ESSAYS IN APPLIED ECONOMETRICS

By

QIONG ZHOU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2009

To my parents, Min Zheng and Yuanguang Zhou

ACKNOWLEDGMENTS

The successful completion of this dissertation was not possible without support of several individuals. Each of these individuals has provided endless support and invaluable suggestions. First and foremost, I am deeply indebted to my chair, Dr. Chunrong Ai, who opened the door of graduate study in the United States for me. Without his incredible support and guidance, I could not have completed this work. I cannot thank him enough for his help through this process. I thank Dr. Wei Shen for bringing me into the new field, strategic management, and for his wonderful guidance on my future career. I thank Dr. Jonathan Hamilton for providing useful suggestions and support throughout my graduate work. I also thank Dr. Joseph Terza for great suggestions. I thank all my friends, here and in China, whose support was crucial during my life in the United States. Finally, I thank my family, my parents, grandparents, aunties and uncles. Their unconditional love and support guided me all my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure page

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ESSAYS IN APPLIED ECONOMETRICS

By

Qiong Zhou

May 2009

Chair: Chunrong Ai
Major: Economics

My research examines three separate studies of applied econometrics. In the first study, I empirically assess the impact of human capital acquisition from the target firm through a merger or an acquisition on post merger CEO turnover in the acquiring firm. Little is known about the effects of a merger or an acquisition on the acquiring firm's management team. The empirical evidence shows that merger is a way to acquire talented human capital, which will change both top management team and board structure of the acquiring firm, and thus result in leadership change in the acquiring firm. Using a sample of 236 mergers during 1996 to 2000 in the US, I find: (1) 46% of CEOs of acquiring firms are replaced within 5 years; 28% leave voluntarily, and 18% are forced to step down; (2) if the target firm's top executives are retained as top executives, the acquiring firm's CEO is more likely to leave; (3) if top executives of the target firm are retained as board directors in the acquiring firm, the acquiring firm's CEO is less likely to leave voluntarily, but no change occurs in the probability of being forced out.

Next, I investigate the finite sample performance of several estimators proposed for the panel data Tobit regression model with individual effects, including the Honoré estimator and the continuously updating GMM estimator. The continuously updating GMM estimator is based on more conditional moment restrictions than the Honoré estimator, and consequently is more

efficient than the Honoré estimator for large samples. My simulation study shows that the continuously updating GMM estimator performs not better, but in most cases worse than the Honoré estimator for small samples. The reason for this finding is that the continuously updating GMM estimator is based on more moment restrictions that require discarding observations. In my design, about seventy percent of observations are discarded. The too few observations lead to an imprecise weighting matrix estimate, which in turn leads to an unreliable updating GMM estimator. This study calls for an alternative estimation method that does not rely on trimming.

In the final study, I propose a maximum likelihood estimator (MLE) for the panel data Tobit regression model with unknown individual effects. To overcome the problem occurred in chapter 2, my proposal is to use log likelihood density function instead of conditional moment restrictions in optimization problem. I suggest to approximate unknown density function of individual effects with a sieve estimator and to estimate finite dimensional unknown parameters and infinite dimensional sieve estimator jointly by applying the method of maximum likelihood estimation. Under some sufficient conditions, I show that (1) the sieve estimator of unknown density function for individual effects is consistent under certain metric; (2) the MLE estimators of the finite dimensional parameters are consistent and asymptotically normally distributed; (3) the estimator for the asymptotic covariance of the parameter is consistent.

CHAPTER 1
HUMAN CAPITAL ACQUISITION AND POST MERGER TURNOVER OF ACQUIRING FIRM'S CEO

Introduction

The past decades have witnessed a great wave of merger and acquisition activities. In the US alone, over 200,000 deals occurred between 1963 and 2007. Much of the existing literature focuses on understanding the organizational changes within the acquiring firm after a merger[1], particularly leadership change. It is important to grasp the possible implications that the merger will have on the top management team.

In the literature on post merger managerial turnover, a large body of studies examine post merger turnover of the target firm's top executives. Three theories have been put forth to explain why some top executives from the target firm leave after the merger. Market discipline theory, for one, argues that the market for corporate control plays an important disciplinary role. Ineffective managers of target firms who performed poorly before the merger are likely to be replaced after the merger (e.g., Walsh and Ellwood, 1991; Martin and McConnel, 1991; Hadlock, Houston, and Ryngaert, 1999; Harford, 2003). This theory is supported by empirical evidence (e.g., Coughlan and Schmidt, 1985; Warner, Watts and Wruck, 1988; Weisbach, 1988; Gibbons and Murphy, 1990; Blackwell, Brickley and Weisbach, 1994).

Relative standing theory or local social status theory, for another, suggests that, if the acquired executives feel inferior, are stripped of status, or locked in a struggle with the acquirers in the merged entity, they will tend to leave post merger (e.g., Hambrick and Cannella, 1993; Lubatkin, Schweiger, and Yaakov, 1999; Cannella and Hambrick, 1993).

[1] For convenience, I will refer to all of the mergers and acquisitions studied as mergers. Below, I discuss the definition of the acquiring firm and the target firm (which applies to all mergers in the data).

A third argument comes from human capital theory. If the cost of human capital investment that the acquired executives have to spend post merger is larger than the potential future returns, the executives from the target firm are likely to leave (e.g., Buchholtz, Ribbens and Houle, 2003).

While it is important to study the fate of the target firm's CEO post merger, it is equally important to observe what happens to the acquiring firm's CEO post merger. Little research has investigated the impact of a merger on the acquiring firm's CEO; the only exception is Lehn and Zhao (2006). They examine the relation between cumulative abnormal returns in the stock market following acquisition announcement and subsequent CEO turnover. They find that if the acquiring firm's CEO makes an acquisition that creates shareholder value, he is expected to be rewarded with extended tenure. In contrast, if the acquiring firm's CEO makes an acquisition that reduces shareholder value, he or she will be punished and replaced. Their argument follows the traditional market discipline perspective of CEO turnover.

To broaden understanding of the impact of a merger on post merger turnover of the acquiring firm's CEO, in this paper I investigate an alternative perspective on how human capital acquisition from the target firm through a merger affects CEO turnover in the acquiring firm post merger.

Bell Atlantic's merger with NYNEX in 1996 supports this view. After the merger, Ivan G. Seidenberg, chairman and chief executive officer of NYNEX, had been retained as vice chairman, president and chief operating officer. Frederic V. Salerno, who was vice chairman of NYNEX, became the new chief financial officer and executive vice president. About one and half years after the closing of the merger, according to the terms of the agreement, Mr. Seidenberg became chief executive officer of the new company and chairman upon Raymond

W. Smith's retirement, who was chairman and chief executive officer of Bell Atlantic at the time of the merger announcement.

One of the motivations for a merger is to acquire talented and skillful top executives from the target firm, especially to obtain potential successors for the current CEO. The top executives who come from the target firm can provide valuable expertise for firm strategy and supply complementary knowledge that is not available within the acquiring firm to help the newly combined firm to survive. In addition, human capital acquisition through merger changes the distribution of power and control within the acquiring firm. With a better potential CEO candidate, the incumbent CEO is easier to replace if his performance drops. Moreover, the acquiring firm would rely less on the incumbent CEO. The new "upper echelon" members could also cause a power struggle within the top management team of the newly combined firm. If the CEO of the acquiring firm is successfully challenged by the acquired top executives and loses control, or fails in a power struggle with the acquired managers from the target firm, he is likely to depart.

To test whether this human capital acquisition theory can explain post merger turnover of the acquiring firm's CEO, I collected a sample of mergers and acquisitions. Of the 236 mergers that occurred from 1996 to 2000, there are 109 or 46% of cases where the CEO of the acquiring firm was replaced within five years of the merger announcement. Of those 109 cases, 42 were forced to depart, and 67 were replaced voluntarily.

I employ both logit and multinomial logit regression models to examine the relationship between post merger turnover of the acquiring firm's CEO and human capital acquisition from the target firm through merger and acquisition activities. Both regression results show that if the target firm's top executive is retained as a top executive in the merged entity after the merger, the acquiring firm's CEO is more likely to leave, either voluntarily or involuntarily. If the target firm's top executives are retained as board directors in the merged firm, the acquiring firm's CEO is less likely to leave voluntarily, but no change occurs in the likelihood of being forced out. No significant association exists between an acquiring firm's board characteristics and the likelihood that the acquiring firm's CEO will be replaced. Finally, the more ownership concentration in the acquiring firm, the less the effect of director acquisition is on post merger CEO turnover.

My study contributes to the literature by further studying the role that a merger plays in post merger CEO turnover in the acquiring firm. Based on this study, another theory about human capital acquisition could be added alongside the market discipline theory to explain CEO turnover in the acquiring firm post merger. Also, the current study suggests that Shleifer and Vishny's (2003) theory of "stock market driven acquisitions" is incomplete. The merger does not only result in market takeover, but also fulfills the acquiring firms' desire for human capital talent.

The paper is organized as follows: Section 2 discusses the hypotheses, Section 3 describes the sample and empirical design uses in the paper, Section 4 discusses the empirical results, and Section 5 concludes the paper.

Human Capital Acquisition

The existing studies find that a firm's good performance is a reflection of a good and efficient top management team, including the CEO, president, chairman of the board, vice-presidents, CFO, COO, and other "upper echelon" executives. The top executives are unique organizational resources (e.g., Daily, Certo, and Dalton, 2000). This unique human capital can have major impacts on organizational actions and performances (e.g., Thompson, 1967). When these resources are aligned with the organizational goal of the acquiring firm, they are much more productive than general labor and are more likely to produce a competitive advantage for the firm (Hitt, Bierman, Shimize, & Kochhar, 2001). Thus, to gain a competitive edge, it is critical to have a talented top management team. However, talented top executives are few and in high demand; good heirs apparent are especially hard to obtain. One way to acquire highly talented and skillful top executives is through merger (e.g., Parons and Baumgartner, 1970; and Pitts, 1976).

Acquiring successful and skillful top managers from the target firm to replace the ineffective incumbent top managers or CEO in the acquiring firm has several advantages. First, an acquired top manager can be a good CEO heir apparent for the acquiring firm, especially when the incumbent CEO of the acquiring firm is close to retirement. Many firms struggle to smooth the power transition process by choosing an heir apparent well in advance of the actual CEO turnover (Wall Street Journal, 1997). Merger provides a way to find a good heir apparent since he can work together with the incumbent CEO as the CEO passes power and control. Second, the acquired executives will provide valuable expertise for firm strategy and supply complementary knowledge that is not available within the acquiring firm. Moreover, merger and acquisition will not only change the acquiring firm's "upper echelon" by acquiring talented executives from the target firm, but also elsewhere in the organization. Such changes in the organization frequently lead to critical problems, difficulties, uncertainties and contingencies. The top management team self-perceived capacity or incapacity for dealing with the critical issue is important for the newly combined firm to survive. Top executives from the target firm who are capable of coping with the new organization's environment could be selected to enter the "upper echelon" and capture power and controls within the firm. Third, such human capital acquisition also brings the incentive to the incumbent top executives to improve the firm's performance. Hence, human capital acquisition can influence post merger turnover of the acquiring firm's CEO through the internal governance mechanism.

The disadvantage of acquiring top executives through merger is that the acquired executives take power and control from incumbent top managers and the CEO, hence exacerbating the power contest between top managers. One outcome of the redistribution of power and control within the firm deriving from human capital acquisition through merger is CEO turnover in the acquiring firm post merger. Many studies find that the competition among top executives plays an important role in CEO dismissal and succession decisions (Boeker, 1992; Cannella & Lubatkin, 1993; Ocasio, 1994; Cannella & Shen, 2001; Shen & Cannella, 2002). Acquired top executives confront significant challenges upon taking office after entering the "upper echelon" of the acquiring firm. To keep the position, the acquired executives might want to obtain more power within the "upper echelon", and have more desire for career advancement and demand better performance. In addition, since acquired top executives are new to the incumbent top executives, they are more likely to have different interests and strategies than the incumbent CEO. Thus, newly joined top executives from the target firm are more likely to challenge the CEO and worsen the power competition. Although such contests are not easily observable, Shen and Cannella (2002) argued that the power contests among top executives will affect the process of the CEO dismissal. When acquired top executives successfully challenge the CEO, the CEO will be dismissed (Preffer, 1981; Sonnenfeld, 1988). Furthermore, when acquired top executives are locked in a struggle with the acquirer, the power contest and interest conflict can harm the firm's performance. A firm's poor performance results not only because of an ineffective CEO, but also due to the power tournament within top management, which normally is the reason for CEO turnover. Therefore, human capital acquisition will affect post merger turnover of the acquiring firm's CEO.

Insofar as both advantages and disadvantages of human capital acquisition will influence post merger turnover of the acquiring firm's CEO, it is an ideal topic to examine how human capital acquisition decisions through merger affect the acquiring firm's CEO replacement. Top executive acquisition from the target firm provides an opportunity to examine the relation between human capital acquisitions through merger and the acquiring firm's CEO turnover after merger. When acquiring firms decide to retain valuable top executives from the target firm, acquired top executives may get positions in the acquiring firms' top management team after merger where they may control the firm's strategy, or they may enter the board of the acquiring firm where they have power to monitor and advise the CEO. Both types of human capital acquisition from the target firm are expected to have some influence on the acquiring firm's incumbent CEO turnover.

Human Capital Acquisition as a Top Team Officer

If a top executive of the target firm is retained as a top team management officer in the acquiring firm after merger, it means that the acquiring firm believes that he is a good manager or potential successor. The firm with a CEO who is going to retire within a few years would like to pass power and control of the firm to a potential successor. With high status and more power within the "upper echelon" of the acquiring firm, the prospective CEO heir would be able to hold the highest position in the firm earlier by subverting the power competition within the top management team, while it is hard for the current CEO to maintain independent control of the firm when there are new top executives acquired from the target firm. On the other hand, acquired top executives from the target firm are believed to be more capable of coping with the newly combined firm compared to the current members of the top management team. Therefore, power competition is going to be escalated by the new competent "upper echelon." With more power contests, the firm's strategy may not be efficient and competitive. For instance, to compete with new members of the "upper echelon," the current CEO may be involved in high risk projects and make wrong decisions, thus the firm's performance will decline, which could lead him to be replaced. Moreover, compared to the firm without potential successors, when a firm has good candidates for CEO succession, the board would be more likely to replace the CEO who does not perform well. As a result, I expect that human capital acquisition of top team officers has a positive effect on post merger turnover of the acquiring firm's CEO.

Hypothesis1: If the target firm's top executive has retained top executives in the newly combined firm after the merger, the acquiring firm's CEO is more likely to leave, either voluntarily or involuntarily.

Human Capital Acquisition as a Board Member

For most of the large mergers, top executives of the target firm, such as the CEO and president who are also shareholders before merger, may become directors on the acquiring firm's board after the deal is completed, especially when the deal is paid for with stock. Therefore, given the context of merger, it is interesting to study the effect of such human capital acquisition on post merger turnover of the acquiring firm's CEO through the internal governance mechanism.

The CEO who is going to retire within a short time is normally older and less aggressive, thus he would like to make a friendly merger and provide a better deal to the target firm's top managers as compared to an ambitious CEO. Thus, top executives of the target firm who have positions on the board of the acquiring firm after the merger could have a good relationship with the incumbent CEO in the acquiring firm. The target firm's top executives would appreciate for being provided a good deal and a position on the board of the acquiring firm after the merger was effected. The current CEO could also benefit from the good relationship since those new directors who are from the top management team of the target firm could provide useful advice about managing the acquired firm and support them in the board. Thus, I expect that if acquiring firms make director acquisition through merger, the likelihood of the acquiring firms CEO being replaced decreases.

Hypothesis2: If the acquiring firm acquired the target firm's top executives and they are listed as board directors in the newly combined firm after the merger, the acquiring firm's CEO is less likely to leave.

Data and Empirical Design

The sample of mergers is obtained from Thomson Financial Securities Data Corporate (SDC) MERGER database, COMPUSTAT, and Disclosure's Compact D SEC database[2]. I begin withdrawing the initial sample from SDC based on the following criteria: (1) the merger or acquisition is announced between January 1, 1996 and December 31, 2000[3]; (2) The transaction occurs in the US; (3) the form of the deal is merger or acquisition; (4) the status of the deal is "completed"; (5) both the acquiring and target firm are publicly traded[4]; (6) the buyer's net sales in the last twelve months is greater than 100 million dollars. These screens yield a candidate sample of 1480 mergers.

In order to ensure that the deal represents "large" investments by acquiring firms, I require that the size of the target firm is at least 30% of the size of the acquiring firm, measured by net assets. The acquiring firm would be more likely to acquire human capital from the target firm

[2] I accessed all databases through the UFL business library.

[3] I ended the period of the sample at 2000, so that I could observe the CEO turnover information within five years following merger announcement.

[4] For deals as form of merger, SDC merger and acquisition dataset has already distinguished between the acquirer and target. So I followed their criterion of which firm is acquirer and which firm is target for merger transaction.

when the deal is large enough, or as a form of merger, especially for the human capital acquisition as directors. 408 deals satisfied this requirement.

Information about CEO turnover of the acquiring firm and top management team retention of the target firm are also required to be available to each acquiring and target firm on the deal's announcement year through five years after the announcement date. The main sources of the information about CEO and top management team are company annual reports and proxy statements in the SEC Edgar database. The SEC Edgar database is also the source of information about firms' governance structures. After searching company annual reports, proxy statements, LexisNexis and news wires, 104 mergers are excluded since CEO turnover information couldn't be identified or because the governance data is incomplete. As a result, this filter reduces the sample to 304 mergers. Moreover, both firms have to be listed on the COMPUSTAT financial statement data for each acquiring and target firm could be collected. The final sample for regression analysis includes 236 mergers made during the sample period.

Definition of CEO Turnover

Turnover is classified into two types. I define "CEO turnover" to include CEO replacement within five years after merger announcement that reported as retired, or replaced but still served on the board, resigned or terminated. For five CEO replacements, which are reported as deceased, I classify the observations as one censored. Therefore, the "no CEO turnover" group includes those transactions where the CEO who announced the merger was still the CEO of the acquiring firm after five years since the announcement, or there was a CEO replacement due to death or poor health. The "CEO turnover" group includes the other transactions where the CEO who announced the merger was replaced within five years after the merger announcement.

The types of CEO turnover are identified by voluntary turnover and forced turnover. "Forced CEO turnover," under the subsample "CEO turnover," is defined as a non-routine CEO replacement, i.e., the CEO was resigned, or terminated within five years of merger. For the other CEO turnover, if the departing CEO was reported as retired within five years of the merger, or if the news reports that the CEO would still serve on the board, as a non-executive chair or vice-chair, then the CEO turnover is classified as a "voluntary turnover." To identify the types of CEO turnover, the acquiring firm's proxy statements are examined for the announcement year and five years after the announcement. Of the 408 deals, CEO turnover type for 104 transactions could not be identified. For 29 deals information about the CEO at the announcement year could not be found, 44 deals have incomplete information about CEO turnover within five years in the dataset, and 31 were acquired by other firms within five years[5]. Therefore, these 104 deals were excluded from the sample.

A dummy variable is created for CEO turnover that takes value of one if the acquiring firm's CEO is replaced within five years of the merger's announcement, and zero otherwise. Also, a categorical variable is created for the turnover type of the acquiring firm's CEO that takes the value of one if the acquiring firm's CEO is replaced voluntarily within five years of the announcement, and two if the acquiring firm's CEO is fired or forced to step down, and zero if there is no turnover.

Table 1-1 presents descriptive statistics for post merger turnover of the acquiring firm's CEO. Panel A of Table 1-1 reports the frequency of "CEO turnover" versus "no CEO turnover" for the sample. Of the 236 mergers in the sample, 127 are not replaced after their respective

[5] Since this paper examines the relation between human capital acquisition and CEO turnover, the effect of human capital acquisition is through internal governance. In addition, this paper wants to examine the different effect of human capital acquisition on the type of CEO turnover. Thus, I exclude those mergers in which CEOs are replaced

acquisitions within five years, 109 are replaced within five years. Within the subsample of 109 observations subject to CEO turnover, 67 are replaced voluntarily within five years, and the remaining 42 are forced to step down within five years after their respective merger. This distribution indicates that close to half of the CEOs in the full sample are subjected to replacement within five years of the mergers, and among them, about sixty percent of CEOs are replaced voluntarily.

Panel B reports the frequency of CEO turnover across different merger announcement year for the full sample of 236 mergers, and for four subsamples. For mergers announced in 1996 and 1999, the probability of post merger CEO turnover in the acquiring firm is higher compared to the mergers announced in 1997, 1998 and 2000, where more than half of CEOs are subjected to replacement within five years. Of mergers announced in 1996, over forty percent - the highest of the five years - of CEOs from the acquiring firm voluntarily stepped down. And of mergers announced in 1999, about thirty percent-the highest among the five years-of CEOs from the acquiring firm were forced to step down.

Human Capital Acquisition

To examine the effect of human capital acquisition from the target firm on turnover of the acquiring firm's CEO post merger, I hand-coded two variables not previously studied: "officer acquisition" and "director acquisition." I first collected the names of top management officers of the target firm reported in the proxy statements from the SEC dataset in the merger announcement year, and then collected names of both top management officers and board directors of the acquiring firm after the merger was effected. Then I checked whether the top

---

by external control market, i.e., takeover or bankruptcy. That is, I collect data that CEOs are only replaced by internal governance.

manager from the target firm got a position in the acquiring firm's top management team or board after the merger was completed.

"Officer acquisition" is defined as a dummy variable[6] which takes the value of one if one or more top managers of the target firm have been acquired as top managers in the acquiring firm after the merger was effected, and takes the value of zero otherwise.

Similarly, "director acquisition" is defined as a dummy variable[7] which takes the value of one if one or more top managers of the target firm have been acquired as board directors in the acquiring firm after the merger was completed, and takes the value of zero otherwise.

Table 1-2 reports the distribution of the human capital acquisition associated with merger for the entire sample and four subsamples. It lists the mean, median, and standard deviation values for both officer acquisition and director acquisition. On average, the probability that the acquiring firm would like to acquire top executives from the target firm as new top managers in the acquiring firm is 0.28, and 0.35 for director acquisition. The probability of the acquiring firm with post merger CEO turnover would like to acquire human capital from the target firm is larger than the acquiring firm without CEO turnover, for both top officer acquisition and director acquisition. The probability of acquiring human capital from the target firm as top managers in the acquiring firm is the highest for acquiring firms with forced CEO turnover (0.40), compared to acquiring firms with voluntary CEO turnover (0.28), firms with CEO turnover regardless of the turnover type (0.33), and firms without CEO turnover (0.24).

[6] First, I collected the number of top managers in the target firm who have retained positions in the acquiring firm and were reported as top management executives. However, in the regression model, I use a dummy variable rather than the number variable, because the existence of human capital acquisition as top officers have effect on CEO subsequent turnover, but the size of such human capital acquisition maybe not important in the model. The effect of human capital acquisition on CEO turnover may not vary over the size of human capital acquisition.

[7] Similar to officer acquisition, I collected the number of top managers in the target firm who become directors on the merged entity's board after the merger was effected, and then scaled it by the acquiring firm's board size to

For director acquisition, namely, acquiring firms acquired top executives from the target firm to be new directors in the acquiring firm post merger; the probabilities of such human capital acquisition are similar for three subsamples with CEO turnover. For instance, it is 0.40 for acquiring firms with forced CEO turnover, 0.38 for acquiring firms with voluntary CEO turnover, and 0.39 for acquiring firms with CEO turnover regardless of the turnover type. However, the probability of director acquisition for acquiring firms without CEO turnover is smaller (0.30).

Empirical Design and Control Variables

To examine the relation between human capital acquisition from the target firm and the probability of the post merger turnover of the acquiring firm's CEO after merger without specifying the turnover type, I estimate the logit model

$$prob(\text{CEO turnover}) = \frac{\exp(x'\beta)}{1+\exp(x'\beta)}$$

where the variable "CEO turnover" is defined as one if the CEO replaced within five years of the merger announcement, and zero if there is no CEO turnover; β is a $k \times 1$ vector of estimated coefficients for CEO turnover; x is a $k \times 1$ vector of explanatory variables which may influence the probability of CEO turnover according to empirical research on CEO turnover and merger issues, including human capital acquisition variables, the acquiring firm's CEO's characteristics, the acquiring and the target firm's performance, and the acquiring firm's corporate governance. I also include four year dummies for the merger announcement years of 1997-2000 to account for aggregate changes over time[8]. Table 1-3 lists the definitions of all variables used in the model.

eliminate the board size effect. Similar results are obtained if I use a scaled number of director acquisition instead of a dummy variable. Thus I used a dummy variable in this paper.

[8] 1996 data is the base year indicated by all year dummies=0.

Table 1-4 provides descriptive statistics of relevant variables for the full sample, including mean, median, standard deviation, minimum and maximum values. Table 1-5 lists the correlations among all variables, including two different dependent variables, the post merger turnover of acquiring firm's CEO, and the turnover types of the acquiring firm's CEO within 5 years after the merger announcement. The correlations, shows that the turnover of the acquiring firm's CEO after merger was positively associated with human capital acquisition but not strong, for both officer and director acquisitions. The turnover type of the acquiring firm's CEO was positively associated with both officer and director acquisitions, but significant for top officer acquisition. In addition, the control variables, CEO age and ownership concentration, was significantly associated with both CEO turnover and CEO turnover type, as might be expected. And post-ROE of the acquiring firm was only significantly associated with CEO turnover type. There is no strong correlation for other control variables. The magnitudes of the correlations do not suggest that multicollinearity is an issue.

CEO’s age

Buchholtz, Ribbens, and Houle (2003) find a significant relation between CEO age and CEO turnover, i.e. the probability of CEO departure will decrease with CEO age until a CEO reaches his/her middle age, and then the probability of CEO turnover will increase. After a merger, CEO’s needs to adjust to a new cast of characters require new investments to build new human capital. At an earlier age, CEOs might lack enough experience to handle the new challenges that arise from merger activities. And, the time for younger CEOs to build enough human capital is too long a wait for the acquiring firm. Thus, the probability of forced turnover may be higher for CEOs who are relatively young. The younger a CEO is, the less important the financial and career security is to him. For younger CEOs, it is less painful to leave a position and a company. Since younger CEOs can relocate relatively easily, the probability of voluntary CEO turnover may be higher for them.

As a CEO grows older, past the middle age and approaching the retirement age, he is less likely to make the new investment since fewer productive years of work are left. The acquiring firm would like to rely more on the new top executives, and decrease its dependence on the older CEO, especially for managing the new part of the combined corporation -- the acquired firm. In addition, merger activity, such a big change within the firm, would inevitably bring some risks. Older CEOs are expected to have less confidence to handle risk since they are less willing to build new human capital to deal with new risks. Thus, firms would like to diffuse their dependence on the current older CEO, and move some dependence to new top executives. Murphy and Zimmerman (1993), and Goyal and Park (2002) report a significant positive relation between CEO age and CEO turnover. Thus, I expect that an older CEO who passes middle age is more likely to retire. On the other hand, older CEOs are closer to their retirement age. With an older CEO, the acquiring firm has more pressure to look for a CEO heir and move power to the heir. Therefore, I expect the probability of voluntary turnover for the older CEO is higher as well.

Therefore, the probability of both voluntary and forced CEO turnover in the acquiring firm after a merger will decrease with age until a CEO reaches middle age, at which point the rate increases. I include both CEO age and CEO age squared at the year of the merger announcement in the analysis to model a curvilinear effect for CEO age.

As seen in Table 1-4, the mean age of CEOs is 53.86 and the median is 54 for the full sample. The age of CEOs with turnover is significantly higher than those without turnover, especially for CEOs with voluntary turnover. The mean age of CEOs without turnover is 51.61, while the mean age of CEOs with turnover is 56.47, 58.39 for voluntary turnover, and 53.4 for forced CEO turnover. Similar results hold for the median age of CEOs.

Firm performance

Pre-merger performance of the acquiring firm: The literature states that a firm's performance commonly implies the CEO's ability. Weisbach (1988), Murphy and Zimmerman (1993), find that the likelihood of CEO turnover is significantly higher when a firm's performance is lower. Post merger turnover of the acquiring firm's CEO after a merger might be due to the acquiring firm's poor performance before the corresponding merger. I expect that the probabilities of both forced and voluntary post merger turnover of the acquiring firm's CEO are inversely related to the acquiring firm's pre-merger performance. I include a measure of the acquiring firm's performance before the merger as an explanatory variable in the model. I calculated industry-adjusted yearly return on common equity (pre-ROE)[9]for one fiscal year prior to the merger announcement in the acquiring firm[10].

Post merger performance of the acquiring firm: I also include a measure of the acquiring firm's performance after the corresponding merger as a control variable to capture the effect of the acquiring firm's post merger performance on post merger turnover of the acquiring firm's CEO. The turnover of the acquiring firm's CEO after merger might be due to the firm's poor performance after the corresponding merger event. I expect that the probability of both forced and voluntary CEO turnover in the acquiring firm to be higher for those acquiring firms

[9] Following Lehn and Zhao (2006), I calculate the industry adjusted accounting performance measures by subtracting the industry median values from firm's corresponding measures' value. The industry median is the median value of the industry portfolio formed by matching the three-digit SIC code from COMPUSTAT.

[10] Return on assets (pre-ROA), and return on average equity (pre-ROAE) are also available in my dataset. They all give similar results as pre-ROE, thus I don't list them in the paper.

who have poor performance after merger. Similarly, industry-adjusted return on common equity (pre-ROE) for the acquiring firm in the year right after the merger announcement are included.

Pre-merger performance of the target firm: Because my main interest is the effect of human capital acquisition from the target firm on post merger turnover of the acquiring firm's CEO, I also include a measure of the target firm's performance before the merger. It is generally believed that a successfully performing firm usually has a competent and effectual top management team. Therefore, if a target firm has great performance before merger, the acquiring firm would be likely to obtain not only the assets of the target firm, but also its good top executives to replace the current top executives who are ineffectual or going to be retired soon in the acquiring firm. The acquired top executives may retain some feeling of centrality and importance, and they would get important positions in the combined enterprise's top management team post merger. One might predict that when target firms have great performance, human capital acquisition through those target firms would be more skillful and powerful which may induce a higher rate of post merger CEO departure in the acquiring firm.

To control the effect of pre-merger performance of the target firm on post merger turnover of the acquiring firm's CEO, I calculate a measure for the industry-adjusted return on common equity (pre-ROE of target) of the target firm in the year before the merger announcement. Table 1-4 reports the mean, median, and standard deviation values for the both acquiring and target firm's performance measures before and after the merger.

Corporate governance variables

Because the effect of human capital acquisition on post merger turnover of the acquiring firm's CEO by internal governance mechanism, I examine whether the relation between human capital acquisition and post merger CEO turnover is related to the characteristics of corporate governance. Specifically, I examine the role the acquiring firm's board characteristics and ownership structure play in the process. All governance data are taken from the acquiring firm's proxy statement that is closest in time to the announcement of the corresponding merger from the SEC database.

Board characteristics: Board size, board independence, and leadership are used to measure a firm's board structure. Board size is defined as the number of directors reported on the board. Board dependence is calculated as the percentage of inside directors on the board during the year of the merger announcement. Inside directors are defined as the board members who are employees, former employees, employee' relatives, attorneys, or accountants. Leadership is a dummy variable to control leadership structure, which takes the value of one if the CEO of the acquiring firm also serves as the chairman on the board, and zero otherwise (e.g. Lehn & Zhao, 2006).

Board size. Yermack (1996) and Jensen (1993) find a significant inverse relation between board size and firm's performance. Small boards are more effective monitors so that can help to improve the CEO's performance. Thus, I expect that the probability of the post merger CEO turnover in acquiring firms with smaller boards should be higher than those with larger boards.

Board independence. The literature argues that dependent directors would decrease the board's monitoring function. CEO turnover is more sensitive to firm's performance when the board is more independent (Weisbach, 1988). Thus, I expect that the less dependent the board, the higher the probability of voluntary CEO turnover. And, one might also expect that the less dependent the board, the higher the probability of forced CEO turnover when the CEO doesn't perform well.

In addition, the literature of the CEO succession issue argues that a seat on the board gives inside directors’ exposure to outside directors and enables them to build social networks and coalitions on the board (Jennings, 1971; Vancil, 1987). This development gives them more power and lends them more confidence with which to succeed the CEO. Therefore, if the number of inside directors on the board is relatively small, I would expect that the firm doesn't have an optional successor for the current CEO, especially for those incumbent CEOs who are close to retirement. Thus, the probability of voluntary CEO turnover from the acquiring firm should be higher for the firm with lower board independence when there is human capital acquisition. Similarly, with lower board dependence, the effective board would procure human capital acquisition from the target firm through merger, who could perform better in the newly combined firm and thus replace the current CEO. If this is the case, under the condition of lower board dependence, the rate of the forced CEO turnover would be higher when there is top officer acquisition.

Leadership structure*:* The literature studies argue that the concentration of decision management and decision control in one individual reduces a firm board's effectiveness (Fama & Jensen, 1993; Jensen, 1993). To improve the efficiency of the board, it is better to separate CEO and chairman positions. Thus, when the time of retirement finally arrives, the incumbent CEO who is also the chairman on the board can be reluctant to leave his position. Or, if the incumbent CEO is far away from retirement, he/she could use his conclusive power to retain his position even when the firm's performance is poor. I expect that a powerful CEO will decrease the likelihood of CEO turnover, both voluntary and forced to step down.

In addition, a CEO who is also the chair of the board may have more power in the firm, and the firm would be more likely to depend on him/her. With new top executives entering, it is easier for them to keep the firm's dependence and power to control the firm. Accordingly, one can expect that with top officer acquisition, the probability of CEO turnover should be lower for firms in which the CEO also serves as the chairman. Acquiring firms in which the CEO also serves as the chairman have less efficient boards; hence the likelihood of CEO turnover should be lower for firms with director acquisition from the target firm.

Table 1-4 lists the mean, median, and standard deviation values of board characteristics measures for the entire sample. The mean value of board size is 11.12 for the entire sample. And, on average, 27% of the directors on the board are insiders. The board size for different turnover types are close (11.38 for firms without turnover, 10.83 for firms with turnover, 11.16 for firms with voluntary turnover and 10.28 for firms with forced turnover.) Similarly, the mean value of inside directors is not significantly different across the subsample. The frequency with which CEOs also serve as chairman on the acquiring firm's board is high; the mean of the dummy variable is 0.69, and the median value is 1 for the full sample. The table reveals no significant difference in leadership across subsamples. The median value of the dummy variable is one for all subsamples, both mergers with different types of turnover and without turnover. The mean value of the leadership dummy is slightly higher (0.76) for acquiring firms with voluntary CEO turnover than those acquiring firms without turnover (0.67) and acquiring firms with forced turnover (0.62).

Ownership structure: The board of directors are also stockholders, thus the ownership structure should play an important role on CEO turnover. I include three variables: ownership concentration, institution ownership, and insider ownership to control the effect of ownership structure on the post merger turnover of the acquiring firm's CEO. Ownership concentration is defined as the percentage of equity held by the five largest stockholders; institution ownership is defined as the percentage of equity held by institutions; and insider ownership is the percentage of equity held by the officers and directors.

Ownership concentration. Economists generally suggest that there is a negative relation between the diffusion of ownership and the stockholders' incentive to monitor top managers' performance if the ownership structure is determined exogenously. Thus, I expect that the more concentrated the ownership, the more incentive the stockholders have to monitor the CEO, the greater the probability of post merger CEO turnover in the acquiring firm.

Furthermore, acquiring firms with more ownership concentration have a more efficient board, thus they have more desire to acquire the valuable human capital from the target firm to succeed the current CEO who is close to retirement or does not perform well. As a result, I expect that with higher ownership concentration, the CEOs of acquiring firms which acquired top officers from target firms face a higher probability of being replaced. On the other hand, with higher ownership concentration, the director acquisition should have less effect on CEO turnover since the board is more efficient.

Institution ownership. Similarly, the board has more incentive to effectively monitor top managers' performance if more equity is held by institutions (Smith, 1996). Thus, I expect that the more ownership held by institutions, the greater the probability of post merger CEO turnover in the acquiring firm.

Insider ownership. Morck, Shleifer and Vishny (1988) present evidence of the relationship between the shareholding of a company's inside directors and the firm's performance. They suggest that there are two conflicting effects of insider ownership: the positive "wealth effect" -- as the number of shares held by the insiders increases, the effect on the wealth of its members from a rise in the market value of the firm increases; and the negative "entrenchment effect" -- as the number of shares held by insiders increases, the likelihood of their being replaced through a proxy fight or takeover declines, and managers have more discretion to pursue their own goals. Therefore, the direction of insider ownership's effect on the probability of CEO turnover is ambiguous.

In Table 1-4, the mean value of ownership concentration, institution ownership and insider ownership, are 31.3%, 45.11%, and 11.12%, respectively. Ownership concentration is significantly higher for acquiring firms with forced CEO turnover than those without CEO turnover or with voluntary CEO turnover. Institution ownership is slightly higher for acquiring firms in which the CEO is forced to step down (47.94) than acquiring firms without CEO turnover (45.09) or with voluntary CEO turnover (43.35). Insider ownership is slightly higher for acquiring firms without CEO turnover (12.11) than acquiring firms in which CEOs are retired(9.71) or forced to depart (10.37), which is consistent with Denis and Sarin's (1997) results and suggests that the negative "entrenchment effect" is larger than the positive "wealth effect."

Empirical Results

Logit Estimates of CEO Turnover

I estimate several logit regression models to test my hypothesis after controlling other explanatory variables associated with CEO turnover, in which the dependent variable is the probability that the acquiring firm's CEO is replaced within five years of the merger announcement. Main independent variables I examined are officer acquisition and director acquisition that proxy the human capital acquisition from the target firm through merger activities. Other control variables includes CEO age, CEO age-squared, board size, board independence, leadership, ownership concentration, institution ownership, insider ownership, acquiring firms' performance before merger, acquiring firms' performance after merger, and target firms' performance before merger as discussed in section 2. Model 1 includes all control variables but not human capital acquisition variables--officer acquisition and director acquisition, where each corporate governance variable is included separately. In model 2, I add two human capital acquisition variables. In model 3-5 besides the independent variables included in model 2, the interaction terms of human capital acquisition and several corporate governance variables are used to capture the joint effect of corporate governance. Model 5 provides the best specification for the logit regression model, which includes two human capital acquisition variables, all control variables discussed in section 2, and the interactions of officer acquisition and director acquisition with leadership, and ownership concentration.

Table 1-6 reports the results from logit models. Each coefficient is estimated as the effect of the independent variables on post merger CEO turnover in the acquiring firm. Standard errors are shown in parentheses. The coefficient on officer acquisition is positive and significant in models 3 and 5. This evidence supports the hypothesis that CEOs of the acquiring firm who made officer acquisitions from the target firm through a merger are more likely to be replaced than CEOs who don't acquire any top executives from the target firm as new top officers in the acquiring firm through merger. For example, in model 5, if the acquiring firm makes officer acquisitions, the odds of the acquiring firm's CEO being replaced within five years after merger announcement (vs. not being replaced) increases by a factor of 5.98[11].

In contrast, the coefficient on director acquisition is negative and significant at the 0.05 level in model 5, indicating that CEOs of the acquiring firm who made director acquisition from the target firm through merger are less likely to be replaced than CEOs who don't acquire any top officers from the target firm as new board directors for the acquiring firm through merger. For example, in model 5, if the acquiring firm makes a director acquisition, the odds of the

[11] To easily interpret the effect of the dependent variable on post merger CEO turnover, I calculated the odds-ratio (OR). Given the formula for the logit regression, *prob*(CEO turnover) = $\exp(x'\beta)/(1+\exp(x'\beta))$, the odds of independent variables are obtained by exponentiate the logit coefficients, $OR=\exp(\beta)$.

acquiring firm's CEO being replaced within five years after merger announcement (vs. not being replaced) decreases by a factor of 0.11.

The coefficients on both CEO age and CEO age-squared are significant in all of the models, indicating that the probability of the acquiring firm's CEO turnover after a merger will decrease with age until a CEO reaches his/her middle age[12], at which point the probability rate of being replaced increases. No significant association exists between the likelihood of post merger turnover of the acquiring firm's CEO and both the acquiring firm and target firm's performance[13].

The coefficients on board characteristics measures, board size, board independence and leadership, are negative, but neither is significant in any models. None of the intersection terms of human capital acquisition and board characteristics is significant in any model, indicating that the relation between human capital acquisition and the likelihood of CEO turnover in the acquiring firm is unrelated to the board characteristics, i.e., acquiring firms choose their board structure either optimally or ineffectually.

The coefficient on ownership concentration is positive, while the coefficient on insider ownership is negative; both are significant. The coefficient on institution ownership is negative but not significant in any model. The evidence shows that higher ownership concentration is associated with a higher probability of CEO turnover, and higher insider ownership is associated with a lower probability of CEO turnover in the acquiring firm while institution ownership has little effect on the probability of CEO turnover. Almost all interaction terms of

[12] For example, in model 5, I calculated that the middle age for CEOs in the acquiring firm is 46 years old.

[13] I replicate the analyses by substituting various measures of accounting performance, industry-adjusted return on assets, return on average assets, and return on average equity. I find that neither before nor after merger accounting performance of acquiring firm and pre-merger accounting performance of target firm is significantly related to the likelihood of CEO turnover in the acquiring firm post merger.

human capital acquisition and ownership structure variables enter the models with insignificant coefficients. Only the interaction of director acquisition and ownership concentration is consistently positive and significant, indicating that with higher ownership concentration, the negative effect of director acquisition from the target firm on post merger CEO turnover in the acquiring firm would decrease.

Multinomial Logit Estimates of CEO Turnover Type

To examine in more detail whether the effect of human capital acquisition from the target firm is different for different types of CEO turnover in the acquiring firm after merger, I estimate the multinomial logit (MNL) model[14]. Considering three possible types of CEO subsequent turnover, this study utilizes a three state MNL model. Using "no CEO turnover" observations as the reference group, the MNL model can be described as follows:

State 0: no CEO turnover

State 1: voluntary CEO turnover

State 2: forced CEO turnover

$$prob(\text{no CEO turnover}) = \frac{1}{1+\exp(x'\beta_1 + x'\beta_2)}$$

$$prob(\text{voluntary CEO turnover}) = \frac{\exp(x'\beta_1)}{1+\exp(x'\beta_1 + x'\beta_2)}$$

$$prob(\text{forced CEO turnover}) = \frac{\exp(x'\beta_2)}{1+\exp(x'\beta_1 + x'\beta_2)}$$

where β_1 is a $k \times 1$ vector of estimated coefficients for voluntary CEO turnover observations, and β_2 is a $k \times 1$ vector of estimated coefficients for forced CEO turnover observations; x is a $k \times 1$ vector of explanatory variables which may influence the probability of voluntary CEO turnover or forced CEO turnover according to the prior empirical research on CEO turnover and merger issues, including human capital acquisition variables, the acquiring firm's CEO's characteristics, the acquiring and the target firm's characteristics, the acquiring firm's corporate governance, and year dummies.

I estimate several multinomial logit models (MNL) to study different types of CEO turnover in the acquiring firm post merger. The dependent variable is types of CEO turnover in the acquiring firm, namely, the acquiring firm's CEO has no turnover, departs voluntarily, or is forced to step down within five years of merger. Independent variables are the same as in logit estimates, which include CEO age, corporate governance and firms' performance. Model 1 includes all independent variables except two human capital acquisition variables, where each corporate governance variable is included separately. In model 2, I add two human capital acquisition variables to model 1. In models 3-5, besides the independent variables included in model 2, the interaction terms of two human capital acquisition variables and several corporate governance variables are used to capture the joint effect of corporate governance. Model 5 provides the best specification for the MNL estimates, which includes all control variables and the interactions of officer acquisition and director acquisition with board dependence and ownership concentration.

Table 1-7 contains the estimation results for models 1 to 5 for the full sample. Each coefficient shown in Table 1-7, whether for "voluntary CEO turnover" or "forced CEO turnover," is interpreted as relative to the omitted outcome "no CEO turnover," and standard error is shown in parentheses. Table 1-7 has two panels: the upper half of the table reports the coefficients for the independent variables of the likelihood of voluntary CEO turnover as relative to no CEO turnover; and the lower half reports the coefficients for the independent

[14] The multinomial probit model gives similar results, thus only the multinomial logit model is discussed in this paper.

variables of the likelihood of forced CEO turnover as relative to no CEO turnover. Overall chi-squares and log likelihood for five models are reported in Table 1-7; all indicate very strong model significance ($p<0.001$).

The coefficient on officer acquisition for voluntary CEO turnover is positive, but only significant in model 5 ($p<.05$) in which I included several interaction terms. To interpret the effect of independent variables better, similarly as in the logit model, I calculate relative risk ratio (RRR) of independent variables on the post merger CEO turnover[15]. For example, in model 5, for acquiring firms with officer acquisition relative to those without officer acquisition, the relative risk for a CEO being replaced voluntarily within five years after merger relative to a CEO who doesn't leave would be expected to increase by a factor of 11.78, given that the other variables in the model are held constant. In other words, acquiring firms who acquired top executives from the target firm as new top managers in the acquiring firm after merger are more likely to replace their CEO voluntarily over no turnover than those who didn't acquire any human capital. The coefficient on the officer acquisition for forced CEO turnover is positive and significant in almost all of the models. The only model in which the coefficient of officer acquisition is not significant is the one where no interaction terms have been included. The relative risk for a CEO being fired as opposed to a CEO who doesn't leave is expected to increase by a factor of 15.44 for acquiring firms with officer acquisition relative to those without officer acquisition, which is larger than the effect of top officer acquisition on voluntary CEO turnover. That is, acquiring firms who acquired top executives from the target firm as new top managers in the acquiring firm after merger are more likely to fire their CEO over no

[15] The RRR of a coefficient indicates how the risk of the outcome falling in the comparison group compared to the risk of the outcome falling in the reference group (in my case, “no CEO turnover” group) changes with the variable in question, which is similar to the odds-ratio in the logit model.

turnover than those who didn't acquire any human capital. The evidence provides support for hypothesis 1 which states that if the target firm's top executives are retained as top executives in the merged entity after the merger, the acquiring firm's CEO is more likely to leave, by both voluntary and forced out.

The coefficient on director acquisition is negative for both voluntary and forced CEO turnover. The negative impact of director acquisition for voluntary CEO turnover is significant in models 4 and 5 in which I included interaction terms of human capital acquisition and ownership concentration and board independence, indicating that acquiring firms who acquired top executives from the target firm as new board directors in the acquiring firm after merger are less likely to replace their CEO voluntarily over no turnover than those who didn't acquire any new top officer. The results show that for acquiring firms with director acquisition relative to those without director acquisition, the relative risk for a CEO being replaced voluntarily within five years after merger relative to a CEO who doesn't leave would be expected to increase by a factor of 0.22 in model 4 and 0.16 in model 5, given the other variables in the model are held constant. For forced CEO turnover, the coefficient on director acquisition is negative but not significant in most of models. It is only significant in model 4, in which I included interaction terms of human capital acquisition and ownership concentration. The results indicate that acquiring firms who acquired top executives from the target firm as new board directors in the acquiring firm after merger are less likely to fire their CEO over no replacement than those who didn't acquire any new director. Thus, for acquiring firms with director acquisition relative to those without director acquisition, the relative risk for a CEO being forced out within five years after merger relative to CEO who doesn't leave would be expected to increase by a factor of 0.1 as shown in model 4, given the other variables in the model are held constant. The insignificant effect of director acquisition could be explained by the optimal board structure in the acquiring firm. Hence, CEOs have little effect on the board structure; new directors from the target firm will monitor the incumbent CEO, and provide advice to strategy making teams to perform efficiently. Therefore, director acquisition through merger is an efficient decision for the acquiring firm, and it has little effect to force CEO to step down.

Table 1-7 also reports influences of other relevant variables. The coefficient for CEO age is consistently negative, and is consistently positive for CEO age squared; both are significant in all 5 models for both voluntary turnover and forced CEO turnover. The results indicate that the effect of CEO age on CEO turnover is curvilinear. In model 5, I find that the likelihood of CEO turnover decreased until the age of 45 for CEOs with voluntary turnover and 48 for CEOs with forced turnover, when it begins to increase. None of the firm's performance variables is significant for voluntary CEO turnover. For forced CEO turnover in the acquiring firm, both acquiring firm and target firm's performance around merger are insignificant; only acquiring firm's ROE before merger is slightly positive and only significant in model 5.

None of the corporate governance variables are significant for voluntary turnover of the acquiring firm's CEO, indicating that voluntary CEO turnover in the acquiring firm is unrelated to corporate governance. The coefficient on ownership concentration is positive and significant for forced CEO turnover in almost all of the models; all other corporate governance variables are not very significant for forced CEO turnover.

When the interaction between corporate governance variables and human capital acquisitions are included in the model, for voluntary CEO turnover, the coefficient for the interaction of director acquisition and ownership concentration is positive and significant ($p<0.01$), and the coefficient for the interaction of director acquisition and board independence is negative and significant ($p<0.1$). For forced CEO turnover, only the coefficient on the interaction of director acquisition and ownership concentration is positive and significant ($p<0.01$); none of the other interaction terms enter the models with significant coefficients.

Conclusion

This study adds a dimension of understanding heretofore overlooked in the literature regarding the effect of a merger on the leadership change in the acquiring firm after the merger, and suggests that one of the motivations for merger is to acquire talented human capital. Based on a sample of 236 completed mergers from 1996 through 2000, the evidence from this study is quite supportive of the view that human capital acquisition from the target firm through merger influences post merger turnover of the acquiring firm's CEO. Talented human capital from the target firm will change both top management team and board structure of the acquiring firm, and thus influence leadership change in the acquiring firm.

If the target firm's top executive is retained as top executive in the newly combined firm after merger, the acquiring firm's CEO is more likely to leave, either voluntarily or involuntarily. In contrast, if top executives of the target firm are retained as new board directors in merged entity post merger, the acquiring firm's CEO is less likely to leave voluntarily, but no change occurs in the probability of being forced out. I find no significant association between the acquiring firm's board characteristics and the probability of CEO turnover. Finally, only ownership concentration has a slightly positive effect on the relation of director acquisition and CEO turnover in acquiring firm after merger; other ownership structure variables are not related to the probability of post merger turnover of the acquiring firm's CEO.

In addition, my study complements Shleifer and Vishny’s (2003) theory of “stock market driven acquisition”. There is an additional reason why there are so many mergers. Shleifer and Vishny's theory states that the irrational overvalued acquiring firm's equity is the reason behind many acquisitions. For instance, the acquiring firm acquires firms with less-valued assets by using overvalued stock. However, this study suggests that acquiring valuable human capital is an important reason to implement merger, although such human capital acquisition through merger may affect the probability that CEOs who make mergers are replaced.

This study represents the first attempt to empirically examine the relation between human capital acquisition from the target firm through merger and post merger leadership change in the acquiring firm. Though I believe that the evidence gained is valuable, there are several limitations. First, I am unable to directly observe and examine the direct human interactions surrounding the process of human capital acquisition on post merger CEO turnover in the acquiring firm. Second, the study is appropriate mainly for mergers or acquisitions when the deal is large enough, and therefore the acquiring firm is more likely to gain human capital acquisition. This argument may be not good for small merger and acquisition activities.

Several extensions of this study could make significant contributions to the literature. First, a study of CEO successorship in the acquiring firm would be helpful. It would be useful to understand whether top executives acquired from the target firm take CEO position after a merger, and if not, who is named as heir apparent and what happens should they fail to succeed. Second, we could better understand the process of leadership change associated with human capital acquisition from the target firm by examining turnover of lower-level top management i.e. president, chairman of the board, vice presidents, CFO and COO. These issues are beyond the scope of this study, but they may provide advance understanding of the effect of a merger on the leadership changes within the acquiring firm.

Table 1-1. Sample distribution and frequency of the CEO turnover

Panel A. Distribution by Turnover Type

Type	No. of M&As	% of sample	% of subsample
Total sample	236	100	
CEO Turnover	109	46	100
Voluntary CEO Turnover	67	28	61
Forced CEO Turnover	42	18	39
No CEO Turnover	127	54	

Panel B. Distribution by Year

Year	No. of M&As	No. of turnover (%)	No. of no turnover (%)	No. of voluntary turnover (%)	No. of forced turnover (%)
1996	37	20 (54%)	17 (46%)	15 (41%)	5 (13%)
1997	47	17 (36%)	30 (64%)	11 (23.53%)	6 (13%)
1998	55	21 (38%)	34 (62%)	13 (23.44%)	8 (15%)
1999	59	33 (56%)	26 (44%)	16 (25.76%)	17 (29%)
2000	38	18 (47%)	20 (53%)	12 (30.95%)	6 (16%)

Note: The sample consists of 236 completed M&As between two public companies during the 1996-2000 period. This table reports the frequency of the CEO turnover across different years and types.

Table 1-2. Descriptive statistics for human capital acquisition

		Officer Acquisition	Director Acquisition
Total Sample	Mean	0.2839	0.3475
	Std. Dev.	0.4518	0.4772
No CEO Turnover	Mean	0.2441	0.3071
	Std. Dev.	0.4313	0.4631
CEO Turnover	Mean	0.3303	0.3945
	Std. Dev.	0.4725	0.4910
Voluntary CEO Turnover	Mean	0.2836	0.3881
	Std. Dev.	0.4541	0.4910
Forced CEO Turnover	Mean	0.4048	0.4048
	Std. Dev.	0.4968	0.4968

Table 1-3. Variable definitions

Data from SDC M&A database, COMPUSTAT and SEC database from 1996 to 2000

Dependent Variable

CEO turnover: 0 if no CEO turnover in the acquiring firm, 1 if voluntary CEO turnover in the acquiring firm, 2 if forced CEO turnover in the acquiring firm.

Independent Variables

Human capital acquisition:

Officer acquisition: 1 if there is one or more top managers of the target firm have retained positions in the merged entity and reported as top management executive after merger effected, and 0 otherwise.

Director acquisition: 1 if there is one or more top managers of the target firm became a director in the merged entity board after merger effected, and 0 otherwise.

Other Control variables:

CEO age: age in years.

CEO age-squared: age^2

Board size: the number of directors reported in board.

Board independence: the percentage of inside directors on the acquiring firm's board at the year of the merger announcement.

CEO/Chairman: 1 if the CEO of the acquiring firm also serves as chairman of the board, 0 otherwise.

Ownership concentration: the percentage of equity held by the five largest blockholders.

Insider ownership: the percentage of equity held by officers and directors.

Institutions ownership: the percentage of equity held by institutions.

Pre-ROE of acquirer: the industry-adjusted yearly return on the common equity of the acquiring firm in the year before the merger announcement.

Post-ROE of acquirer: the industry-adjusted yearly return on the common equity of the acquiring firm in the year right after the merger announcement.

Pre-ROE of target: the industry-adjusted yearly return on the common equity of the target firm in the year before the merger announcement.

Interaction terms (with officer acquisition or director acquisition)

Officer acquisition× Board independence

Officer acquisition× CEO/Chairman

Officer acquisition× Ownership concentration

Director acquisition× Board independence

Director acquisition× CEO/Chairman

Director acquisition× Ownership concentration

Table 1-4. Explanatory variable descriptive statistics for the total sample

Variable	Mean	Std. Dev.	Median	Min	Max
CEO age	53.8559	7.6289	54	37	83
Board size	11.1229	4.5455	10	2	28
Board independence	0.2701	0.1751	0.25	0	1
CEO/Chairman	0.6864	0.4649	1	0	1
Ownership concentration	31.2995	28.1485	24.34	0	99.99
Institution ownership	45.1051	27.2018	48.92	0	99.86
Insider ownership	11.1199	16.4384	4.125	0	99.99
Pre-ROE of acquirer	5.5724	63.7349	3.16	-146.435	908.5188
Post-ROE of acquirer	7.7055	133.8818	1.874	-427.347	1858.282
Pre-ROE of target	-28.7566	402.4704	0	-6170.87	195.7014

Table 1-5. Correlations

	Turnover	Turnover type	Officer acquisition	Director acquisition	CEO age	Board size	Board independence	CEO/ Chairman
Turnover	1							
Turnover type	0.9019*	1						
Officer acquisition	0.0953	0.1243*	1					
Director acquisition	0.0915	0.0876	0.5274*	1				
CEO age	0.3197*	0.1932*	-0.0189	0.0734	1			
Board size	-0.0607	-0.0824	-0.0306	0.1568*	0.0408	1		
Board independence	0.0321	0.0192	0.1156*	-0.0057	0.0746	-0.3108*	1	
CEO/Chairman	0.0399	-0.0078	0.1015	0.0520	0.1799*	-0.0481	-0.0428	1
Ownership concentration	0.1384*	0.1987*	0.797	0.1046	0.1165*	-0.2182*	0.0798	0.0083
Institution ownership	0.0005	0.0246	0.1067	0.0310	0.0125	-0.0580	-0.0605	0.2476*
Insider ownership	-0.0653	-0.0531	-0.0884	-0.0623	0.0858	-0.1740*	0.0794	-0.0698
Pre-ROE of acquirer	0.0550	0.0807	0.1174*	0.0727	0.0275	0.0084	0.1550*	-0.1275*
Post-ROE of acquirer	0.0864	0.1342*	-0.0098	-0.0456	0.0739	-0.0241	0.0183	-0.0442
Pre-ROE of target	0.0607	0.0520	0.0444	0.0449	0.0351	-0.1246*	0.0612	-0.0467

	Ownership concentration	Institution ownership	Pre-ROE of acquirer	Post-ROE of acquirer	Pre-ROE of target
Ownership concentration	1				
Institution ownership	0.1879*	1			
Insider ownership	0.3476*	-0.1075*			
Pre-ROE of acquirer	-0.0142	0.0844	1		
Post-ROE of acquirer	0.0226	0.0155	0.0479	1	
Pre-ROE of target	0.0614	0.1177*	0.0151	0.0158	1

Table 1-6. Results of Logit Regression Model

	1	2	3	4	5
Officer acquisition		0.4745	1.4689*	1.0472	1.7878*
		(0.3935)	(0.8488)	(0.6791)	(1.0279)
Director acquisition		-0.0589	-0.7411	-1.6849**	-2.226**
		(0.3758)	(0.7586)	(0.6609)	(0.9726)
CEO age	-0.7902**	-0.8051**	-0.7723**	-0.7618**	-0.732**
	(0.3121)	(0.3144)	(0.3189)	(0.3258)	(0.3307)
CEO age^2	0.0084***	0.0085***	0.0083***	0.0082***	0.0080**
	(0.0030)	(0.0030)	(0.0030)	(0.0031)	(0.0031)
Board size	-0.0105	-0.0122	-0.0226	-0.0099	-0.0176
	(0.0364)	(0.0373)	(0.0385)	(0.0392)	(0.0403)
Board independence	-0.3124	-0.4347	-0.4971	-0.6756	-0.6877
	(0.9325)	(0.9405)	(0.9523)	(0.9621)	(0.9714)
CEO/Chairman	-0.0631	-0.1266	-0.1144	-0.0542	-0.0477
	(0.3345)	(0.3395)	(0.4185)	(0.3495)	(0.4207)
Ownership concentration	0.0204***	0.0195***	0.0198***	0.0094	0.0091
	(0.0063)	(0.0064)	(0.0064)	(0.0081)	(0.0081)
Institution ownership	-0.0046	-0.0052	-0.0051	-0.0063	-0.006
	(0.0062)	(0.0062)	(0.0063)	(0.0065)	(0.0066)
Insider ownership	-0.0215**	-0.0207*	-0.0219**	-0.023**	-0.0228*
	(0.0108)	(0.0108)	(0.0110)	(0.0117)	(0.0117)
Pre-ROE of acquirer	0.0029	0.0024	0.0023	0.0023	0.0021
	(0.0026)	(0.0026)	(0.0028)	(0.0026)	(0.0027)
Post-ROE of acquirer	0.0018	0.0018	0.0019	0.0018	0.002
	(0.0021)	(0.0020)	(0.0021)	(0.0020)	(0.0021)
Pre-ROE of target	0.001	0.0008	0.0006	0.0008	0.0007
	(0.0037)	(0.0027)	(0.0019)	(0.0023)	(0.0018)
Officer acquisition × CEO/Chairman			-1.2765		-1.0625
			(0.9624)		(1.0623)
Director acquisition × CEO/Chairman			0.8778		0.8041
			(0.8683)		(0.9824)
Officer acquisition × ownership concentration				-0.0194	-0.0162
				(0.0171)	(0.0169)
Director acquisition × ownership concentration				0.0508***	0.048***
				(0.0169)	(0.0166)
Constant	17.896**	18.303**	17.479**	17.242**	16.46*
	(8.1066)	(8.1704)	(8.2943)	(8.4658)	(8.6012)
Model Chi2	57.07***	58.80***	60.69***	70.77***	71.86***
Log Likelihood	-134.361	-133.497	-132.55	-127.51	-126.968

*, ** and *** indicate significance at 10%, 5% and 1% levels, respectively.

Table 1-7. Results of multinomial logistic regression model

	1	2	3	4	5
Equation 1: Voluntary CEO Turnover					
Officer acquisition		0.1853	0.2589	0.7728	2.4661**
		(0.4656)	(1.1125)	(0.7709)	(1.1953)
Director acquisition		0.0442	-0.3388	-1.5131**	-1.8500*
		(0.4356)	(0.8818)	(0.7275)	(1.0973)
CEO age	-0.7894**	-0.8079**	-0.8262**	-0.7647**	-0.7646**
	(0.3507)	(0.3507)	(0.3569)	(0.3625)	(0.3712)
CEO age-squared	0.0086***	0.0088***	0.009***	0.0085**	0.0085**
	(0.0033)	(0.0033)	(0.0034)	(0.0034)	(0.0035)
Board size	0.0086	0.0074	0.0063	0.0119	0.0105
	(0.0416)	(0.0426)	(0.0442)	(0.0442)	(0.0449)
Board independence	0.0818	0.0279	-0.0427	-0.1772	1.3122
	(1.0561)	(1.0696)	(1.0820)	(1.0879)	(1.4145)
CEO/Chairman	0.1478	0.1173	-0.034	0.174	0.2898
	(0.3972)	(0.4014)	(0.4861)	(0.4095)	(0.4174)
Ownership concentration	0.0146*	0.0138*	0.0146*	0.0034	-0.0004
	(0.0075)	(0.0076)	(0.0076)	(0.0097)	(0.0098)
Institution ownership	-0.0083	-0.0085	-0.0084	-0.0101	-0.0103
	(0.0071)	(0.0072)	(0.0072)	(0.0074)	(0.0077)
Insider ownership	-0.0178	-0.0172	-0.017	-0.0194	-0.0175
	(0.0123)	(0.0124)	(0.0124)	(0.0132)	(0.0130)
Pre-ROE of acquirer	0.0014	0.0012	0.0015	0.001	0.0024
	(0.0038)	(0.0038)	(0.0041)	(0.0038)	(0.0039)
Post-ROE of acquirer	-0.0004	-0.0005	-0.0005	-0.0009	-0.001
	(0.0032)	(0.0033)	(0.0033)	(0.0036)	(0.0037)
Pre-ROE of target	0.0102	0.0106	0.0135	0.0114	0.012
	(0.0072)	(0.0074)	(0.0083)	(0.0076)	(0.0077)
Officer acquisition ×CEO/Chairman			-0.1134		
			(1.2291)		
Director acquisition ×CEO/Chairman			0.5326		
			(1.0026)		
Officer acquisition × ownership concentration				-0.0215	-0.0236
				(0.0199)	(0.0203)
Director acquisition × ownership concentration				0.0528***	0.0590***
				(0.0192)	(0.0198)

Table 1-7. Continued

Officer acquisition × board independence					1.3008
					(3.4754)
Director acquisition × board independence					-6.3018*
					(3.6288)
Constant	16.6395*	17.1595*	17.6774*	16.1104*	15.8288
	(9.2408)	(9.2400)	(9.3901)	(9.5460)	(9.7505)
Equation 2: Forced CEO Turnover					
Officer acquisition		0.7799	2.5457**	1.6189*	2.7367**
		(0.5088)	(1.0997)	(0.8953)	(1.2689)
Director acquisition		-0.2155	-1.379	-2.2698**	-1.1546
		(0.4977)	(1.0632)	(0.9460)	(1.2561)
CEO age	-0.6853*	-0.7244**	-0.6596*	-0.6774*	-0.6770*
	(0.3629)	(0.3659)	(0.3760)	(0.3797)	(0.3874)
CEO age-squared	0.0071**	0.0075**	0.0069*	0.0071**	0.0071*
	(0.0034)	(0.0035)	(0.0036)	(0.0036)	(0.0037)
Board size	-0.0454	-0.047	-0.0672	-0.0497	-0.0748
	(0.0531)	(0.0535)	(0.0563)	(0.0572)	(0.0607)
Board independence	-1.0406	-1.1603	-1.2324	-1.6118	0.6635
	(1.3438)	(1.3544)	(1.3956)	(1.4019)	(1.6575)
CEO/Chairman	-0.329	-0.4341	-0.2348	-0.3645	-0.3392
	(0.4384)	(0.4470)	(0.5789)	(0.4588)	(0.4691)
Ownership concentration	0.0270***	0.0263***	0.0279***	0.0172*	0.0143
	(0.0076)	(0.0078)	(0.0080)	(0.0102)	(0.0103)
Institution ownership	0.0002	-0.0012	-0.0012	-0.0017	-0.0015
	(0.0079)	(0.0080)	(0.0082)	(0.0084)	(0.0086)
Insider ownership	-0.0223	-0.021	-0.0249*	-0.0244*	-0.0233
	(0.0139)	(0.0139)	(0.0145)	(0.0147)	(0.0146)
Pre-ROE of acquirer	0.0037	0.0032	0.0029	0.0032	0.0069**
	(0.0027)	(0.0027)	(0.0028)	(0.0027)	(0.0032)
Post-ROE of acquirer	0.0034	0.0031	0.0038	0.003	0.0028
	(0.0031)	(0.0031)	(0.0033)	(0.0031)	(0.0031)
Pre-ROE of target	0.0002	0.0001	0.0001	0.0002	0.0001
	(0.0011)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Officer acquisition ×CEO/Chairman			-2.3608*		
			(1.2604)		

Table 1-7. Continued

Director acquisition ×CEO/Chairman			1.4673		
			(1.2005)		
Officer acquisition × ownership concentration				-0.0255	-0.0267
				(0.0200)	(0.0206)
Director acquisition × ownership concentration				0.0575***	0.0710***
				(0.0207)	(0.0223)
Officer acquisition × board independence					-6.4915
					(4.0871)
Director acquisition × board independence					-3.6506
					(3.1609)
Constant	15.314	16.3031*	14.6181	15.2263	15.1151
	(9.5271)	(9.6190)	(9.8955)	(9.9759)	(10.1600)
Pseudo R-Square	3.173***	3.179***	3.188***	3.207***	3.227***
Model Chi2	81.768	84.348	88.792	97.347	106.853
Log likelihood	-194.672	-193.382	-190.532	-186.882	-182.129

*, ** and *** indicate significance at 10%, 5% and 1% levels, respectively.

CHAPTER 2
ESTIMATION OF CENSORED REGRESSION MODEL: A SIMULATION STUDY

Introduction

Censoring is common in econometric applications. Censoring usually occurs for two reasons. First, censoring is the result of individual rational behavior subject to a non-negativity constraint. For instance, in empirical analysis of individual labor supply, hours worked are positive if and only if the individual chooses to participate in the labor force (e.g. Heckman, 1979, 1980). Likewise, the observed consumption of meat, for example, is positive if and only if the individual chooses to consume meat. Firm's R&D expenditure is positive if and only if the firm engages in R&D activity. Second, censoring is the result of survey design. For example, in data collection, it is common practice to top and/or bottom code the income variable. In this case, the empirical analysis of income requires a model to deal with censoring data problem (see Solon, 1992, 1999; Zimmerman, 1992; Ashenfelter and Zimmerman, 1997, for application).

A popular model to deal with censoring data problem is Tobit model. The standard approach for estimating the Tobit model is maximum likelihood estimation. For panel data with censoring data problem, a natural choice of models is the panel data Tobit model with individual effect. However, estimation of the panel data Tobit model, when the individual effect is allowed to correlate with the explanatory variables arbitrarily, is nontrivial and difficult. Under some conditions, Honoré (1992) derives a conditional moment restriction which requires discarding part of observations. Based on this conditional moment restriction, Honoré proposes a consistent and asymptotically normally distributed estimator. His estimator, however, is not efficient since it does not use all moment restrictions implied by the conditional moment restriction. To increase the efficiency of the estimates, one could exploit other moment restrictions by applying the two-step GMM, or the continuously updating GMM, or the empirical likelihood estimator (ELE). Asymptotically, the latter three estimators are equivalent and are more efficient, at least not worse, than the Honoré estimator.

The main objective of this paper is to study two issues. First, I consider the efficiency of these estimators in large samples. It is interesting to see how much efficiency can be gained by the last three estimators relative to the Honoré estimator when sample size is moderate or large. Second, we argue the small sample properties of all four estimators, particularly the relative performance of the Honoré estimator vs. the other three estimators.

The remainder of the paper is organized as follows. Section 2 sets up the model and describes the four estimators. The virtues and drawbacks of the estimators are discussed. Section 3 studies the finite sample performances of these estimators via a Monte Carlo study and discusses the Monte Carlo results. Section 4 concludes.

Model and Estimators

The standard model for censored dependent variable is Tobit model proposed by Tobin (1956). The standard approach for extracting the rich information in panel data is to use the individual specific effect (see Hsiao, 1986). Thus, a natural model for analyzing a panel dataset containing censored dependent variables is the panel data Tobit regression model with individual effects given by

$$\begin{aligned} y_{it}^* &= \alpha_i + x_{it}'\beta_0 + \varepsilon_{it}, \\ y_{it} &= \max\left\{0, y_{it}^*\right\}, i = 1, 2, ..., N; t = 1, 2, ..., T \end{aligned} \qquad (2\text{-}1)$$

where i denotes the individual, t denotes time; y_{it}^* denotes the latent dependent variable; y_{it} denotes the observed dependent variable; x_{it} denotes the k-dimensional column-vector of explanatory variables; α_i denotes the unobserved individual specific effect; β_0 denotes the true value of the unknown parameter (column-) vector to be estimated; and ε_{it} denotes the error term. The error term ε_{it} is usually assumed to be normally distributed; and in this case the model is called type I Tobit model (see Amemiya, 1985, for other types of Tobit models).

Estimation of Tobit model, particularly the version of model (2-1), can be difficult, depending on the assumptions imposed on the error term ε_{it} and the individual effect α_i. If we assume the individual effect in model (2-1) is random, where both the error term and the individual effect are assumed to be uncorrelated with the explanatory variables and normally distributed, then the standard maximum likelihood estimator (MLE) of model (2-1) is consistent and asymptotically normal.

The normally distributed error term, though commonly imposed in empirical analysis, is difficult to justify. Moreover, the assumption that the individual effect is independent of the explanatory variables is equally difficult to justify. Without independence between the individual effect and the explanatory variables, estimation of model (2-1), requires using the fixed effect Tobit model, which is difficult even if the normality assumption on the error term is maintained. Without both the independence between the individual effect and the explanatory variables and the parametric specification of the error term density, estimation of model (2-1) is even more difficult.

The difficulty arises because the individual effect α_i enters the model nonlinearly and the simple time-differencing approach widely used in linear panel data models does not work here. To see this, notice that, for any period t, at the true value β_0, the residual value is

$$y_{it} - x_{it}'\beta_0 = \max\left\{y_{it}^* - x_{it}'\beta_0, -x_{it}'\beta_0\right\} = \max\left\{\alpha_i + \varepsilon_{it}, -x_{it}'\beta_0\right\} \tag{2-2}$$

which clearly is censored at $-x_{it}'\beta_0$. Thus, for any two periods t and s, at true value β_0, applying simple time differencing, we obtain:

$$\left(y_{it} - x_{it}'\beta_0\right) - \left(y_{is} - x_{is}'\beta_0\right) = \max\left\{\alpha_i + \varepsilon_{it}, -x_{it}'\beta_0\right\} - \max\left\{\alpha_i + \varepsilon_{is}, -x_{is}'\beta_0\right\} \tag{2-3}$$

The unobserved individual effect α_i is obviously not removed here. But this does not necessarily mean that standard regression techniques do not yield consistent estimates for the model parameter. Standard regression techniques still produce a consistent estimate if the right hand side of (2-3) is uncorrelated with the explanatory variables or some instrumental variables[1]. Unfortunately, the right hand side of (2-3) is correlated with the explanatory variables and it is hard to find instrumental variables.

To overcome this problem, Honoré (1992) suggests a clever transformation of the model so that the explanatory variables are uncorrelated with the time-differenced residuals. The intuition of his approach is that the linearity of the model holds for those individuals whose dependent variables are not censored in both periods and are not far apart. Applying simple time differencing to those individuals would preserve the zero correlation between the explanatory variables and the residuals. Specifically, define the artificially censored residuals as

$$e_t(y_{it} - x_{it}'\beta_0) = \max\{y_{it} - x_{it}'\beta_0, -x_{is}'\beta_0\} = \max\{\alpha_i + \varepsilon_{it}, -x_{it}'\beta_0, -x_{is}'\beta_0\}$$

$$e_s(y_{is} - x_{is}'\beta_0) = \max\{y_{is} - x_{is}'\beta_0, -x_{it}'\beta_0\} = \max\{\alpha_i + \varepsilon_{is}, -x_{is}'\beta_0, -x_{it}'\beta_0\}$$

The following assumption is imposed:

- *Assumption: The error terms ε_{it} and ε_{is}, conditional on $(x_{it}, x_{is}, \alpha_i)$, are identically distributed.*

Under the above assumption, it is easy to show that

$$E\{q(e_t(y_{it} - x_{it}'\beta_0)) - q(e_s(y_{is} - x_{is}'\beta_0)) | x_{it}, x_{is}\} = 0 \tag{2-4}$$

holds for any function $q(\cdot)$. To see exactly which observations are used to determine the parameters, denotes $\Delta x_i = x_{it} - x_{is}$. For the sake of arguments, suppose that $\Delta x_i{}'\beta_0 \geq 0$. Consider four cases: (i) the dependent variable is not censored in both periods; (ii) the dependent variable is censored in both periods; (iii) the dependent variable is censored in period t, but not in period s; and (iv) the dependent variable is not censored in period t but censored in period s.

[1] Therefore, the conditional mean given the explanatory variables is zero.

Case (i): $y_{it} > 0, y_{is} > 0$.

In this case, we have

$$e(y_{is} - x_{is}'\beta_0) = y_{is} - x_{is}'\beta_0$$

$$e(y_{it} - x_{it}'\beta_0) = \begin{cases} -x_{is}'\beta_0, \text{ if } y_{it} \le x_{it}'\beta_0 \\ y_{it} - x_{it}'\beta_0, \text{ if } y_{it} > x_{it}'\beta_0 \end{cases}.$$

So trimming occurs if $y_{it} \le \Delta x_i'\beta_0$. Regardless of whether trimming is used or not, observations in this case are used to determine the model parameters.

Case (ii): $y_{it} = 0, y_{is} = 0$.

In this case, it is straightforward to show that

$$e(y_{it} - x_{it}'\beta_0) = e(y_{is} - x_{is}'\beta_0)$$

Hence, observations like these are not used in conditional moment restrictions (2-4) to determine the model parameters.

Case (iii): $y_{it} = 0, y_{is} > 0$.

In this case, we have

$$e(y_{it} - x_{it}'\beta_0) = -x_{is}'\beta_0$$

$$e(y_{is} - x_{is}'\beta_0) = y_{is} - x_{is}'\beta_0$$

Observations like these are used to determine the model parameters.

Case (iv): $y_{it} > 0, y_{is} = 0$

In this case, we have

$$e(y_{is} - x_{is}'\beta_0) = -x_{it}'\beta_0$$

$$e(y_{it} - x_{it}'\beta_0) = \begin{cases} -x_{is}'\beta_0, \text{ if } y_{it} \le x_{it}'\beta_0 \\ y_{it} - x_{it}'\beta_0, \text{ if } y_{it} > x_{it}'\beta_0 \end{cases}.$$

Clearly, trimming occurs when $y_{it} \le \Delta x_i'\beta_0$. When trimming occurs, the observation cannot be used to determine the model parameters.

To summarize, the conditional moment functions (2-4) only uses observations where the dependent variable is not censored in both periods as well as observations where the dependent variable is censored in one period but not in the other period and where no trimming occurs. Figure 2-1 and Figure 2-2 show the set of observations that has been discarded in estimation.

Now, we present several estimators discussed in the literature which use conditional moment restrictions (2-4) to obtain consistent estimations for panel data Tobit model with fixed effect.

Honoré Estimation

Based on the conditional moment restrictions (2-4), Honoré (1992) proposes four estimators. In this study, we focus on the last estimator, which has a convex objective function[2] and the first order condition that coincides with some unconditional moment restrictions (which given in (2.5) in Honoré, 1992). By using his setting for the case of two time periods (i.e., $s=1, t=2$), we have the following estimator:

$$\hat{\beta}^H = \arg\min_{\beta} \sum_{i=1}^{N} \sum_{s<t} \gamma(y_{it},\ y_{is},\ \Delta x_i'\beta) \qquad (2\text{-}5)$$

where N is the number of observations, and

$$\delta = \Delta x_i'\beta, \quad \gamma(y_{it},\ y_{is},\ \delta) = \begin{cases} \tfrac{1}{2} y_1^2 + \delta y_1 - y_1 y_2, \text{ if } \delta \le -y_2 \\ \tfrac{1}{2}(y_2 + \delta - y_1)^2, \text{ if } -y_2 < \delta \le y_1 \\ \tfrac{1}{2} y_2^2 - \delta y_2 - y_1 y_2, \text{ if } y_2 < \delta \end{cases}$$

Under sufficient conditions, Honoré shows that $\hat{\beta}^H$ is consistent and asymptotically normally distributed:

$$\sqrt{N}(\hat{\beta}^H - \beta_0) \xrightarrow{\text{d}} N(0, \Gamma^{-1} V \Gamma^{-1})$$

where

[2]Thus it is straightforward to derive the limit distribution of the estimator. The corresponding conditional moment restriction will be referred as "smooth" conditional moment restriction.

$$\Gamma = \tfrac{1}{N}\sum_{i=1}^{N} 1\left\{-y_{i2} < \Delta x_i'\hat{\beta}^H < y_{i1}\right\}\Delta x \Delta x',$$

$$V = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} y_{i2}^2 1\left\{y_{i1} < \Delta x_i'\hat{\hat{\beta}}^H\right\} + y_{i1}^2 1\left\{\Delta x_i'\beta^H < -y_{i2}\right\} \\ +(y_{i1} + y_{i2} - \Delta x_i'\hat{\hat{\beta}}_0)^2 1\left\{-y_{i2} < \Delta x_i'\beta^H < y_{i1}\right\}\end{bmatrix}\Delta x \Delta x'.$$

GMM Estimation

Noting that the Honoré estimator does not use all information implied by conditional moment restrictions (2-4), Ai and Li (2006) suggest that other moment restrictions implied by conditional moment restrictions (2-4) can be used to increase efficiency. Specifically, their idea is to use a series of basis functions[3] to approximate the arbitrary function $q(\cdot)$. For some integer k_1, let

$$q(z) = (q_1(z), q_2(z), .\;.\, q_{k_1}(z))'$$

denote known basis functions that approximate any square integrable function of z, where $z = e(y_{it} - x_{it}{}'\beta_0),\ t = 1, 2, ..., T$. For some integer k_2, let

$$p(x_i) = (p_1(x_{it}, x_{is}), p_2(x_{it}, x_{is}), ..., p_{k_2}(x_{it}, x_{is}))'$$

denote known basis functions that approximate any square integrable function of x_i, where $x_i = (x_{it}, x_{is})$.

Condition (2-4) implies the following unconditional moment restrictions:

$$E\left\{q(e(y_{it} - x_{it}'\beta_0)) - q(e(y_{is} - x_{is}'\beta_0)) \otimes p(x_i)\right\} = 0 \qquad (2\text{-}6)$$

where $\otimes$ denotes the Kronecker product. Obviously, equation (2-6) includes $k_1 \times k_2$ moment functions.

Let $g(y_i, x_i, \beta)$ denote the $k_1 \times k_2$-dimensional column-vector of moment functions in

[3]Alternatively, we could approximate the conditional mean functions (2-4) by using the fitted value from the regression $q(z)$ on $p(x_i)$. More details are presented in Monte Carlo study.

(2-6):

$$g(y_i, x_i, \beta) = vec\left\{q(e(y_{it} - x_{it}'\beta_0)) - q(e(y_{is} - x_{is}'\beta_0)) \otimes p(x_i)\right\}$$

where $s = 1, 2, ..., T-1$,and $t = s+1, ..., T$. Denote

$$\hat{g}(\beta) = \tfrac{1}{N}\sum_{i=1}^{N} g(y_i, x_i, \beta),\ \hat{\Omega}(\beta) = \tfrac{1}{N}\sum_{i=1}^{N} g(y_i, x_i, \beta) g(y_i, x_i, \beta)'$$

Also, let $\tilde{\beta}$ be the initial estimator obtained by

$$\tilde{\beta} = \arg\min_{\beta} \hat{g}(\beta)'\hat{W}^{-1}\hat{g}(\beta),$$

where $\hat{W}$ is a random weighting matrix, and it is positive definite.

Hansen's (1982) best 2-step GMM estimator is

$$\tilde{\beta}^{GMM} = \arg\min_{\beta} \hat{g}(\beta)'\hat{\Omega}^{-1}(\tilde{\beta})\hat{g}(\beta) \qquad (2\text{-}7)$$

Altonji and Segal (1996), however, document that, although Hansen's 2-step GMM estimator has desirable large sample properties, it has poor finite sample performance. Hansen, Heaton and Yaron (1996) propose a continuously updated best GMM estimator which has smaller bias than 2-step GMM,

$$\tilde{\beta}^{update} = \arg\min_{\beta} \hat{g}(\beta)'\hat{\Omega}^{-1}(\beta)\hat{g}(\beta) \qquad (2\text{-}8)$$

The continuously updating GMM is analogous to the 2-step GMM except that the criterion function in (2-8) is simultaneously minimized over β in both $\hat{\Omega}^{-1}(\beta)$ and $g(\beta)$. Newey and Smith (2004) show that the continuously updating GMM has better finite sample properties.

Empirical Likelihood Estimation

Alternatively, the empirical likelihood approach could be applied to conditional moment restrictions (2-4) (Qin and Lawless, 1994; Kitamura, Tripathi and Ahn, 2004; Donald, Imbens, and Newey, 2004). The idea here is to treat the joint density $f(y, x)$ as unknown and estimate it together with the model parameter β_0. Specifically, let π_i denote the probability of (y_i, x_i). Then the model parameter can be estimated by the following constrained maximum likelihood estimation:

$$\begin{aligned} \hat{\beta}^{EL,1} &= \arg\max_{\beta,\pi} \sum_{i=1}^{N} \ln \pi_i \\ \text{subject to } &\sum_{i=1}^{N} \pi_i = 1,\ \sum_{i=1}^{N} \pi_i g(y_i, x_i, \beta) = 0,\ \pi_{\mathrm{i}} \geq 0. \end{aligned} \tag{2-9}$$

Since the constrained optimization (2-9) is hard to implement, the estimator can be difficult to compute. Notice that by applying the Lagrange approach, it is straightforward to show that the maximum likelihood estimator solves:

$$\hat{\beta}^{EL} = \arg\min_{\beta \in \mathrm{B}} \max_{\lambda \in \hat{\Lambda}(\beta)} \sum_{i=1}^{N} \ln(1 + \lambda' g(y_i, x_i, \beta)) \tag{2-10}$$

where λ is a vector of Lagrange multipliers. $\hat{\beta}^{EL}$ is the empirical likelihood estimator. Newey and Smith (2004) also show that for quadratic $p(x_i)$, $\tilde{\beta}^{update}$ is identical to $\hat{\beta}^{EL,1}$ in their theorem 2.1.

The five estimators presented above, though all derived from the same conditional moment restrictions (2-4), have different computational advantages and disadvantages. Honoré's estimator is the easiest to compute since his criterion function is globally convex. Hansen's best GMM estimator has desirable large sample properties, but its finite sample performance is poor. The continuously updating GMM estimator is more difficult to compute since the criterion function is not globally convex. Moreover, the weighting matrix can be singular for certain values of parameters, especially when sample size is small relative to the number of moment functions in the conditional moment restrictions. When this happens, the iteration process that searches for the estimator would stop. The empirical likelihood estimator $\hat{\beta}^{EL,1}$ is also difficult to compute because the constraints may not be satisfied by all parameters. The advantage is that the implied probability π_{i} is also estimated. The $\hat{\beta}^{EL}$ version of the empirical likelihood estimator is relatively easier to compute since it does not force the constraints to be satisfied. The advantage of the last three estimators is that they are asymptotically equivalent and superior to the Honoré estimator up to the second order (See Newey and Smith, 2004). The disadvantage is that all three estimators are hard to compute.

Monte Carlo Experiments

To analyze the interesting issues mentioned above, a Monte Carlo study is conducted. In all the four designs, I consider two time periods (i.e. $T = 2$) and include two explanatory variables. For each individual and each period, denote $x_{it} = (x_{1it}, x_{2it})'$, where $t = 1, 2$, with $x_{1it} = \alpha_i + \eta_{it}$ [4]. The true value of the parameter is $\beta_0 = (1,1)'$. All results presented in this paper are from 1000 replications of model (2-1).

Design 1

In design 1, the random variables $\alpha_i, \eta_{i1}, \eta_{i2}$, and x_{2it} ($t = 1, 2$) are independent of each other and are drawn from the standardized chi-squared distribution[5] of degree of freedom three (with mean zero and variance one)[6]. Conditional on α_i, the error terms ε_{1i} and ε_{2i}, are independent and follow the normal distribution $N(0, \frac{1}{2} + \frac{1}{2}\alpha_i^2)$ [7].

Implementing GMM and EL estimators requires the specified form of the moment functions. In the simulation study, I choose the polynomial basis functions which are given by

$$q(z) = (z, z^2, z^3, ...)'_{k_1 \times 1}$$

and

[4] Allow for correlation between covariates and individual effect would give more general results.

[5] Normally distributed explanatory variables may give too optimistic conclusions from Monte Carlo's (Chesher, 1995).

[6] Data is generated by Matlab 7.0.4. By using "seed=11", we assure that all 1000 replications are the same for different sample size N and (k_1, k_2), so that it is meaningful to compare the results.

[7] Without loss of generality, we allow the heteroskedasticity by letting the variance of the error terms depends on the fixed effect.

$$p(x_i) = (x_{1i2} - x_{1i1}, x_{2i2} - x_{2i1}, x_{1i2}, x_{2i2}, 1, x_{1i2} - x_{1i1}, \ldots)'_{k_2 \times 1}$$

where $q(z)$ is a k_1-dimensional column-vector of polynomial basis functions of z, and $p(x_i)$ is a k_2-dimensional column-vector of polynomial basis functions of x_i. For instance, when $k_1 = 1$ and $k_2 = 2$, the unconditional moment restrictions (2-5) can be written as

$$E[g(y_i, x_i, \beta)] = E\{[e_2(y_{i2} - x_{i2}'\beta_0) - e_1(y_{i1} - x_{i1}'\beta_0)] \otimes (x_{i2} - x_{i1})'\} = 0 \qquad (2\text{-}11)$$

In the appendix, I show that when $k_1 = 1$ and $k_2 = 2$, the moment functions (2-6) are the same moment functions used in the Honoré model[8]. For other k_1 and k_2, the moment functions are constructed analogously. By changing (k_1, k_2) different numbers of conditional moment functions can be used to estimate the model.

To evaluate the small sample properties, we consider N=200. For each sample we compute estimators when (k_1, k_2) takes on the following values: (1, 2), (1, 5), (1, 9), (1, 15), (2, 5), (2, 9), (2, 15), (3, 5), (3, 9), (3, 15)[9]. To evaluate efficiency in the large sample, we consider N=500 and N=1000[10].

Design 2

Design 2 differs from design 1 only in the distribution of the regressors. The random variables $\alpha_i, \eta_{i1}, \eta_{i2}$, and x_{2it} ($t = 1, 2$) are independent from each other and are drawn from the standardized chi-squared distribution of degree of freedom one (with mean zero and

[8] In order to compare with the Honoré estimator, $g(y_i, x_i, \beta)$ used in the continuously updating GMM should include the moment functions used in the Honoré model. Based on Honoré's design, I show in appendix A that $k_1 = 1$ and $k_2 = 2$ gives exactly the same moment functions as those of the Honoré model.

[9] Considering the sensitivity to the choice of polynomial, we use only second order self-power series for $k_2 = 9$, and all second order power series for $k_2 = 15$.

[10] For N=200, 500, 1000, when (k_1, k_2) is large, there are some observations cannot obtain converged GMM estimator, thus we can't include them in our sample. Therefore, we discard the H

Honoré's results obtained from these bad observations, and generate additional number of samples for both the Honoré estimator and the updating GMM to ensure the comparability.

variance one). By decreasing the degree of freedom of the chi-squared distribution, design 2 generates samples with small variance.

Design 3

In design 3, I modify design 1 by increasing the noise in the error terms. Conditional on α_i, the error terms ε_{1i} and ε_{2i} are independent and follow the normal distribution $N(0,1+\alpha_i^2)$.

Design 4

Instead of using the transformed unconditional moment functions (2-5), in design 4, I use the conditional moment functions (2-4), where the conditional mean of $q(z_t)-q(z_s)$ given the regressor is replaced by the fitted value of $q(z)$, i.e.,

$$E\left\{q(e_t(y_{it}-x_{it}'\beta_0))-q(e_s(y_{is}-x_{is}'\beta_0))\middle|x_{it},x_{is}\right\}=\hat{q}(z,z^2,...z^{k_1})$$

where $\hat{q}(z,z^2,z^3,...z^{k_1})$ is the fitted value of $q(z)$ by applying the regression of $q(z)$ on $p(x_i)$ given $\beta=\hat{\beta}^H$. However, because too many observations are discarded in our designs, the weighted matrix $W=E(qq'|x)$ is more likely to be singular, especially when sample size is relatively small to the number of moment functions in the conditional moment restrictions. When it happens, we cannot take the inverse of the weighted matrix, which in turn leads to unreliable estimates.

Table 2-1 shows the fraction of censored observations and the fraction of discarded observations. Clearly, in our designs, about 75% of observations have censored dependent variables. In design 1, when the sample size is 200, about 63% of observations are discarded, only 37% of observations are used to determine the model parameters, i.e. only 75 of 200 observations are used. In design 2, about 68% of observations are not used to determine the model parameters. In design 3, fewer observations (around 58%) are discarded when we increase the noise in the error terms.

I compare two estimators for β_0: the Honoré estimator $\hat{\beta}^H$ and the continuously updating GMM estimator $\hat{\beta}^{update}$. For the continuously updating GMM estimates, the Honoré estimator is used as the starting value for the simulation to make the comparison easier. The simulation studies for EL estimator and 2-step GMM are also implemented. For EL estimators, the computations, however, are very difficult and time consuming. Considering the asymptotic equivalence of $\hat{\beta}^{update}$ and $\hat{\beta}^{EL,1}$, and that Newey and Smith (2004) already provide the results for the smooth moment functions[11], I focus on $\hat{\beta}^{update}$ in the paper. For 2-step GMM, I use the Charlier, Melenberg, and Soest (2000) method, set $\hat{\beta}^H$ as the starting point and go one Newton - Raphson step towards GMM[12]. 2-step GMM results are not presented in the paper due to its poor performance.

The results for the estimator are presented in Table 2-2 to 2-7[13]. Table 2-2 and 2-3 shows the results from design 1, Table 2-4 and 2-5 shows the results from design 2, and Table 2-6 and 2-7 shows the results from design 3. Tables have the contents as follows: the true value of the parameters (True), estimated mean bias (Mean_Bias), standard deviation (Std.), median bias (Median_Bias), root mean squared error (RMSE)[14], inter-quartile of the parameters (the difference between 75th and 25th percentiles of the sample) (IQR), mean absolute deviation of the parameters (MAD), and the difference between the maximum and the minimum value of the parameter (Range).

---

[11] Our moment functions are not continuously differentiable; therefore, the results should contribute to the literature by either confirming or refuting the existing results. However, due to the computational difficulty, we couldn't get the results by now; the study should be done in the future.

[12] This yields an estimator $\hat{\beta} = \hat{\beta}^H - \frac{g(\hat{\beta}^H)}{\partial g(\hat{\beta}^H)/\partial \beta}$, which is asymptotically equivalent to 2-step GMM.

[13] All the results reported in this paper were performed by Matlab 7.0.4

[14] A RMSE value closer to 0 indicates a better fit. This statistic is also known as the fit standard error and the standard error of the regression. RMSE measures the average mismatch between each data point and the model.

In all three designs with different sample size, when the continuously updating GMM estimate uses the moment functions (2-6) (where $k_1 = 1$ and $k_2 = 2$), which are exactly the same moment functions used by the Honoré estimate, the continuously updating GMM estimator is identical to the Honoré estimator.

In design 1, for a small sample size of 200, the performance of the updating GMM is worse than the Honoré estimator in most cases of that more moment functions are used. Only in case of 5 moment functions are used (i.e. $k_1 = 1$ and $k_2 = 5$), the updating GMM performs slightly better. When a large sample size of 500 is applied, in terms of RMSE, Std., IQR, MAD and Range, the performance of the updating GMM is slightly better if more but not too many moment functions are used. The mean bias and median bias, however, are still greater than that of the Honoré estimator. If too many number of moment functions, relative to the sample size, are used in estimates, for instance (2, 15), (3, 9) or (3, 15), the Honoré estimator outperforms the continuously updating GMM. If the sample size is further increased to 1000, the Monte Carlo evidence suggests that the updating GMM performs better except the case of $k_1 = 3$ and $k_2 = 15$.

Similar to design 1, when the large sample size of 500 and 1000 is applied in design 2, the RMSE, Std., IQR, and MAD are smaller than those of the Honoré estimator in most cases. And when the sample size is set to be 200, the updating GMM performs worse than the Honoré estimator if more moment functions are used.

In design 3, the performance of the updating GMM is improved. For a sample size of 200, the updating GMM performs better than the Honoré estimator in both the case of ($k_1 = 1$, $k_2 = 5$) and ($k_1 = 1$, $k_2 = 9$). And when the sample size is set to be 500 or 1000, the results of the updating GMM are better than that of the Honoré estimator.

To summarize, the performance of the continuously updating GMM is bad in all designs. This may be because all estimators we studied in this paper are based on the conditional moment restrictions (2-4) which require discarding part of the sample. As we described in section 2, the conditional moment functions in (2-4) only uses observations where the dependent variable is not censored in both periods as well as observations where the dependent variable is censored in one period but not in the other period and where no trimming occurs. Other observations would result in zero in the left hand side of (2-4), therefore, these observations make zero contribution in the estimation. In all three designs, over 60% observations are discarded by Honoré's artificially censored residuals. With too few observations can be used in the estimation, the weighted matrix used in GMM estimator would be imprecise estimated and more likely to be singular. Thus the computation of the inverse of weighted matrix is difficult in this case. For the 2-step GMM estimates, it's hard to compute the inverse of weighted matrix in (2-7) since it is close to zero given the Honoré estimator, and the results are very poor. The computation difficulty and time consuming of the EL estimates would be caused by similar reason. Although the updating GMM estimator has been successfully computed, we suggest that the sample discarding problem leads to unreliable estimates. This finding is supported by the improvement of the performance of the continuously updating GMM estimator in design 3. Compared to design 1, the performance of the continuously updating GMM is improved. It is because fewer observations are discarded in design 3 (see Table 2-1), which is made possible by imposing more heteroskedasticity in the error terms.

Although the continuously updating GMM behaves not well in all designs, there is some change between different designs. Comparing the results of design 1, 2, and 3, the performance of the continuously updating GMM seems depend on the variance imposed on the repressors and error terms. According to our settings, the sample generated in design 1 has larger variance than the sample generated in design 2, and the sample from design 3 have more heteroskedasticity than design 1. However, the continuously updating GMM performs the best in design 3, and the worst in design 2. This may suggest that the continuously updating GMM behaves better under heteroskedasticity in the finite sample.

Conclusion

This paper studies finite sample performance of several estimators proposed for the panel data Tobit regression model with fixed effect, including the Honoré estimator, the continuously updated best GMM estimator, and the empirical likelihood estimator. The continuously updated best GMM estimator and the empirical likelihood estimator could use more moment restrictions than the Honoré estimator and consequently are more efficient than the Honoré estimator in large samples.

We compare the continuously updating GMM estimator, which has similar finite sample performance as the empirical likelihood estimator, and the Honoré estimator via a Monte Carlo simulation study. By using moderate sample sizes, the results show that the updating GMM estimator outperforms the Honoré estimator with an appropriate number of moment functions. And with large sample sizes, e.g. N=1000, even using large numbers of moment functions, the updating GMM performs slightly better than the Honoré. However, the updating GMM estimator performs worse than the Honoré estimator for most cases when the sample size is small (e.g. N=200). Therefore, we suggest that increasing the number of moment functions used in estimation does not automatically lead to a large increase in efficiency, unless the sample size is very large relative to the number of moment functions used. This is may be because all estimators studied in this paper are based on the moment restrictions (2-4) which require discarding observations. And in our designs, close to seventy percent of observations are discarded. Having too many discarded observations would lead to an imprecise weighting matrix estimate, which consequently leads to an unreliable best GMM estimator.

It would therefore be interesting to do another simulation study with fewer discarded observations by changing the data generated process. For example, we could decrease the fraction of discarded observations to be less than 15% by adding a positive constant to dependent variable. Our study also suggests an alternative estimation method (such as conditional maximum likelihood estimation) that does not rely on trimming for the panel data Tobit regression model with individual effects.

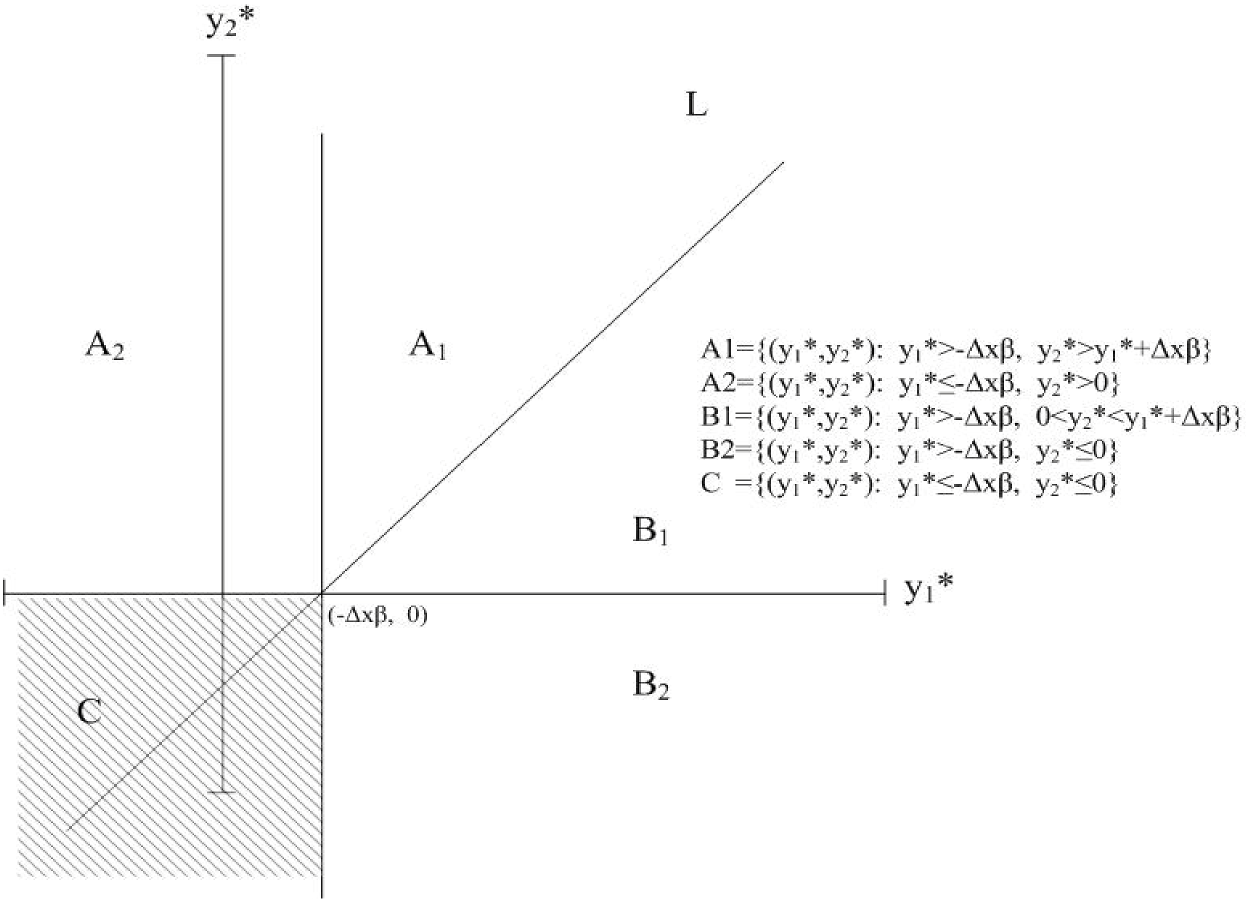

Figure 2-1[15]. Discarding observations when $\Delta x_i'\beta_0 \le 0$

---

[15]Observations in the set C, i.e. $(y_1^*, y_2^*) \in C$, are discarded.

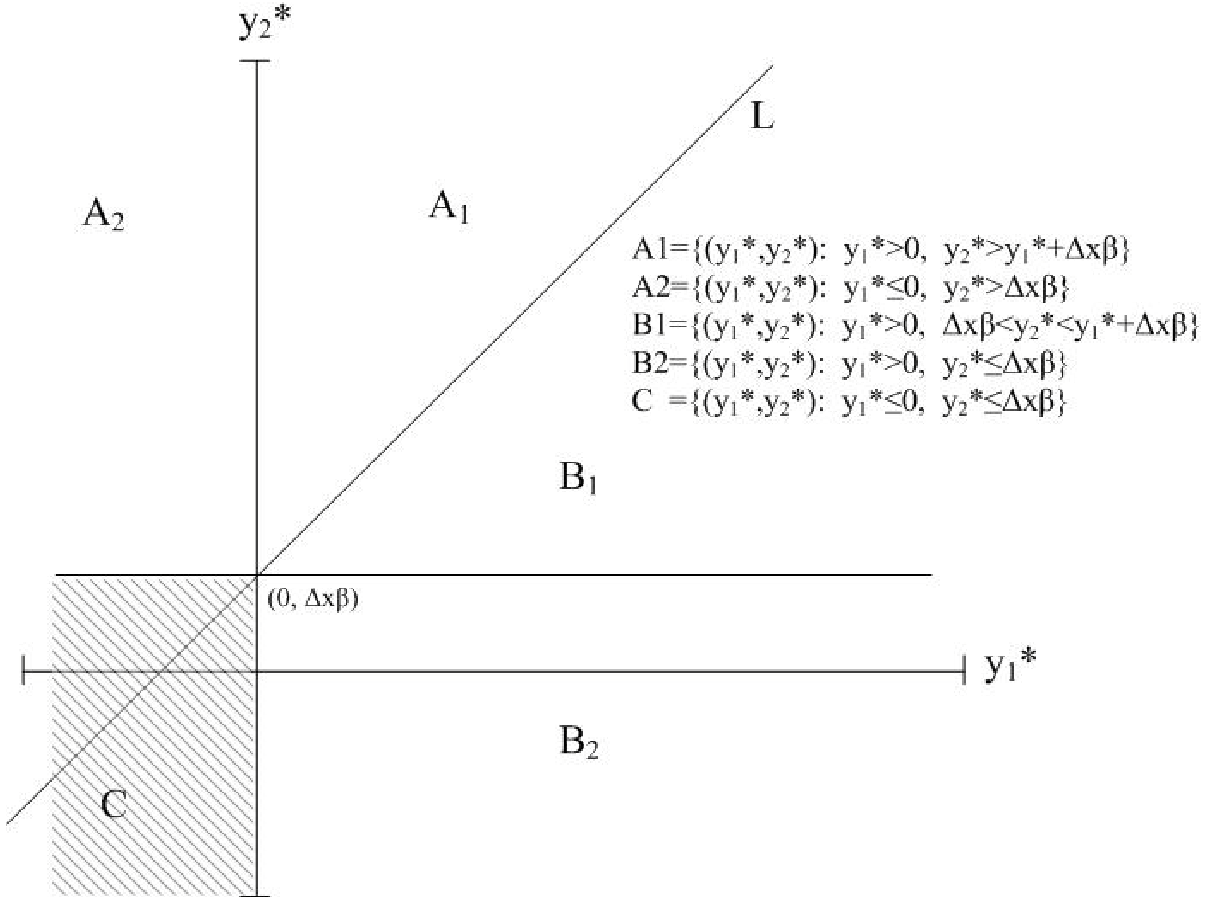

Figure 2-2. Discarding observations when $\Delta x_i' \beta_0 \geq 0$

Table 2-1. Censored observations and discarded observations

Design	N	Observations with censored dependent variable[16]		Discarded observations[17]	
		Number of observations[18]	Fractions	Number of observations	Fractions
1	200	143.772	71.89%	125.605	62.8%
	500	359.372	71.87%	314.282	62.86%
	1000	719.762	71.98%	629.451	62.95%
2	200	152.1470	76.07%	134.4740	67.237%
	500	380.9140	76.18%	337.6330	67.53%
	1000	762.2740	76.23%	675.4030	67.54%
3	200	145.6350	72.82%	116.9160	58.46%
	500	363.9830	72.8%	292.8280	58.57%
	1000	729.106	72.91%	585.794	58.58%

[16] We count the observations whose dependent variable is censored in at least one period.

[17] Discarded observations are those observations whose artificially censored residuals for different periods equal. Therefore, these observations are not used to determine the model parameters.

[18] This column gives the mean of the number of observations with censored dependent variable for 1000 replications.

Table 2-2. Monte Carlo study for the Honoré and the updating GMM estimator beta1 in design 1

N		(k1, k2)	True	Mean_Bias	Std.	Median_Bias	RMSE	IQR	MAD	Range
	Honoré	N/A	1	0.01566	0.16224	0.00857	0.16316	0.21037	0.12651	1.11470
		(1,2)	1	0.01566	0.16224	0.00857	0.16316	0.21037	0.12651	1.11470
		(1,5)	1	0.01262	0.15771	0.00005	0.15837	0.20261	0.12240	1.34870
		(1,9)	1	0.02031	0.14475	0.01115	0.14632	0.18757	0.11344	0.99730
		(1,15)	1	0.02582	0.16644	0.01070	0.16860	0.22817	0.13286	1.11110
200	Updating GMM	(2,5)	1	0.03742	0.18202	0.00830	0.18602	0.21322	0.13604	1.76120
		(2,9)	1	0.03557	0.18126	0.01845	0.18491	0.22419	0.13881	1.61810
		(2,15)	1	0.03606	0.19232	0.01570	0.19587	0.24563	0.15039	1.37580
		(3,5)	1	0.03463	0.17940	0.01800	0.18290	0.21909	0.13893	1.35850
		(3,9)	1	0.04032	0.18746	0.02485	0.19195	0.23382	0.14546	1.48590
		(3,15)	1	0.04010	0.20686	0.02725	0.16095	0.26412	0.16212	1.52260
	Honoré	N/A	1	0.00619	0.09712	0.00540	0.09742	0.13753	0.07760	0.61960
		(1,2)	1	0.00619	0.09712	0.00540	0.09742	0.13753	0.07760	0.61960
		(1,5)	1	0.00572	0.09359	0.00080	0.09387	0.12139	0.07388	0.64790
		(1,9)	1	0.00204	0.07884	0.00100	0.07895	0.10525	0.06316	0.49347
		(1,15)	1	0.00480	0.08834	-0.00240	0.08857	0.12056	0.07095	0.58115
500	Updating GMM	(2,5)	1	0.00523	0.08807	-0.00042	0.08832	0.11061	0.06768	0.80576
		(2,9)	1	0.00607	0.08563	0.00280	0.08594	0.11049	0.06716	0.55816
		(2,15)	1	0.01005	0.10374	0.00280	0.10434	0.13394	0.08222	0.83296
		(3,5)	1	0.00840	0.09043	0.00125	0.09091	0.12087	0.07176	0.60708
		(3,9)	1	0.01004	0.09853	0.00255	0.09915	0.12574	0.07792	0.71632
		(3,15)	1	0.01454	0.12182	0.00520	0.12290	0.16693	0.09854	0.74848
	Honoré	N/A	1	0.00293	0.06826	0.00031	0.06839	0.08389	0.05313	0.52887
		(1,2)	1	0.00293	0.06826	0.00031	0.06839	0.08389	0.05313	0.52887
		(1,5)	1	0.00417	0.06733	0.00131	0.06756	0.08629	0.05262	0.51924
		(1,9)	1	0.00418	0.05966	-0.00008	0.05988	0.07565	0.04669	0.40893
		(1,15)	1	0.00515	0.06173	0.00081	0.06201	0.07666	0.04792	0.39703
1000	Updating GMM	(2,5)	1	0.00483	0.06062	0.00095	0.06088	0.07663	0.04728	0.48997
		(2,9)	1	0.00489	0.06011	0.00020	0.06037	0.07478	0.04673	0.47840
		(2,15)	1	0.00625	0.06451	0.00350	0.06488	0.08605	0.05095	0.48489
		(3,5)	1	0.00524	0.06320	-0.00036	0.06348	0.07989	0.04932	0.45097
		(3,9)	1	0.00549	0.06315	-0.00028	0.06346	0.08093	0.04908	0.50132
		(3,15)	1	0.00702	0.07193	0.00680	0.07235	0.09506	0.05693	0.44601

Table 2-3. Monte Carlo study for the Honoré and the updating GMM estimator beta2 in design 1

N		(k1, k2)	True	Mean_Bias	Std.	Median_Bias	RMSE	IQR	MAD	Range
	Honoré	N/A	1	0.01281	0.16016	0.00478	0.16083	0.21009	0.12564	1.12700
		(1,2)	1	0.01281	0.16016	0.00478	0.16083	0.21009	0.12564	1.12700
		(1,5)	1	0.01301	0.15963	0.00020	0.16032	0.21094	0.12460	1.09030
		(1,9)	1	0.02890	0.18571	0.01860	0.18814	0.21452	0.14108	1.33140
		(1,15)	1	0.03245	0.17156	0.01940	0.17478	0.21811	0.13390	1.27080
200	Updating GMM	(2,5)	1	0.02555	0.17029	0.01090	0.17237	0.20223	0.12703	1.66270
		(2,9)	1	0.03646	0.20241	0.01440	0.20588	0.24845	0.15248	1.95860
		(2,15)	1	0.03226	0.19938	0.01290	0.20218	0.26614	0.15787	1.24740
		(3,5)	1	0.03316	0.17794	0.01305	0.18119	0.22429	0.13742	1.23580
		(3,9)	1	0.03505	0.20773	0.01545	0.21088	0.25558	0.15995	1.74390
		(3,15)	1	0.04091	0.20219	0.01945	0.20650	0.25625	0.15513	1.62920
	Honoré	N/A	1	0.01066	0.10032	0.00450	0.10098	0.12986	0.07873	0.68958
		(1,2)	1	0.01066	0.10032	0.00450	0.10098	0.12986	0.07873	0.68958
		(1,5)	1	0.00996	0.09609	0.00755	0.09670	0.12320	0.07501	0.70364
		(1,9)	1	0.00762	0.10078	-0.00131	0.10117	0.12197	0.07830	0.74675
		(1,15)	1	0.00809	0.09345	0.00155	0.09389	0.12024	0.07274	0.82581
500	Updating GMM	(2,5)	1	0.00785	0.08889	0.00225	0.08933	0.10800	0.06839	0.74863
		(2,9)	1	0.00778	0.09650	-0.00062	0.09691	0.11967	0.07491	0.73062
		(2,15)	1	0.00913	0.10556	0.00660	0.10607	0.13218	0.08245	0.91925
		(3,5)	1	0.00845	0.09332	-0.00187	0.09380	0.12012	0.07241	0.66668
		(3,9)	1	0.01271	0.11207	0.00555	0.11291	0.13642	0.08562	0.83672
		(3,15)	1	0.01901	0.12444	0.00775	0.12581	0.15312	0.09654	0.92630
	Honoré	N/A	1	0.00071	0.06708	-0.00036	0.06715	0.09011	0.05317	0.42117
		(1,2)	1	0.00071	0.06708	-0.00036	0.06715	0.09011	0.05317	0.42117
		(1,5)	1	0.00071	0.06708	-0.00036	0.06715	0.09011	0.05317	0.42117
		(1,9)	1	0.00063	0.06502	0.00028	0.06508	0.08544	0.05118	0.41175
		(1,15)	1	0.00076	0.06733	0.00033	0.06740	0.09175	0.05384	0.39473
1000	Updating GMM	(2,5)	1	0.00060	0.05960	0.00046	0.05966	0.07730	0.04681	0.42425
		(2,9)	1	0.00205	0.05775	0.00125	0.05785	0.07275	0.04504	0.42096
		(2,15)	1	0.00233	0.06201	-0.0005	0.06211	0.07624	0.04833	0.42557
		(3,5)	1	0.00089	0.06202	-0.00354	0.06209	0.07726	0.04847	0.39849
		(3,9)	1	0.00117	0.05831	0.00024	0.05838	0.07763	0.04594	0.41712
		(3,15)	1	0.00145	0.06530	-0.00058	0.06538	0.08471	0.05175	0.42924

Table 2-4. Monte Carlo study for the Honoré and the updating GMM estimator beta1 in design 2

N		(k1, k2)	True	Mean_Bias	Std.	Median_Bias	RMSE	IQR	MAD	Range
200	Honoré	N/A	1	0.02998	0.18898	0.01907	0.19134	0.23054	0.14240	1.44960
	Updating GMM	(1,2)	1	0.02998	0.18898	0.01907	0.19134	0.23054	0.14240	1.44960
		(1,5)	1	0.03065	0.20247	0.00593	0.20478	0.21275	0.14067	2.83450
		(1,9)	1	0.02984	0.17060	0.00522	0.17320	0.19748	0.12970	1.28480
		(1,15)	1	0.04104	0.20058	0.02040	0.20474	0.24450	0.15367	1.89440
		(2,5)	1	0.05342	0.21362	0.01393	0.22021	0.22016	0.15400	2.11350
		(2,9)	1	0.05576	0.20982	0.02299	0.21711	0.24190	0.15726	1.76930
		(3,5)	1	0.05539	0.21860	0.01549	0.22552	0.23207	0.15777	2.05480
		(3,9)	1	0.06043	0.23220	0.02142	0.25422	0.25203	0.15812	2.23670
500	Honoré	N/A	1	0.01079	0.11218	0.00130	0.11270	0.14488	0.08860	0.70040
	Updating GMM	(1,2)	1	0.01079	0.11218	0.00130	0.11270	0.14488	0.08860	0.70040
		(1,5)	1	0.01017	0.10747	0.00238	0.10795	0.13620	0.08321	1.00720
		(1,9)	1	0.01006	0.11063	-0.00200	0.11109	0.13209	0.08458	1.07900
		(1,15)	1	0.00825	0.10226	-0.00320	0.10260	0.12741	0.07871	0.87156
		(2,5)	1	0.01287	0.09629	0.00177	0.09715	0.10938	0.07210	0.79344
		(2,9)	1	0.01613	0.10389	0.00652	0.10514	0.12770	0.07846	0.95202
		(3,5)	1	0.02087	0.10551	0.00759	0.10756	0.11342	0.07827	0.74586
		(3,9)	1	0.02393	0.11021	0.00802	0.11274	0.12912	0.07992	0.98921
1000	Honoré	N/A	1	0.00486	0.07223	0.00102	0.09773	0.05681	0.50255	0.07240
	Updating GMM	(1,2)	1	0.00486	0.07223	0.00102	0.09773	0.05681	0.50255	0.07240
		(1,5)	1	0.00579	0.07042	-0.00128	0.08913	0.05538	0.52004	0.07066
		(1,9)	1	0.00578	0.06976	0.00025	0.09176	0.05493	0.43490	0.07000
		(1,15)	1	0.00542	0.06440	0.00113	0.08738	0.05105	0.43706	0.06463
		(2,5)	1	0.00503	0.05955	0.00082	0.07714	0.04634	0.44918	0.05976
		(2,9)	1	0.00618	0.05935	0.00163	0.07805	0.04671	0.41927	0.05967
		(3,5)	1	0.00692	0.06207	0.00164	0.08108	0.04854	0.43335	0.06246
		(3,9)	1	0.01119	0.07085	0.00560	0.08955	0.05512	0.45700	0.07173

Table 2-5. Monte Carlo study for the Honoré and the updating GMM estimator beta2 in design 2

N		(k1, k2)	True	Mean_Bias	Std.	Median_Bias	RMSE	IQR	MAD	Range
	Honoré	N/A	1	0.02127	0.20196	-0.00875	0.20307	0.21788	0.14604	2.46390
		(1,2)	1	0.02127	0.20196	-0.00875	0.20307	0.21788	0.14604	2.46390
		(1,5)	1	0.02063	0.18644	0.00124	0.18758	0.21198	0.13921	1.42360
		(1,9)	1	0.04671	0.30548	0.00575	0.30903	0.23216	0.16631	7.05220
200	Updating GMM	(1,15)	1	0.03861	0.20708	0.00904	0.21065	0.23562	0.15198	2.14310
		(2,5)	1	0.04802	0.21730	0.01179	0.22255	0.22026	0.15502	2.17410
		(2,9)	1	0.05514	0.23158	0.01143	0.23806	0.25752	0.16870	2.05110
		(3,5)	1	0.04436	0.21002	0.01561	0.21466	0.22812	0.15629	1.80400
		(3,9)	1	0.05631	0.24145	0.01753	0.24222	0.26811	0.17023	2.27300
	Honoré	N/A	1	0.01276	0.10464	0.00635	0.10541	0.13734	0.08175	0.69287
		(1,2)	1	0.01276	0.10464	0.00635	0.10541	0.13734	0.08175	0.69287
		(1,5)	1	0.01233	0.10630	0.00237	0.10702	0.13121	0.08147	0.86188
		(1,9)	1	0.01426	0.11732	0.00093	0.11819	0.13958	0.08849	1.02940
500	Updating GMM	(1,15)	1	0.01154	0.10060	0.00126	0.10126	0.12267	0.07843	0.74408
		(2,5)	1	0.01077	0.09768	0.00064	0.09827	0.11696	0.07352	0.87717
		(2,9)	1	0.01906	0.11788	0.00466	0.11941	0.13344	0.08814	0.89096
		(3,5)	1	0.01737	0.10736	0.00570	0.10876	0.12249	0.07977	0.91179
		(3,9)	1	0.01986	0.12191	0.00607	0.12641	0.13972	0.09021	0.94951
	Honoré	N/A	1	0.00671	0.07745	0.00018	0.10387	0.06100	0.62988	0.07773
		(1,2)	1	0.00671	0.07745	0.00018	0.10387	0.06100	0.62988	0.07773
		(1,5)	1	0.00621	0.07254	0.00202	0.09725	0.05751	0.53026	0.07280
		(1,9)	1	0.00674	0.07412	0.00127	0.10023	0.05902	0.57015	0.07442
1000	Updating GMM	(1,15)	1	0.00386	0.06254	0.00414	0.08518	0.04955	0.44351	0.06265
		(2,5)	1	0.00387	0.06086	0.00097	0.08366	0.04870	0.47413	0.06097
		(2,9)	1	0.00519	0.06908	0.00162	0.08988	0.05476	0.46083	0.06927
		(3,5)	1	0.00576	0.06415	0.00329	0.08494	0.05014	0.44782	0.06440
		(3,9)	1	0.01057	0.07358	0.00839	0.10124	0.05833	0.55048	0.07433

Table 2-6. Monte Carlo study for the Honoré and the updating GMM estimator beta1 in design 3

N		(k1, k2)	True	Mean_Bias	Std.	Median_Bias	RMSE	IQR	MAD	Range
200	Honoré	N/A	1	0.02692	0.21813	0.01612	0.28028	0.16857	1.52820	0.21979
	Updating GMM	(1,2)	1	0.02692	0.21813	0.01612	0.28028	0.16857	1.52820	0.21979
		(1,5)	1	0.02235	0.21030	0.00210	0.26299	0.16118	1.92000	0.21149
		(1,9)	1	0.023493	0.17836	0.01393	0.23908	0.14094	1.36550	0.17990
		(2,5)	1	0.04660	0.22372	0.01313	0.26899	0.17152	2.03680	0.22853
		(2,9)	1	0.04734	0.22876	0.02610	0.28051	0.17436	1.78480	0.23361
		(3,5)	1	0.05391	0.23819	0.03137	0.28050	0.18060	2.20900	0.24422
		(3,9)	1	0.06095	0.25246	0.04329	0.30248	0.19326	2.05370	0.25972
500	Honoré	N/A	1	0.01802	0.15001	0.00437	0.19284	0.11802	0.98186	0.15109
	Updating GMM	(1,2)	1	0.01802	0.15001	0.00437	0.19284	0.11802	0.98186	0.15109
		(1,5)	1	0.01660	0.14282	0.00241	0.18003	0.11046	1.34580	0.14378
		(1,9)	1	0.01201	0.11559	0.00158	0.14894	0.08971	0.93100	0.11622
		(2,5)	1	0.02536	0.14052	0.00207	0.15747	0.10463	1.00190	0.14279
		(2,9)	1	0.01797	0.12201	0.01073	0.15153	0.09466	0.83111	0.12333
		(3,5)	1	0.02378	0.13482	0.00625	0.15827	0.10281	0.97060	0.13690
1000	Honoré	N/A	1	0.00332	0.09022	-0.00220	0.11314	0.07029	0.65966	0.09028
	Updating GMM	(1,2)	1	0.00332	0.09022	-0.00220	0.11314	0.07029	0.65966	0.09028
		(1,5)	1	0.00560	0.08953	0.00093	0.11541	0.06972	0.64668	0.08970
		(1,9)	1	0.00502	0.07627	-0.00056	0.09701	0.05989	0.49129	0.07643
		(2,5)	1	0.00672	0.08392	-0.00031	0.10559	0.06512	0.62865	0.08418
		(2,9)	1	0.00573	0.07817	-0.00029	0.09935	0.06127	0.58770	0.07838
		(3,5)	1	0.00651	0.08569	0.00077	0.10765	0.06675	0.56499	0.08593

Table 2-7. Monte Carlo study for the Honoré and the updating GMM estimator beta2 in design 3

N		(k1, k2)	True	Mean_Bias	Std.	Median_Bias	RMSE	IQR	MAD	Range
200	Honoré	N/A	1	0.02151	0.21566	0.00341	0.27877	0.16892	1.50100	0.21673
	Updating GMM	(1,2)	1	0.02151	0.21566	0.00341	0.27877	0.16892	1.50100	0.21673
		(1,5)	1	0.02129	0.21159	-0.00091	0.26777	0.16552	1.53640	0.21266
		(1,9)	1	0.03233	0.23275	0.01531	0.29232	0.17990	1.69800	0.23499
		(2,5)	1	0.02918	0.22665	0.01122	0.26313	0.17128	2.01830	0.22853
		(2,9)	1	0.04647	0.26001	0.02214	0.32298	0.19748	2.51290	0.26414
		(3,5)	1	0.04283	0.25517	0.01293	0.29934	0.19167	1.92540	0.25874
		(3,9)	1	0.05274	0.27588	0.01126	0.33927	0.21268	2.08460	0.28088
500	Honoré	N/A	1	0.01902	0.13937	0.00823	0.17314	0.10781	0.95435	0.14066
	Updating GMM	(1,2)	1	0.01902	0.13937	0.00823	0.17314	0.10781	0.95435	0.14066
		(1,5)	1	0.01779	0.13873	0.00533	0.16853	0.10541	1.06870	0.13987
		(1,9)	1	0.01636	0.14855	-0.00062	0.17259	0.11208	1.35880	0.14945
		(2,5)	1	0.01985	0.13646	0.00589	0.16029	0.10190	1.23500	0.13790
		(2,9)	1	0.02239	0.15455	0.00510	0.18469	0.11665	1.07980	0.15617
		(3,5)	1	0.02006	0.13821	0.00502	0.15623	0.10403	0.96338	0.13966
1000	Honoré	N/A	1	0.00173	0.08862	-0.00037	0.11819	0.07031	0.57762	0.08864
	Updating GMM	(1,2)	1	0.00173	0.08862	-0.00037	0.11819	0.07031	0.57762	0.08864
		(1,5)	1	0.00178	0.08542	0.00167	0.11444	0.06727	0.55091	0.08544
		(1,9)	1	0.00183	0.08777	-0.00040	0.11928	0.06987	0.51560	0.08779
		(2,5)	1	0.00379	0.07774	-0.00093	0.09637	0.06077	0.59907	0.07783
		(2,9)	1	0.00315	0.08269	0.00217	0.10576	0.06495	0.55477	0.08275
		(3,5)	1	0.00227	0.07854	-0.00136	0.10428	0.06150	0.58639	0.07858

Table 2-8. A 200 Observations Sample

y_1	y_2	Er dif[19]	Er1[20]	Er2[21]	$-x_1'\hat{\beta}^H$	$-x_2'\hat{\beta}^H$
0.7284	1.3859	-0.7284	1.2296	0.5012	0.5012	-3.4312
0	0.6889	0.6888	0.7528	1.4416	-0.1409	0.7528
0	0	0	1.9336	1.9336	-0.3042	1.9336
0	2.6306	1.9842	0.2096	2.1938	0.2096	-0.4369
0.6219	4.4099	-0.1107	-0.6867	-0.7974	-1.3086	-5.2073
2.0017	0	0	1.7496	1.7496	-3.278	1.7496
0	0	0	2.164	2.164	0.3354	2.164
0	0.4988	0	2.6909	2.6909	2.6909	-2.6105
1.7654	2.4429	-1.3214	0.2747	-1.0466	-1.4907	-3.4895
0	0	0	2.1041	2.1041	1.064	2.1041
0	0	0	0.6630	0.6630	-0.0808	0.6630
0	0	0	2.1257	2.1257	1.0786	2.1257
0	0	0	2.6812	2.6812	2.6812	2.5731
9.9717	5.3846	-2.8503	3.9703	1.12	-6.0014	-4.2646
0	0	0	1.7309	1.7309	0.0923	1.7309
5.3358	4.2358	-2.1433	3.383	1.2397	-1.9529	-2.9962
2.5881	4.4284	1.0858	2.6011	3.6869	0.0130	-0.7414
6.7975	0	0	1.3346	1.3346	-7.6071	1.3346
0	0	0	1.9118	1.9118	1.9118	-0.2117
0	0	0	1.5491	1.5491	-0.1226	1.5491
7.3974	0.3186	-3.1876	3.2382	0.0506	-4.1591	-0.2680
0	0	0	3.1937	3.1937	2.8143	3.1937
0	3.5761	0	1.4059	1.4059	1.4059	-2.3176
0.3108	2.8436	-0.3108	1.1429	0.8322	0.83215	-2.3319
0	0	0	0.608	0.608	-0.5082	0.608
4.9005	0.6691	-0.1689	0.5783	0.4094	-4.3222	-0.2597
4.7071	2.9592	-0.0669	1.0871	1.0202	-3.62	-1.9391
5.8177	3.2097	-4.7453	4.0499	-0.6955	-1.7679	-3.9052
1.9682	0	-1.2692	2.0434	0.7743	0.0752	0.7743
2.4877	0	-0.0914	1.4182	1.3268	-1.0695	1.3268
4.0126	2.2597	0.6668	2.2837	2.9505	-1.7289	0.6909
0	0	0	1.0257	1.0257	0.2397	1.0257
0	0	0	0.9603	0.9603	0.6988	0.9603
1.561	0.0145	-0.6443	-0.2890	-0.9333	-1.8499	-0.9478
0	0	0	1.5998	1.5998	1.5998	0.83104
3.1159	8.9506	1.055	2.1875	3.2424	-0.9284	-5.7082
2.3776	0.9126	0.9126	0.1957	1.1083	-2.3481	0.1957
0	0.2306	0.2306	1.7671	1.9977	0.76913	1.7671

[19] This is the difference between the artificially censored residuals of period t and s.

[20] This column gives the artificially censored residuals for period 1 given the Honoré estimator.

[21] This column gives the artificially censored residuals for period 2 given the Honoré estimator.

CHAPTER 3

MAXIMUM LIKELIHOOD ESTIMATION OF PANEL DATA TOBIT MODEL

Introduction

The censored regression model is an important and interesting econometric model in econometric applications. A natural model for analyzing panel data containing censored dependent variable is the panel data Tobit model with individual effects.

We focus on the following Tobit model:

$$\begin{aligned} y_{it}^* &= \alpha_i + x_{it}'\beta_0 + \varepsilon_{it}, \\ d_{it} &= 1\{\alpha_i + x_{it}'\beta_0 + \varepsilon_{it}\} \\ y_{it} &= \begin{cases} y_{it}^*, & \text{if } d_{it} = 1 \\ 0, & \text{if } d_{it} = 0 \end{cases}, i = 1, 2, ..., N; t = 1, 2, ..., T \end{aligned} \tag{3-1}$$

where i denotes the individual, t denotes time; y_{it}^* denotes the latent dependent variable; y_{it} denotes the observed dependent variable; x_{it} denotes the k-dimensional vector of time-variant explanatory variables; α_i denotes the unobserved individual specific effects; β_0 denotes the true value of the unknown parameter vector to be estimated; ε_{it} denotes the error term, and d_{it} denotes a indicator function.

Estimation of panel data Tobit model (3-1) can be difficult, depending on the assumptions imposed on the error term and the individual effect. For assumption about the error term's distribution - in empirical work - we usually assume that ε_{it} follows the normal distribution $N(0, \sigma_0^2)$ [1]; and for different time periods, conditional on the explanatory variables and individual effects, we assume that the error terms are identically distributed. For the individual effect, when it is allowed to correlate with explanatory variables arbitrarily, the estimation of

[1] For simplicity, we assume that σ_0^2 is fixed.

model (3-1) is nontrivial and difficult since it enters the model nonlinearly and thus simple time differencing cannot remove it.

Several estimators have been proposed for Tobit model (3-1). Chamberlain (1984) uses a parametric specified model to derive his estimator. His estimator could provide some interesting quantities such as marginal effect of covariates on the observed censored variable. However when the parametric model is misspecified, Chamberlain's estimator will, in general, be asymptotically biased. To overcome this problem, Honoré (1992) offers a semiparametric model which avoids assumption of the distribution. Under some conditions, Honoré's estimator is consistent and asymptotically normally distributed. The drawback of his estimator is that the estimator is not efficient since it does not use all moment restrictions. To increase the efficiency of the estimates, the two-step GMM, the continuously updating GMM, and the empirical likelihood estimator (ELE) exploit other moment restrictions. Asymptotically, the latter three estimators are equivalent and are more efficient, at least not worse than Honoré's estimator. But, all three estimators are hard to compute due to the non-smooth problem and require discarding observations. The latter three approaches rely on artificial censoring to restore the zero correlation between the explanatory variables and the time-differenced artificially censored residuals. However, when most individual observations are discarded, as we discussed in chapter 2, these estimators have poor finite sample performance. Thus, it is imperative to find an alternative estimator that does not rely on trimming but is still consistent and asymptotically efficient.

In this paper, we consider the conditional maximum likelihood estimation (MLE) for model (3-1), given the normality assumption on the error term. Notice that we don't use conditional moment restrictions (2-4) in log-likelihood function, which means that conditional

MLE does not rely on trimming. However, there is risk in implementing the MLE method due to the presence of unobserved individual effects. Since we don't know the distribution of individual effects, the misspecification of the parametric form of the density that defines the log likelihood function can seriously bias estimate of model parameters. To overcome the problem of unknown density, we consider the semiparametric estimation proposed by Gallant and Nychka (1987) which can consistently approximate an unknown density function.

The main objective of this article is to implement Gallant and Nychka's (1987) semiparametric estimation to consistently approximate the unknown density for individual effects and then propose a consistent MLE estimator for unknown parameters in model (3.1) and unknown density of individual effects, while also deriving the asymptotic distribution for model parameters in (3.1) by using Shen's (1997) technique for sieve estimation.

The rest of the paper is organized as follows. Section 2 formally introduces the settings for maximum likelihood estimator. Section 3 shows the consistency of the MLE estimator and computes its convergence rate. Section 4 derives the $\sqrt{n}$-asymptotically normality of the MLE estimator for unknown parameters in the model (3.1). Section 5 provides a consistent covariance estimator for the model parameters estimator, and Section 6 concludes the paper. All technical proofs are presented in the Appendix.

MLE Estimator

For simplicity, we will drop the subscript i, and only consider two periods (i.e. T=2) in this study. Denote $x=(x_1', x_2')'$, and $y=(y_1, y_2)'$. Suppose that $\{(y, x), i=1,2,...,n\}$ is a sample of observations, and is drawn from the true density $f_0(y|x)$ with support of Y×X, where Y is a subset of R^{d_y} and X is a compact subset of R^{d_x}. Let z denote all observed time-invariant explanatory variables including x. Suppose that the error term ε_t is independent of z and α, conditional on α_i, the error terms ε_1 and ε_2 are independent. Also assume that conditional on z, α has an unknown density function $h_0(\alpha \mid \text{z})$. Let Φ and φ denote the standard normal probability distribution and density function respectively.

For each individual i, it is easy to compute that the conditional joint density of y given (x,α) is:

$$f(y|x,\ \alpha) = \prod_{t=1}^{2}\left[(1-d_t)(1-\Phi(\frac{\alpha + x_t'\beta_0}{\sigma_0})) + d_t(\frac{1}{\sigma_0}\varphi(\frac{y_t - \alpha - x_t'\beta_0}{\sigma_0}))\right]$$

Hence, we can write the conditional density function of y given z as:

$$f(y|z,\theta_0) = \int\prod_{t=1}^{2}\left[(1-d_t)(1-\Phi(\frac{\alpha + x_t'\beta_0}{\sigma_0})) + d_t(\frac{1}{\sigma_0}\varphi(\frac{y_t - \alpha - x_t'\beta_0}{\sigma_0}))\right]h_0(\alpha|z)d\alpha \qquad (3\text{-}2)$$

where θ_0 denotes the unknown coefficients β_0, σ_0, h_0 [2]. Then the log likelihood function is

$$\log f(y|z,\theta_0) = \log\int\prod_{t=1}^{2}\left[(1-d_t)(1-\Phi(\frac{\alpha + x_t'\beta_0}{\sigma_0})) + d_t(\frac{1}{\sigma_0}\varphi(\frac{y_t - \alpha - x_t'\beta_0}{\sigma_0}))\right]h_0(\alpha|z)d\alpha \qquad (3\text{-}3)$$

Therefore, the parameters of interest $\theta_0 = (\beta_0, \sigma_0, h_0)$, contains a vector of finite dimensional unknown parameters $(\beta_0, \sigma_0) \in \mathrm{E}$, where $\beta_0 \in \mathrm{B} \subset R^{d_\beta}, \sigma_0 \in R$, and an infinite dimensional parameters of density $h_0 \in \mathrm{H}$. The parameter space H is the set of admissible density functions and defined by Gallant and Nychka (1987). For simplicity, through the paper, denote $\theta = (\beta, \sigma, h) \in \Theta$, and $\Theta = \mathrm{E} \times \mathrm{H}$.

Under the assumption that Tobit model (3-1) identifies θ_0, we propose to estimate θ_0 by conditional MLE. Heuristically, if the function form $h(\alpha|z)$, the conditional density of α given z

, were known, then the function form of the conditional density of y given z, $f(y|z,\theta)$ would be known. The true value θ_0 solves the following constrained problem:

$$\begin{aligned}&\max_{\theta} E\{\log f(y|z,\theta)\}\\&\text{subject to} \int h(\alpha|z)d\alpha,\ h(\alpha|z)\geq 0.\end{aligned} \tag{3-4}$$

The true value of θ_0 could then be estimated by maximizing the sample analog of (3-4).

However, such an optimization procedure (3-4) is difficult to implement because the specified form of $h(\alpha|z)$, and therefore $f(y|z,\theta)$ is, in fact, unknown. Moreover, the density function $h(\alpha|z)$ is infinitly dimensional and it is impossible to be estimated from finite data points when the space H is too large. Often, optimization over a large parameter space leads to inconsistency or roughness. To overcome the problem, we use sieve approximation for the unknown density function $h(\alpha|z)$ introduced by Gallant and Nychka (1987). Their idea is to use a smoothness assumption to obtain an analytically tractable series representation for the unknown density $h(\alpha|z)$ by replacing H with a sieve space H_k. More specifically, the optimization is carried out within a sieve space H_k, which is a computable and finite-dimensional compact parameter space that is dense in the original space H as k increases. Let H_k be a sequence of approximating spaces to H (not necessarily a subset of H), denoted as a sieve, in the sense that for any $h \in \mathrm{H}$, there exists $h_k \in \mathrm{H}_k$ such that

$$\left\| h_k(\alpha|z) - h(\alpha|z) \right\|_2 \to 0 \text{ as } k \to \infty.$$

[2] It is important to see that (3-3) does not depend on the individual effect α. α has been integrated out.

Following Gallant and Nychka (1987), the common choice of $h(\alpha|z)$ is a series expansion $h_k(\alpha|z) = \varphi^2(\alpha)P_k^{\,2}(z,\alpha)$. Let Λ^{γ} denote a Hőlder space of order γ (see Ai and Chen for an exact definition of the Hőlder space, 2003). Obviously, $h(\alpha|z)$ is restricted to

$$H_k = \begin{Bmatrix} h_k(\alpha|z) = \varphi^2(\alpha)P_k^{\,2}(z,\alpha) \in \mathrm{H}_k \subset \Lambda^{\gamma}(\mathrm{Z}), k = 0,1,..., \text{ and} \\ \left\| h_k(\alpha|z) \right\|_2 = \sqrt{\left| \int h_k(\alpha|z)^2 d\alpha \right|} \le \mathrm{C}_0 \end{Bmatrix} \tag{3-5}$$

We propose to take $P_k(\alpha) = \exp(\frac{1}{2}(\pi_0 + \pi_1\alpha + \pi_2\alpha^2 + ... + \pi_k\alpha^k))$, therefore

$$h_k(\alpha|z) = \varphi^2(\alpha)\exp(\pi_0 + \pi_1\alpha + \pi_2\alpha^2 + ... + \pi_k\alpha^k)$$

Obviously $\pi_0,...,\pi_k$ are functions of z, i.e. $\pi_0 = \pi_0(z),...,\pi_k = \pi_k(z)$, for $k = 1,2,...$. Let $p_{k_1}(z)$ denote a vector of known basis functions of degree k_1 that can approximate any square integrable function of z arbitrarily well, and ζ_j denote a vector of coefficients. Thus, for each arbitrarily fixed α, $j = 1,2,...,k$, $\pi_j(z)$ can be approximated by $p_{k_1}(z)'\zeta_j$ for some vector of coefficients ζ_j as $k_1 \to \infty$, i.e.

$$\pi_j(z) = \left(p_1(z),. \;\; . p_{k_1}(z)\right)' \begin{pmatrix} \zeta_{j1} \\ . \\ . \\ \zeta_{jk_1} \end{pmatrix} = \sum_{s=1}^{k_1} p_s(z)\zeta_{js} \tag{3-6}$$

The above proposed specification form for $h_k(\alpha|z)$ has considerable computational advantages. First, $h_k(\alpha|z)$ is guaranteed to be non-negative everywhere in our settings. Next, we need to consider the density restriction $\int h_k(\alpha|z)d\alpha = 1$. To ensure that $\int h_k(\alpha|z)d\alpha = 1$, we solve

$$\pi_0 = -\ln\int \varphi^2(\alpha)\exp(\pi_1\alpha + \pi_2\alpha^2 + ... + \pi_k\alpha^k)d\alpha$$

Substituting π_0 back into $h_k(\alpha|z)$, we obtain

$$h_k(\alpha|z)=\frac{\varphi^2(\alpha)\exp(\pi_1\alpha+\pi_2\alpha^2+...+\pi_k\alpha^k)}{\int\varphi^2(\alpha)\exp(\pi_1\alpha+\pi_2\alpha^2+...+\pi_k\alpha^k)d\alpha} \tag{3-7}$$

where π_j takes the form (3-6). Clearly, the proposed form (3-7) of $h_k(\alpha|z)$ satisfies the density restrictions in the maximum problem (3-4).

Now, the semiparametric approximation to the density function $f(y|x)$ with $h(\alpha|z)$ replaced by a sieve estimator $h_k(\alpha|z)$ is:

$$f_k(y|z,\theta)=\int\prod_{t=1}^{2}\left[(1-d_t)(1-\Phi(\frac{\alpha+x_t'\beta}{\sigma}))+d_t(\frac{1}{\sigma}\varphi(\frac{y_t-\alpha-x_t'\beta}{\sigma}))\right]h_k(\alpha|z)d\alpha$$

Denote $\Theta_k=\mathrm{E}\times\mathrm{H}_k$ which is a finite dimensional approximation to Θ. To summarize, the conditional MLE estimator of θ_0 maximizes the sample analog of the parametric version of (3-4) with $h_k(\alpha|z)$ restricted to the sieve space H_k:

$$\hat{\theta}=\arg\max_{\theta\in\Theta_k}\sum_{i=1}^{n}\ln\left\{\int\prod_{t=1}^{2}\left[(1-d_t)(1-\Phi(\frac{\alpha+x_t'\beta}{\sigma}))+d_t(\frac{1}{\sigma}\varphi(\frac{y_t-\alpha-x_t'\beta}{\sigma}))\right]h_k(\alpha|z)d\alpha\right\} \tag{3-8}$$

The advantage of the conditional MLE procedure (3-8) is that it is natural and easy to implement. Once $h\in\mathrm{H}$ is replaced by $h_k\in\mathrm{H}_k$, the estimation problem effectively becomes a parametric one; hence commonly used econometric software packages can be used to compute the estimator $\hat{\theta}$. In addition, it is important to see that (3-8) does not depend on conditional moment restrictions, so we don't need to discard observations that are not satisfied by the conditional moment restrictions and worry about the trimming problem[3].

[3] As we discussed in Chapter 2, when most of individual observations cannot satisfy conditional moment restrictions (2-4) and we have to discard them in estimation, the estimation will a give bad performance estimator.

In the following sections, we show the consistency of β, σ and h_k, and derive the asymptotic distribution for β. Our main results are

$$\sqrt{N}(\hat{\beta}-\beta_0) \xrightarrow{\text{A}} N(0, V^{-1})$$
$$\|\hat{\sigma}-\sigma_0\|_2 = O_p(n^{-\frac{1}{4}})$$
$$\|\hat{h}_k - h_0\|_2 = \sqrt{\int (\hat{h}_k(\alpha) - h_0(\alpha))^2 \mu_0(\alpha) d\alpha} = O_p(n^{-\frac{1}{4}})$$

Consistency

We begin by introducing additional notation to aid the exposition. Denote $N(\varepsilon, \Theta_k, \|\cdot\|_s)$ as the minimal number of ε-radius covering balls of Θ_k under the metric $\|\cdot\|_s$, it measures the size of Θ_k .The estimator we consider is the one $\hat{\theta} \in \Theta_k$ from (3-8). Let $l(y, \theta) = \ln f_k(y|z, \theta)$, and define

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} l(y, \theta)$$

$$L(\theta) = E\{l(y, \theta)\}$$

Let $\|\cdot\|_s$ denote the pseudo metric on $\Theta = \mathrm{E} \times \mathrm{H}$ given by

$$\|\theta - \theta_0\|_s = \|\beta - \beta_0\|_E + \|\sigma - \sigma_0\|_E + \|h - h_0\|_2 \tag{3-9}$$

where $\|\cdot\|_E$ denotes the Euclidean norm, and $\|h - h_0\|_2 = \sqrt{\int [h - h_0]^2 d\alpha}$.

In this section, we first present sufficient conditions and apply the results of Gallant and Nychka (1987) to obtain consistency of the MLE estimator $\hat{\theta}$ for θ_0 under the metric $\|\cdot\|_s$. We then establish the convergence rate under a weaker metric $\|\cdot\|$. The convergence rate result will be used in Section 4 to establish the asymptotic normality for $\hat{\beta}$.

- *Assumption 1: (a) For each t, conditional on α, $\{(y_{it}, x_{it}),\ \text{i=1,2,...,n}\}$ are i.i.d, and satisfy the Tobit model (3-1);(b) $E(y_t^p)$ exists and finite for some $p > 2$;*
- *Assumption 2: (a) X is compact with nonempty interior; (b) the support of z is bounded;*
- *Assumption 3: (a)The support of α is a subset of the finite interval $[\alpha_1, \alpha_2]$; (b) $h_0(\alpha|z)$ is bounded and bounded away from zero;*
- *Assumption 4: the error term is ε_{it} independent of z and α, for each t, ε_{it} are i.i.d. error, and follows the normal distribution $N(0, \sigma_0^2)$;*
- *Assumption 5: $\beta_0 \in \mathrm{B} \subset R^{d\beta}$ is compact with nonempty interior*;
- *Assumption 6: There exists finite constants σ_1 and σ_2 such that $0 < \sigma_1 \le \sigma_0 \le \sigma_2 < \infty$;*
- *Assumption 7: $0 < \Phi_1 \le \Phi(\frac{\alpha + x_t'\beta}{\sigma}) \le \Phi_2 < 1$ holds for all x, $\beta \in \mathrm{B}$, and $\sigma_0 \in [\sigma_1, \sigma_2]$;*
- *Assumption 8: $\theta_0 = (\beta_0, \sigma_0, h_0) \in \Theta$ is the only solution of $\theta \in \Theta$ for maximum problem (3-4).*

Assumption 1 rules out time series observations. Assumption 1(b) is needed for consistency proof. This condition is common and requires the existence of moments. Assumption 2 is a typical condition imposed for sieve estimation, and can always be satisfied by discarding large values of the regressors. Assumption 3(a) restricts the support of individual effects α and Assumption 3(b) bounds the unknown density $h_0(\alpha|z)$ away from zero. Assumption 4 is commonly imposed in empirical work for the error term. Assumption 5 is commonly imposed in the literature for compactness of parameters. This condition is needed to prove consistency. Assumption 6 requires that σ is bounded and bounded away from zero. Assumption 7 basically requires normal distribution $\Phi(\frac{\alpha + x_t'\beta}{\sigma})$ is bounded and is needed to prove consistency as well as derive the asymptotical normality. And Assumption 8 is an identification condition[4] .

To prove consistency, first we need to verify the compactness which is commonly imposed in the nonparametric and semiparametric econometrics literature. Clearly, following assumptions 5 and 6, E is compact in the topology generated by the norm $\|\beta\|_E + \|\sigma\|_E$ Then the compactness is satisfied when the infinite-dimensional parameter space H consists of bounded and smooth functions. Recall that the definition of H follows by Gallant and Nychka (1987), by applying the results of their Theorem 1; the closure of H is compact under the norm $\|h\|_2$.Thus, the closure of $\Theta = \mathrm{E} \times \mathrm{H}$ is compact under the norm $\|\theta\|_s$.

Next, the denseness condition requires that the sieve spaces Θ_k approximate the parameter space Θ. Obviously, in our settings, $\mathrm{H}_{k-1} \subset \mathrm{H}_k$. Applying Theorem 2 of Gallant and Nychka(1987), we obtain that $\bigcup_{k=1}^{\infty} \mathrm{H}_k$ is dense in the closure of H with respect to $\|h\|_2$. Clearly, by the definition of $\Theta_k = \mathrm{E} \times \mathrm{H}_k$, $\bigcup_{k=1}^{\infty} \Theta_k$ is dense in the closure of Θ with respect to $\|\theta\|_s$.

Recall that in Tobit model (3-1), variables are generated according to nonlinear regression. To show that the convergence is uniform, we need stochastic dominant and Lipschitz condition.

Rewrite the log likelihood function $l(y|z,\theta)$ as

$$l(y,\theta) = \ln f_k(y|z,\theta) = \ln \int g(y|z,\beta,\sigma) h_k(\alpha|z) d\alpha$$

where

[4] Gallant and Nychka(1987) gave the proof of identification condition for the example of the sample selection problem in section 4. Their technique could be used to prove the identification condition in our settings.

$$g(y|z,\beta,\sigma) = g_{00}(y|z,\beta,\sigma) + g_{01}(y|z,\beta,\sigma) + g_{10}(y|z,\beta,\sigma) + g_{11}(y|z,\beta,\sigma)$$

$$g_{00}(y|z,\beta,\sigma) = (1-d_1)(1-d_2)(1-\Phi(\frac{\alpha + x_1'\beta}{\sigma}))(1-\Phi(\frac{\alpha + x_2'\beta}{\sigma}))$$

$$g_{01}(y|z,\beta,\sigma) = (1-d_1)d_2(1-\Phi(\frac{\alpha + x_1'\beta}{\sigma}))(\frac{1}{\sigma}\varphi(\frac{y_2 - \alpha - x_2'\beta}{\sigma}))$$

$$g_{10}(y|z,\beta,\sigma) = d_1(1-d_2)(\frac{1}{\sigma}\varphi(\frac{y_1 - \alpha - x_1'\beta}{\sigma}))(1-\Phi(\frac{\alpha + x_2'\beta}{\sigma}))$$

$$g_{11}(y|z,\beta,\sigma) = d_1 d_2(\frac{1}{\sigma}\varphi(\frac{y_1 - \alpha - x_1'\beta}{\sigma}))(\frac{1}{\sigma}\varphi(\frac{y_2 - \alpha - x_2'\beta}{\sigma}))$$

Following by Assumption 1-8, in the appendix we show that $l(y,\theta)$ is stochastically dominated and Lipschitz continuous. From Newey (1991), with envelope condition, Lipschitz condition and compactness condition, we have $\sup |L_n(\theta) - L(\theta)| = o_p(1)$ over $\theta \in \Theta_k$, and $L(\theta)$ is continuous in θ. The denseness condition and the continuity of $L(\theta)$ ensure that $L_n(\theta) \to L(\theta)$ uniformly over $\theta \in \Theta$.

Applying Theorem 0 of Gallant and Nychka (1987), we have:

- **Theorem 1.** Suppose that θ are obtained by maximizing (3-8), and Assumption1-8 hold, we have $\|\hat{\theta} - \theta_0\|_s = o_p(1)$

Theorem 1 is proved in the appendix.

Theorem 1 provides a consistency result under the pseudo metric $\|\cdot\|_s$, but it is not enough to establish the asymptotic normality of $\hat{\beta}$. It is well-known that to derive the asymptotic distribution, we typically need the convergence rate of the estimator θ to be faster than $n^{-1/4}$. The convergence rate tells us the curvature of the criterion function.

We now introduce another (weaker) metric $\|\cdot\|$ and compute the convergence rate of the estimator $\hat{\theta}$ under this metric. Let $E(\cdot)$ and $\text{var}(\cdot)$ denote the expectation and variance. To find the convergence rate ε_n, we need to show that $\|\theta_1 - \theta_2\| = o_p(\varepsilon_n)$ for any $\theta_1, \theta_2 \in \Theta_k$.

Suppose that the parameter space Θ_k is connected in the sense that for any two points $\theta_1, \theta_2 \in \Theta_k$, there exists a continuous path $\{\theta(\tau) = \theta_0 + \tau(\theta - \theta_0) \in \Theta_k : \tau \in [0,1]\}$. Thus, $\theta(0) = \theta_0, \quad \theta(1) = \theta$. Clearly, in our settings, $l(y, \theta_0 + \tau(\theta - \theta_0))$ is continuously differentiable at $\tau = 0$. Denote the directional derivative of $l(y, \theta)$ at θ_0 by:

$$l'(\theta_0)[\theta - \theta_0] = \frac{\partial l(y, \theta_0 + \tau(\theta - \theta_0))}{\partial \tau}\Big|_{\tau=0}$$

Notice that

$$\begin{aligned} &\frac{\partial l(y, \theta_0 + \tau(\theta - \theta_0))}{\partial \tau}\Big|_{\tau=0} \\ &= \frac{\partial \ln f_k(y|z, \theta_0)}{\partial \beta'}(\beta - \beta_0) + \frac{\partial \ln f_k(y|z, \theta_0)}{\partial \sigma}(\sigma - \sigma_0) + \frac{\partial \ln f_k(y|z, \theta_0)}{\partial h}[h_k - h_0] \end{aligned}$$

Now, applying a Taylor expansion, for all $\theta \in \Theta_k$ and all z, we could write

$$l(y, \theta) - l(y, \theta_0) = l'(\theta_0)[\theta - \theta_0] + r(\theta, \theta_0) \qquad \text{(3-10)}$$

where $r(\theta, \theta_0)$ is the remainder term.

Hence, the average Kullback-Leibler information is:

$$K(\theta_0, \theta) = E\{l(z, \theta_0)\} - E\{l(z, \theta)\} = -\frac{1}{2}\frac{\partial^2 E[l(z, \theta_0 + \tau(\theta - \theta_0))]}{\partial \tau^2}\Big|_{\tau=0} - R(\theta, \theta_0)$$

since $E\{l'(\theta_0)[\theta - \theta_0]\} = 0$. Notice that the equality

$$E\left\{\left[\frac{\partial[l(y, \theta_0 + \tau(\theta - \theta_0))]}{\partial \tau}\Big|_{\tau=0}\right]^2\right\} + \frac{\partial^2 E[l(y, \theta_0 + \tau(\theta - \theta_0))]}{\partial \tau^2}\Big|_{\tau=0} = 0$$

holds for any θ.

For any $\theta_1, \theta_2 \in \Theta_k$, define the metric $\|\cdot\|$ as:

$$\|\theta-\theta_0\|^2 = E\left\{[l'(\theta_0)[\theta-\theta_0]]^2\right\}$$
$$= E\left\{\left[\frac{\partial \ln f_k(y|z,\theta_0)}{\partial \beta'}(\beta-\beta_0)+\frac{\partial \ln f_k(y|z,\theta_0)}{\partial \sigma}(\sigma-\sigma_0)+\frac{\partial \ln f_k(y|z,\theta_0)}{\partial h}[h_k-h_0]\right]^2\right\} \tag{3-11}$$

and the fact that

$$|l(y,\theta_0)-l(y,\theta)| \le c(z)\|\theta_0-\theta\|_s$$

where $\|\theta-\theta_0\|_s = \|\theta-\theta_0\|_2$, imply that

$$\|\theta-\theta_0\| \le c\|\theta-\theta_0\|_s$$

Notice that the remainder $r(\theta_0,\theta)$ is in the order of $\|\cdot\|_2^3$.Thus, in the neighborhood of θ_0,

$\|\theta-\theta_0\|^2$ is equivalent to the average Kullback-Leibler information $E\{l(y,\theta_0)-l(y,\theta)\}$

We now present additional conditions for computing the convergence rates under the metric $\|\cdot\|$:

- *Assumption 9:* $\ln N(\varepsilon,\Theta_k,\|\cdot\|_s) \le const \times kk_1 \times \ln\frac{1}{\varepsilon}$ *;*
- *Assumption 10:* $kk_1 = o_p(n^v)$ *for some* $0<v<1$*;*
- *Assumption 11: for any* $\theta\in\Theta$ *, there exists* $\theta_k \in \Theta_k$ *satisfying* $\|\theta-\theta_k\|_s = o_p(kk_1^{-\xi}) = o_p(n^{-\xi v})$.

Notice that under condition 10, we have the expected criterion difference reduced to the Kullback-Leilber pseudo distance:

$$K(\theta_0,\theta) = E\{l(z,\theta)-l(z,\theta_0)\} = o_p(n^{-2\xi v}).$$

Hence, applying Theorem 1 of Shen and Wong (1994), we obtain the following Lemma:

- **Lemma:** Under Assumptions 1-11, we have

$$\left\|\hat{\theta}-\theta_0\right\| = o_p(\max(n^{-(1-v)/3}, n^{-\xi v}))$$

The lemma is proved in the Appendix.

Asymptotic Distribution

Having computed the convergence rate of the proposed MLE estimator under $\|\cdot\|$, we now derive the asymptotic distribution of $e=(\beta,\sigma)$. For simplicity, we assume $d_\beta = 2$.

We first introduce some notations. Denote $f(\theta)=\lambda' e$ for any fixed and nonzero $\lambda \in R^3$. To study the asymptotic distribution, we discuss linear approximations of the criterion difference by the corresponding derivatives and the degree of smoothness of function $f(\theta)$.

Let $\bar{V}$ denote the closure of the space spanned by $\Theta - \{\theta_0\}$ under the metric $\|\cdot\|$. Suppose that $\|\cdot\|$ induces an inner product $\langle \cdot,\cdot \rangle$ on $\bar{V}$, then for any $v_1, v_2 \in \bar{V}$, $(\bar{V}, \|\cdot\|)$ is a Hilbert space with the inner product:

$$\langle v_1, v_2 \rangle = E\left\{\frac{dl(z,\theta_0)}{d\theta}[v_1] \times \frac{dl(z,\theta_0)}{d\theta}[v_2]\right\} \tag{3-12}$$

where

$$\frac{dl(z,\theta_0)}{d\theta}[v] = \frac{dl(z,\theta_0+\tau v)}{d\theta}\Big|_{\tau=0}$$

By the results in Van der Vaart (1991) and Shen (1997), $f(\theta)=\lambda' e$ must be bounded (i.e., $\sup_{0\neq\theta-\theta_0\in\bar{V}} \frac{f(\theta)-f(\theta_0)}{\|\theta-\theta_0\|} < \infty$) so that it could be estimated at the $\sqrt{n}$ -rate. Thus, we need to show that $f(\theta)=\lambda' e$ is bounded. In addition, by the Riesz representation theorem, there exists $v^* \in \bar{V}$ such that, for any $\theta \in \Theta$, we have

$$f(\theta)-f(\theta_0)=f'(\theta_0)[\theta-\theta_0]=\lambda'\left(e-e_0\right)=\left\langle v^*,\theta-\theta_0\right\rangle \tag{3-13}$$

Hence, the asymptotic distribution of $\hat{\beta}-\beta_0$ is the same as the asymptotic distribution of $\left\langle\hat{\theta}-\theta_0,v^*\right\rangle$.

By definition,

$$\begin{aligned}\left\|v^*\right\| &= \sup_{0\neq\theta-\theta_0\in\bar{V}}\frac{\left|f(\theta)-f(\theta_0)\right|^2}{\left\|\theta-\theta_0\right\|^2}\\ &=\sup\frac{\lambda'(e-e_0)(e-e_0)'\lambda}{\left\|\theta-\theta_0\right\|^2}\\ &=\sup\ \lambda'\left(E\left\{D_w(z)'D_w(z)\right\}\right)^{-1}\lambda\end{aligned} \tag{3-14}$$

where $D_w(z)=\frac{dl(y,\theta_0)}{de'}-\frac{dl(y,\theta_0)}{dh}[w]$. It suffices to find w^* which minimize $E\left\{D_w(z)'D_w(z)\right\}$.

Denote $\bar{W}=\mathrm{H}-\left\{h_0\right\}$, and $\bar{V}=\mathrm{E}\times\bar{W}$. For each component e_j, $j=1,2,3$, let $w_j^*\in W$ denote the solution to

$$\min\left\{\left[\frac{dl(y,\theta_0)}{de_j}-\frac{dl(y,\theta_0)}{dh}[w_j]\right]^2\right\}, \tag{3-15}$$

Define $w^*=(w_1^*,w_2^*,w_3^*)$, $\frac{dl(y,\theta_0)}{dh}[w^*]=(\frac{dl(y,\theta_0)}{dh}[w_1^*],\frac{dl(y,\theta_0)}{dh}[w_2^*],\frac{dl(y,\theta_0)}{dh}[w_3^*])$, we have

$$D_{w^*}(z)=\frac{dl(y,\theta_0)}{de'}-\frac{dl(y,\theta_0)}{dh}[w^*] \tag{3-16}$$

Hence, $\left\|v^*\right\|^2=\lambda'\left(E\left\{D_{w^*}(z)'D_{w^*}(z)\right\}\right)^{-1}\lambda$ and $v^*=\lambda^{-1}\left\|v^*\right\|^2(1,-w^*)=(v_e^*,v_h^*)\in\bar{V}$ with $v_e^*=\left(E\left\{D_{w^*}(z)'D_{w^*}(z)\right\}\right)^{-1}\lambda$, $v_h^*=-w^*\times v_e^*$.

Moreover, (3-14) implies that $\mathrm{E}\left\{D_{w^*}(z)'D_{w^*}(z)\right\}$ has to be finite positive-definite in order for $f(\theta)=\lambda'\beta$ to be bounded.

Recall that

$$\frac{\partial l(y,\theta_0)}{\partial\beta'}=\frac{\int\frac{\partial g}{\partial\beta'}h_0 d\alpha}{\int g(y,\beta_0,\sigma_0)h_0 d\alpha},\quad \frac{\partial l(y,\theta_0)}{\partial\sigma}=\frac{\int\frac{\partial g}{\partial\sigma}h_0 d\alpha}{\int g(y,\beta_0,\sigma_0)h_0 d\alpha}$$

and

$$\frac{dl(y,\theta_0)}{dh}[w^*]=\frac{\int g(y,\beta_0,\sigma_0)w^* d\alpha}{\int g(y,\beta_0,\sigma_0)h_0 d\alpha}$$

Thus

$$\begin{aligned}D_{w^*}(z)&=\frac{dl(y,\theta_0)}{de'}-\frac{dl(y,\theta_0)}{dh}[w^*]\\&=\frac{\int\left[(\frac{\partial g}{\partial e'})h_0-g(y,\beta_0,\sigma_0)w^*\right]d\alpha}{f_k(y,\theta_0|x)}\end{aligned}$$

implies that

$$\begin{aligned}&E\left\{D_{w^*}(z)'D_{w^*}(z)\right\}=E\left\{\left[\frac{\int\left[(\frac{\partial g}{\partial e'})h_0-g(y,e_0)w^*\right]d\alpha}{f_k(y,\theta_0|x)}\right]'\left[\frac{\int\left[(\frac{\partial g}{\partial e'})h_0-g(y,e_0)w^*\right]d\alpha}{f_k(y,\theta_0|x)}\right]\right\}\\&=\mathrm{E}\left\{\left[\int((\frac{\partial g}{\partial e'})h_0-g(y,e_0)w^*)d\alpha\right]'\left[\int((\frac{\partial g}{\partial e'})h_0-g(y,e_0)w^*)d\alpha\right]\left(f_k(y,\theta_0|x)\right)^{-2}\right\}\\&=\mathrm{E}\left\{\left[\frac{\partial f_k}{\partial e'}-\frac{\partial f_k}{\partial h}w^*\right]'\left[\frac{\partial f_k}{\partial e'}-\frac{\partial f_k}{\partial h}w^*\right]\left(f_k(y,\theta_0|x)\right)^{-2}\right\}\\&=V\end{aligned}\qquad(3\text{-}17)$$

Recall that $f_k(y,\theta_0|x) \neq 0$, therefore we need $\mathrm{E}\left\{\left[\frac{\partial f_k}{\partial e'} - \frac{\partial f_k}{\partial h} w^*\right]'\left[\frac{\partial f_k}{\partial e'} - \frac{\partial f_k}{\partial h} w^*\right]\right\}$ is finite positive-definite.

Let $Q_n(q) = n^{-1/2}\sum_{i=1}^{n}\left[q(z_i) - E\left(q(z_i)\right)\right]$ denote the empirical process induced by any function $q(\cdot)$. And let $\varepsilon_n = o_p(n^{-1/2})$.

- *Assumption 12. (a)* $\mathrm{E}\left\{\left[\frac{\partial f_k}{\partial e'} - \frac{\partial f_k}{\partial h} w^*\right]'\left[\frac{\partial f_k}{\partial e'} - \frac{\partial f_k}{\partial h} w^*\right]\right\}$ is finite positive-definite*; (b)* $\beta_0 \in \mathrm{int}(E)$*;*
- *Assumption 13. There is a* $v_k^* = (v_e^*, v_{h_k}^*) \in \Theta_k - \{\theta_0\}, v_{h_k}^* = -\Pi_k w^* \times v_e^*$ *such that* $\left\|v_k^* - v^*\right\| = o_p(n^{-1/4})$*;*
- *Assumption 14.* $\xi \geq \frac{1}{4}$ *and* $v \leq \frac{1}{4}$.

Assumption 12(a) is a local identification condition for β_0, this condition cannot be relaxed. This condition must be satisfied for the estimated finite dimensional parameter β to be $\sqrt{n}$-consistent. However, it is hard to verify it in practice since we don't know the true value of the model Assumption 12(b) used to satisfy that β_0 is an interior point of E. Assumption 13 is needed due to the presence of unknown h_0. We simply assume that the same sieve space H_k approximates the space $W = H - \{h_0\}$ well[5]. Assumption 14 is required to ensure that the rate of MLE estimator θ in our setting converges to the true value θ_0 under the norm is at least $n^{-1/4}$.

[5] Theorem 2 could be proved even if v_h^* is approximated by any other sieve spaces, possibly different from H_k.

Theorem 2. Under Assumptions 1-12, $\sqrt{n}(\hat{\beta}-\beta_0) \to N(0,V^{-1})$ where

$$V = \mathrm{E}\left\{D_{w^*}(z)' D_{w^*}(z)\right\}$$

Theorem 2 shows that under sufficient conditions, the proposed conditional MLE estimator of the parametric component β_0 is $\sqrt{n}$-consistent and asymptotically normally distributed. Now we compute the variance for the proposed MLE estimator of the parametric component β_0.

Covariance Estimator

The asymptotic distribution derived above can be used for statistical inference on the parametric component (β_0, σ_0). To conduct the statistical inference, we need a consistent and easy to compute estimator of the covariance matrix. In this section, we provide one such estimator by consistently estimating V^{-1} given in section 4.

The only thing we need to estimate is w^* given (3-17). For each e_j, $j = 1,2,3$, we approximate w_j^* by $\hat{w}_j^*$, which is the solution to the minimization problem (3-15). Let $\widehat{w}^* = (\widehat{w}_1^*, \widehat{w}_2^*, \widehat{w}_3^*)$, then the estimator of V^{-1} is

$$\widehat{V}^{-1} = \frac{1}{n}\sum_{i=1}^{n}\left[\int\left((\frac{\partial g(\hat{e})}{\partial e'})\hat{h} - g(\hat{e})\widehat{w}^*\right)d\alpha\right]' \times \left[\int g(\hat{e})\hat{h}d\alpha\right]^{-2} \times \left[\int\left((\frac{\partial g(\hat{e})}{\partial e'})\hat{h} - g(\hat{e})\widehat{w}^*\right)d\alpha\right]$$

- **Theorem 3**. Under Assumption 1-12, we have $\hat{V} - V = o_p(1)$.

Conclusion

In this paper, we propose a semiparametric conditional maximum likelihood estimator for the panel data fixed effect Tobit regression model (3-1). We present some sufficient conditions, under which we show that the MLE estimator of the density function for individual effects is consistent and derive the convergence rate. We also show that the estimated finite dimensional parameter is $\sqrt{n}$ consistent and asymptotically normally distributed. In addition, we provide a consistent estimator of its asymptotic covariance matrix.

MLE estimation uses smooth objective function by integral over individual effect instead of using the conditional moment restrictions (2-4). Therefore, conditional MLE estimation doesn't need to discard observations. The MLE estimator proposed in this paper should be more efficient than Honoré's estimator, the two-step GMM, the continuously updating GMM, and the empirical likelihood estimator (ELE). In addition, not only can the model parameters be consistently estimated by MLE estimation, but also the distributions can be consistently estimated. As we showed in Section 3, we provide a consistent estimator for the density function of unknown individual effects. Also, we provide a consistent estimator for the error term's variance which will tell us the distribution for the error term. Moreover we can compute the estimated density function for y given x. The disadvantage of the conditional MLE estimator is that we need to know the parametric specification form of the error term's distribution. The normally distributed error term, though commonly imposed in empirical analysis, however, is difficult to justify. When the normality distribution assumption is not true, in general, the MLE estimator will be biased. In addition, the computation of MLE estimation is too complicated.

There are several limitations in the current study. It is unclear whether more efficiency advantages truly exist in finite samples. It would therefore be interesting to do a simulation study about the finite sample performance of the MLE estimator similar to the one we did in Chapter 2. However, due to the double integral, simulation for MLE cannot be implemented in MATLAB. Second, in this study, we assume that the identification condition is satisfied. It is possible to show that the identification holds for our specified model. The proof of identification will provide a missing piece in the literature. In the future, these two parts should be finished and added to the paper.

APPENDIX
PROOF FOR CHAPTER 3

PROOF. (Theorem 1). Let c denote a generic constant that may have different values in different expressions.

To show envelope condition and Lipschitz continuous, we will discuss four cases.

Case 1: $d_1 = 0, d_2 = 0$.

Hence,

$$l(y,\theta) = \ln f_{00}(y|z,\theta) = \ln \int g_{00}(y|z,\beta,\sigma) * h_k(\alpha|z) d\alpha$$

Notice that, because the explanatory variables and the unknown coefficients are all bounded following by assumption 1-7, we have

$$2\ln(1-\Phi_2) \le \ln f_{00}(y|z,\theta) \le 2\ln(1-\Phi_1),$$

hence, $\ln f_{00}(y|z,\theta)$ is bounded and has bounded derivatives with respect to β and σ. The directional derivative with respect to the unknown density h is

$$\frac{\partial \ln f_{00}(y|z,\theta)}{\partial h}[\nabla h] = \frac{\int (1-\Phi(\frac{\alpha + x_1'\beta}{\sigma}))(1-\Phi(\frac{\alpha + x_2'\beta}{\sigma}))\nabla h(\alpha|z) d\alpha}{\int (1-\Phi(\frac{\alpha + x_1'\beta}{\sigma}))(1-\Phi(\frac{\alpha + x_2'\beta}{\sigma}))h(\alpha|z) d\alpha},$$

and the fact that

$$1-\Phi_2 \le 1-\Phi(\frac{\alpha + x_t'\beta}{\sigma_0}) \le 1-\Phi_1$$

imply that $\frac{\partial \ln f_{00}(y|z,\theta)}{\partial h}$ is bounded by $c\|\nabla h(\alpha|z)\|_2$. So $\ln f_{00}(y|z,\theta)$ is stochastically dominated and Lipschitz continuous.

Case 2: $d_1 = 0, d_2 = 1$.

In this case, we have

$$l(y,\theta)=\ln f_{01}(y|z,\theta)=\ln\int g_{01}(y|z,\beta,\sigma)*h_k(\alpha|z)d\alpha$$

Notice that for all $\alpha\in[\alpha_1,\alpha_2]$,

$$\exp(-\frac{\max(|u-\alpha_1|,|u-\alpha_2|)^2}{2\sigma^2})\le\exp(-\frac{(u-\alpha)^2}{2\sigma^2})\le 1$$

We obtain

$$\int\frac{1-\Phi_2}{\sigma_2}\varphi(\frac{\max(|y_2-x_2'\beta-\alpha_1|,|y_2-x_2'\beta-\alpha_2|)}{\sigma_1}))h_k(\alpha|z)d\alpha\le f_{01}(y|z,\theta)\le\int\frac{1-\Phi_1}{\sigma_1}h_k(\alpha|z)d\alpha$$

It follows that $\ln f_{01}(y|z,\theta)$ is stochastically dominated if $E(y_2^p)$ exists and is finite for some $p>2$.

Some calculations give

$$\frac{\partial f_{01}(y|z,\theta)}{\partial\beta}=\frac{-x_1}{\sigma}\int\varphi(\frac{\alpha+x_1'\beta}{\sigma})(\frac{1}{\sigma}\varphi(\frac{y_2-\alpha-x_2'\beta}{\sigma}))h(\alpha|z)d\alpha$$
$$+\frac{x_2}{\sigma}\int(1-\Phi(\frac{\alpha+x_1'\beta}{\sigma}))(\frac{y_2-\alpha-x_2'\beta}{\sigma^2}\varphi(\frac{y_2-\alpha-x_2'\beta}{\sigma}))h(\alpha|z)d\alpha$$

which is bounded by

$$(\frac{|x_1|}{\sigma_1(1-\Phi_1)}+\frac{|x_2||y_2|+|x_2|+c}{\sigma_1^2})\int(1-\Phi(\frac{\alpha+x_1'\beta}{\sigma}))(\frac{1}{\sigma}\varphi(\frac{y_2-\alpha-x_2'\beta}{\sigma}))h(\alpha|z)d\alpha$$
$$=\frac{|x_1|+|x_2||y_2|+|x_2|+c}{\sigma_1^2(1-\Phi_1)}f_{01}(y|z,\theta)$$

for some constant c. Thus, the derivative of $l(y,\theta)$ with respect to β is stochastically dominated by

$$\left|\frac{\partial\ln f_{01}(y|z,\theta)}{\partial\beta}\right|\le\frac{|x_1|+|x_2||y_2|+|x_2|+c}{\sigma_1^2(1-\Phi_1)}$$

Similarly, we find

$$\left|\frac{\partial f_{01}(y|z,\theta)}{\partial\sigma}\right| = \int \frac{\alpha + x_1'\beta}{\sigma^2}\varphi(\frac{\alpha + x_1'\beta}{\sigma})(\frac{1}{\sigma}\varphi(\frac{y_2-\alpha-x_2'\beta}{\sigma}))h(\alpha|z)d\alpha$$

$$+\int(1-\Phi(\frac{\alpha + x_1'\beta}{\sigma}))(\frac{(y_2-\alpha-x_2'\beta)^2-\sigma}{\sigma^4}\varphi(\frac{y_2-\alpha-x_2'\beta}{\sigma}))h(\alpha|z)d\alpha$$

is bounded by $\frac{|x_1|+|y_2|^2+|x_2|^2+c}{\sigma_1^3(1-\Phi_1)} f_{01}(y|z,\theta)$. Hence, the derivative of $l(y,\theta)$ with respect to σ

is stochastically dominated by

$$\left|\frac{\partial \ln f_{01}(y|z,\theta)}{\partial\sigma}\right| \le \frac{|x_1|+|y_2|^2+|x_2|^2+c}{\sigma_1^3(1-\Phi_1)}$$

For the directional directive, we have

$$\frac{\partial \ln f_{01}(y|z,\theta)}{\partial h}[\nabla h] = \frac{\int(1-\Phi(\frac{\alpha + x_1'\beta}{\sigma}))(\frac{1}{\sigma}\varphi(\frac{y_2-\alpha-x_2'\beta}{\sigma}))\nabla h(\alpha|z)d\alpha}{f_{01}(y|z,\theta)}$$

which is bounded by $c\|\nabla h(\alpha|z)\|_2$.

Therefore, $\ln f_{01}(y|z,\theta)$ is stochastically dominated and Lipschitz continuous.

Case 3: $d_1 = 1, d_2 = 0$.

Similarly,

$$l(y|z,\theta) = \ln f_{10}(y|z,\theta) = \ln\int g_{10}(y|z,\beta,\sigma)h_k(\alpha|z)d\alpha$$

Using similar line of arguments, we can show that

$$\int\frac{1-\Phi_2}{\sigma_2}\varphi(\frac{\max(|y_1-x_1'\beta-\alpha_1|,|y_1-x_1'\beta-\alpha_2|)}{\sigma_1}))h_k(\alpha|z)d\alpha \le f_{10}(y|z,\theta) \le \int\frac{1-\Phi_1}{\sigma_1}h_k(\alpha|z)d\alpha$$

and

$$\frac{\partial \ln f_{10}(y|z,\theta)}{\partial\beta} \le \frac{|x_2|+|x_1||y_1|+|x_1|+c}{\sigma_1^2(1-\Phi_1)}$$

Thus, $\ln f_{10}(y|z,\theta)$ is stochastically dominated if $E(y_1^p)$ exists and is finite for some $p>2$, and Lipchize continuous holds.

Case 4: $d_1=1, d_2=1$.

$$l(y|z,\theta)=\ln f_{11}(y|z,\theta)=\ln\int g_{11}(y|z,\beta,\sigma)h_k(\alpha|z)d\alpha$$

Using similar line of arguments, we can show that $\ln f_{11}(y|z,\theta)$ are stochastically dominated and Lipchize continuous.

Therefore, the above discussions imply that:

(a) there exists a measurable function $c_1(y,z)$ with $E[c_1(y,z)^p]<\infty$ for some $p>2$ such that $|l(y,\theta)|\le c_1(y,z)$ for all $\theta\in\Theta_k$ and y,z .

(b) for all z and $\theta_1,\theta_2\in\Theta_k$, there exists a measurable function $c_2(y,z)$ with $E[c_1(y,z)^2]<\infty$ such that $|l(y,\theta_1)-l(y,\theta_2)|\le c_2(y,z)\|\theta_1-\theta_2\|_s$

Theorem 1 now follows from application of Theorem 0 of Gallant and Nychka (1987).

Q.E.D.

PROOF. (Lemma 1). Given $\|\theta-\theta_0\|=\sqrt{E\left\{\left[\frac{\partial\left[l(y,\theta_0+\tau(\theta-\theta_0))\right]}{\partial\tau}\Big|_{\tau=0}\right]^2\right\}}$, the Condition C1 of Shen and Wong (1994) is true with $\alpha=1$.

Furthermore, after some manipulations, we know Condition C2 of Shen and Wong (1994) holds with $\beta=1$.

$$
\begin{aligned}
&\sup\ \operatorname{var}\{l(y,\theta_0)-l(y,\theta)\} \\
&= E\left\{[l(y,\theta_0)-l(y,\theta)]^2\right\} - \left[E\{l(y,\theta_0)-l(y,\theta)\}\right]^2 \\
&\le E\left\{[l(y,\theta_0)-l(y,\theta)]^2\right\} \\
&= E\left\{c^2{}_2(y)\|\theta_1-\theta_2\|_s^2\right\} \\
&\le 2A_2\varepsilon^2
\end{aligned}
$$

To show Condition C3 of Shen and Wong (1994), we have to calculate the metric entropy function. Let $B_\delta(\theta)$ be the same as defined in Ossiander (1987). Notice that

$$
\begin{aligned}
&E\left\{\sup{}_{B_\delta(\theta)}[l(y,\theta_0)-l(y,\theta)]^2\right\} \\
&\le E\left\{c^2(z)\|\theta_1-\theta_2\|_s^2\right\} \\
&\le C\delta^2
\end{aligned}
$$

for some constant $C>0$.

Hence, let $F_k=\{l(y,\theta_0)-l(y,\theta):\theta\in\Theta_k\}$, for some constant $A_3>0$, and all small $\varepsilon>0$, applying assumption 7 and the inequality $\ln(x)\le x-1$ for all $x>0$, we have

$$
\begin{aligned}
\ln N(\varepsilon,F_k) &\le \ln N(\varepsilon,\Theta_k,\|\cdot\|_s) \\
&\le A_3\times kk_1\times\ln\varepsilon^{-1} \\
&\le A_3\times kk_1\times(\varepsilon^{-1}-1) \\
&\le A_3\times kk_1\times\varepsilon^{-1}
\end{aligned}
$$

Then condition C3 holds with $r_0 = v/2$, $r=1$. Assume $\zeta_k=1$, the lemma now follows from application of theorem 1 of Shen and Wong (1994).

Q.E.D.

PROOF. (Theorem 2). Recall the notation $L_n(\theta)$ and $L(\theta)$ introduced in Section 3. To simplify notation, denote $u^*=\pm v^*$, $u_k^*=\Pi_k u^*$ Denote

$$
\begin{aligned}
\phi(\tau) &= L(\theta_0+\tau u_k^*), \\
\phi_n(\tau) &= L_n(\hat{\theta}+\tau u_k^*),
\end{aligned}
$$

The fact that $\hat{\theta}$ is the solution for (3-8) implies that $\phi_n(\tau)$ is maximized at $\tau = 0$, and θ_0 is the solution for (3-4) implies that $\phi(\tau)$ is maximized at $\tau = 0$. Hence, we have the following first order condition:

$$\begin{aligned} \phi'(\tau) &= E\left\{\frac{\partial l(\theta_0 + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right\} = 0, \\ \phi_n'(\tau) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial l(\hat{\theta} + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0} = 0 \end{aligned} \qquad (3\text{-}18)$$

Since $\hat{\theta}$ converges to θ_0 in probability, applying the empirical process theorem, we have

$$Q_n\left(\frac{\partial l(\hat{\theta} + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right) - Q_n\left(\frac{\partial l(\theta_0 + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right) = o_p(n^{-1/2}) \qquad (3\text{-}19)$$

Substituting (3-18) into (3-19), it follows that,

$$E\left\{\frac{\partial l(\hat{\theta} + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right\} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial l(\theta_0 + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right) + o_p(n^{-1/2})$$

Applying a Taylor expansion around θ_0 for $E\left\{\frac{\partial l(\hat{\theta} + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right\}$, we obtain

$$\begin{aligned} & E\left\{\frac{\partial l(\hat{\theta} + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right\} \\ &= E\left\{\frac{\partial l(\theta_0 + \tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right\} + E\left\{\frac{\partial^2 l(\hat{\theta} + \tau u_k^* + s(\hat{\theta} - \theta_0))}{\partial \tau \partial s}\Big|_{\tau=0,s=0}\right\} + r_2(\hat{\theta} - \theta_0) \\ &= E\left\{\frac{\partial^2 l(\hat{\theta} + \tau u_k^* + s(\hat{\theta} - \theta_0))}{\partial \tau \partial s}\Big|_{\tau=0,s=0}\right\} + r_2(\hat{\theta} - \theta_0) \end{aligned}$$

Note that the remainder term $r_2(\hat{\theta} - \theta_0)$ is dominated by $\left\|\hat{\theta} - \theta_0\right\|^2$, i.e

$$r_2(\hat{\theta} - \theta_0) \le \left\|\hat{\theta} - \theta_0\right\|^2 = o_p(n^{-1/2})$$

Also, given the definition of $\langle \cdot, \cdot \rangle$ in (3-12), we have,

$$E\left\{\frac{\partial l(\hat{\theta}+\tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right\}$$

$$=E\left\{\frac{\partial^2 l(\hat{\theta}+\tau u_k^*+s(\hat{\theta}-\theta_0))}{\partial \tau \partial s}\Big|_{\tau=0,s=0}\right\}+r_2(\hat{\theta}-\theta_0)$$

$$=\left\langle \hat{\theta}-\theta_0, u_k^* \right\rangle + o_p(n^{-1/2})$$

Notice that, follows from assumption 11, and $u^* = \pm v^*$, $u_k^* = \Pi_k u^*$, we have

$$\left\langle \hat{\theta}-\theta_0, u_k^* \right\rangle = \left\langle \hat{\theta}-\theta_0, v^* \right\rangle$$

$$E\left\{\frac{\partial l(\hat{\theta}+\tau u_k^*)}{\partial \tau}\Big|_{\tau=0}\right\} = E\left\{\frac{\partial l(\hat{\theta}+\tau v^*)}{\partial \tau}\Big|_{\tau=0}\right\} + o_p(n^{-1/2})$$

Hence, we have,

$$\left\langle \hat{\theta}-\theta_0, v^* \right\rangle = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial l(\theta_0+\tau v^*)}{\partial \tau}\Big|_{\tau=0}\right) + o_p(n^{-1/2})$$

Since $\widehat{\beta}-\beta_0 = \left\langle \hat{\theta}-\theta_0, v^* \right\rangle$, we obtain Theorem 3 by applying a standard central limit theorem.

Q.E.D.

PROOF. (Theorem 3). To prove the consistency of the estimated covariance matrix, we need to show that $\widehat{w}^*$ converges to w^* uniformly over W. Applying an argument similar to that used to prove theorem 1, notice that the fact of $\hat{\theta}$ converge to the true value θ_0 under $\| \ \|_s$ at the rate $o_p(n^{-1/4})$, we can show that, for $j = 1,2,3$, $\frac{dl(y,\theta_0)}{de_j} - \frac{dl(y,\theta_0)}{dh}[w_j{}^*]$ satisfies stochastic dominant and the Lipschitz condition in $\theta \in \Theta_k$, and $\hat{w}_j{}^* \in \left\{ v_h \in W : \|v_h\|_s \le c < \infty \right\}$. Therefore, we could obtain $\|\hat{w}^* - w^*\|_s = o_p(1)$. The consistency of $\hat{V}_0$ now follows immediately.

Q.E.D.

LIST OF REFERENCES

AI, C. (1997): "A Seminonparametric Maximum Likelihood Estimator," *Econometrica*, 65, 933-964.

AI, C., AND LI, Q.(2005): *Handbook of Panel Data*, in press.

AI, C. (2007): "Seminonparametric Maximum Likelihood Estimation of Conditional Moment Restriction Models," *International Economic Review*, 48, 1093-1118.

AI, C AND CHEN, X (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795-1843.

AI, C AND CHEN, X (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables," *Journal of Econometrics*, 141, 5-43.

ALTONJI, J.G., AND SEGAL, L.M. (1996): "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Economics and Business Statistics*, 14, 353-366.

AMEMIYA, T. (1985): *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

BECKER, G. S. (1964): *Human Capital*, New York: Columbia University Press.

BLACKWELL, DAVID W., JAMES A. BRICKLEY, AND MICHAEL S. WEISBACH (1994): "Accounting information and internal evaluation: Evidence from Texan banks," *Journal of Accounting and Economics*, 17, 331-359.

BOEKER, W. (1992): "Power and managerial turnover: Scapegoating at the top," *Administrative Science Quarterly*, 37: 400-421.

BOROKHOVICH, K.A., PARRINO, R., AND TRAPANI, T. (1996): "Outside directors and CEO selection," *Journal of Financial and Quantitative Analysis*, 31(3): 337-355.

BRICKLY, J.A., COLES, JEFFREY L., AND JARRELL, G. (1997): "Leadership structure: Separating the CEO and the chairman of the board," *Journal of Corporate Finance*, 7: 43-66.

BUCHHOLTZ, A. K., AND RIBBENS, B. (1994): "Role of Chief Executive Officers in Takeover Resistance: Effects of CEO Incentives and Individual Characteristics," *Academy of Management Journal*, 37, 554-579.

BUCHHOLTZ, A. K., BARBARA A. RIBBENS, AND IRENE T. HOULE (2003): "The Role of Human Capital in Postacquisition CEO Departure," *Academy of Management Journal*, 46, 506-514.

BURMAN, P., CHEN, AND K.W. (1989): "Nonparametric Estimation of A Regression Function," *Annals of Statistics*, 17, 1567--1596.

CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

CANNELLA, A., AND HAMBRICK, D. (1993): "Effects of executive departure on the performance of acquired firms," *Strategic Management Journal*, 14(summer special issue): 137-152.

CANNELLA, A. A. JR., AND HAMBRICK, D. (1993): "Relative standing: A framework for understanding departures of acquired executives," *Academy of Management Journal*, 36: 733-762.

CANNELLA, A.A., JR., AND LUBATKIN, M.H. (1993): "Succession as a sociopolitical process," *Academy of Management Journal*, 36: 763-793.

CANNELLA, A.A., JR., AND SHEN W. (2001): "So close and yet so far: Promotion versus exit for CEO heir apparent," *Academy of Management Journal*, 44: 252-270

CANNELLA, A.A., JR., AND SHEN W. (2002): "Power dynamics within top management and their impacts on CEO dismissal followed by inside succession," *Academy of Management Journal*, 45: 1195-1206

CHARLIER, E., MELENBERG, B. AND SOEST, A.V. (2000): "Estimation of a Censored Regression Panel Data Model Using Conditional Moment Restrictions Efficiently", *Journal of Econometrics*, 95 25-56.

CHEN, X., AND SHEN, X. (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66, 289-314.

CHEN, X., LINTON, O. AND VAN KEILEGOM, I. (2003): "Estimation of Seminonparametric Models when the Criterion function is not Smooth," *Econometrica*, 71, 1583-1600.

CRAGG J.G. (1983): "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 51, 751-764.

COUGHLANM, ANNE T., AND RONALD M. SCHMIDT (1985): "Executive Compensation, Management Turnover, and Firm Performance: An Empirical Investigation," *Journal of Accounting and Economics*, 7, 43-66.

DAILY, CERTO, AND DALTON (2000): "A decade of corporate women: Some progress in the boardroom, non in the executive suite," *Strategic Management Journal*, 20: 93-99.

DENIS, DAVID J., DENIS, DIANE K., AND SARIN, A. (1997): "Ownership structure and top executive turnover," *Journal of Financial Economics*, 45: 193--221.

DONALD, S.G., IMBENS, G.W., NEWEY, AND W.K. (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55-93.

FAMA, EUGENE, AND MICHAEL JENSEN (1983): “Separation of Ownership and Control,” *Journal of Law and Economics*, 26, 301-325.

FARRELL, K.A., AND WHIDBEE, D.A. (2000): “The consequences of forced CEO succession for outside directors,” *Journal of business*, 73: 597-627.

FENTON, V., AND GALLANT, A.R. (1996): "Convergence Rate of SNP Density Estimators," *Econometrica*, 64, 719-727.

GALLANT, A. RONALD AND NYCHKA, W. DOUGLAS (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55(2), 363-390.

GAMSON, W.A., AND SCOTCH, N.A. (1964): “Scapegoating in baseball,” *American Journal of Sociology*, 70: 69-72.

GEMNELLA, A., AND HAMBRICK, D. (1993): “Effects of Executive Departure on the Performance of Acquired Firms,” *Strategic Management Journal*, 14 (summer special issue), 137-152.

GIBBONS, ROBERT S., AND KEVIN J MURPHY (1990): “Relative Performance Evaluation for Chief Executive Officers,” *Industrial and Labor Relations Review*, 43, 30S-51S.

G. GOFFEE, L. LOWENSTEIN, AND S. ROSE-ACKERMAN (EDS.) (1988): *Knights, Raiders, and Targets: The Impact of the Hostile Takeover*, New York: Oxford University Press.

GOYAL, V.K., AND PARK, CHUL W. (2002): “Board leadership structure and CEO turnover,” *Journal of Corporate Finance*, 8: 49-66.

HADLOCK, CHARLES, JOEL HOUSTON, AND MICHAEL RYNGAERT (1999): “The Role of Managerial Incentives in Bank Acquisitions,” *Journal of Banking and Finance*, 23, 221-249.

HAMBRICK, D., AND CANNELLA, A. (1993): "Relative Standing: A Framework for Understanding Departures of Acquired Executives," *Academy of Management Journal*, 36, 733-742.

HAMBRICK, D., AND MASON, P. (1984): "Upper Echelons: The Organization as A Reflection of Its Top Managers," *Academy of Management Review*, 9, 193-206.

HANSEN, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

HANSEN, L.P., HEATON, J., AND YARON, A. (1996): “Finite-sample Properties of Some Alternative GMM Estimators,” *Journal of Business and Economic Statistics*, 14, 262—280.

HARFORD, JARRAD (2003): “Takeover Bids and Target Director's Incentives: Retention, Experience and Settling-up,” *Journal of Financial Economics*, 69, 51-83.

HECKMAN, J.J., AND T.E. MACURDY (1980): "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies*, 47, 47-74.

HEGARTY, W.HARVEY, AND KRUG, JEFFREY A. (2001): “Predicting who stays and leaves after an acquisition: A study of Top managers in multinational firms,” *Strategic Management Journal*, 22(2): 185-196.

HITT, M.A., LEONARD BIERMAN, KATSUHIKO SHIMIZU, AND RAHUL KOCHHAR (2001): “Direct and Moderating Effects of Human Capital on Strategy and Performance in Professional Service Firms: A Resource-based Perspective,” *Academy of Management Journal*, 44, 1, 13-27.

HONORÉ B. (1992): “Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-565.

HONORÉ B., AND HU, L. (2004): "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," *Journal of Econometrics*, 122, 293-316.

HSIAO, C. (1986): *Analysis of Panel Data*, Cambridge: Cambridge University Press.

HUSON, MARK R., PARRIONA, ROBERT, AND STARKS, LURA T. (2001): “Internal monitoring mechanisms and CEO turnover: a long-term perspective,” *Journal of Finance*: 2265-2297

JENNINGS, E.E. (1971): *Routes to the executive suite*, New York: McGraw-Hill.

JENSEN, MICHAEL C. (1993): “The Modern Industry Revolution, Exit, and the Failure of Internal Control Systems,” *Journal of Finance*, 48, 831-880.

KALLINIKOS, JANNIS (1984): *Control and influence relationships in multinational corporations: the subsidiary's viewpoint*, Uppsala.

KITAMURA, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, 69, 1661-1672.

KITAMURA, Y., TRIPATHI, G. AND H. AHN (2004): "Empirical Likelihood-based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667-1714.

KOR. YASEMIN Y. (2006): “Direct and interaction effects of top management team and board compositions on R&D investment strategy,” *Strategic Management Journal*, 27(11): 1081-1099.

LEHN, KENNETH M., AND MENGXIN ZHAO (2006): "CEO Turnover after Acquisitions: Are Bad Bidders Fired?" *Journal of Finance*, LXI (4), 1759-1811.

LEE, S. Y. D., AND ALEXANDER, J. A. (1998): "Using CEO Succession to Integrate Acquired Organizations: A Contingency Approach," *British Journal of Management*, 9, 181-197.

LUBATKIN, M., SCHWEIGER, D., AND YAAKOV, W. (1999): Top Management Turnover in Related M&A's: An Additional Test of the Theory of Relative Standing," *Journal of Management*, 25(1), 55-73.

MARTIN, KENNETH J., AND JOHN J. MCCONNELL (1991): "Corporate Performance, Corporate Takeovers, and Management Turnover," *Journal of Finance*, 46, 671-687.

MORCK, R., SHLEIFER, A., AND VISHNY, R.W. (1988): "Characteristics of targets of hostile and friendly takeovers. In A.A. Auerbach, Corporate takeovers: Causes and consequences": 101-129. Chicago: University of Chicago Press.

MURPHY, KEVIN J., AND JEROLD L. ZIMMERMAN (1993): "Financial Performance Surrounding CEO Turnover," *Journal of Accounting and Economics*, 16, 273-315.

MICHAEL S. (1988): "Outside Directors and CEO Turnover," *Journal of Financial Economics*, 20, 431-460.

NEWEY, WHITNEY K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047--1070.

NEWEY, WHITNEY K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59(4), 1161-1167.

NEWEY, WHITNEY K. (1994): "A Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.

NEWEY, WHITNEY K. (1997): "Convergence rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168

NEWEY, WHITNEY K. AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

NEWEY, WHITNEY K. AND POWELL, JAMES L. (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565-1578.

OCASIO, W.(1994): "Political dynamics and the circulation of power: CEO succession in US industrial corporations, 1960-1990." *Administrative Science Quarterly*, 39: 285-312.

OSSIANDER, M. (1987): "A central limit theorem under metric entropy with L_2 bracketing," *Annals of Probability*, 15(3), 879-919.Owen, A. B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237-249.

PARONS, R.Q., AND BAUMGARTNER, J.S. (1970): *Anatomy of a merger: How to sell your company*, PrenticeHall: Englewood Clifs, NJ.

PFEFFER, JEFFREY, AND GERALD R. SALANCIK (2003): *The External Control of Organizations: A Resource Dependence Perspective*, Stanford: Stanford University Press.

PITTS, ROBERT A. (1976): “Diversification strategies and organizational policies of large diversified firms,” *Journal of Economics and Business*, 28: 181-188.

PREFFER, J. (1981): *Power in organizations*, Boston: Pitman.

QIN, J. AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300-325.

ROSEN, S. (1987): “Human Capital.” *In The New Palgrave: A Dictionary of Economics*, Vol.2, ed. by J. Eatwell, M. Milgate, and P. Newman, New York: Macmillan, 681-690.

SCOONES, DAVID, AND DAN BERNHARDT (1998): “Promotion, Turnover, and Discretionary Human Capital Acquisition,” *Journal of Labor Economics*, 16(1), 122-141.

SHEN, X. AND WONG W.H. (1994): "Convergence Rate of Sieve Estimates," *The Annals of Statistics*, 22(2), 580-615.

SHEN XIONGTONG (1997): "On Method of Sieves and Penalization," *The Annals of Statistics*, 25(6), 2555-2591.

SHLEIFER, A., AND VISHNY, ROBERT W. (2003): “Stock market driven acquisitions,” *Journal of Financial Economics*, 70, 295--311.

SIEHL, C, SMITH, D., AND OMURA, A. (1990): "After the Merger: Should Executives Stay or Go?" *Academy of Management Executive*, 4(1), 50-60.

SMITH, M. P. (1996): “Shareholder activism by institutional investors: Evidence from CalPERS,” *Journal of Finance*, 51: 227-252.

SONNENFELD, J. (1988): *The hero's farewell: What happens when CEOs retire*, New York: Oxford University Press.

THOMPSON, J.D. (1967): *Organizations in action*, New York: McGraw-Hill.

VANCIL, R.M. (1987): *Passing the baton*, Boston: Harvard University Press.

VAN DER VAART, A. (1991): "On Differentiable Functionals," *The Annals of Statistics*, 19, 178-204

VIRANY, B., TUSHMAN, M., AND ROMANELLI, E. (1992): "Executive Succession and Organization Outcomes in Turbulent Environments: An Organizational Learning Approach," *Organization Science*, 3, 72-91.

WALKING, R., AND LONG, M. (1984): "Agency Theory, Managerial Welfare, and Takeover Bid Resistance," *Rand Journal of Economics*, 15, 54-68.

Wall Street Journal, (1997): *CEOs have less say in designating successors*, June 24: B1, B2.

WALSH, J. (1988): "Top Management Turnover Following Mergers and Acquisitions," *Strategic Management Journal*, 9, 173-183.

WALSH, J. (1989): "Doing A Deal: Merger and Acquisition Negotiations and Their Impact upon Target Company Top Management Turnover," *Strategic Management Journal*, 10, 307-322.

WALSH, J., AND ELLWOOD, J. (1991): "Mergers, Acquisitions, and the Pruning of Managerial Deadwood: An Examination of the Market for Corporate Control," *Strategic Management Journal*, 12, 201-218.

WARNER, JEROLD B., ROSS L. WATTS, AND KAREN H. WRUCK (1988): “Stock Prices and Top Management Changes,” *Journal of Financial Economics*, 20, 461-492.

WEISBACH, MICHAEL S. (1988): “Outside directors and CEO turnover,” *Journal of Financial Economics*, 20: 431-460.

WILLIAMSON, O. (1985): *The Economic Institutions of Capitalism*, New York: Free Press.

WONG, H.W., AND SEVERINI (1991): "On Maximum Likelihood Estimation in Infinite Dimensional Parameter Spaces," *The Annals of Statistics*, 19, 603-632

YERMACK, DAVID (1996): “Higher market valuation of companies with a small board of directors,” *Journal of Financial Economics*, 40: 185-211.

BIOGRAPHICAL SKETCH

Qiong Zhou was born in November 1981, in China. She received her Bachelor of Arts degree in economics and Bachelor of Science degree in mathematics in June 2003 from Huazhong University of Science and Technology. She started graduate school at the University of Florida in 2003. She got her M.A. in economics in 2005 and Ph.D. in economics in 2009 from the Department of Economics, Warrington College of Business Administration. In 2009, she began employment at Shanghai University of Finance and Economics. Her research interests include econometrics, strategic management, industrial organization, and applied economics.