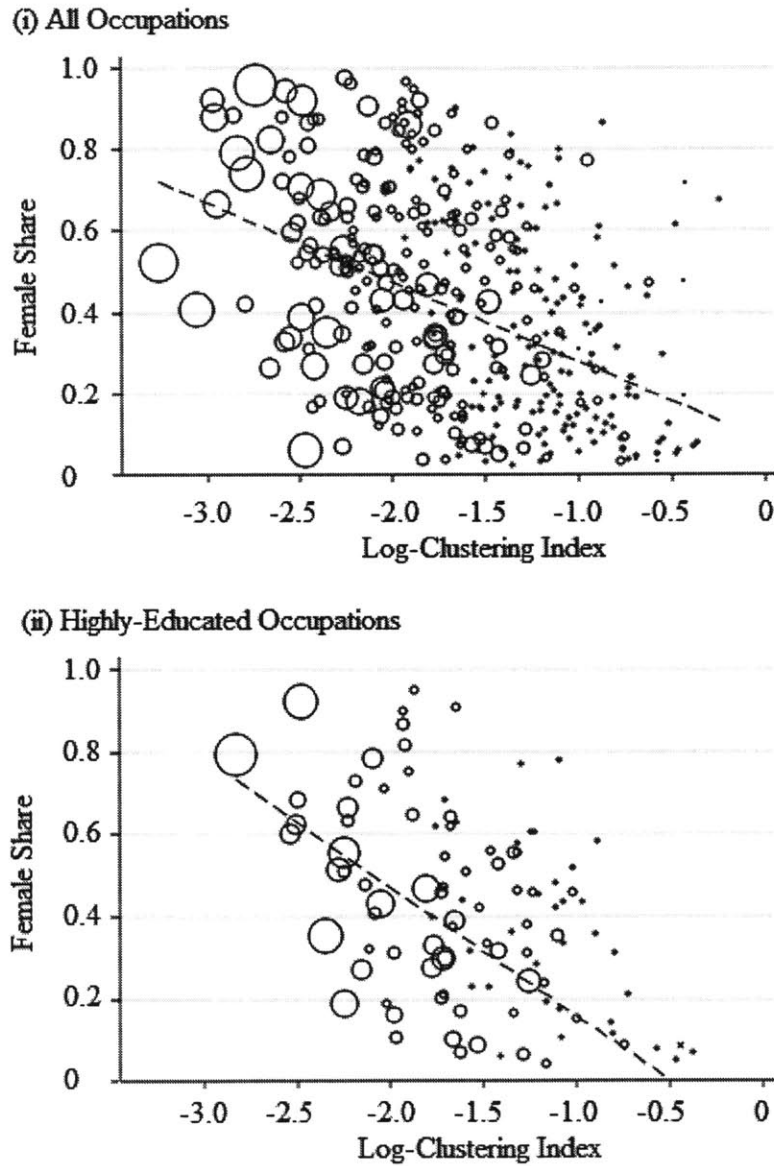
Notes. Circle size denotes number of workers in the occupation.
Scatterplot (i) excludes extraction, construction, production, farming, and
military occupations. Scatterplot (ii) excludes occupations in which the
majority of workers do not hold bachelor's degrees.

FIGURE 2

Median Wages by Sex, Education, Age, and Occupational Clustering
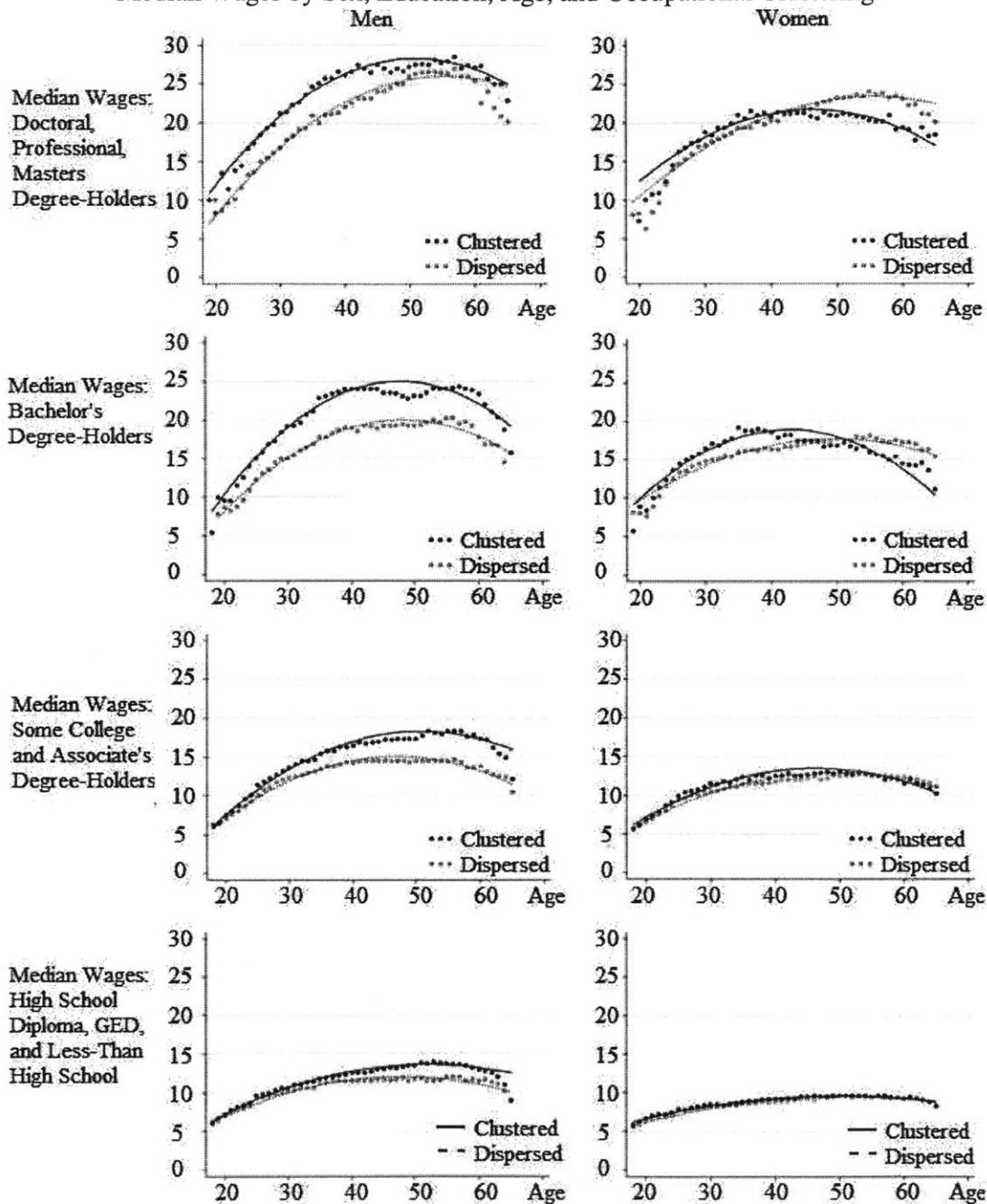
Notes. Lines represent predicted median wages using quantile regression with linear and quadratic age terms.

FIGURE 3

Delayed Marriage: Share of Men and Women
Who are Non-Never Married, by Age, Occupation Type, and Education
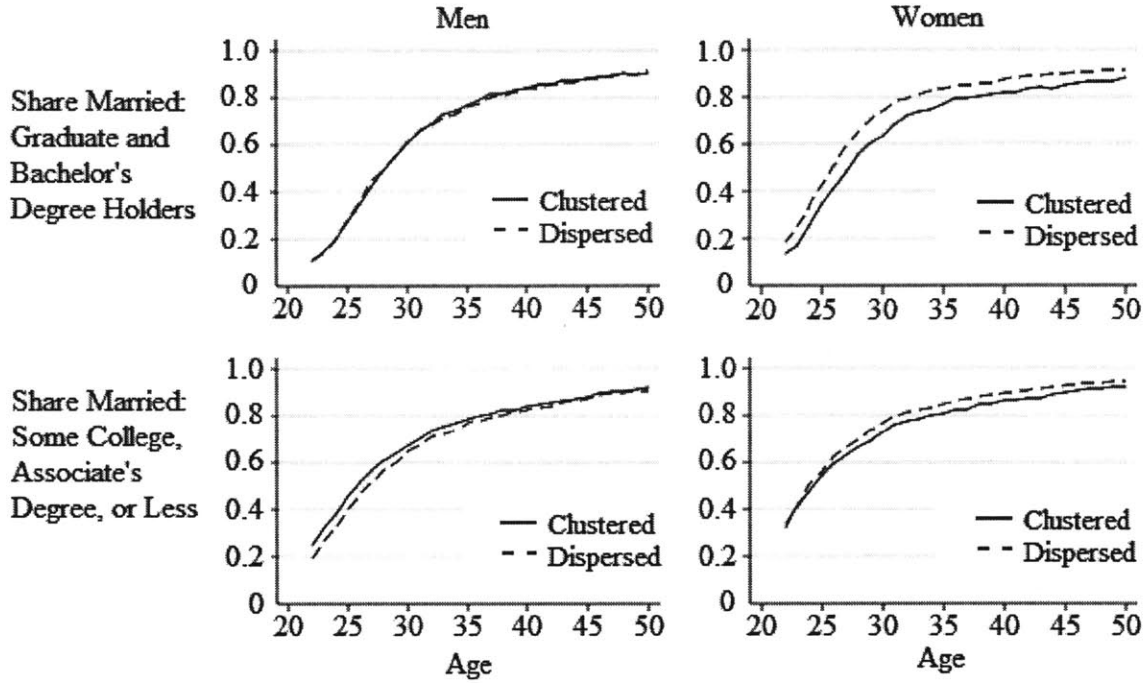
Notes. Shares are calculated separately at each year of age. Standard errors are at most 0.0057.

FIGURE 4
Share of Men and Women who are Divorced (and Not Re-Married),
by Occupational Clustering and Education
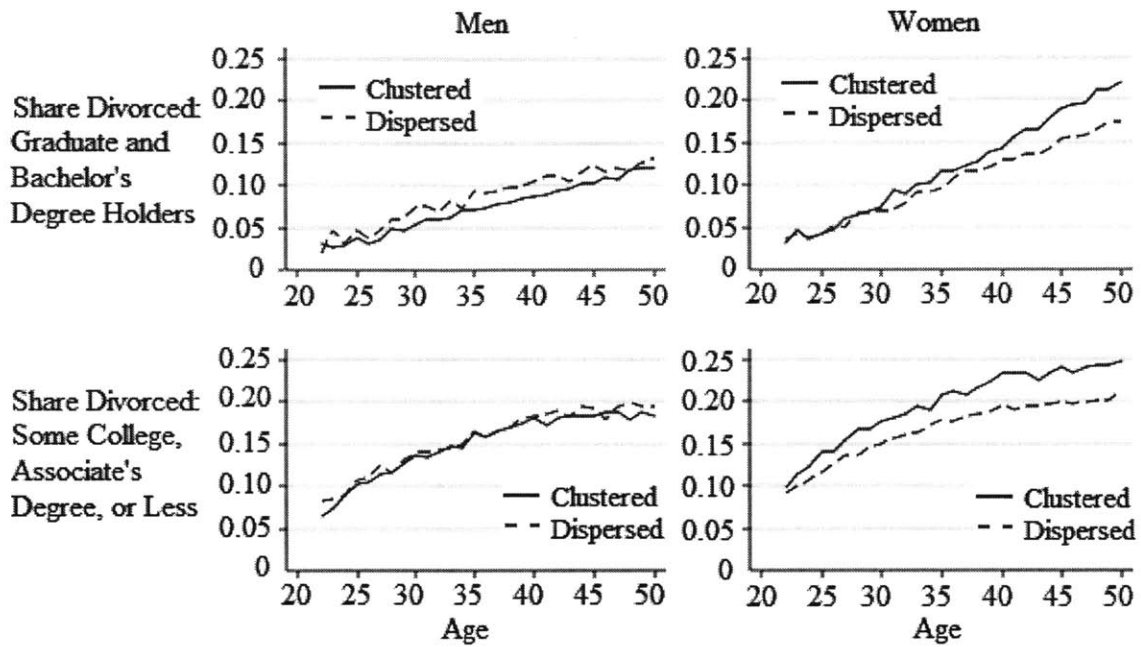
Notes. Shares calculated separately at each year of age. Standard errors are at most 0.0043.

Chapter 3

Do Agents Game Their Agents' Behavior? Evidence from Sales Managers[1]

The constraints on principals' ability to induce efficient behavior through their economic agents are the defining determinants of economic organization. In the classic principal-agent model, a principal (e.g. the firm, which provides employment and a compensation plan) contracts directly with its agents (e.g. the worker, who provides productive effort). In practice, profit-maximizing principals are far-removed from rank-and-file agents. For example, shareholders of publicly traded firms rely on a long chain of intermediary executives and managers to set and monitor workers' employment practices on their behalf. As such, models invoking a profit-maximizing firm implicitly assume the interests of their intermediary agents, even if they are not identical to those of shareholders, are sufficiently aligned that their ultimate employment practices also approximate profit-maximizing behavior.

Although it is well-known that agents' incentive plans may encourage activities that are inconsistent with the interests of principals, evidence of such "gaming" draws almost exclusively from the top and bottom of organizations (e.g. CEOs and rank-and-file workers; see Murphy 1999 and Lazear and Oyer 2009 for reviews). However, little empirical work examines how misaligned managerial incentives propagate to subordinates. While this issue is difficult to examine, it is also important, since managers are the intermediary agents responsible for making decisions on behalf of the "Firm." Indeed, early organizational researchers dismissed profit maximization as the chief motive governing managerial decision making. Based on their observations, they concluded managers are imperfect and self-interested coordinators of economic activity, that firms should not be treated as monolithic, and that the inability of organizational hierarchies to coordinate activities efficiently determines firm structure, governance, and scope (classic studies include Baumol 1959; Chandler 1977; Coase 1937; Crozier 1964; Cyert and March 1963; Penrose 1959; Simon 1957, 1964; and Williamson 1963, 1967).[2]

---

[2] This emphasis might be attributed to the three common strategies for acquiring data--using publicly-available accounting data, company-researcher data use agreements for single-firm studies, or sports statistics for athletes.

In this paper, I theorize why firms delegate authority to intermediary managers, identify the misuse of authority and incentives, and describe how sales organizations attempt to control the costs of managerial gaming affecting salespeople. I propose firms delegate authority over staffing and incentive decisions to immediate managers (even though sales are observed by the firm) because sales managers accumulate private information allowing them to distinguish salespeoples' ability from exogenous circustances affecting performance, such as market conditions, product quality, or territory quality.[3] Managers' private information allows them to screen and incentivize salespeople more efficiently than would a firm that conditions these decisions on sales figures alone. Although managerial incentives generally align managers' decisions with profit maximization, authority and quotas provide an opportunity and a motive for managers to shift sales to their desired measurement periods through staffing and incentive decisions that affect subordinates.

The model yields the hypothesis that quotas distort managerial incentives to make decisions that are consistent with the interest of the firm. Intuitively, the model captures the institutional features that allows decisions motivated by the manager's personal interests to be identified: (i) that sales managers' have a unique interest in the marginal sales that meet their quotas, (ii) their quota attainments are determined by the cumulative credited business of their subordinates, and (iii) that managers can affect the timing of sales through staffing and incentive decisions affecting subordinates. The model yields the hypotheses that managers will be more likely to forgo terminating poor performing subordinates (Hypothesis 1) and will provide downward quota adjustments (Hypothesis 2) when the managers are on the cusp of meeting a quota, compared to when they are not. The model captures the instituional account—that managers' decisions effectively "pull in" sales into the current fiscal year at the expense of future sales and contrary to the interests of the firm.

I test this hypothesis using a novel and uniquely well-suited data featuring salespeople working at firms that subscribe to an on-demand (over "the cloud") sales performance management service. The data include longitudinal detail on the hierarchical positions, incentive plans, performance, and pay of 7,492 sales managers and their 61,092 immediate subordinates in 244 firms. I parametrically estimate the formal model, distinguishing the turnover and quota adjustments of salespeople whose sales are critical for the manager to meet a quota (the quasi-experimental "treatment") with salespeople working under managers who would or would not meet a quota anyway (the controls). As such, the identification strategy uses sales of a subordinates' peers as an exogenous source of variation affecting whether a subordinate's sales will be crucial for the manager to meet a quota, and the sales of a subordinate as a "treatment bubble." This allows distortions in subordinates' staffing and incentives to be causally attributed to managers' personal interests, thereby addressing a key challenge for empirical agency research. I estimate 13-15% of quota adjustments and retentions among poor performers are explained by the managers' unique personal interest in meeting a quota.

[FIGURE 1]

To illustrate a puzzling implication of gaming by intermediary agents, Figure 1 shows that the cumulative sales of the manager's subordinates often just reach the manager's quota. Indeed, for

---

[3] For brevity, I use "sales" to refer to performance measures. In the data, performance measures also include presale, support, and renewal activities.

both rank-and-file salespeople and also their managers, there are nearly four times as many quotas surpassed within 5% as there are quotas missed within 5%. This figure excludes the 7% of instances that managers' quotas are the sum of subordinates' quotas, and so this pattern cannot be explained by cumulatively-aligned quotas alone.

The model offers an explanation for the delegation of authority in organizational hierarchies, with results providing evidence that intermediary agents extract rents on their private information on personnel quality and market conditions. This enables managers to pull in sales to reach 100% quota through decisions affecting subordinates. I discuss how sales organizations use a variety of rules and practices to prevent such gaming from becoming too costly.

I.   Sales Management, Weak Monitoring, and Gaming in Sales

A. Background

Like rank-and-file salespeople, sales managers receive variable pay that can vary widely depending on measured performance. Unlike rank-and-file salespeople, sales managers' performance is measured largely by the cumulative sales of their subordinates. In the data, mean annual variable pay is about one-half of base pay for both managers and non-managers.

Variable pay includes commissions and bonuses. Their rates depend on quota attainment. Quotas are specific thresholds at which workers typically receive a discrete bonus (in an archetypal quota-bonus plan) and/or begin to earn commissions on marginal sales (in an archetypal commission plan). Quotas are generally set in advance of a measurement period. However, organizations typically allow managers to adjust subordinates' quotas for reasons outside of the salespersons' control.

Salespeople who exceed their quotas typically become eligible progressive bonuses or accelerators that increase the rate at which commissions are paid. Salespeople who consistently exceed their quotas may also be considered for promotions, transfers, superior leads, or superior accounts. Making quotas and other discrete benchmarks also confer prestige, influence, and symbolic rewards (Larkin 2009). Salespeople who do not meet quotas typically earn a base pay, which reduces risk borne by salespeople and provides income to new recruits. Guaranteed income is often temporary and may be phased-out or drawn from future variable pay (in the case of "draws").

B. Why Organizations Use Quotas

Based primarily on interviews, I find three main reasons sales compensation plan designers use quotas. First, quotas focus variable pay around marginal effort. Some sales positions, such as account managers with an account renewal quota, can achieve some large share of their quota with relatively little effort. In the classical agency model, this is similar to the property that a firm can capture maximum rents by paying an agent's participation constraint for the first-best effort. The firm may have a better understanding of this level of effort than it does the marginal cost of this effort, which is needed under a linear incentive plan (*i.e.* a piecerate or pure commission plan).

Second, quotas communicate minimum acceptable performance. Consistently missing a quota is generally understood to be a ground for dismissal. Terminating a salesperson for performance

reasons when that salesperson consistently makes quota may be interpreted as symptomatic of poor communication by management. In most sales settings, sales people consume territories, sales leads, and support resources that are inherently valuable. As such, firms are generally not willing to retain a salesperson indefinitely at low pay for low performance. Interviewees explicitly referred to retaining a poor performing sales representative in terms of the opportunity cost of replacing them with a productive new recruit. In a classical agency framework, this is analogous to a firm communicating that it is incentive compatible for a firm to replace an agent for performance beneath a threshold.

Third, managers and plan designers widely believe that quotas have a behavioral effect on workers that tends to boost performance. Managers explain the importance of communicating how much sales people should expect to make if they make their quota. I interpret these reports to signify that managers use quotas to invoke loss aversion, thereby eliciting greater effort for less pay when below a quota threshold.[4]

C. How Incentives Affect the Timing of Sales Activities

Outside of executives and sales, variable pay is often a small or negligible component of compensation. One reason is that incentive contracts encourage workers to "game" plans by engaging in activities that are correlated with performance measures but contrary to the interests of the firm.[5] Examples abound of how misaligned incentives prompt undesired behaviors. At the top of organizations, executives adapt accounting procedures, accrual procedures, and voluntary disclosures to maximize bonus rewards (Aboody and Kasznik 2000, Healy 1985, Yermack 1997). At the bottom of organizations, several studies show how seemingly innocuous pay-for-performance schemes have backfired.[6]

Despite their ubiquity in sales, empirical work shows misaligned and nonlinear incentives distort sales activities as well. Using data from an enterprise software vendor, Larkin (2007) finds that accelerating quarterly commissions lead salespeople to use discretionary discounts to concentrate transactions into fewer quarters, costing the employer an estimated 6-8% of revenues. Using Compustat data, Oyer (1998) exploits variation in fiscal years by company and within industry to show that manufacturing firms' sales rise in the fourth quarter. He interprets this finding as consistent with the incentive effects of annual quotas, although the firm-level public disclosure

---

[4] Heath, Larrick, and Wu (1999) discuss how performance goals are analogous to reference points in Prospect Theory. Psychologists typically find explicit targets can improve measured performance when they are challenging, specific, attainable, and supported by coaching and other practices (for reviews, see Shinkle 2012, Steel and Konig 2006). Although this literature is not conclusive, sales plan designers widely report their belief that quotas have psychological importance.

[5] Interviewees often view "gaming" as pejorative, noting that it is the responsibility of the firm to design incentives and for salespeople to pursue them. I use this term as shorthand.

[6] Job Training Partnership Act training agencies manipulate the timing of students' graduation dates to boost the share of graduates with jobs (Courty and Marschke 1997, 2003). In lending, the desire to avoid appearing to have poorly assessed borrowers' risks led bank loan officers to fail to disclose bad news to their supervisors (Hertzberg, Liberti, and Paravisini 2007). After implementing a per-passenger commission, bus drivers in Chile had a higher incidence of traffic accidents than prior to the implementation or compared with a competing bus company that did not pay by commission (Johnson, Reiley, and Muñoz 2011). Baker (1992), Ethiraj and Levinthal (2009), Feltham and Xie (1994), Holmstrom and Milgrom (1991), and Kerr (1974) provide other examples.

64

data do not permit a direct test or analysis of gaming at any level of the organization. Incentives to boost the size of subscriptions led account managers at Dun & Bradstreet to overstate their clients' historical usage, spurring millions of dollars in lawsuits (Roberts 1989).

Salespeople describe several additional practices used to shift credit across measurement periods. Salespeople may boost sales figures by enticing distribution channels to place large orders to keep as inventory, a practice referred to as "channel stuffing." Salespeople may delay closing deals until future measurement periods, a practice referred to as "sandbagging." Salespeople and managers may exchange credit for sales across measurement periods. Salespeople and managers may misrepresent the quality of their territory to affect the sales forecasts used as the bases of their quotas. Sales managers can provide incentives, called "SPIFs," directly to salespeople at downstream firms who sell their products.

Employer "ratcheting," or the practice of moving quota thresholds based on past performance, also provides an incentive to manipulate the size and timing of sales. Organizational researchers conducting fieldwork have long recognized the pervasiveness of restricting output to avoid quota increases.[7] Murphy (2000) shows that firms that set managerial quotas according to internal standards (such as a budget or past performance) have less-variable bonuses and smoother earnings than those that use external standards that cannot be gamed. Leone, Misra, and Zimmerman (2006) find evidence of dynamic sales quota ratcheting in a Fortune 500 firm. They note that quotas rise with over-performance more readily than they fall with under-performance. Asymmetric ratcheting further compresses incentives around meeting quota by weakening the benefits of missing quotas (because quotas are unlikely to fall) and weakening incentives to exceed quotas (because quotas are likely to rise).

D. How and Why Firms Monitor Sales Activities

Sales hierarchies devote considerable resources to identifying and retaining salespeople who exceed quotas. In the data, turnover is 47% per year, and sales performance is highly skewed.[8] However, in many cases it is difficult for the firm to attribute sales numbers to the skill of a salesperson, rather than exogenous factors such as the quality of the product, territory, or market conditions. As such, managers play a large role in identifying and retaining high performers.

Early organizational research emphasizes how managerial behaviors depart from profit maximization. Coase (1937), Penrose (1959), and Williamson (1967) invoke diminishing returns to management and the alienation of managerial interests to explain the limited growth of firms. Simon (1957) argues managers "satisfice," adopting decisions that meet some non-maximizing acceptability threshold. Cyert and March (1963) argue managers possess neither the motives nor the cognitive means to make profit-maximizing decisions, and managers' private information allows them to pursue tangential objectives. Crozier (1964) argues that hierarchies use

---

[7] For example, this practice has been labeled "soldiering" by Frederick Winslow Taylor (1912), "targeting a bogey" by Elton Mayo (Roethlisberger and Dickson 1939), and "quota restriction" by Donald Roy (1953).

[8] The sales industry often cites the "80-20 rule," the rule-of-thumb that 80% of sales are made by 20% of the salesforce. In the data, this slightly exagerates the variation in sales performance at most firms. Prior to controlling for tenure, about 25% of salespeople are responsible for 75% of sales at the median firm. The variation is greatest for sales representatives at enterprise software companies and narrowest for engineers and support occupations.

impartiality as a pretense for the centralization and consolidation of organizational power. Chandler (1977) argues that managerial hierarchies are independent sources of power, permanence, and continued growth. Baumol (1959), Gordon (1961), and Williamson (1963) interpret profit as a constraint to which manager's other goals--such as job security, influence, prestige, and advancement--may be pursued.

Recent advances in agency theory incorporate the role of supervisors in reducing gaming (see Gibbons 2005 or Miller 2005 for a review). Monitoring allows firms to condition employment and payment on agents' inputs (such as effort) and discourages opportunism. To prevent opportunism and politicking among managers, firms may use bureaucratic rules and internal auditing. When performance measures are not contractible, firms may commit to subjective awards by relying on its reputation or by delegating subjective awards to an impartial supervisor. Subjective bonuses have other challenges, however; managers use evaluations to distribute performance rewards as they see fit, potentially eliciting cognitive biases, influence activities, and perceptions of unfairness. Conyon and He (2004), using evidence from CEO compensation committees, find three-tier agency models are better able to explain decision-making among executive compensation committees, compared to managerial power and collusion models.[9]

Social psychologists have examined why workers and managers pursue an organization's interests, even when such activities are unrewarded, and the circumstances under which individuals are willing to harm the organizations or other individuals to pursue personal goals. Bennett and Robinson (2000) and Berry et al. (2007) find that individuals who commit deviant behaviors harming individuals are also more likely to commit deviant behaviors harming the organization. Organizational researchers also offer explanations why the solutions offered by the standard agency theory may not work in practice. Neihoff and Moorman (1993) find that monitoring reduces organizational commitment behavior, and Gneezy and Rustichini (2000) find pecuniary penalties reduces guilt for breaking norms. Nickerson and Zenger (2008) hypothesize that perceptions of unfairness, envy, and dissatisfaction with relative pay restrict efficiency in settings such as sales, where individuals are highly motivated by pecuniary rewards. Studies by Benford and Snow (2000), Kaplan (2008), and Obloj and Sengul (2012) suggest managers will learn to frame their activities within new organizational initiatives, and learn to game their plans.

Salespeople are stereotypically driven by incentives rather than loyalty to their firms, as reflected by high turnover rates and high variation in compensation that typifies sales settings. Perhaps due to these reasons, information is rarely observed or communicated perfectly in sales hierarchies. For example, sales managers and their subordinates learn how difficult it is to sell a given product in a given territory, while other functions (potentially sales operations, the CFO, or marketing) use past performance, subjective reports, or other sources of information to produce

---

[9]   For theoretical treatment on contracting or monitoring effort, see Akerlof and Yellen 1986, Calvo and Wellisz 1978, Harris and Raviv 1978, Joseph and Thevaranjan 1998, Shapiro and Stiglitz 1984. On subjective bonuses, see Baker, Gibbons, and Murphy (1994) or Bull (1987). On bureaucratic rules and internal monitoring, see Baliga (1999), Kofman and Lawaree (1993), Milgrom (1988), Milgrom and Roberts (1998), Suzuki (2011), or Tirole (1986, 1992). On delegating awards to impartial supervisors, see Strauz (1997). On evaluation biases and gamesmanship, see Kaplan and Henderson (2005), Milkovich and Newman (2011: 374-5), or Prendergast and Topel (1992). Wu (2011) finds a Chinese newspaper reduced opportunism by centralizing authority, but did so at the expense of managerial incentives to support subordinates.

forecasts. One purpose of these forecasts are to set quotas, making subordinates' reports potentially unreliable.

While sales managers are responsible for monitoring gaming behavior, nonlinearities in their incentive plans encourage activities not perfectly aligned with profit maximization. Moreover, managers typically focus their effort on negotiating and closing deals—the point at which there is the greatest opportunity to affect the timing of measured performance.

As such, sales organizations also rely on a variety of reporting practices to reduce information asymmetries, maintain incentive alignment, and discourage gaming. Customer relationship management (CRM) tools allow salespeople to report progress on their sales pipeline and share information regarding how clients' purchasing decisions are made. These too may be gamed; interviewees report that subordinates may misrepresent the status of intermediary sales activities to avoid interventions by managers, whose desired closing date for sales may conflict with their own, or to avoid others from expropriating their client relationships and accounts. Moon and Mentzer (1999), in a study of a sales organization, found salespeople grossly misrepresented forecasts and the state of their sales pipeline, which they believed were used to adjust quotas.[10]

Some firms use subjective bonuses or promotions to reward managers perceived to be acting primarily on the interests of the firm (for evidence outside sales, see Cappelli and Conyon 2011). Some firms restrict managers' staffing, incentive, and pricing decisions, requiring large decisions to be approved or reviewed by sales operations or superiors. Indeed, the effort and expense firms dedicate to designing plans, monitoring activities, and improving coordination suggest that hierarchical coordination is indeed costly.

II.   Managerial Quotas' Effect on Staffing and Incentives

This study focuses on the manager's interest to achieve the discrete pay, recognition, and job security associated with quota attainment. While firms also desire that their managers meet and surpass their quotas, this study examines managerial behaviors that (i) affect the staffing and incentives of subordinates, (ii) are encouraged by annual quotas, and (iii) are consistent with the interests of the manager, and not the interests of the firm or subordinates. Specifically, this study examines the timing of termination decisions and subordinate quota adjustments.[11]

---

[10]   Outside of sales, Forbes, Lederman, and Tombe (2012) provide an example of how workers and organizations can game monitoring devices. They find evidence that airlines misreport the length of flight delays around a 15-minute threshold, so that their flights are officially counted as on-time. They find misreporting is most-pronounced among airlines providing incentives for delays within 15-minutes.

[11]   Perhaps the best-identified example of the effect of manager's incentive plans on their subordinates' employment and performance is Bandiera, Barankay, and Rasul's (2007) field experiment on supervisors of fruit-pickers. Introducing a piece rate improved efficiency by leading supervisors to be more selective in whom they recruit and leads them to focus effort on assisting the most productive workers. Because workers consist of migrants who lived on the farm and their work consists of picking fruit, recruitment costs are negligible, the difficulty of the task is relatively well-known and homogenous, workers require minimal training, and there are no quotas. These characteristics distinguish fruit-pickers from the complex sales settings that distort incentives among sales managers.

The model connects the relevant institutional details of sales to an agency framework, and serves three purposes. First, it explains why firms delegate authority over staffing and incentive decisions to managers, even though the performance measure (sales) is observable to firms. Second, it illustrates how sales quotas distort the timing of staffing and incentive decisions. Third, the model yields the strategy for identifying decisions motivated by the unique interests of the manager.

A. Managerial Quotas and Subordinates' Staffing

Developing a sales team is among a sales manager's chief responsibilities. This involves recruiting, training, assisting, and disciplining subordinates. Because managers' performance is measured primarily through the cumulative sales of their subordinates, their plans incentivize them to build productive teams. However, the incentives provided by quotas affect the timing of staffing decisions.

The intuition follows. Hiring and training new salespeople consumes time. In sales settings involving complex products or services, the typical applicant for a sales position is intensively screened, recruits receive training and/or a shadowing period, and newly-trained salespeople are given several months to develop skills and establish a "sales pipeline" beginning with initial leads and ending with a purchase order (and potentially installation, renewal, and support). Salespeople refer to this as the "ramp up" period. The sales industry often uses twice the length of the sales cycle as a heuristic for the ramp up time. For business-to-business sales, which constitute the majority of sales in these data, interviewees suggest a typical ramp up time would be six to twelve months. This estimate is consistent with results presented in the next section.

For this reason, replacing a poor-performing but experienced salesperson with a new recruit is an investment involving the substitution of present sales with greater expected future sales as the new recruit is hired and ramped up; in Oyer's (1998) framework, retaining a poor performer is a way for managers to "pull in" sales from future measurement periods. Therefore, annual quotas create incentives for managers to retain poor-performing subordinates who would otherwise be terminated.

A natural question arising from agency theory is: Why do firms delegate termination decisions to managers, rather than specifying termination criteria in contracts? Based on interviews, I propose that it is very difficult for firms to translate sales figures into a claim about the quality of a salesperson, particularly when sales people are covering different products, territories, or functions. As such, manager's chief responsibilities include: (i) selecting, mentoring, and screening a sales force, and (ii) developing an awareness of the qualities of sales territories and products. I interpret these responsibilities to signify that managers are responsible for screening for high ability salespeople, which cannot be distiniguished from sales figures alone due to luck. Here, I use the term "ability" to include the personal characteristics that contribute to a salespersons' sustained performance. I use the term "luck" to include exogenous factors beyond a salesperson's control, including the quality of the product, the quality of the territory, the quality of the leads, and so on. The model then captures the following insight: Firms delegate authority for termination decisions to managers to use their private knowledge of salespeoples' ability and

luck, allowing them to accelerate screening for new hires when exogenous factors affecting performance make it difficult for the firm to do so using sales figures alone.[12]

To analyze the decision to employ a supervisor and the effects of a sales quota on a supervisor's staffing decisions, first consider the following firm-worker model where the firm observes production (sales) but not ability or luck directly. For now, I abstract from the wage and effort decisions, and the only choice is the firm's decision to retain or replace a worker, implicitly at the worker's reservation rate. The employment relationship in the model may be thought of as a firm filling a unique job tied to a valuable asset; I abstract from the firm's cost of acquiring that asset and the external competition that would lead the firm to adopt a reservation level of profitability for that asset and otherwise replace the worker. For example, for a medical device manufacturer, the asset may be the exclusive ability to sell a proprietary product to hospitals in California, which is tied to a specific sales job; for a newspaper, the asset may be an existing set of advertisers, who are assigned to an account manager.[13]

Firms are risk-neutral, there are infinite periods, and firms discount future periods at δ. In period $t=1$, the firm hires a worker, and then production occurs with the firm observing output. In periods $t>1$, the firm chooses an action $A_r^f \in \{\text{retain, replace}\}$, and then production occurs with the firm observing output. The production of worker i in period t is

$$y_{it} = r_{it} + \alpha_i + \epsilon_{it} \tag{1}$$

where $r_{it} \in \{0,1\}$ denotes whether the worker is "ramped up," $\alpha_i \in \{0,1\}$ denotes the worker's period-invariant ability, and $\epsilon_{it} \in \{0, 1\}$ denotes the worker's period-specific luck. Let $\Pr(\alpha_i=1)=\Pr(\epsilon_{it}=1)=0.5$, and $r_{it}=0$ in the worker's first period of employment and $r_{it}=1$ thereafter if that worker is ever retained. A ramped up worker may be thought of as a worker with accumulated firm- and client-specific human capital and a mature sales pipeline. Crucially, suppose the firm observes y_{it} and r_{it}, but does not observe α_i or ϵ_{it}; the firm observes performance but not the worker's ability or luck directly. It can be shown that, for $\delta \in (0.5,1)$, the net present value (NPV) of a new recruit exceeds the NPV of a revealed low ability worker and the firm replaces a worker if and only if production is $y_{it} = 0$ for a new recruit or $y_{it} = 1$ for a ramped-up worker. [14] These are the two conditions in which the worker is revealed to be low ability.[15]

---

[12] Analytically, the value of the supervisor in this setup most-closely resembles Harris and Raviv's (1978) model in which a firm is willing to pay to contract on a risk-neutral worker's effort rather than output. The strategic manipulation of information to affect decision-making has long traditions in organizational theory and decision theory (see especially Barnard 1938; Crozier 1964; Cyert and March 1963), with agency theory giving increasing attention to incorporating bureaucratic rules and politicking behavior into (see Gibbons, Matouschek, and Roberts 2012; or Tirole 1986, 1992; for a review).

[13] Here, assets might be used interchangeably with "resources" in the resource-based view parlance; see Barney (1991), Rumelt (1984), or Wernerfelt (1984). That is, these assets are inherently valuable, firm-specific, and not easily imitated by competitors.

[14] This decision rule, in which a firm conditions terminations on both performance experience, is also consistent with the tendency for firms to raise quotas as a new recruit is ramped up.

For $\delta < 0.5$, the NPV of a ramped-up low ability worker is greater than the NPV of a new recruit of unknown ability such that the firm retains all incumbents regardless of ability. Therefore, for the rest of this model I impose the parameter restriction $\delta \in (0.5,1)$. In this case, the NPV of workers are:

$$v_N = 1 + \delta(0.5v_H + 0.5v_L) \tag{2}$$

$$v_H = 2.5(1 - \delta)^{-1} \tag{3}$$

$$v_L = 0.5v_N + 0.5(1.5 + \delta v_L) \tag{4}$$

where v_N denotes the NPV of a new recruit, v_H denotes the NPV of an experienced high ability worker, and v_L denotes the NPV of a low ability worker. For $\delta \in (0.5,1)$, $v_N > v_L$ and the firm replaces revealed low ability workers with new recruits. By substitution, the NPV of a new worker is $v_N = (1-0.25\delta-0.125\delta^2)(1-1.75\delta+0.75\delta^2)^{-1}$.

Now, suppose instead that the firm may choose to delegate the replacement decision to a supervisor. I assume the supervisor is risk-neutral and may be paid a reservation rate from any surplus the supervisor generates (thereby allowing us to abstract from supervisor turnover). The crucial assumption is that the supervisor observes (but cannot verify) productive inputs α_i and ϵ_{it}.[16]

Formally, if the firm does not to hire a supervisor, play proceeds as above. If the firm hires a supervisor, in period $t=1$, the supervisor hires a worker, production occurs, and then the supervisor observes α_i, ϵ_{it}, and r. In periods $t>1$, the supervisor chooses an action $A_r^f \in \{$retain, replace$\}$, production occurs, and then the supervisor observes α_i, ϵ_{it}, and r. Let v^s denote the NPV of a subordinate with a supervisor.

The supervisor's contract specifies a piecerate and bonus for production exceeding a quota, Q_i, normalized to zero. Define the supervisor's total quota attainment in the event the subordinate is retained as

$$Q_i^H \equiv Q_i^L + y_i \tag{5}$$

where $Q_i^L \in [-5,5]$ is a random exogenous variable denoting what the quota attainment would be without subordinate i's contribution, and y_{it} is the subordinate's contribution. Let the supervisor's contract take the form $w_{it} = cS_{it} + b_{it}$ where $c \geq 0$ is the rent-sharing (commission) rate, $S(\delta) \equiv v_N^S$

---

[15] Note that, by construction, the firm's posterior belief regarding the worker's ability is equal to the anterior belief for "medium" performance, $\Pr(\alpha_i=1|r_{it}=0, y_{it}=1) = \Pr(\alpha_i=1|r_{it}=1, y_{it}=2)=0.5$. This simplifies the exposition.

[16] In practice, firms' performance evaluation procedures and customer relationship management tools may be an example of how the firm may try to verify α_i and ϵ_{it}. Such evaluations are notoriously unreliable and subject to bias, with meaningful subjective assessments of subordinates' abilities eschewed by managers in the interest of avoiding conflict (Bretz, Milkovich, and Read 1989). Baliga (1999) shows that firms may choose to hire a supervisor even in the prescence of non-verifiable information and the potential for collusion.

− v_N s the surplus created when the firm hires the supervisor, and b_{it} is a bonus, where b_{it} = B ≥ 0 if $Q_i^L + y_{it} ≥ Q_i$ and zero otherwise.

The firm can make terminating low ability subordinates strictly incentive compatible for all values Q_i^L by choosing an arbitrarily-small c and by setting B=0, *i.e.*, by eliminating the incentives for meeting quota and paying a linear piecerate. To see this, note that when the manager replaces low ability workers, the firm's NPV of new, high ability, and low ability workers are:

By substitution, the value of the supervised new worker is $v_N^S = (1 - 0.25δ)(1 - 1.15δ + 0.5δ^2)^{-1}$, which is greater than the value of the unsupervised worker, v_N, yielding $S(δ)>0$ for $δ ∈ (0.5, 1)$. Intuitively, the rent-sharing rule makes it incentive compatible for the supervisor to use the private information of $α_i$ and $ε_{it}$ to accelerate the screening process for new workers.[17]

Although the model predicts firms may induce supervisors to terminate low ability subordinates by providing simple linear incentives, such plans are exceedingly rare in practice. Rather, incentive plans routinely feature quotas and other target-based goals. Unfortunately, there is no standard, empirically-established modeling technique yielding quotas and discrete incentives in incomplete contracts (for a theoretical discussion, see Frankel 2011, Kim 1997, Levin 2003, Oyer 2000, Park 1995, Steel and Konig 2006). Rather, consider the manager's decision whether to replace an experienced (r_{it} = 1), low ability ($α_i$) worker when a bonus for meeting a quota (B>0) is treated as exogenous.

For $B > 4cS(δ)$, contrary to the desires of the firm, managers retain low ability workers when Q_i^L ∈ [-2, 1]. That is, when the bonus is sufficiently high, rent-sharing is sufficiently small, and the future is sufficiently discounted, quotas lead managers on the cusp of meeting a quota to pull in sales by retaining experienced, low ability subordinates.[18] Intuitively, because the experienced worker enjoys r_{it}=1 with certainty whereas the new recruit enjoys $α_t$=1 with probability 0.5, the immediate production for an experienced, low ability worker first-order stochastically dominates production for a new recruit.

Figure 2 illustrates how quotas distort incentives to screen low ability workers.

[FIGURE 2]

---

[17]   The model features firms that learn the ability of the worker over time. This is because $ε_{it}$ is an independent and random draw in each period. For firms, the more troubling (but plausible) scenario is that $ε_{it}$ is serially-correlated within a position; since individuals are employed at positions over time, firms may misattribute sustained high performance to the individual rather than the position (e.g. a good territory). If managers enjoy private information regarding the value of the firm's assets, this could allow managers to extract sustained rents on this information by pursuing personal objectives (such as pulling in sales, as above, or also collusion with subordinates, nepotism, or so on).

[18]   Note that an implication of the model is that gaming around quotas could be reduced by boosting local linear incentives around the threshold and reducing the descrete benefits of achieving it.

The model's predictions are driven by managerial incentives to meet a quota. For identification purposes, this prediction is particularly powerful because the treatment concerns an interior range of quota attainment values. However, perhaps the larger distortion emerges from commission accelerators, which make pay convex in sales and provide incentives to concentrate sales into fewer measurement periods. Although the prediction that managers push out sales by replacing poor performers when they are far below quota may be unintuitive, interviews suggest this is common when the manager's job is secure and marginal incentives are weak; interviewed sales managers referred to the general practice of concentrating losses in a single measurement period (including making staffing changes) as "taking a bath."

Drawing from interviews and from the model they inform, I hypothesize that annual quotas lead managers to refrain from terminating poor-performing subordinates late in measurement periods if the poor performer's sales are needed for the manager to meet quota, and to terminate poor performers late in the measurement period if doing so is unlikely to affect whether the manager will meet quota.

Hypothesis 1:   Managers will delay terminating poor performing subordinates until future measurement periods if and only if those subordinates are essential to the manager's quota achievement (*i.e.*, subordinates are "quota-critical").

This hypothesis requires four specific hypothesis tests. For months in the fourth fiscal quarter, I hypothesize that turnover of quota-critical poor performers will be lower than (H1a) turnover among poor performers whose manager *would not have* met quota anyway, and (H1b) turnover among poor performers whose manager *would have* met quota anyway. Furthermore, I examine whether these foregone turnovers are delayed until the following fiscal year. Specifically, in the month following the annual measurement period (the "thirteenth" fiscal month), I hypothesize the turnover of quota-critical poor performers will be higher than (H1c) turnover among subordinates whose managers did not meet quota, and (H1d) turnover among managers who did meet quota, but would not have without the subordinate's credit.

A. Managerial Quotas and Subordinates' Incentives

The second hypothesis concerns the alignment of subordinates' incentives with managers' interests. In particular, I examine whether subordinates are more likely to receive quota adjustments in the fourth quarter when the manager is in the neighborhood of reaching a quota.

Quota adjustments allow firms to adapt incentives to unforeseen circumstances. Otherwise, salespeople who are far below their quota for reasons unrelated to their prior effort may suffer from weak marginal incentives late in measurement periods, prompting them to quit or hoard sales until the next measurement period. To discourage salespeople from hoarding nearly-closed deals (commonly referred to as "sandbagging"), managers are typically instructed to reduce quotas only for reasons outside a worker's control, to communicate how circumstances meet predetermined criteria for quota adjustments with subordinates, and be mindful of adjustments' reputational effects.

In context of the model, I interpret this as evidence that managers observe α_i and ϵ_{it} prior to production, are delegated authority to choose $A_q^S \in \{$keep quota, reduce quota$\}$, and that firms

want managers to choose to keep quota for $\{\epsilon_{it} = 1, \alpha_i = 1\}$, to reduce quota for $\{\epsilon_{it} = 0, \alpha_i = 1\}$, and to fire the worker (as before) for $\alpha_i = 0$. The quota adjustment is a decision that creates surplus if the quota is kept for the output corresponding to $\{\epsilon_{it} = 0, \alpha_i = 1\}$ and is reduced for the output corresponding to $\{\epsilon_{it} = 1, \alpha_i = 0\}$. Intuitively, quotas specify an output level, but the key assumption is that firms cannot contract quota adjustments on α_i or ϵ_{it}, just as they could not for terminations. However, the immediate productivity boosts only depend on y_{it}; a downward quota adjustment may boost marginal incentives when production is low either due to bad luck or lower prior production due to low ability.

Although firms may immediately boost marginal incentives by adjusting quotas for all subordinates far below their quotas late in measurement periods, in practice, downward adjustments are not routine. This is because downward adjustments are also implicitly costly, as they distort incentives and harm morale if adjustments are anticipated or not viewed as exogenous. Otherwise, salespeople may ratchet effort, sandbag sales, and misrepresent forecasts early in measurement cycles in anticipation of adjustments. Subordinates may also interpret downward adjustments as "selective intervention," and evidence of favoritism, entitlements, or managerial opportunism. These beliefs may impair the firm's or the manager's ability to commit to future quota adjustments only for exogenous and pre-defined circumstances (Foss 2003).[19]

I interpret this to signify that a profit-maximizing firm would like to condition a quota on ϵ_{it} and credibly commit to it in advance, thereby avoiding moral hazard and sandbagging if downward adjustments are expected even if $\epsilon_{it} = 1$. However, because firms observe only y_{it} and not ϵ_{it}, they delegate quota adjustment decisions to a supervisor. Because contracts specify y_{it} and the quota's effect on workers' effort effective for both $\{\epsilon_{it} = 0, \alpha = 1\}$ and $\{\epsilon_{it} = 1, \alpha = 0\}$, supervisors on the cusp of making a quota retain poor performers and provide a downward quota adjustment. As such, providing a downward quota adjustment for a $\{\epsilon_{it} = 0, \alpha = 1\}$ worker is also tantamount to pulling in sales, as it boosts immediate incentives but invites future moral hazard as subordinates sandbag effort in anticipation of quota reductions.

The second hypothesis examines whether managers on the cusp of meeting their quotas are more likely to reduce subordinates' quotas.

Hypothesis 2:   Subordinates are more likely to have their quotas adjusted when their manager's ultimate quota attainment will be near the quota threshold.

While a quota adjustment is a relatively interpretable and standardized outcome variable, quota adjustments are arguably not the main way managers shift subordinates' incentives. Managers may also shift incentives by changing subordinates' implicit commission rates. For example, discretionary bonus pools may be distributed through targeted commissions and bonuses, through incentives for channels, through rewards for selling a certain product or to a certain client ("bounties"), through tournaments, or through other means. For salespeople greatly

---

[19] Indeed, firms recognize that quotas can become out of reach for subordinates, but rather than adjusting quotas downward, may tolerate the sandbagging and hoarding to avoid the reputational costs of downward quota adjustments when the exogenous circumstances cannot be verified.

exceeding their quotas, a downward adjustment can boost marginal incentives by moving the salesperson to a higher commission rate tier.

The identification strategy requires distinguishing subordinates who managers believe necessary to meet their quota. As such, the model relies on calculating each parameter in $Q_i \equiv Q_i^L + y_i$, and examining the turnover and quota adjustments among poor performing subordinates for whom $Q_i^H > 100\%$ and $Q_i^L > 100\%$.

III. Data

A. Description of the Data

Data come from a firm that offers on-demand (over "the cloud") sales performance management (SPM) software. Data include how 22 million transactions are credited to 7,492 sales managers and their 61,092 immediate subordinates in 244 client firms. Client firms upload their compensation plans to a server, and then the clients' salespeople log in to the server (e.g. through a computer or smart phone) to report credited transactions and track their progress toward quotas or other benchmarks. The service is designed to make incentives and real-time performance transparent to salespeople and their managers, to calculate and automate compensation, to enable monitoring, to produce an audit-trail, and to promote flexibility in adapting compensation plans. Data begin in January 2008 and end in October 2011, although not all firms are represented throughout this period. No one firm represents more than 13% of workers or worker-months. Table 1 provides descriptive statistics.

[TABLE 1]

For each worker within a position, relevant data include a unique worker identifier (linkable if a person changes positions), a job title, a position title, the parent position, and the compensation plan. Position and parent position identifiers allow the construction of longitudinal organizational hierarchies, which also determine performance monitoring and other privileges within the SPM software. The most common job titles among workers with one level of subordinates are `territory manager, :sales director,' `regional director,' `regional manager,' `sales engineer manager,' and `regional vice president.' Each transaction includes a timestamp, the share of credit for each worker credited on the transaction, the amount and currency of the order, and the incentive plan to which the credit pertains. Each credited transaction further specifies associated quotas, commissions, and bonuses. Pay calculations from credited business are checked against a summary of terminal payments for each worker's payroll period. Likewise, annual quota attainment calculations made from quota thresholds and credited business are checked against the SPM software's transaction-level rolling measure for percent quota attainment, which was updated in real-time and made visible to workers and managers via the client software's virtual dashboard. Because calculated pay may be automatically linked to payroll, forecasts, and audit reports, it is unusually incentive compatible to enter plans and transactions accurately.

The data are unusually rich in that they allow a large number of workers' pay and performance to be tracked longitudinally and in a fashion that is reliable and standardized across firms. Because data come from an on-demand SPM software service, they largely avoid the selection dilemmas presented by data from single firms that opt-in as research sites. Likewise, this also helps address external validity concerns inherent to single-firm studies. However, the data also feature

74

limitations. Analytically, the chief limitation is that employment is subject to truncation; for example, some salespeople enter the data when the company subscribes to the service or are still employed when the data end. Because it is optional for system administrators to input incumbents' hire dates, tenure cannot always be determined for workers employed prior to the subscription to the SPM service. Although not essential to the analysis, the data offer two descriptive limitations. First, non-monetary rewards are not reported in the system; these may include prizes with great pecuniary or psychic value. Second, the data do not include education, demographics, or other descriptives sometimes available in empirical personnel research.

Compared to single-firm data, it is also difficult to identify and describe the institutional detail underlying data collected by on-demand SPM software. Single-firm studies typically complement quantitative results by interviewing workers whose experience is largely-similar to those in the data (Ichniowski and Shaw 2009 refer to this process as "insider econometrics").

B. Bringing the Data to the Model

The model yields the prediction that the manager will be less likely to terminate poor performing subordinates and more likely to provide downward quota adjustments when the subordinate's sales are needed for the manager to meet a quota. To do so, the data must distinguish subordinates within the "treatment bubble," whose sales are needed for the manager to make quota. In the model, subordinates inside the treatment bubble are those for whom $Q_i^H > 100\%$ and $Q_i^L < 100\%$; those below the treatment bubble are those for whom $Q_i^H < 100\%$ and $Q_i^L < 100\%$ (the manager would miss quota anyway), and those above the treatment bubble are those for whom $Q_i^H > 100\%$ and $Q_i^L > 100\%$ (the manager would make the quota anyway).

First, I estimate Equation 1, $y_{it} = r_{it} + \alpha_i + \epsilon_{it}$, the production given the productive inputs of being "ramped up," ability, and temporarily lucky. Second, I estimate terms in Equation 5, $Q_i^L + y_{it}$, the quota attainment absent the subordinate's contribution plus the subordinates contribution. By definition, this yields Q_i^H.

<u>Subordinate's Production y_{it}.</u> To estimate the subordinates's contribution $y_{it} = r_{it} + \alpha_i + \epsilon_{it}$, I estimate an OLS spline regression for the total monthly business credited as a function of tenure (the "ramp up," r_{it}), a worker fixed effect (the worker ability, α_{it}), and the noise term (the worker's period-specific luck, ϵ_{it}). I use a spline regression with quarterly knots because the functional relationship between sales and tenure as a compromise between the weak assumptions of month fixed effects and power of linear approximations. I do this for each standard job classification within a firm. The regression takes the form

$$\ln(y_{it}) = \underbrace{\beta_0 + \beta_1 M_{3it} + \beta_2 M_{6it} + \beta_3 M_{9it} + \beta_4 M_{12it}}_{\substack{\text{Month Spline Terms} \\ (\text{ramp up } r_{it})}} + \underbrace{\alpha_i}_{\substack{\text{worker f.e.} \\ (\text{ability } \alpha_i)}} + \underbrace{\varepsilon_{it}}_{\substack{\text{residual} \\ (\text{luck } \varepsilon_{it})}}$$

where i is the individual salesperson, t is the month, $\ln(y_{it})$ is the natural logarithm of the salesperson i's total credited business in month j,[20] M_3, M_6, M_9, and M_{12} respectively denote months into the first, second, third, and fourth quarters, are zero for months prior to their respective quarters, and are three for months following their respective quarters. M_{13} is zero for months prior to one year tenure, and otherwise denotes months greater than twelve. α_i is an individual fixed effect. I perform this regression separately for 834 standard job category-firm combinations that collectively feature 71,001 employment spells and 679,523 employee-months with credited transactions. Table 2 presents these regressions at the industry (rather than firm-job) level of aggregation. The worker fixed effect α_i estimates worker-invariant characteristics, while the residual ϵ_{it} represents the idiosyncratic noise affecting measured performance in a given month. The empirical distributions of both α_i and ϵ_{it} are approximately normal.

[TABLE 2]

First, consistent with the model's assumption regarding the ramp-up period, Table 2 shows that workers become more productive with tenure. Credited business generally rises most rapidly for new recruits and then decelerates, presumably as salespeople accumulate basic skills and develop a sales pipeline. Recall that this is a necessary condition to yield the model's predictions; intuitively, managers may expect an experienced, poor-performing incumbent with a mature pipeline to outperform a new recruit.

Second, Table 2 reports variation in workers' performance α_i controlling for tenure. The hypotheses concern workers who are revealed to be low ability by sustained low performance; for the remainder of the paper, I denote workers in the bottom quartile as "poor performers." Intuitively, these are the workers who have the bottom-quartile credited sales for workers of their job classification and tenure. Even if this method misclassifies the ability of some workers due to unobserved factors, it may offer a better-approximation of the supervisor's beliefs. Because poor performers have higher turnover, workers with bottom quartile α_i values represent 25% of workers but only 18% of worker-months.

Lastly, Table 2 reports variation in a workers' period-specific "luck" ϵ_{it}. Depending on the industry, the worker fixed effect explains 70%-88% of the variation in measured performance after controlling for tenure. Because this table aggregates to the industry level, it overstates the variation explained by worker effects when the regression model is run for job classifications within firms.

<u>Conditional managerial quota attainments, Q_i^H and Q_i^L.</u> Following Equation 5, define $Q_i^H \equiv Q_i^L + y_i$, where Q_i^L is the expected quota attainment if the poor performing subordinate turns over in the fourth fiscal quarter (months 10 - 12), y_i is the subordinate's contribution in the fourth quarter,[21] and their sum is the manager's expected quota attainment if the subordinate is retained in the fourth quarter.

To estimate Q_i^L, I recalculate the manager's quota attainment under the counterfactual scenario that subordinate i's sales in the fourth quarter are set to zero. This includes actual sales for the

---

[20]   Specifically, $\ln(y_{it})$ is the logged sum of the salesperson's split order credit-- the credit value of transactions multiplied by the share credited to the individual salesperson. The split order credit is the elemental unit of measured performance.

[21]   Note that y_i is the sum of the three y_{it} corresponding to months of the fourth quarter.

full fiscal year of all of the subordinate's peers under the same manager, plus the actual sales for fiscal months 1 - 9 for subordinate i.

To estimate Q_i^H, first I re-estimate the firm-job regressions in equation 8 with fiscal month fixed effects. Exponentiating the predicted values yields a prediction for sales credit in months of the fourth quarter that adjusts for fiscal month effects and subordinate i's prior performance. These values are then summed to estimate predicted sales in the fourth quarter. This prediction is strictly positive (following from exponentiating the logged dependent variable), and the difference between the logged predicted sales and the logged actual sales credit for those who stay is approximately normal with mean zero. Then $Q_i^H = Q_i^L + \hat{y}_{i10} + \hat{y}_{i11} + \hat{y}_{i12}$, where $\hat{y}_{im}$ is the exponentiated predicted values from the regression in month m. Note that I use the anterior predicted sales rather than actual sales even when actual sales are revealed in the data. I do this because the independent variable of interest is the manager's expectations, and because actual sales aren't observed for workers who turn over.[22]

Peer effects, the manager's private information, and other factors introduce measurement error into Q_i^L and Q_i^H. However, the power of the test is that it occurs at an interval, and so measurement error would lead to attenuation bias, reducing the likelihood of finding a significant result.

Defining Turnover. The data treat turnover events as any severance in employment for the purpose of recordkeeping within the SPM software. As such, turnover events include quits, layoffs, and fires. Turnover events also include internal job transfers to a position not covered by the SPM software, but do not include transfers to a position covered by or added to the SPM software. For example, a salesperson who transfers within sales or who is promoted to a sales manager is likely to remain in the data (and therefore not be counted as a termination). However, a salesperson who transfers out of sales may drop out of the data and be counted as a turnover event. Data include 38,159 turnover events and 15,695 internal transfers. Turnover is highly periodic, with peaks at the end of measurement cycles.

The reasons for turnover (terminations or quits) are immaterial in this setting, since it distinguishes the likelihood of turnover inside the "treatment bubble" where the manager has a personal interest in retaining the worker. In doing so, the identification strategy also solves a problem for research on turnover, since the distinction between terminations and quits is often blurred; workers who would be fired are often counselled to quit, and workers who quit may have instead been persuaded by enticements to stay.

IV. Results

A. Results for Managerial Quotas and Subordinates' Staffing

To test Hypothesis 1, I test whether the turnover of "quota-critical" poor-performing subordinates in the fourth quarter (*i.e.* $Q_i^L < 100\% < Q_i^H$) is lower than the turnover for poor performers for managers whom the model predicts would or would not meet a quota anyway (*i.e.* $100\% < Q_i^L < Q_i^H$ or $Q_i^L < Q_i^H < 100\%$). To test the specific hypothesis that turnover is delayed

---

[22]   As a check, I compare predicted sales against actual sales among retained subordinates, and confirm that actual log-sales are approximately normally distributed with mean zero about predicted log-sales.

for a month, rather than foregone, I also test whether the turnover of quota-critical poor performers in Month 13 is higher than non-quota critical poor performers.

[TABLE 3]

Consistent with Hypothesis 1, Table 3 shows that turnover is 5.6% when the manager reaches the quota threshold if and only if the subordinate's credited transactions are counted, substantially lower than the 18.6% turnover rate when the manager does not meet quota (H1a) and the 22.2% turnover rate when the manager would meet quota anyway (H1b). Both differences are significant with $p < 0.01$. These results suggest managers forego terminating poor performers late in the measurement period when these subordinates are needed for the manager to meet quota. Table 3 also shows that "thirteenth month" turnover among poor-performing subordinates is substantially higher when the manager met quota but would not have without the subordinates' fourth quarter credited transactions, compared to managers who did not meet quota (H1c) or would have met quota anyway (H1d). Taken together, results suggest managers delay terminating poor performers until the following fiscal year when doing so would reduce their likelihood of making an annual quota.

Although the result does not hold for non-poor performers, this result should be treated with caution. Most non-poor performers are making their quotas, suggesting that their performance is sufficient to avoid termination. The highest performers are likely to delay quitting until the end of the year to collect their bonus pay.[23]

To provide an estimate for how much turnover among poor performing salespeople is foregone because the manager is near a quota threshold, I restrict the sample to the poor performing salespeople whose sales are not necessary for their managers to meet quota and perform a logistic regression predicting the likelihood a subordinate will turn over as a linear function of the manager's and subordinate's quota attainment. I compare this predicted likelihood to the actual turnover rates among salespeople whose sales were necessary for their managers to meet quota, and estimate that total actual fourth quarter turnover among poor performers is 15% (s.e.: 1%) less than it would be if turnover among these "quota-critical" salespeople followed the same linear trend.

[FIGURE 3]

To examine turnover rates in greater detail, Figure 3 presents fourth quarter turnover rates within 5% bins of the expected quota attainments when the subordinate turns over, Q_i^L, rather than retained, Q_i^H. Fourth quarter turnover of poor performers drops sharply when the model predicts the manager will barely make quota. The low turnover among poor-performing subordinates whose managers who (barely) didn't need the subordinate to meet quota suggests managers may be conservative in terminations.

---

[23] The data are generally consistent with this account. Only 31% of quotas among poor performers are met, and the median variable pay of poor performers (including team incentives) is only 4% of fourth quarter salary. In contrast, 60% of quotas among non-poor performers are met, 80% of non-poor performers receive variable pay in the fourth quarter, and the median variable pay among non-poor performers is 60% of base salary. For this reason, the primary incentive for improving performance among non-poor performers may be increasing variable pay, while the primary incentive among poor performers may be avoiding dismissal.

Next, I perform a falsification test against effects and common shocks occurring at the level of the industry, the level of the firm, and the level of the manager's peers. To do so, I compare turnover rates of subordinates' against the quota attainment of a manager who shares the same upper-manager as their own direct manager. Intuitively, if managers are thought of as the "parent" position in the organizational hierarchy, subordinates may be compared against the quota attainment of the manager filling their "uncle" position.

[FIGURE 4]

Figure 4 presents two series of data. The hollow markers show the turnover rates at the uncle's quota attainment levels for the universe of their "niece/nephew" subordinates. The solid markers restrict the sample to subordinates outside the treatment bubble of the parent position. This is important, since the parent's and uncle's quota attainments are not independent. The hollow series shows that the turnover among poor performers in the neighborhood of the uncle's quota exhibits a small dip, which is not as pronounced as it is around the parent's quota. Turnover within 10% of the uncle's quota is 2.4% (with a standard error of 0.3%) lower than it is within 50% of the uncle's quota. The restriction introduced in the solid markers eliminates statistically significant effects in the neighborhood of the uncle's quota threshold, suggesting the dip among the hollow markers is caused by correlation in the uncle and parent's quota attainment.

Results show turnover among poor performers is lower late in annual measurement periods when poor performers' credited transactions would affect whether the manager would meet quota, and greater in the following month. More broadly, results suggest managers delay terminating poor performers until the following fiscal year when doing so is likely to affect their perceived likelihood of making quota. By showing that the parent position's quota attainment predicts turnover net of the "uncle" position's quota attainment, I rule out common shocks at the level of industry, firm, or division within the firm (at the next level of the organizational hierarchy).

B. Results for Managerial Quotas and Subordinates' Incentives

To test the hypothesis that managers' quotas affect their subordinates' incentives, I test whether subordinates are more likely to have their quotas adjusted just prior to the end of the fiscal year when the manager's sales are projected be near quota.

I run six logistic regressions to evaluate the likelihood of quota adjustments over months of the fourth quarter, as a function of the manager's and subordinate's expected quota attainment. The manager's expected quota attainment is calculated as Q_i^H above; it is the actual quota attainment through the third quarter plus the sum of the predicted attainment of the subordinates. The subordinate's expected quota attainment is sales as a percent of quota, including actual quota attainment through the third quarter plus the expected fourth quarter quota attainment. These regressions take the form

$$\ln \left(\frac{\hat{p}}{1 - \hat{p}} \right) = \hat{\beta}_0 + \mathbf{X}'_m \hat{\beta}^m + \mathbf{X}'_s \hat{\beta}^s$$

79

where $\hat{p}$ is the logit-estimated probability for the outcome of interest, X_m is a dummy vector indicating the 5% quota attainment bin of the manager and X_s is a dummy vector indicating the 5% quota attainment bin of the subordinate. This regression is performed for the six outcomes of interest corresponding to {Upward Adjustment, Downward Adjustment} × {Month 10, Month 11, Month 12}.

[TABLE 4]

[FIGURE 5]

Table 4 shows the logistic regression results for downward adjustments in Months 10, 11, and 12. Figure 5 illustrates the regression's estimated probabilities at the manager's and subordinate's expected quota attainments, holding the other's constant at their mean, for both upward and downward adjustments for each month.

Before discussing the main results, some trends deserve note. First, downward adjustments are more common than upward adjustments. Upward adjustments, while potentially bad for morale, may adjust for an exogenous circumstance that boosts performance measures (*e.g.* a product launch) or may be implemented as a penalty (too many clients canceled sales). Second, the probability that quotas are adjusted downward in months ten and eleven declines slightly as the manager's expected quota attainment increases. This slight downward trend may be because when a manager has a high quota attainment, it suggests the subordinate's peers are doing well, and that the subordinate is responsible for poor sales and should not be awarded a downward quota adjustment. Third, the probability that quotas are adjusted downward in months ten and eleven are more-sensitive to the quota attainment of the subordinate. One explanation is that managers interpret greatly-surpassed quotas as a signal the quota is already too low (consistent with ratcheting models). Fourth, quota adjustments in months ten and eleventh are more likely to place subordinates in the neighborhood of ultimately reaching their quota, rather than far above or far below. Fifth, quota adjustments are both more-rare and less-sensitive to the manager's and subordinate's quota attainment in the twelfth fiscal month than they are in the tenth and eleventh fiscal months. Lastly, in addition to these general trends, subordinates throughout the distribution receive both downward and upward quota adjustments, including downward adjustments for salespeople expected to surpass their quotas and upward adjustments for salespeople expected to miss their quotas. This may be because adjusting one subordinate's quota may also require adjusting comparable peers' quotas as well, in the interest of fairness.

Table 4 and Figure 5 lend support for the main prediction. The regression estimates that the likelihood a subordinate receives a quota adjustment is significantly lower (p < 0.05) in each of the six 5% intervals of the manager's quotas between 75-90% and 110-125%, compared to the interval 100-105%, for both fiscal months 10 and 11. The jump in the probability of an adjustment in the neighborhood of quota attainment does not appear in the twelfth fiscal month. I do not find strong evidence that upward adjustments are more likely in the neighborhood of either a subordinate's or manager's quota threshold.

To provide an intuition of the magnitude, I restrict the sample to workers whose managers are not within 10% of making quota, re-estimate the likelihood of receiving a downward quota adjustment as a linear function of the manager's ultimate quota attainment, and compare the estimated likelihood of receiving a quota adjustment against the actual likelihood of receiving a quota adjustment among workers whose managers were within 10% of the quota threshold.

These estimates suggest 13% (s.e.: 1%) of downward quota adjustments in months 10 and 11 are explained by the jump in quota adjustments among managers within 10% of making quota.

The likelihood of quota adjustments in the tenth and eleventh months of the fiscal year rise in the neighborhood of making quota, but not more pronounced for those who barely-met a quota compared to those who barely-miss it. As such, subordinates are more-likely to receive downward adjustments when their managers will not meet a quota anyway. One interpretation is that quota adjustment is a blunt instrument for boosting sales. Channel stuffing, discounting, credit-trading, sales hoarding, and reallocating effort (both in quantity and toward deals near completion) may offer more-immediate and precise ways of making a quota.

V.   Solutions to Managerial Gaming?

Results suggest sales functions' turnover and quota adjustments are different than what they would be in the absence of managers' discrete interest in making quotas. Presumably, there are profit-maximizing criteria for terminating poor performers and adjusting quotas, and these are distinct from manager's interest in the marginal sale associated with his or her quota attainment. Moreover, this study tests for a narrow, relatively well-identified gaming behavior. Managers' decisions routinely shift the timing of sales, including the timing and implementation of product pricing, product promotions, and sales campaigns. Managers also govern the allocation of resources across presale, final negotiation, and support activities. Interviews suggest that pricing, channel stuffing, and gaming discretionary bonuses may be more costly to the firm. As such, the gaming behavior described in this paper should be construed as a part of broader behaviors elicited when incentives are strong and information is poor.

Agency and organizational theory provide some guidance as to how managerial gaming may be mitigated. Any potential solution to gaming around quota thresholds should be weighted against the three chief benefits described in the review section: incentivizing marginal effort, communicating minimum acceptable performance for retention, and invoking loss aversion.

With those in mind, the model calls attention to specific levers for reducing managerial gaming. First, there are levers for smoothing incentives around quota thresholds:

- Locally-linear incentives. Firms may use locally-linear incentives around quota thresholds and eliminate bonuses for 100% quota attainment (setting $B = 0$). This implicitly deemphasizes the quota itself. However, further research is required to ascertain whether these practices also diminish quotas' benefits.

- Quota obfuscation. In practice, distance to quota (Q_i^L) is often obfuscated by complex crediting procedures, cancelled sales, or credit clawbacks. In principle, this uncertainty would smooth expected pay around expected sales; Ederer, Holden, and Meyer (2012) explore a theoretical model in which obfuscating performance measures reduces gaming. However, due to morale and incentive concerns, plan designers generally eschew complex and nontransparent plans; Englmaier, Roider, and Sunde (2012) provide evidence that obfuscating piecerates dilutes incentive effects.

Second, there are levers around the firm's discovery of salespeoples' luck:

- Bureaucratic rules and monitoring. If exogenous circumstances affecting a sales performance x_t could be codified and verified to the firm *ex post*, then firms can deter managerial gaming through bureaucratic decision rules and monitoring. In the model, this is not possible because the firm does not observe x_t. In practice, organizations do try to deter gaming through formal procuedures, including performance appraisals, bureaucratic rules, complaints, audits, and monitoring (*e.g.* through CRM tools). Decisions that could be motivated by opportunism, such as large discounts, may require approval or review by a senior manager. Firms may also require managers to justify termination and quota adjustment decisions; these criteria may be unverifiable (almost got a big client, but was unlucky) or verifiable (poor sales were due to unfavorable exchange rates). However, such rules can reduce flexibility and raise bureaucratic costs. Neihoff and Moorman (1993) and Gneezy and Rustichini (2000) provide evidence that monitoring and pecuniary penalties reduce organizational citizenship behaviors.

- Improving forecasts. If exogenous circumstances affecting sales performance x_t could be contracted upon in advance, then the firm would not need to rely on a manager's discretion. Indeed, forecasts are used to set sales quotas for this purpose. However, accurately forecasting sales is notoriously difficult, particularly when selling new products. Some organizations condition forecasts or quotas on easily-observed future circumstances, such as exchange rates in commodities. Forecasting using past performance can also distort incentives through the ratchet effect, and salespeople can game the customer relationship management (CRM) tools used to generate forecasts (Moon and Mentzer 1999).

- *Relative performance measures*. If salespeople are performing the same tasks and are subject to common shocks, firms may use plans that reward "relative" performance. Sales tournaments are a common application. However, these plans are also challenging to implement: exogenous shocks may be specific to an individual salesperson, relative performance measures discourage teamwork and diffusion of best selling practices, and anticipated tournaments encourage sandbagging.

- *Equalizing opportunities*. If there is little accumulation of position-specific human capital, firms can reorganize work to create comparable jobs. For example, rather than assigning a salesperson all leads from a narrow product or territory, an organization could distribute shares of leads from a wide range of products or territories, or rotate jobs, making sales performance comparable across salespeople.

Third, there are levers around aligning managerial incentives directly:

- *Conditioning managerial quotas*. Firms may adjust managerial quotas downward when they make new hires (reducing Q_i by r when the manager employs a new hire). While this would eliminate a manager's incentive to retain a poor-performing subordinate, it would also reduce incentives for retaining capable salespeople.

- *Promoting altruism*. Firms may attempt to make managers interested directly in the well-being of the firm, aligning managers' interests with those of the firm (making the manager's utility that of the firm, like that in Akerlof and Kranton's (1986) model of

identity). In industries where high variable pay is common, salespeople are notoriously driven by commissions rather than by company loyalty. Behavioral research may offer insight for promoting organizational citizenship behavior and mitigating gaming (see, for example, Podsakoff et al. 2009 for a review).[24]

Indeed, each theoretical solution has real world analogues. Together, they provide insight as to the conditions in which specific levers are effective.

The model also provides guidance as to how the characteristics of the selling environment determine the structure of sales hierarchies. The model predicts that firms would not hire a supervisor if it could observe workers' person-specific luck, x_i. Indeed, in settings where the solutions above are easy to implement (such as inside sales representatives within call centers, where performance is easy to measure and may be compared against workers subject to the same exogenous factors affecting performance), there are a higher ratio of first-line salespeople to first-line managers, and these managers are given less discretion to retain a poor performer. As such, cheaper and more-effective monitoring technology may not only improve incentives, but it may also flatten bureaucracies.

VI. Conclusion

This study examines the imperfect ability of organizational hierarchies to motivate intermediary managers to act on the organization's behalf. It does so by exploiting the institutional features that (i) that sales managers' have a unique interest in the marginal sales that meet their quotas, (ii) their quota attainments are determined by the cumulative credited business of their subordinates, and (iii) that managers can affect the timing of sales through staffing and incentive decisions affecting subordinates. Because an incumbent poor performer with accumulated job-specific skills and an established sales pipeline is likely to be more productive in the short term than a new recruit, sales managers may pull in sales by retaining poor performers until the following fiscal year. Because incentives for salespeople far below quota are weak, adjustments that make quotas "within reach" allow managers to boost immediate sales, while risking future moral hazard and disillusionment as adjustments become expected and viewed as entitlements. Following the model, I propose firms delegate authority for these decisions to managers due to their private knowledge of salespeoples' abilities and the sales environment, which is difficult for the firm to observe and verify. Although piecerates encourage mangers to use this information to improve the screening and motivation of the salesforce, discrete quotas and sales goals provide managers an incentive to game the staffing and incentives of subordinates. This study makes three main contributions.

First, this study makes a methodological contribution to empirical organizational research. I connect an agency model to predictions within a "treatment bubble," yielding a novel strategy for identifying the interests of individuals within organizations. I execute this strategy within a novel class of data, generated by on-demand sales performance management software. This service creates standardized, reliable, longitudinal, cross-firm data, enabling the study of compensation plans and performance for a large number of workers while mitigating the habitual selection issues and external validity concerns of studies in which participating firms "opt-in."

---

[24] The SAS Institute and Apple retail stores are known for high performance and low turnover among frontline salespeople without commissions.

Second, this study contributes a theory for the delegation of authority in organizational hierarchies, inspired by the institutional details of managing sales. Using a firm-supervisor-worker model, I hypothesize that firms use managerial incentives to make terminating low ability workers incentive compatible, since a firm cannot fully contract on a worker's ability or the difficulty of the selling environment. I show how firms attempt to use the many theoretical solutions to managing gaming, including the practices of monitoring, reporting, incentives, bureaucratic rules, job rotations, and attempts to get workers to identify with the firm. Moreover, the model helps explain why organizations in which relative performance appraisal is easy (such as call centers), hierarchies tend to be flatter and immediate managers have less discretion. I show how organizational and agency theory can inform under what conditions these practices may be effective.

Third, by highlighting the case of sales managers, this study contributes well-identified evidence for a specific failure of agency intermediation. While both classic and recent theoretical work emphasizes misalignment of managerial incentives in determining the structure of the firm, there has been scant evidence misaligned incentives propagate within organizational hierarchies. As such, this study corroborates the classic hypothesis that the interests of managers, who are the intermediary agents responsible for perfoming the activities of the firm, are inconsistent with profit maximization.

Works Cited

Aboody, David, and Ron Kasznik. 2000. "CEO stock option awards and the timing of corporate voluntary disclosures." *Journal of Accounting and Economics*, 29: 73-100.

Akerlof, George, and Janet Yellen. 1986. *Efficiency Wage Models of the Labor Market*. Orlando, Florida: Academic Press.

Arrow, Kenneth. 1998. "What Does Economics Have to Say About Racial Discrimination." *Journal of Economic Perspectives*, 12(2): 91-100.

Baker, George. 1992. "Incentive Contracts and Performance Measurement." *Journal of Political Economy*, 100(3): 598-614.

Baker, George, Robert Gibbons, and Kevin Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, 109: 1125-56.

Baliga, Sandeep. 1999. "Monitoring and Collusion with 'Soft' Information." *Journal of Law, Economics, and Organization*. 15(2): 434-40.

Bandiera, Oriana, Iwan Barankay, and Imran Rasu. 2007. "Incentives for Managers and Inequality Among Workers: Evidence from a Firm-Level Field Experiment." *Quarterly Journal of Economics*, 122(2): 729-73.

Baumol, William. 1957. "Speculation, Profitability, and Stability." *Review of Economics and Statistics*, 39: 263-71.

Barney, J.B. "Firm Resources and Sustained Competitive Advantage." *Journal of Management*, 17(1): 99-120.

Basu, Amiya, Rajiv Lal, V. Srinivasan, and Richard Staelin. 1985. "Salesforce Compensation

Plans: An Agency Theoretic Perspective." *Marketing Science*, 4: 267-91.

Becker, Gary. 1957. *The Economics of Discrimination*. Chicago, University of Chicago Press.

Bennett, Rebecca, and Sandra Robinson. 2000. "Development of a Measure of Workplace Deviance." *Journal of Applied Psychology*, 85(3): 349-60.

Benford, R. and D. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology*, 26: 611-39.

Berry, Christopher, Deniz ONes, and Paul Sackett. 2007. "Interpersonal Deviance, Organizational Deviance, and Their Common Correlates: A Review and Meta-Analysis." *Journal of Applied Psychology*, 92(2): 410-24.

Bull, Clive. 1987. "The Existence of Self-Enforcing Implicit Contracts." *Quarterly Journal of Economics*, 102(1): 147-59.

Benford, R. and D. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology*, 26: 611-39.

Bretz, Robert, George Milkovich, and Walter Read. 1989. "Comparing the Performance Appraisal Practices in Large Firms with the Directions in Research Literature: Learning More and More About Less and Less." CAHRS Working Paper.

Calvo, G., and S. Wellisz. 1978. "Supervision, Loss of Control and the Optimal Size of the Firm." *Journal of Political Economy*, 86: 943-52.

Cappelli, Peter, and Martin Conyon. 2011. "What Do Performance Appraisals Do? " Working Paper.

Chandler, Alfred. 1977. *The Visible Hand*. Cambridge, MA: Harvard University Press.

Coase, Ronald. 1937. "The Nature of the Firm." *Economica*, 4(16): 386-407.

Conyon, Martin, and Lerong He. 2004. "Compensation Committees and CEO Compensation Incentives in US Entrepreneurial Firms." *Journal of Management Accounting Research*. 16(1): 35-56.

Courty, Pascal, and Gerald Marschke. 1997 "Measuring Government Performance: Lessons from a Federal Job-Training Program." *American Economic Review*, 87(2): 383-88.

____. 2004. "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics*, 22(1): 23-56.

Crozier, Michael. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.

Cyert, Richard, and James March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ.

Depken II, Craig. 2000. "Wage Disparity and Team Productivity: Evidence from Major League Baseball." *Economics Letters*, 67(1): 87-92.

Eckles, David, and Martin Halek. 2007. "The Problem of Asymmetric Information: A Simulation of How Insurance Markets Can Be Inefficient." *Risk Management and Insurance Review*, 10: 93-105.

Englemaier, Florian, Andreas Roider, and Uwe Sunde. 2012. "The Role of Salience in Performance Schemes: Evidence from a Field Experiment." IZA Discussion Paper No. 6448.

Ethiraj, Sendil, and Daniel Levinthal. 2009. "Hoping for A to Z While Rewarding Only A: Complex Organizations and Multiple Goals." *Organization Science*, 20: 4-21.

Feltham, G.A. and J. Xie. 1994. "Performance Measure Congruity and Diversity in Multi-Task Principal/ Agent Relations." *The Accounting Review*, 69: 429-53.

Frankel, Alexander. 1991. "Discounted Quotas." Working Paper.

Forbes, Silke, Mara Lederman, and Trevor Tombe. 2012. "Quality Disclosure Programs with Thresholds: Misreporting, Gaming, and Employee Incentives." Working paper.

Foss, Nicolai. 2003. "Selective Intervention and Internal Hybrids: Interpreting and Learning from the Rise and Decline of the Oticon Spaghetti Organization." *Organization Science*, 14: 331-49.

Gibbons, Robert. 2005. "Incentives Between Firms (and Within)." *Management Science* 51: 2-17.

Gneezy, Uri, and Aldo Rustichini. 2000. "A Fine is a Price." *Journal of Legal Studies* 20: 1-18.

Gordon, Robert.A. 1961. *Business Leadership in the Large Corporation.* Berkeley.

Healy, Paul M. 1985. "The Effect of Bonus Schemes on Accounting Decisions." *Journal of Accounting and Economics*, 7: 85-107.

Heath, Chip, Richard Larrick, and George Wu. 1999. "Goals as Reference Points." *Cognitive Psychology*, 38: 79-109.

Hertzberg, Andrew, Jose Maria Liberti, and Daniel Paravisini. 2010. "Information and Incentives Inside the Firm: Evidence from Loan Officer Rotation." *The Journal of Finance*, 65(3): 795-828.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics*, 10(1): 74-91.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multi-Task Principal-Agent Analyses: Linear Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization*, 7: 24-52.

Holtausen, Robert, David Larker, and Richard Sloan. 1995. "Annual Bonus Schemes and the Manipulation of Earnings." *Journal of Accounting and Economics*, 19(1): 29-74.

Johnson, Ryan, David Reiley, and Juan Carlos Muñoz. 2011. "'The War for the Fare': How Bus Driver Compensation Affects Bus System Performance." NBER Working Paper.

Kerr, Steven. 1975. "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Journal*, 18(4): 769-83.

Kaplan, Sarah, and Rebecca Henderson. 2005. "Inertia and Incentives: Bridging Economics and Organizational Theory." *Organization Science*, 16(5): 509-21.

Kaplan, Sarah. 2008. "Framing Contests: Strategy Making Under Uncertainty." *Organization Science* 19(5): 729-52.

Kim, Son Ku. 1997. "Limited Liability and Bonus Contracts." *Journal of Economics and Management Strategy*, 6: 899-913.

Kofman, Fred, and Jacques Lawarree. 1993. "Collusion in Hierarchical Agency." *Econometrica*, 61: 629-56.

Larkin, Ian. 2006. "The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software." Working Paper.

____. 2009. "Paying $30,000 for a Gold Star: An Empirical Investigation into the Value of Peer Recognition to Software Salespeople." Working Paper.

Larkin, Ian, and Stephen Leider. 2012. "Incentive Schemes, Sorting, and Behavioral Biases of Employees: Experimental Evidence." *American Economic Journal: Microeconomics*, 4(2): 184-214.

Leone, Andrew, Sanjog Misra, and Jerold Zimmerman. 2004. "Investigating Dynamic Sales Quotas." Working Paper.

Levin, Jonathan. 2003. "Relational Incentive Contracts." *American Economic Review*, 93(3): 835-57.

Miller, Gary. 2005. "Solutions to Principal-Agent Problems in Firms." In C. Menard and M. Shirley eds. *Handbook of New Institutional Economics*, 349-70. Springer.

Moon, Mark, and John Mentzer. 1999. "Improving Salesforce Forecasting." *The Journal of Business Forecasting*. 18(2): 7-12.

Murphy, Kevin. 1999. "Executive Compensation." in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3, North Holland.

Murphy, Kevin. 2000. "Performance Stnadards in Incentive Contracts." *Journal of Accounting and Economics*. 30(3): 245-78.

Neihoff, Brian, and Robert Moorman. 1993. "Justice as a Mediator of the Relationship Between Methods of Monitoring and Organizational Citizenship Behavior." *Academy of Management Journal*, 36(3): 527-56.

Nickerson, Jack, and Todd Zenger. 2008. "Envy, Comparison Costs, and the Economic Theory of the Firm." *Strategic Management Journal*, 29(13): 1371-94.

Obloj, Tomasz, and Metin Sengul. 2012. "Incentive Life-Cycles: Learning and the Division of Value in Firms." *Administrative Science Quarterly*, 57: 305-47.

Oyer, Paul. 1998. "Fiscal Year Ends and Non-Linear Incentive Contracts: The Effect on Business Seasonality." *Quarterly Journal of Economics*, 113: 149-85.

____. 2000. "A Theory of Sales Quotas with Limited Liability and Rent Sharing." *Journal of Labor Economics*, 18(3): 405-26.

Park, Eun-Soo. 1995. "Incentive Contracting under Limited Liability." *Journal of Economics and Management Strategy*, 4: 477-90.

Penrose, Edith. 1959. *The Theory of the Growth of the Firm*. New York: John Wiley and Sons.

Podsakoff, Nathan, et al. 2009. "Individual and organizational-level consequences of organizational citizenship behaviors: A meta-analysis." *Journal of Applied Psychology*, 94(1): 122-41.

Pope, Devin, and Maurice Schweitzer. 2011. "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes." *American Economic Review*, 129-57.

Post, Richard J., and Kenneth E. Goodpaster. 1984. "H.J. Heinz Company: The Administration of Policy." In Kenneth E. Goodpaster, ed., *Ethics in Management*. Boston: Harvard Business

School.

Prendergast, Canice and R. Topel. 1992. "The Costs of Favoritism." Mimeo.

Roberts, Johnnie. 1989. "Credit Squeeze: Dun & Bradstreet Faces Flap Over How It Sells Reports on Business." *Wall Street Journal*, March 2.

Roethlisberger, Fritz, and William Dickson. 1939. *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago.* Cambridge, MA: Harvard University Press.

Roy, Donald. 1952. "Quota Restriction and Goldbricking in a Machine Shop." *American Journal of Sociology*, 67(2): 427-42.

Rumelt, Richard. 1984. "Towards a Strategic Theory of the Firm." In R.B. Lamb ed. *Competitive Strategic Management.* Englewood Cliffs, NJ: Prentice Hall.

Shapiro, Carl, and Joseph Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, 74(3): 433-44.

Shavell, Steven. 1979. "On Moral Hazard and Insurance." *Quarterly Journal of Economics*, 93(4): 541-62.

Shinkle, George. 2012. "Organizational Aspirations, Reference Points, and Goals: Building on the Past and Aiming for the Future." *Journal of Management*, 38(1): 415-55.

Simon, Herbert. 1957. "A Behavioral Model of Rational Choice," in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting.* New York: Wiley.

Strausz, Roland. 1997. "Collusion and Renegotiation in a Principal-Supervisor-Agent Relationship." *Scandinavian Journal of Economics*, 99(4): 497-18.

Steel, Piers, and Cornelius König. 2006. "Integrating Theories of Motivation." *Academy of Management Review*, 31(4): 889-913.

Suzuki, Yutaka. 2012. "Collusive Supervision and Organization Design in a Three-Tier Agency Model with a Continuum of Types." Working Paper.

Taylor, Fredrick Winslow. 1912. *The Principles of Scientific Management.* Harper and Brothers: New York, NY.

Tirole, Jean. 1986. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics, and Organization*, 2: 181-214.

_____. 1992. "Collussion and the Theory of Organizations," in Jean-Jacques Laffont, ed., *Advances in Economic Theory* Vol. 1. Cambridge: Cambridge University Press.

Williamson, Oliver. 1963. "Managerial Discretion and Business Behavior." *American Economics Review*, 53(5): 1032-57.

_____. 1967. "Hierarchical Control and Optimal Firm Size." *Journal of Political Economy*, 123-38.

Wu, Yanhui. 2011. "Authority, Incentives, and Performance: Theory and Evidence from a Chinese Newspaper." Working Paper.

Yermack, David. 1997. "Good Timing: CEO Stock Option Awards and Company News Announcements." *Journal of Finance*, 52: 449-76.

Tables and Figures

TABLE 1— Descriptive Statistics by Industry

	Information & Enterprise Software	Scientific, Professional, & Technical Services	Manu-facturing	Other
	(1)	(2)	(3)	(4)
a. Basic Descriptives				
Firms	56	38	62	88
Firm-Months	1,032	679	1,165	1,693
Persons	18,662	14,448	22,297	13,920
Person-Months	286k	237k	330k	160k
b. Standardized Jobs				
Sales Managers	1,960	1,658	2,139	1,883
Sales Reps	4,028	5,255	6,829	3,686
Account Managers	2,227	2,239	2,174	2,332
Other	12,159	8,207	12,238	6,689
c. Orders & Payments				
Transactions	23m	22m	74m	37m
Commissions	15m	14m	39m	30m
Bonus Payments	49k	47k	176k	86k
Other V. Payments	14k	6k	54k	116k
Var. Pay (USD)	2,277m	1,462m	3,729m	1,523m
d. Key Variables				
Turnover Events	8,968	3,170	9,346	7,909
Comp Plans	1,877	1,052	1,842	1,126
Quotas	28,702	21,087	32,006	19,175
EE-Fiscal Years	35,703	30,472	44,461	23,982

NOTE – Firms include all those for which data are complete. "Other" industries include retail trade (14 firms), wholesale trade (13), administrative support (12), and finance and insurance (10). The suffix "m" denotes the count is in millions. Data include 89 private and 133 public companies and subsidiaries thereof. Sales managers include persons with reporting subordinates. Credited transactions include only unique order-worker combinations, and do not count, for example, annuity tails or redundant individual/team credits. Other variable payments include, for example, draws that are arguably part of base pay.

TABLE 2— OLS Spline Regression Predicting Log-Credited Business of Subordinate Employees with Employee Fixed Effects, by Industry

	Information & Enterprise Software	Scientific, Professional, & Technical Services	Manu-facturing	Other
	(1)	(2)	(3)	(4)
Months 1 - 3 (M_3)	0.281*	0.132*	0.268*	0.217*
	(0.014)	(0.015)	(0.011)	(0.016)
Months 4 - 6 (M_6)	0.157*	0.125*	0.162*	0.166*
	(0.006)	(0.006)	(0.005)	(0.007)
Months 7 - 9 (M_9)	0.136*	0.020*	0.070*	0.102*
	(0.004)	(0.003)	(0.003)	(0.004)
Months 10-12 (M_{12})	0.109*	0.018*	0.063*	0.0747*
	(0.003)	(0.003)	(0.002)	(0.003)
Months 13+ (M_{13})	-0.099*	0.002	0.009*	0.048*
	(0.002)	(0.001)	(0.001)	(0.002)
Constant	9.077*	6.766*	8.771*	6.846*
	(0.030)	(0.029)	(0.023)	(0.034)
SD of α_i	4.794	6.359	3.999	5.503
SD of ε_{it}	3.066	2.415	2.522	2.429
Share of var($\ln y_{it}$) explained by α_i	0.710	0.874	0.715	0.837
Fixed Effects (EEs)	11,397	10,961	8,314	7,864
Obs (EE-Months)	141,493	91,373	168,101	69,383

NOTE – * : $p < 0.01$. These regressions are performed at the level of four industries, while distinguish poor performers use regressions at the level of 834 job-firm combinations, where employee fixed-effects are normalized such that employee fixed effects for jobs-in-firms have a mean of zero. Regression results are for workers without subordinates who ultimately had greater than six months tenure and who had at least four peers with their firm-job category.

TABLE 3— Monthly Turnover by Estimated Quota Attainment With and Without Subordinate's Credits, for Poor and Non-Poor Performing Subordinates

		"Misses Quota Anyway"	"Makes Quota Conditionally"	"Makes Quota Anyway"
Parent's Q Attainment		$Q_i^H < 100\%$	$Q_i^H \geq 100\%$	$Q_i^H \geq 100\%$
Discounted Q Attainment		$Q_i^L < 100\%$	$Q_i^L < 100\%$	$Q_i^L \geq 100\%$
Outcome of Parent's Q Attainment wrt. Sub.:		(1)	(2)	(3)

a. Poor Performing Subordinates

		(1)	(2)	(3)
Fourth Quarter	mean	18.6%	5.6%	22.2%
	s.e.	(0.69%)	(0.36%)	(0.35%)
	n	3,114	3,972	13,853
"Month 13"	mean	7.4%	14.4%	4.1%
	s.e.	(0.53%)	(0.61%)	(0.19%)
	n	2,427	3,306	10,137

b. All Other Subsordinates

|		(1)	(2)	(3)
Fourth Quarter	mean	10.4%	11.1%	18.8%
	s.e.	(0.30%)	(0.32%)	(0.16%)
	n	9,857	9,133	56,887
"Month 13"	mean	3.7%	4.7%	1.5%
	s.e.	(0.22%)	(0.25%)	(0.05%)
	n	6,946	6,862	45,358

NOTE – Turnover rates are for subordinates in the fourth quarter and the first month of the year following the fiscal year for which the manager's quota attainments and the subordinate-adjusted quota attainments are reported. "Month 13" results include only cases where the manager-worker pair matches across fiscal years (*i.e.* neither the manager nor subordinate were transferred, the manager did not turnover, and the subordinate did not turn over in the previous quarter).

TABLE 4— Logit Predicting Subordinate's Quota Adjustment by Manager's and Subordinate's Realized Quota Attainment, for Fiscal Months 10 - 12

	Month 10 (1)		Month 11 (2)		Month 12 (3)	
	β^m	β^s	β^m	β^s	β^m	β^s
75-80%	-0.775**	-0.318	-0.530*	0.006	0.754**	0.215
	(0.210)	(0.178)	(0.209)	(0.151)	(0.263)	(0.196)
80-85%	-0.379**	-0.525**	-0.152	-0.317*	0.221	0.104
	(0.141)	(0.157)	(0.143)	(0.139)	(0.273)	(0.183)
85-90%	-0.615**	-0.180	-0.543**	-0.206	0.228	-0.271
	(0.147)	(0.133)	(0.160)	(0.130)	(0.231)	(0.194)
90-95%	-0.325**	-0.425**	-0.472**	-0.346**	0.153	0.0451
	(0.119)	(0.128)	(0.141)	(0.116)	(0.227)	(0.157)
95-100%	-0.233	0.212*	0.094	0.025	0.366	0.0216
	(0.119)	(0.108)	(0.116)	(0.106)	(0.249)	(0.161)
100-105%	ref	ref	ref	ref	ref	ref
105-110%	0.009	0.184*	0.055	-0.057	0.252	-0.103
	(0.0821)	(0.088)	(0.091)	(0.085)	(0.202)	(0.133)
110-115%	-0.360**	-0.190	-0.410**	-0.381**	0.769**	0.0243
	(0.090)	(0.097)	(0.104)	(0.094)	(0.198)	(0.134)
115-120%	-0.440**	-0.217*	-0.327**	-0.523**	0.361	-0.087
	(0.084)	(0.091)	(0.092)	(0.088)	(0.198)	(0.127)
120-125%	-0.523**	-0.274**	-0.522**	-0.506**	0.222	-0.183
	(0.098)	(0.104)	(0.106)	(0.098)	(0.208)	(0.141)
Exterior 5% F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.096*		-2.213*		-2.923*	
	(0.070)		(0.075)		(0.158)	
Obs.	41,665		41,090		38,228	

NOTE — ** : $p < 0.01$, * : $p < 0.05$. Each logit regression spans two columns. Includes manager-subordinate pairs in which both ultimate quota attainments are within 50% of quota. Exterior controls include 5% quota attainment bins (shown in the accompanying figure). Standard errors clustered by the manager-worker pair.

FIGURE 1. : Histogram of Quota Attainment, Salespeople and their Immediate Managers

a. Managers   b. Subordinates
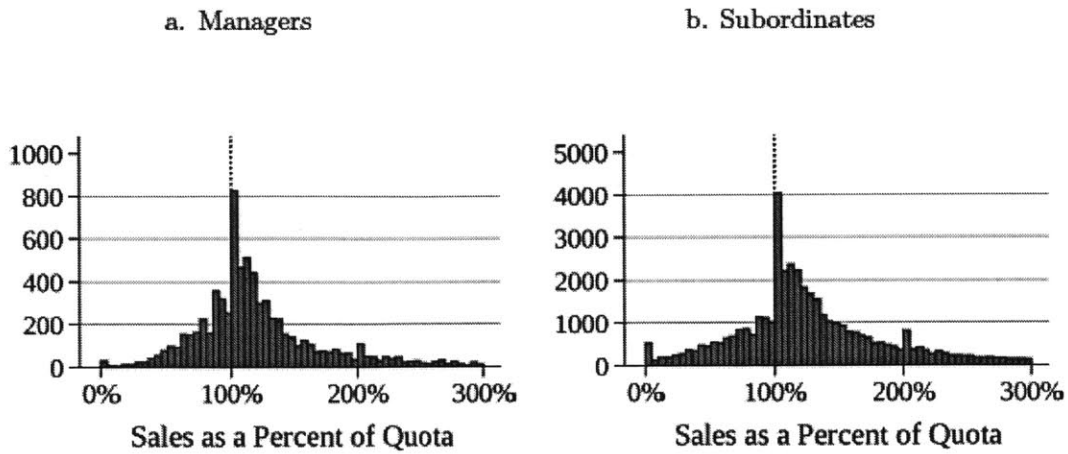
NOTE – Histograms represent counts of realized total quota attainment (at 5% intervals) at the end of annual measurement periods among salespeople and sales managers employed over the entirety of the measurement period. All panels corresponding to managers include only plans with orders that are credited to at least four subordinates, who themselves do not have subordinates, and whose subordinates' sum of quotas are not within 5% of the manager's quota. The quartiles of the sum of subordinates' quotas are 50%, 100%, and 200% of the manager's quota, and only 7% of the sum of the subordinates quotas are exactly the manager's quota.

FIGURE 2. : Expected Payoffs: Retaining and Terminating Low Ability Workers
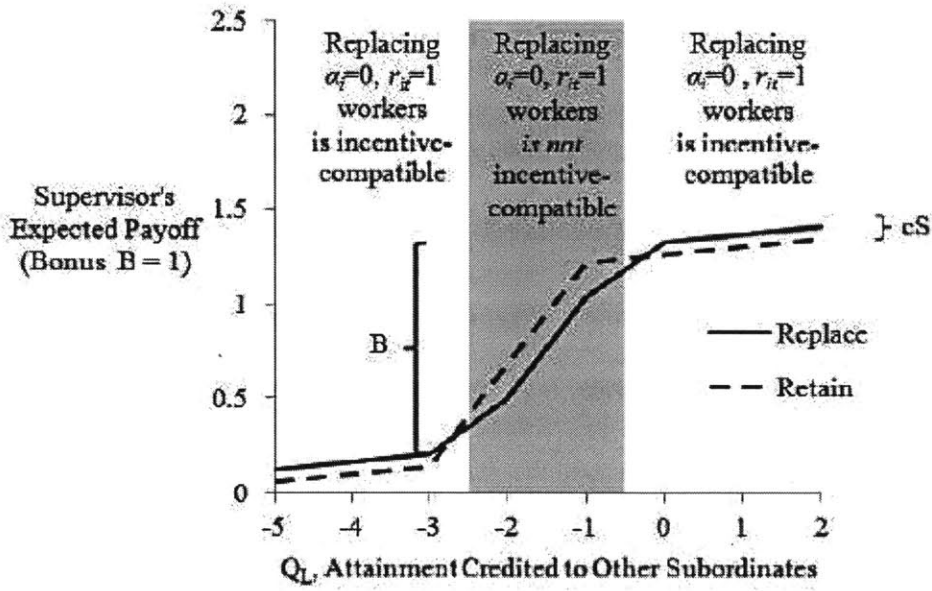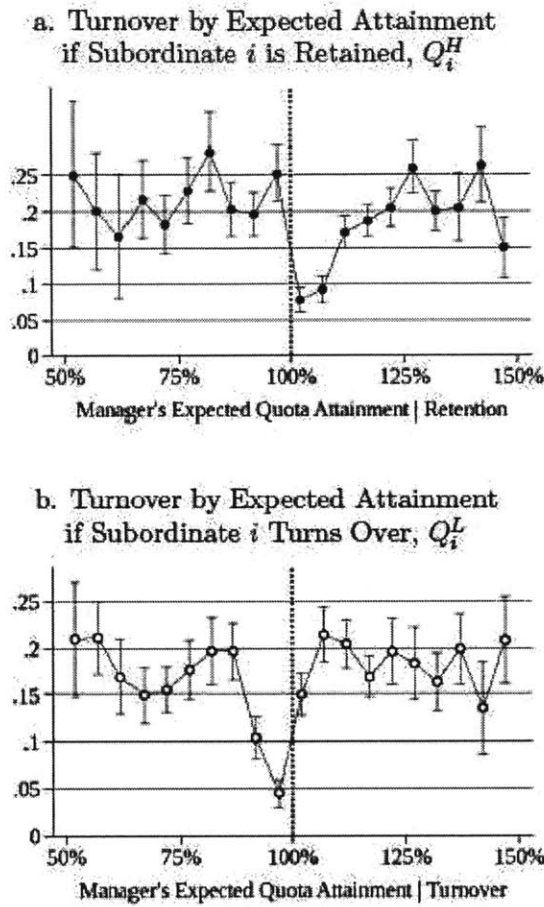

94

FIGURE 3. : Turnover of Poor-Performing Subordinates at Managers' Quota Thresholds

a. Turnover by Expected Attainment
if Subordinate i is Retained, Q_i^H

b. Turnover by Expected Attainment
if Subordinate i Turns Over, Q_i^L

NOTE – Quota attainment is discretized into 5%-wide
bins. Whiskers represent 95% confidence intervals.

FIGURE 4. : Turnover of Poor-Performing Subordinates at Uncle-Position's Quota Thresholds
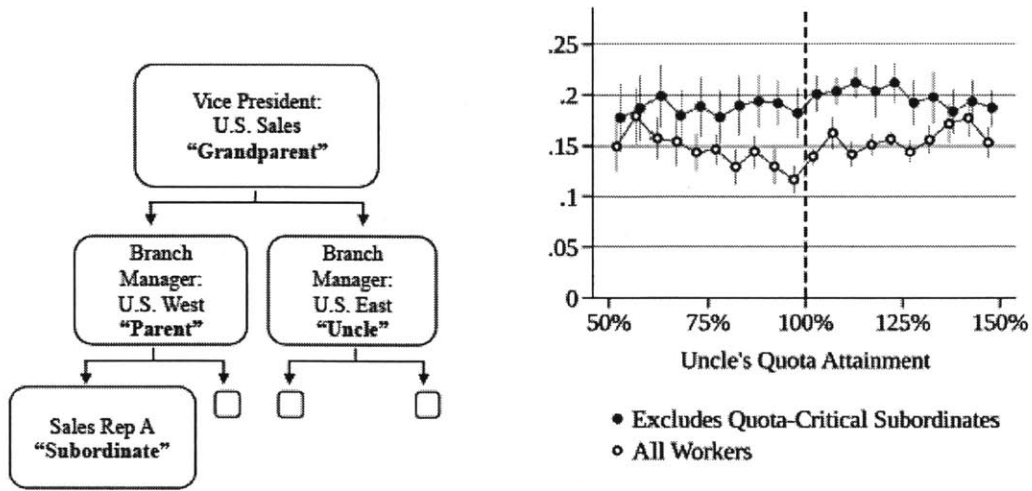
NOTE – Among rank-and-file subordinates, I use "uncle" to denote a randomly-selected manager who reports to the subordinate's parent's parent ("grandparent"). Fewer than 1% of all subordinates have no "uncles."Solid markers omit workers whose sales were predicted to be critical to meeting the parent position's quota. Whiskers represent 95% intervals.

FIGURE 5. : Predicted Probabilities a Subordinate's Quota is Adjusted, by Manager's and Subordinate's Own Ultimate Quota Attainment
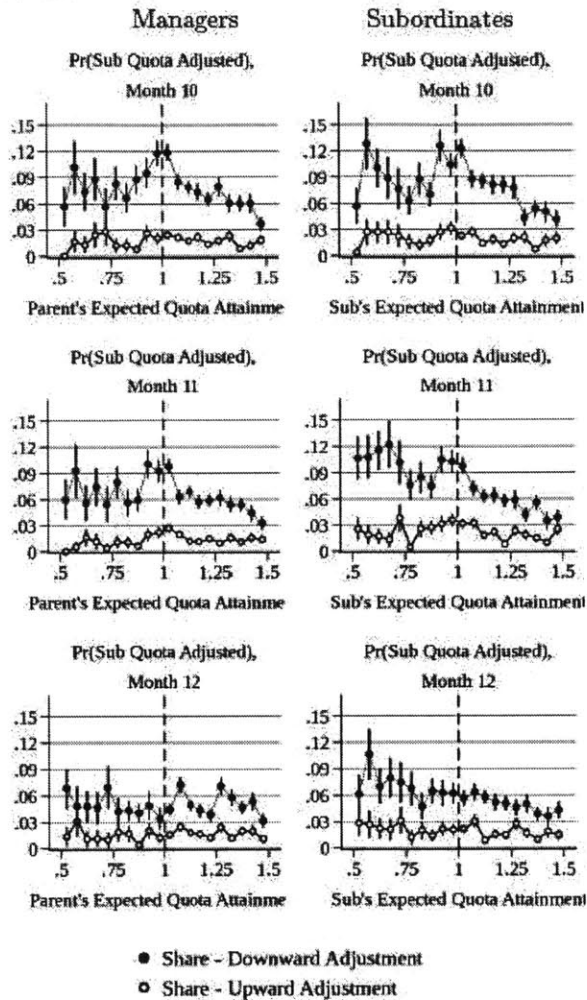
NOTE – Predicted probabilities by manager's quota are at mean values for subordinate quota attainment bins, and predicted probabilities by the subordinate's quota threshold are at the mean values for manager quota attainment bins. Quota attainment bins are post-adjustment. Whiskers represent standard errors.