EXPECTED FEES FOR MANAGED FUTURES, OPTIMAL FEEDER CATTLE PRICE SLIDES, AND AGGREGATE vs. DISAGGREGATE DATA IN MEASURING SCHOOL EFFECTIVENESS

By

FRANCISCA G-C. RICHTER

Bachelor of Science
Pontificia Universidad Catolica del Peru
Lima, Peru
1995

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1997

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2000

EXPECTED FEES FOR MANAGED FUTURES,

OPTIMAL FEEDER CATTLE PRICE SLIDES,

AND AGGREGATE vs. DISAGGREGATE

DATA IN MEASURING SCHOOL

EFFECTIVENESS

Thesis Approved:

____
Thesis Adviser

____

____

____

____
Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to the many people that helped me during these years of study. I sincerely thank my advisor, Dr. B. Wade Brorsen for his patience as well as his clever and always optimistic guidance throughout this project. I also wish to thank my other committee members, Dr. Leroy Folks, Dr. Harry Mapp, and Dr. Brian Adam for their help, encouragement, and recommendations.

My appreciation also goes to Dr. Barry K. Moser. His teachings on Linear Models, as well as his suggestions allowed a simpler derivation of estimators in the school effectiveness essay.

Finally, all my gratitude goes to my husband Hanz, and my parents, Ramón and Josefa García-Cobián; to them I owe and offer this project.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Paper I

Estimating Fees for Managed Futures:

A Continuous-Time Model with a Knockout Feature

Estimating Fees for Managed Futures:

A Continuous-Time Model with a Knockout Feature

Abstract

Past research regarding incentive fees based on high-water marks has developed models for the specific characteristics of hedge funds. These theoretical models have used either discrete time or a Black-Scholes type differential equation. However, for managed futures, high-watermarks are measured more frequently than for hedge funds, so a continuous-time model for managed futures may be appropriate. A knockout feature is added to our continuous model which is something unique to managed futures although it could also have some relevance to hedge funds. The procedures allow deriving the distribution function for the fund's survival time, which has not been derived in past research. The distribution of the maximum until ruin is derived as well, and used to provide an estimate of expected incentive fees. An estimate of the expected fixed fee is also obtained. The model shows that the expected incentive fee would be maximized if all funds were invested in margins, but for total fees to be maximized in the presence of a knockout feature, less than half of the funds should be invested. This is precisely what fund managers do. This result suggests that designing a fund with incentive fees only may cause fund managers to adopt the highest leverage, and thus, highest risk possible.

Key Words: hedge funds, managed futures, incentive fee, high-water marks, ruin.

Estimating Fees for Managed Futures:

A Continuous-Time Model with a Knockout Feature

1. Introduction

Commodity funds are investment partnerships that pool money from investors to trade primarily futures and options on futures. Hedge funds are similar, but usually concentrate on stocks. Hedge funds use short selling, derivatives, and other strategies that a mutual fund is not permitted to use. Besides a fixed management fee of around one to three percent of returns, managers of commodity and hedge funds are paid an incentive fee. It is usually paid at the end of each period (typically a month for commodity funds and a quarter or year for hedge funds), and it consists of a fraction (around one fifth) of returns above the maximum managed capital reached in the previous periods.

Incentive fees seem to better align the manager's objective with that of investors, but the effect of this fee on managers' actions is still unclear. For example, commodity fund managers do not invest all the available capital and sometimes will not accept new money when performing well. Commodity fund managers usually invest less than half of all available funds in futures margins and keep the rest in U.S. Treasury Bills. Our aim is to better understand both fees so as to measure the benefits they provide to managers and investors. Past research in this area cannot explain why commodity fund managers invest only part of all available capital (Grinblatt and Titman; Goetzmann, Ingersoll, and Ross; Carpenter). One important difference between hedge and commodity funds is that managed capital in many commodity funds is not allowed to drop beyond a certain fixed value, usually a fraction of the initial capital. If this happens, the fund is considered ruined. We will refer to this as the *knockout* feature of the contract. A similar rule characterizes a down-and-out option, where the option becomes worthless if the asset price ever falls to or below a given barrier. Most hedge funds do not have a formal knockout rule; perhaps because most of them are considerably less risky than commodity funds. Investors could, however, follow a type of stop-loss rule, which would be equivalent to the knockout feature if they were allowed to withdraw funds on a short notice. Another analogy with option theory is worth noting. The incentive fee can be viewed as the payoff of a path-dependent option, in particular, a lookback option. In fact, Goetzmann, Ingersoll,and Ross take advantage of this analogy and obtain an expression for valuing a hedge fund contract based on the Black-Scholes model for options. Grinblatt and Titman's attempt to solve the manager's investment problem assuming that he can hedge the fee in his personal portfolio, leads to the manager opting for increasing the variance of managed capital to infinity. Carpenter argues that the manager cannot hedge the fee in his account since shorting securities that he purchases on his client's behalf is a breach of fiduciary duty. She develops a dynamic optimal trading strategy that maximizes a fund manager's expected utility of terminal wealth. In her model, when the asset value decreases and bankruptcy approaches, the manager increases portfolio volatility up to infinity.

However Brown, Goetzmann, and Park, find that hedge fund managers do not behave according to what would be expected by the theory. On the contrary, they find that managers do not increase portfolio variance when performing poorly. They argue that a possible explanation to this is that fund managers engage in a trade-off between maximizing the option-like feature of their contract and avoiding ruin because of the high cost involved in closure. However, ruin, made explicit as a knockout feature, is not considered in the previous models. Ruin will not only reduce the size of the incentive fee by reducing the life of the fund, but it will also lead to not receiving the fixed fee beyond the time of ruin.

The purpose of this paper is to develop a model that will account for this knockout feature in order to analyze its influence on the incentive fee, fixed fee, and therefore management performance. The continuous-time approach allows obtaining complete distributions for the life of a fund and the maximum until ruin. The distribution of the life of a fund is not found in the literature since the knockout feature has not been modeled previously. For the sake of simplicity, it is assumed that the incentive fee is paid all at once and at the end, eliminating in this way any possible periodic withdrawal. The expected fixed fee will be higher as the probability of ruin becomes smaller, since it is an increasing function of the life of the fund. Graphs are included at the end that show that the knockout feature's effect on the incentive fee alone, does not explain why commodity fund managers prefer to invest only a fraction of total capital. However, once the fixed fee is considered, results indicate that managers may be better off by investing less than half of total equity in margins, when the chance of ruin is high. This result corresponds to their actual behavior. As they increase investment, the volatility of managed capital increases, ruin is more likely, and therefore total fees are reduced.

2. Continuous Models for the Life of a Fund and the Incentive Fee

We will assume that capital, managed by a commodity or hedge fund behaves according to a Brownian motion process. A Brownian motion is chosen to model managed capital rather than a geometric Brownian motion to ease the derivation of the distributional forms. The fact that there is still no consensus on what is the process that drives returns in commodity funds, (Clark) supports our decision in that respect. Let $\{X(t), t \geq 0\}$ be the process, managed capital follows. This Brownian motion is characterized by:

1. Initial capital is known and will be denoted by x_0.
2. The mean and variance parameters of the process, μ and σ^2, are also known and are both positive.
3. The fund will stop operating as soon as one of the following events occur:
    a. Managed capital falls below a fixed amount $a < x_0$, or
    b. t^* units of time have passed since the process started[1].
4. The incentive fee that the manager will receive consists of a fraction of the maximum reached by the process during its life minus the initial capital.

Since we are interested in estimating this incentive fee as well as the average life of the fund we will proceed as follows:

1. Obtain the distribution of T, the time at which a fund stops operating, i.e., the life of the fund.

---

[1] Notice we can always let t^* be big enough so as to relax this restriction and have as the only reason for closure going below a. However, the restriction is necessary since without it, the probability of the life of a fund being infinite would be positive and we would not have a proper distribution, i.e., one that integrates to one in the open interval $(-\infty, \infty)$.

2. Obtain the distribution of M_T, the maximum up to this stopping time T, find the expected value of M_T and therefore, have the expected incentive fee as a function of parameters x_0, μ, σ^2, and t^*.

3. The Probability Distribution of *T*, the Life of a Fund

Let $\{X(t), t \geq 0\}$ be a Brownian motion process with mean and variance parameters $\mu > 0$ and $\sigma^2 > 0$ starting at x_0. Let T_a be the time the process reaches $a < x_0$, if it ever does. Under certain conditions, the distribution of this random variable is known to be inverse Gaussian and as we shall see, it will be useful in deriving the distribution of the life of a fund. The life of a fund, T can be defined as follows:

$$T = \begin{cases} T_a & \text{if } T_a \in (0, t^*) \\ t^* & \text{if } (T_a > t^* \text{ and } T_a < \infty) \text{ or } (T_a = \infty) \end{cases}, \tag{1}$$

where T_a is the time at which managed capital would fall to \$$a$ and t^* is the longest period of (fixed) time a fund will operate. The distribution of the first hitting time, T_a, has been thoroughly studied and its derivation as well as its properties can be found in Chhikara and Folks. Since we are considering a process with positive drift, there is a chance that the process will never hit a, which is below x_0. We will call this event $[T_a = \infty]$ and

$P(T_a = \infty) = 1 - \exp(2\mu(a - x_0)/\sigma^2)$.

Excluding this possibility, i.e., given that a will be reached at some finite point in time, the distribution of the time the process first hits a is inverse Gaussian with parameters $-\nu$ and λ, which are functions of the parameters of the process: x_0, μ, σ^2, and a. The notation commonly used is the following:

$$T_a \mid (T_a < \infty) \sim IG(-\nu, \lambda), \quad \text{where } \nu = \frac{(a - x_0)}{\mu}, \ \lambda = \frac{(a - x_0)^2}{\sigma^2}.$$

In order to simplify notation we will refer to $T_a \mid (T_a < \infty)$ by T_a^c where c refers to the conditional nature of this random variable. The cumulative distribution function of T_a^c can be expressed in terms of the parameters of the process as follows:

$$F_{T_a^c}(t; \mu, \sigma^2, x_0, a) = \Phi\left(\frac{\mu t - x_0 + a}{\sigma\sqrt{t}}\right) + \exp(2\mu(x_0 - a)/\sigma^2)\Phi\left(\frac{-\mu t - x_0 + a}{\sigma\sqrt{t}}\right) \quad (2)$$

where $\mu, \sigma^2 > 0, \quad x_0 > a > 0$.

Knowing this we can go back to derive the distribution of the life of a fund. We will start by obtaining the probability that the fund stops operating at T_a. We want $P(T < t)$ for $t \in (0, t^*)$.

By definition of T, and since $t^* < \infty$, the following holds:

$$\begin{aligned} P(T < t) &= P(T_a < t \quad \text{and} \quad T_a < \infty) \\ &= P(T_a^c < t)\, P(T_a < \infty) \\ &= F_{T_a^c}(t)\, P(T_a < \infty) \end{aligned}$$

Thus, the cumulative distribution function (CDF) of the life of a fund is

$$F_T(t) = \begin{cases} e^{2\mu(a - x_0)/\sigma^2}\Phi\left(\frac{\mu t - x_0 + a}{\sigma\sqrt{t}}\right) + \Phi\left(\frac{-\mu t - x_0 + a}{\sigma\sqrt{t}}\right) & \text{if} \quad t \in (0, t^*) \\ 1 & \text{if} \quad t = t^* \end{cases} \quad (3)$$

From this, we can obtain the probability that the fund stops operating at t^*:

$$P(T = t^*) = 1 - e^{2\mu(a - x_0)/\sigma^2}\Phi\left(\frac{\mu t - x_0 + a}{\sigma\sqrt{t}}\right) - \Phi\left(\frac{-\mu t - x_0 + a}{\sigma\sqrt{t}}\right).$$

Another way to obtain this last result is deriving it directly from the definition of *T*. Since $\left[T_a > t^* \text{ and } T_a < \infty\right]$ and $\left[T_a = \infty\right]$ are disjoint events, the following holds:

$$P\left(T = t^*\right) = P\left(T_a > t^* \text{ and } T_a < \infty\right) + P\left(T_a = \infty\right)$$

Multiplying and dividing the first term of the right hand side by $P\left(T_a < \infty\right)$, allows us to express $P\left(T = t^*\right)$ in terms of the CDF of T_a^c (which is known) in the following way:

$$\begin{aligned} P\left(T = t^*\right) &= P\left(T_a > t^* \text{ and } T_a < \infty\right)\frac{P\left(T_a < \infty\right)}{P\left(T_a < \infty\right)} + P\left(T_a = \infty\right) \\ &= P\left(T_a^c > t^*\right)P\left(T_a < \infty\right) + P\left(T_a = \infty\right) \\ &= \left(1 - F_{T_a^c}(t^*)\right)P\left(T_a < \infty\right) + P\left(T_a = \infty\right) \end{aligned}$$

Since $P\left(T_a < \infty\right) + P\left(T_a = \infty\right) = 1$, we can multiply through and further simplify the expression to obtain $P\left(T = t^*\right) = 1 - F_{T_a^c}(t^*)P\left(T_a < \infty\right)$.

4. The Probability Distribution of M_T, the Maximum up to Time *T*

Our next aim is to derive the distribution of the maximum value reached by managed capital during the life of a certain fund. Since the incentive fee is a fraction of the difference between this maximum and the initial capital, such a fee could be estimated by $\alpha\left(E\left(M_T\right) - x_0\right)$ where $E\left(M_T\right)$ is the expected value of the maximum, x_0 the initial capital and α, a number between zero and one. In what follows, the distribution of M_T will be obtained based on its relationship with another functional of a Brownian motion whose distribution and properties have already been derived and summarized by Borodin and Salminen (p.233). This functional represents the time a process takes to exit a given interval, assuming that the initial point is contained in the interval. Denote by $T_{a,m}$ the first exit time from the interval *(a, m)* of the Brownian Motion process with positive drift and with parameters previously defined; its PDF is given by[2]

$$P\left(T_{a,m} \in dt\right) = e^{-\frac{\mu^2 t}{2\sigma^2}} \left(e^{\frac{\mu(a-x_0)}{\sigma^2}} ss\left(t; \frac{m-x_0}{\sigma}, \frac{m-a}{\sigma}\right) + e^{\frac{\mu(m-x_0)}{\sigma^2}} ss\left(t; \frac{x_0-a}{\sigma}, \frac{m-a}{\sigma}\right) \right) dt \qquad (4)$$

where the function $ss(t;u,v)$ is defined as

$$ss(t;u,v) = \sum_{-\infty}^{\infty} \frac{v-u+2kv}{\sqrt{2\pi}\, t^{3/2}} \exp\left(-\left(v-u+2kv\right)^2 / 2t\right), \qquad u < v$$

The first term of the density function is the probability of exiting at *(t, dt)* through *a* and the second term, the probability of exiting at *(t, dt)* but through *m*. To obtain the PDF of M_T, notice that the following equivalence holds:

$$\begin{aligned} \left[M_T > m\right] &\equiv \left[\text{max imum up to} T \text{ is greater than } m\right] \\ &\equiv \left[m \text{ is reached before } \min\{t_a, t^*\}\right] \\ &\equiv \left[m \text{ is reached before } a \text{ and } t^*\right] \\ &\equiv \left[\text{exit through interval } (a,m) \text{ occurs through } m \text{ and before } t^*\right] \\ &\equiv \left[T_{a,m} < t^* \text{ and } X(T_{a,m}) = m\right] \end{aligned} \qquad (5)$$

where t_a is the time it takes to reach *a*. From this, $P\left(M_T > m\right)$ can be expressed as $P\left(T_{a,m} < t^* \text{ and } X\left(T_{a,m}\right) = m\right)$ for all $m>x_0$. This probability can be obtained by integrating the second term of equation (4). When $m=x_0$, the equivalence in (5) does not make much sense since at the initial point, the process is already at the upper limit of the interval. We

[2] When $\sigma^2 = 1$, the density below becomes the one that Borodin and Salminen present in their handbook.

can say that the probability that the process exits (a, x_0) before t^* and that it does so from x_0, is one, therefore, $P(M_T < x_0) = 0$ as expected. Finally, we can express the CDF of M_T, the maximum capital reached by a fund during its life as

$$F_{M_T}(m) = \begin{cases} 1 - \int_0^{t^*} e^{-\frac{\mu^2 t}{2\sigma^2} + \frac{\mu(m-x_0)}{\sigma^2}} ss\left(t; \frac{x_0 - a}{\sigma}, \frac{m-a}{\sigma}\right) dt & if\ m > x_0 \\ 0 & if\ m = x_0 \end{cases} \tag{6}$$

Since M_T is a positive random variable, its expected value may be obtained by integrating its CDF in the following way:

$$E(M_T) = x_0 + \int_{x_0}^{\infty} \left(1 - F_{M_T}(m)\right) dm \tag{7}$$

$$= x_0 + \int_{x_0}^{\infty} \int_0^{t^*} e^{-\frac{\mu^2 t}{2\sigma^2} + \frac{\mu(m-x_0)}{\sigma^2}} ss\left(t; \frac{x_0 - a}{\sigma}, \frac{m-a}{\sigma}\right) dt\ dm\ . \tag{8}$$

So once parameters μ, σ^2, and x_0 are estimated from the process followed by managed capital, the expected incentive fee can be estimated as

$$\hat{E}(\text{Incentive fee}) = \alpha \int_{x_0}^{\infty} \int_0^{t^*} e^{-\frac{\hat{\mu}^2 t}{2\hat{\sigma}^2} + \frac{\hat{\mu}(m-x_0)}{\hat{\sigma}^2}} ss\left(t; \frac{x_0 - a}{\hat{\sigma}}, \frac{m-a}{\hat{\sigma}}\right) dt\ dm \tag{9}$$

where α, as before, is the fraction of the maximum reached by managed capital over its initial value. Slutsky's theorem guarantees this estimator to be consistent as long as $\hat{\mu}$ and $\hat{\sigma}$ are consistent for μ and σ respectively.

5. The Expected Fixed Fee

Besides the incentive fee, fund managers receive a fixed fee, usually paid at the end of each month. It consists of a percentage of the funds available at the end of the month, and it is paid as long as the fund operates.

Mathematically we could define it as $\beta \sum_{i=1}^{\|T\|} X(i)$, where β is the monthly rate and $\|T\|$ is the greatest integer smaller than or equal to *T*, the life of the fund in months. However, two difficulties arise when taking the expected value of the fixed fee.

First, with discrete payments we would have to calculate $E(X(\|T\|))$. So, to make the problem tractable we assume continuous payments. Now the fixed fee can be represented as $\lim_{m\to\infty} \frac{\beta}{m} \sum_{i=1}^{mT} X(i/m)$, since the bigger *m* is, the closer $X(\| mT \| / m)$ is to $X(T)$.

Still, there is one more difficulty with this approach. When taking the expected value, we will need to know the distribution of *X* given that ruin occurs at time *T*. But the mean process of managed funds given ruin at *T* is unknown. What is known is the mean process *after* reaching the point of ruin *a*, if the fund were to continue operating. It would be the mean of a Brownian motion process starting at *a* and having a positive slope equal to μ. We call this process $X_r(t)$.

Then, let us redefine the fixed fee in the following way:

$$ff = \lim_{m\to\infty} \frac{\beta}{m} \left[\sum_{i=1}^{mt^*} X(i/m) - \sum_{i=mT}^{mt^*} X_r(i/m) \right] \tag{10}$$

This is the total fixed fee without a knockout rule minus the forgone fixed fee due to ruin. Taking expectations we arrive at

$$E(ff) = E_T\left[E_X(ff \mid T = T)\right]$$

$$= E_T\left[\lim_{m\to\infty}\frac{\beta}{m}\left[\sum_{i=1}^{mt^*}(x_0 + \mu\, i/m) - \sum_{i=mT}^{mt^*}\left(a + \mu\left((i - mT)/m\right)\right)\right]\right]$$

$$= \beta E_T\left[x_0 t^* + 0.5\mu\left(t^*\right)^2 - a\left(t^* - T\right) - 0.5\mu\left(t^* - T\right)^2\right] \tag{11}$$

The term in brackets is the area under the mean process of a Brownian motion starting at x_0, with mean μ, minus the area under the mean process of a Brownian motion starting at time T , position *a*, with the same mean parameter, up to time t^*.

Since the distribution of the life of the fund has already been derived (recall it was a mixture of a discrete and continuous distribution), numerical integration is performed in Matlab to obtain the estimated fixed fee. The sum of estimated fixed and incentive fees constitutes the manager's total income.

6. Interpretation of Numerical Results

The distribution of the life of a fund is given so that it is possible to know how likely a fund is to survive in the presence of the knockout feature. Figure 1 shows how the mean and variance of the fund are related to the probability of ruin. Initial capital was assumed to be $1,000,000 and the knockout point, $500 thousand. It is clear that a higher mean will move capital away from the knockout value but as variance increases, the chance of ruin increases as well. If the fund manager is fired due to ruin, he loses both, fees and reputation.

The expected incentive fee, which is a fixed percentage of the expected maximum until ruin, was plotted against the proportion of equity invested in margins (figure 2). This proportion will determine the size of the mean and variance parameters of the process. Due to the complexity of the expression for the expected maximum (equation 8), numerical integration was performed to calculate the double integral. Gaussian quadrature with 96 points was used for this purpose (Davis and Rabinowitz). The computer program was implemented in Matlab.

The expected maximum is definitely an increasing function of equity invested, although it increases at a decreasing rate. This outcome suggests that fund managers would maximize their expected incentive fee by investing all available capital. Figure 3 shows how the expected fixed fee behaves for processes with different mean parameters. When fixed fees are added to incentive fees, for processes with a small mean parameter (1.5-2), the maximum total fee is reached by investing less than all available capital (figure 4). As the mean increases, ruin becomes less likely, and managers are better off investing all available capital. For the fixed fee, the fraction β is set at 0.03/12 referring to an annual rate of 0.03, and the incentive fee percentage is 20.

Managers may also want to avoid ruin in order to maintain their reputation, since this is precisely what preserves the opportunity of managing funds later.

7. Conclusions

Past research on incentive fees cannot explain why futures fund managers do not invest all their capital. A model is developed that considers a knockout feature, and both incentive and fixed fees. With incentive fees alone, fund managers would maximize fees by investing all of their capital. With fixed plus incentive fees, however, and when ruin is more likely, fund managers are better off by investing only part of their capital.

Thus, the effect of the knockout rule on total fees can explain why managers do not invest all their capital. Our results suggest that designing a hedge fund with incentive fees only would cause managers to adopt the highest leverage attainable.

Another reason for the knockout rule could be a desire to terminate funds that have a negative drift before all capital is lost. Also, it could be that since fund managers are encouraged to take higher risk due to the incentive fee, the knockout rule would keep them from taking extremely high risk. If this were the case, the knockout rule would be protecting the investor's interests.

The knockout feature in our model only determines the end of the fund whereas in reality, ruin implies loss of reputation to the manager that will definitely influence future income. Considering the loss of reputational capital is likely to reinforce our finding that managers are better off keeping part of their funds in Treasury Bills.

Our simple model was able to replicate manager's behavior by considering the effects of the knockout feature on the sum of fixed and incentive fees. This suggests that future research should consider incorporating the knockout rule into more elaborated models that include withdrawals, random benchmarks, and dynamic leverage adjustments, but that have not yet been able to explain manager's behavior.

References

Borodin, A.N. and P. Salminen. *Handbook of Brownian Motion-Facts and Formulae*. Basel, Germany: Birkhäuser 1996.

S.J. Brown, W. N. Goetzmann, and J. Park. "Conditions for Survival: Changing Risk and the Performance of Hedge Fund Managers and CTAs." Working Paper, NYU Stern School of Business, November 1997.

J.N. Carpenter. "The Optimal Dynamic Investment Policy for a Fund Manager Compensated with an Incentive Fee." Working Paper, NYU Stern School of Business, October, 1998.

R.S. Chhikara and J.L. Folks. *The Inverse Gaussian Distribution.* New York: Marcel Dekker, Inc. 1989

P. Clarke. "Managed Futures: Distributional Characteristics, Return Seasonalities and Implications for Performance Evaluation." Presented at the Fall Research Seminar, Chicago Board of Trade, December, 1996.

P. Davis and P. Rabinowitz. "Additional Abscissas and Weights for Gaussian Quadratures of High Order: Values for n=64, 80, and 96." *Journal of the National Bureau of Standards*, 60(1958): 613-14.

W. N. Goetzmann, J. Ingersoll Jr., and S. Ross. "High Water Marks." Working Paper, Yale School of Management, July 1997.

M. Grinblatt and S. Titman. "Adverse Risk Incentives and the Design of Performance-Based Contracts." *Management Science*, 35(1989): 807-822

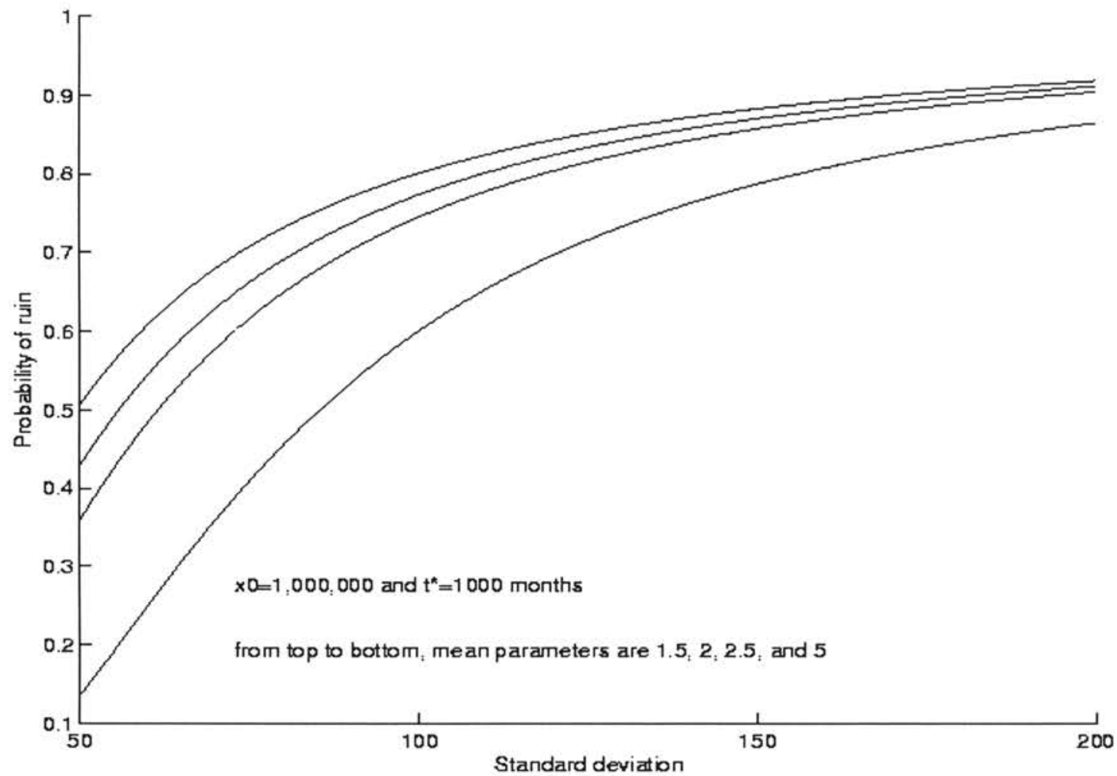

Figure 1. Probability of ruin vs. standard deviation for various mean values

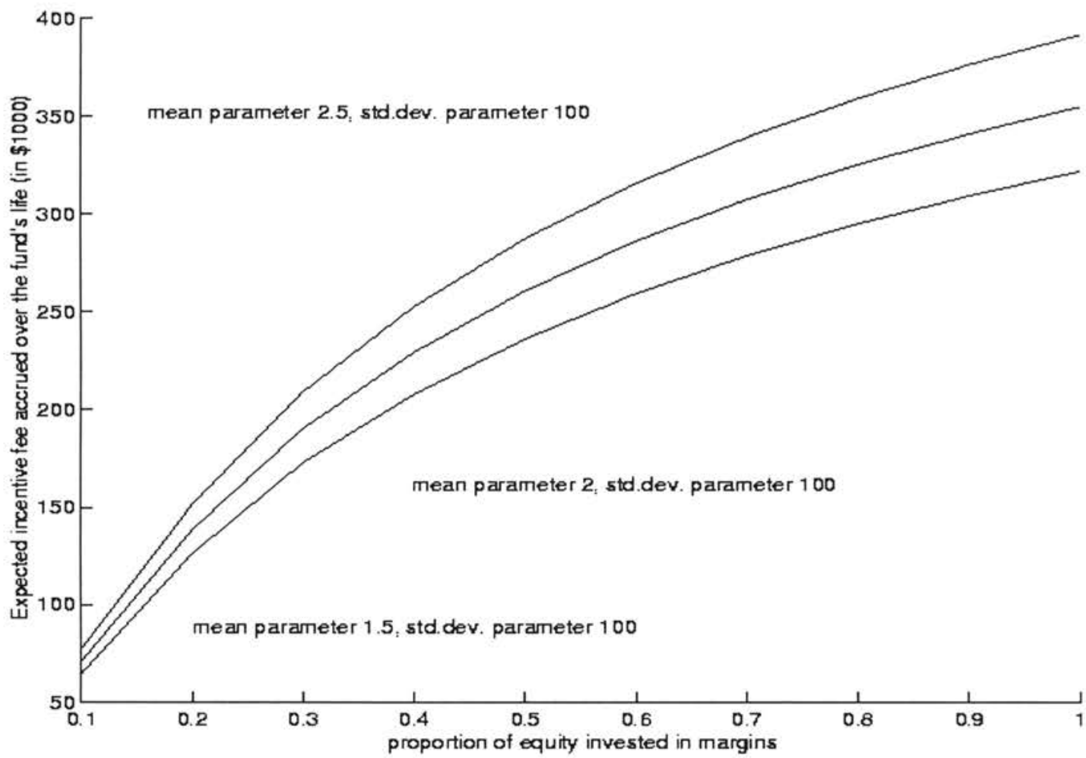

Figure 2. Expected incentive fee vs. leverage for various mean values

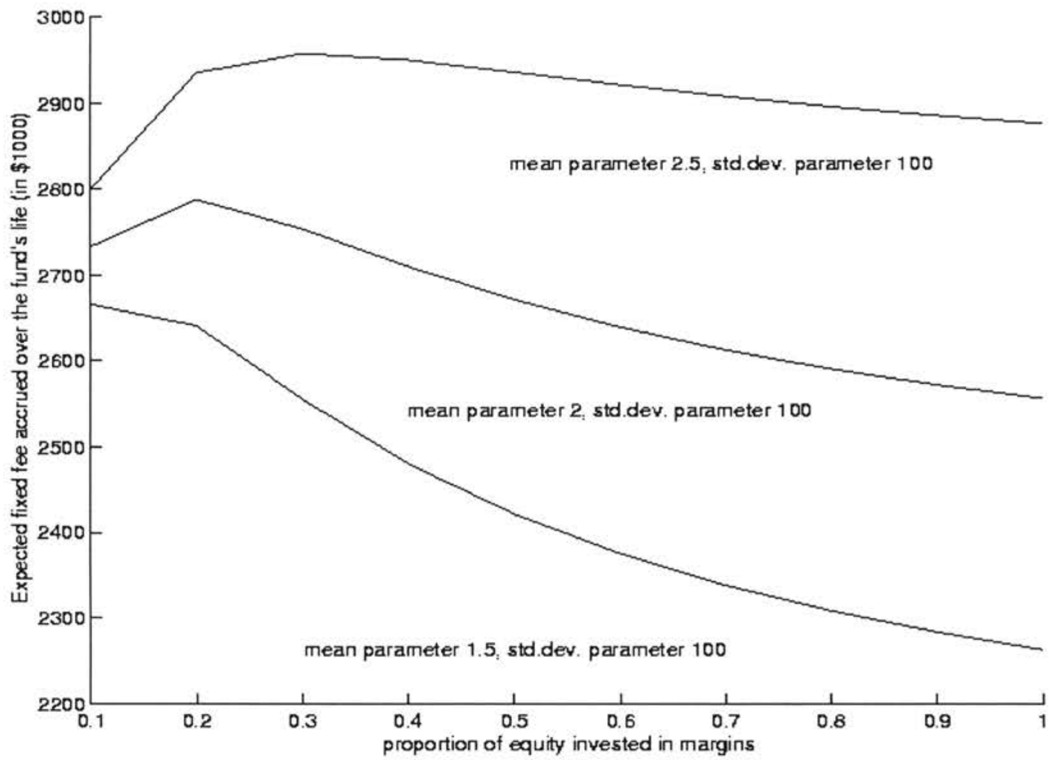

Figure 3. Expected fixed fee vs. leverage for various mean values

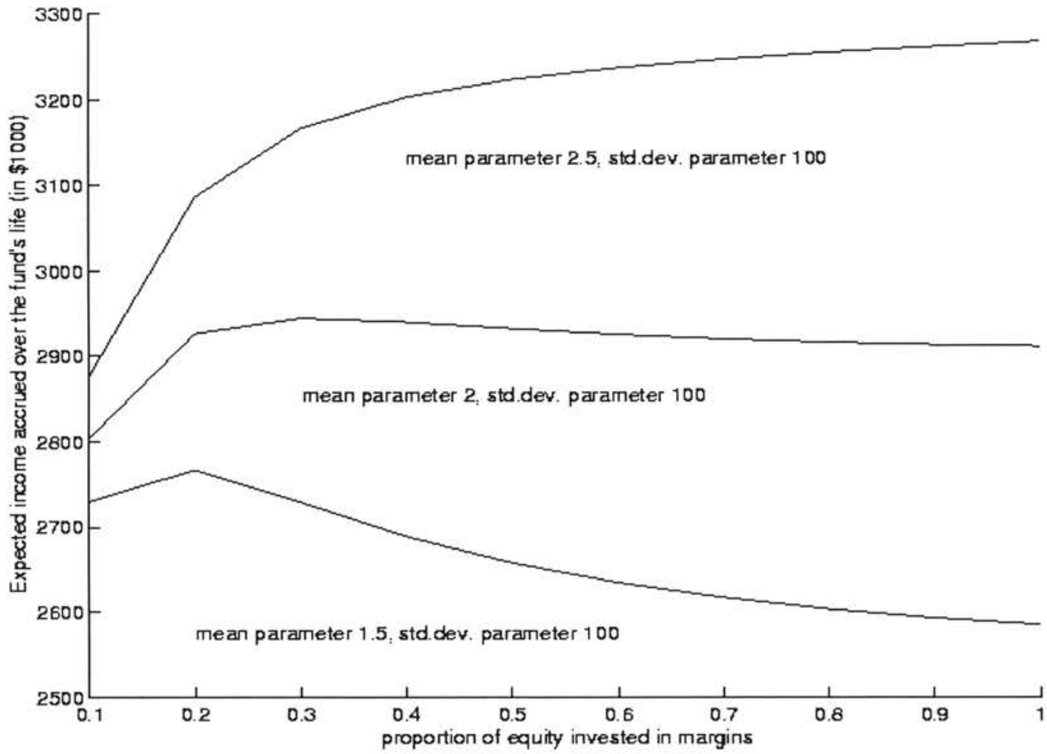

Figure 4. A fund manager's expected income in the presence of a knockout feature

Paper II

Feeder Cattle Price Slides

Feeder Cattle Price Slides

Abstract

A theoretical model is developed to explain the economics of determining price slides for feeder cattle. The contract is viewed as a dynamic game with continuous strategies where buyer and seller are the players. We determine the value of the slide that guarantees unbiased estimate of cattle weight. An empirical model using Superior Livestock Auction (SLA) data shows that price slides used are smaller than those needed to cause the producer to give unbiased estimates of weight. Consistent with the model's predictions, producers slightly underestimate cattle weights.

Key Words: asymmetric information, feeder cattle, price slide, game theory

Feeder Cattle Price Slides

1. Introduction

Feeder cattle prices normally decrease as cattle weights increase. Also, a given buyer only wants cattle that are within some weight range. Thus, feeder cattle weight is critical in determining price. Estimating the weight of cattle can be difficult for both buyers and sellers. This is especially true when cattle are sold for future delivery. In many private treaty sales the buyer never sees the cattle before purchase (although an order buyer might). In video auctions, the buyer only sees the cattle on a television screen. Thus, the seller is often better able to estimate weights than the buyer. Since sellers and buyers have asymmetric information about cattle weights, contracts need to be structured to provide sellers with an incentive to accurately estimate average delivery weights.

The usual approach to dealing with uncertain weight is to adjust the original contract price by a "price slide." The price slide (sometimes called a one-way slide) specifies the rate at which the contract price will be reduced when the average delivered weight is greater than the weight established in the contract plus a specified tolerance. With a one-way slide, no adjustment is made to the contract price if delivered cattle weigh less than the specified limit. Suppose, for example, that a producer estimates average delivered weight at 500 lbs. The producer could sell cattle at $70/cwt. with a price slide of 10 cents per cwt. for each pound of actual average weight over 520 lbs. If cattle average 530 lbs. at delivery, then $1/cwt (10 cents/cwt./lbs.×10 lbs.) is deducted from the contract price, i.e., from the $70/cwt. If, however, actual average weight is 515 lbs., no adjustment is made from the contract price.[1] The one-way price slide is an implicit option and therefore the value of the option should be reflected in the price.

Superior Livestock Auction (SLA) currently sells over a million head a year which is more than any other auction in the United States. Feeder cattle sold through SLA are sold with a price slide. Most private treaty sales also use a price slide, though price slides are not used in traditional auctions. The interest of producers in the topic is demonstrated by two extension articles (Bailey and Holmgren; and Prevatt) on price slides. However, no research has yet been done in support of extension efforts.

A contract has four essential variables: the contract price (base price), the price slide, the allowable weight difference (weight tolerance), and the estimated cattle average delivery weight (base weight). Bailey and Holmgren argued that sellers may obtain higher contract price offers if they select small allowable weight differences (or weight tolerances) and large price slides. Other important elements of the contract are time to delivery and cattle weight variability. Characteristics such as breed, sex, lot size, condition location, and frame size are also likely to be considered when setting the contract price.

Bailey, Brorsen, and Fawson found the surprising result that time to delivery has a positive effect on prices at Superior Livestock Auction (SLA), while other empirical studies on cash forward contracting have consistently found that forward contract prices decrease as time to delivery increases (e.g., Brorsen, Coombs, and Anderson; Elam). The

[1] Prevatt refers to the compensation scheme in (1) as a one-way slide. If buyer and seller had symmetric information, a two-way slide could be used where premiums are paid if cattle are lighter than expected, but one-way slides are most commonly used.

positive relationship between time to delivery and the contract price could be due to the implicit option created by the price slide.

In this paper, a theoretical model is developed to explain the economics of determining price slides for feeder cattle that will encourage sellers to accurately estimate cattle weight. The contract between buyer and seller is viewed as a dynamic game with continuous strategies where buyer and seller are the players. If, as in reality, the seller is to set the value of the price slide, necessary conditions for (subgame perfect) equilibrium can be obtained. It is also possible to determine the value of the slide (as an exogenous variable) so that equilibrium is reached when the seller gives an unbiased estimate of cattle weight. In other words, optimal values of the slide are obtained so that it is in the seller's best interest to give an accurate estimate of the cattle's weight. Our research shows that either price slides should be set higher than Prevatt suggests, or the weight tolerance needs to be lower. The model's predictions are compared to actual SLA observations. The slides used in SLA are smaller than those needed to give the producer an incentive not to underestimate weight. Consistent with the model's predictions, producers slightly underestimate cattle weights.

2. Analytical Model

Consider a feeder cattle buyer who contracts with a seller for future delivery of cattle at a price per cwt. established at the time of the contract (this is the contract price, p_0). The seller estimates the average weight of the cattle to be sold, called the base weight (y_0), and sets the price slide ($\gamma > 0$). To this the buyer responds by offering a contract price (p_0) per cwt. that maintains expected utility at zero (to simplify the model, perfect competition is assumed so that neither buyer nor seller are able to make profits). The seller then decides either to accept or reject the contract.

The price slide modifies the contract price in the following way:

$$p(y; y_0, p_0, \gamma, \delta) = \begin{cases} p_0 - \gamma(y - y_0 - \delta) & \text{if } y \geq y_0 + \delta \\ p_0 & \text{if } y < y_0 + \delta \end{cases} \tag{1}$$

The tolerance in feeder cattle weight estimation error is known as weight tolerance and is represented by $\delta > 0$. We assume that it is pre-established so that neither buyer nor seller can decide upon its value[2]. The delivery weight is given by y, and $p(y)$ is the price actually paid per cwt. at the time of delivery, when the average weight is finally revealed to buyer and seller. The payment $p(y)$ is a compensation scheme which penalizes the seller if delivered weights are greater than $y_0+\delta$. Compensation schemes of this type are used in many real world contractual relationships where asymmetric information exists (e.g., Phlips; Harris and Raviv).

Let r_s and s_b be the seller's payoff and the buyer's share of the cattle's value in total dollars, respectively:

$$r_s = \begin{cases} (p_0 - \gamma(y - y_0 - \delta))y & \text{if } y \geq y_0 + \delta \\ p_0 y & \text{if } y < y_0 + \delta \end{cases} \tag{2}$$

$$s_b = \begin{cases} (\nu(y,z) - p_0 + \gamma(y - y_0 - \delta))y & \text{if } y \geq y_0 + \delta \\ (\nu(y,z) - p_0)y & \text{if } y < y_0 + \delta \end{cases} \tag{3}$$

Here, $v(y,z)$ is the value per cwt. of the cattle when weight is known and z is a vector of other relevant variables. So, if weight were known with certainty, the buyer would be paying the real market value for cattle and his share would be zero. We assume that $v(y,z)$ is a decreasing function of y since heavier cattle are normally worth less per pound than lighter cattle. The utility of buyer and seller will depend on s_b and r_s respectively.

The contract can be viewed as a two-person dynamic game with continuous strategies that will be approached using 'backwards induction' (Gibbons; Fundenberg and Tirole). The stages of the game are as follows:

1. The seller offers an estimate of the weight y_0 and the price slide γ.
2. The buyer offers a price per cwt., the contract price p_0.
3. The seller either accepts or rejects the offer.

Assume the seller accepts the contract at stage 3. This implies that the buyer offered a contract price that, given the values of y_0 and γ (fixed for the buyer), maximizes the seller's utility while keeping the buyer's utility at its reservation level (which we will assume to be zero). So the seller knows the problem with which the buyer is confronted. If he could solve the buyer's problem, that is obtain the buyer's best response function $p_0^*(y_0,\gamma)$ that guarantees the buyer his reservation utility, the seller would be able to select the optimum values for weight and price slide, y_0^* and γ^*, so that $p_0^*(y_0^*,\gamma^*)$

---

[2] In practice, the weight tolerance is also a choice variable of the seller. For a time, the SLA did fix the weight tolerance, but quit since it was unpopular. In order to keep the model as simple as possible, weight tolerance is assumed fixed. Weight tolerances observed in the SLA do vary little for a given weight range, so the assumption is reasonable.

maximized his own utility. The existence of $p_0^*(y_0,\gamma)$, y_0^*, and γ^* would guarantee sub-game perfection and the problem would be solved.

Unfortunately, although the seller knows the general problem faced by the buyer, he is not able to 'rationally guess' the buyer's subjective probability distribution of cattle weights (note that the distribution of weights is crucial in calculating expected profit or utility). Most likely the buyer will choose a distribution of weights based on the information given by the seller (y_0 and γ). With this probability distribution, the buyer is to obtain his best response function $p_0^*(y_0,\gamma)$. Assuming the distribution is normal, we obtain a necessary condition for optimally selecting the price slide that suggests the price slide should be bigger than the market's weight discount and not equal to it as has been proposed by Prevatt. Closed form solutions for the optimal slide problem, however, would require estimating two more parameters: the buyer's estimate of the cattle's mean weight and variance (the empirical section, however, does use regression to estimate the buyer's subjective mean and variance of cattle weights). With less equations than variables, solutions are no longer possible, but future research could focus on trying to determine general characteristics of the buyer's subjective distribution function that would allow for equilibrium solutions to the problem.

Still, with further assumptions on the probability distribution of weights used by buyer and seller, values for the slide are found *as if determined exogenously*, such that the seller has no incentive to give an erroneous estimate of the cattle's weight.

2.1. A lower bound for the price slide

The buyer's problem is to find p_0 that makes his expected utility from the transaction equal to his reservation utility, i.e.:

$$\int_{y_{\min}}^{y_{\max}} u_b(r_b(p_0;y_0,\gamma))f_b(y;y_0,\gamma)dy = 0\,, \tag{4}$$

where y_0 and γ are taken as constants, u_b is the buyer's utility function, and f_b is the buyer's subjective density function of cattle weights. Assume the solution to the buyer's problem is the best response function $p_0^*(y_0,\gamma)$; then the seller's problem is to find y_0 and γ that satisfy

$$\max_{y_0,\gamma} \int_{y_{\min}}^{y_{\max}} u_s(r_s(p_0^*,y_0,\gamma))f_s(y;y_0,\gamma)dy\,, \tag{5}$$

where u_s is the seller's utility as a function of his payoff and f_s is the density function of cattle weights according to the seller's beliefs. Assuming risk neutrality we can directly substitute (3) into (4) and express the buyer's problem as

$$\int_{y_{\min}}^{y_0+\delta}(\nu(y)-p_0)y\,f_b(y)dy \quad + \quad \int_{y_0+\delta}^{y_{\max}}(\nu(y)-p_0+\gamma(y-y_0-\delta))y\,f_b(y)dy \quad = 0\,,$$

or

$$\int_{y_{\min}}^{y_{\max}}\nu(y)y\,f_b(y)dy \quad + \quad \int_{y_0+\delta}^{y_{\max}}\gamma(y-y_0-\delta)y\,f_b(y)dy \quad = p_0\int_{y_{\min}}^{y_{\max}} y\,f_b(y)dy\,.$$

Rearranging, we have that the contract price should satisfy

$$p_0 = \frac{E_b(\nu(y)y)+\gamma\int_{y_0+\delta}^{y_{\max}}(y-y_0-\delta)yf_b(y)dy}{E_b(y)}\,, \tag{6}$$

where E_b denotes the expectation with respect to the buyer's density function of cattle weights.

This result indicates that the contract price is the buyer's expected value of cattle per cwt. plus the discount per cwt. the buyer expects due to the slide. In other words, the buyer includes the expected discount in the contract price.

To analyze the seller's problem, one more assumption is made; the seller's distribution of weights is assumed normal with unspecified parameters μ and σ^2. So, by substituting equation (2) into (5), the seller's problem can be expressed as:

$$\max_{y_0,\gamma} \quad E(r_s) = p_0^* \,(y_0,\gamma)\,\mu - \frac{\gamma}{\sigma\sqrt{2\pi}} \int_{y_0+\delta}^{y_{max}} (y - y_0 - \delta)\, y \exp(-\frac{(y-\mu)^2}{2\sigma^2}) dy \,. \qquad (7)$$

The first order conditions are $\partial E(r_s)/\partial y_0 = 0$, and $\partial E(r_s)/\partial \gamma = 0$. Thus a necessary condition is:

$$\partial E(r_s)/\partial y_0 = \mu \,\partial p_0^* / \partial y_0 + \gamma \int_{y_0+\delta}^{y_{max}} \tfrac{1}{\sigma\sqrt{2\pi}}\, y \exp(-\tfrac{(y-\mu)^2}{2\sigma^2})\, dy = 0\,.$$

Note that the integrand in this expression is the one for the expected value of *y*, although the integral is computed over only part of the range. Thus, realizing that the following inequality holds

$$\tfrac{1}{\sigma\sqrt{2\pi}} \int_{y_0+\delta}^{y_{max}} y \exp(-\tfrac{(y-\mu)^2}{2\sigma^2})\, dy < \tfrac{1}{\sigma\sqrt{2\pi}} \int_{y_{min}}^{y_{max}} y \exp(-\tfrac{(y-\mu)^2}{2\sigma^2})\, dy = \mu\,,$$

we can rearrange the first order condition as follows:

$$-\frac{(\partial p_0^* / \partial y_0)}{\gamma} = \frac{\dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{y_0+\delta}^{y_{max}} y \exp(-\frac{(y-\mu)^2}{2\sigma^2}) dy}{\mu} < 1\,.$$

Therefore, we have as a necessary (but not sufficient) condition for optimality that

$$\gamma > -(\partial p_0 * / \partial y_0). \tag{8}$$

Equation (8) says that the slide should be set above the absolute value of the slope of the buyer's best response function. Since $p_0^*(y_0, \gamma)$ is expected to be very close to the real value of the cattle, $v(y)$, equation (8) also suggests that the slide should be greater than the market's weight discount and not equal to it as extension articles have suggested.

2.2. A price slide that provides incentives for unbiased estimates of cattle weight

Now let us further assume that the slide could be determined by a 'supervising entity' in order to promote fair contracts. Rather than letting the seller set the price slide value to his own convenience, we would like to set the value of the slide so that equilibrium is reached when the seller gives an unbiased estimate of the weight.

The seller's revenue in equation (7) is still maximized but now, γ is set exogenously. Now the seller has only the base weight as a choice variable, and thus only one first-order condition: $\partial E(r_s) / \partial y_0 = 0$. This condition gives base weight y_0 as a function of the price slide. If we want the base weight to be the real mean weight, then impose $y_0 \equiv \mu$, and solve for γ^{**}. The price slide γ^{**} is the value of the slide that makes the seller accurately estimate cattle weight. Since the buyer knows that the price slide is not a variable for the seller any more, he will take the seller's estimate of weight as the mean of his own subjective probability distribution of weights; i.e., $\mu_b \equiv y_0$. The variance of weights is assumed to be equal for buyer and seller, σ^2.

With γ^{**}, the seller optimizes revenue (by satisfying his first order condition) by letting his real estimate of the weight μ be the base weight y_0. The value of the slide derived by proceeding in this way, and after simplification is:

$$\gamma^{**} = \frac{\int_{y_{\min}}^{y_{\max}} \nu(y) y \left(\frac{1}{\mu} - \frac{y-\mu}{\sigma^2}\right) \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) dy}{\int_{\mu+\delta}^{y_{\max}} (y-\mu-\delta) y \left(\frac{1}{\mu} + \frac{y-\mu}{\sigma^2}\right) \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) dy + (\mu+\delta)\exp\left(-\frac{\delta^2}{2\sigma^2}\right)}, \tag{9}$$

where y_{min} and y_{max} are $-\infty$ and ∞ respectively in theory, but for interpretation purposes they can be understood as realistic lower and upper bounds for the weight of the animal since the probability of y values beyond those limits can be considered negligible. The derivation of equation (9) is given in the appendix. It should be noticed that this slide is an increasing function of the weight tolerance, so the smaller the weight tolerance, the smaller the slide needed to guarantee unbiased estimates of weight. Equation (9) above will be used to interpret results in the following section.

4. Empirical Models

In this section we take an empirical approach to better understand feeder cattle contracts and check the findings from the previous section. First we test whether base weights are unbiased predictors of actual weights. Then, using regression analysis we obtain estimated mean and variance equations of cattle weight at delivery based on market characteristics, delivery time, and the information the buyer can access: base weight, price slide, and weight tolerance. We assume the buyer can use these equations to obtain his subjective distribution of cattle weight, and thus the expected discount applied to the price per cwt. of cattle as well as the base price.

Since the model suggests that the contract price be a function of the discount due to the slide as well as market and cattle characteristics (equation 6), a second regression is performed that regresses the contract price on the expected discount and other characteristics. This price regression allows us to check how the slide, through the expected discount, affects the price-time to delivery relationship. Finally, the slide required for sellers to provide an unbiased estimate of the weight (γ^{**} given in equation (9)) is obtained for the data and compared to slides actually used.

The data used in this section are actual Superior Livestock Auction data for the 1987-1989 period (3688 observations) and the 1993-1994 period (2364 observations). The data contain information on lot characteristics, contract prices, base weights, and other relevant variables needed to estimate the models for cattle weighing not more than 900 pounds. The weight tolerance for the 1993-1994 data is constant at 10 lbs. and thus, does not enter as a variable in the regressions. In what follows, the 1987-1989 period will be referred to as period 1 and the 1993-1994 period, as period 2.

3.1. Test for unbiasedness

If sellers' estimates of average delivered weights are unbiased, the mean of the difference between actual and estimated delivery weights should be zero. This hypothesis is tested using a paired differences t-test.

The t-ratio of the paired t-test is 8.15 for period 1 and 7.24 for period 2. These values indicate that the actual and estimated weights are significantly different at the 5 percent level, so sellers underestimate average weights. Table 1 shows that the bias was small (5.65 lbs. and 5.37 for periods 1 and 2 respectively). Since the raw data deviate from normality we also test the bias non-parametrically with the sign test, and confirm that it is statistically different from zero at the 5% level.

3.2. Weight bias and weight variability

Recall the buyer is to propose a contract price. According to equation (6), the contract price should be a function of the expected discount and the expected market value of the cattle. To construct these expectations, the buyer needs to assume a probability distribution of weights. If he assumes normality, it is necessary to estimate weight mean and variance based on the information received from the seller and some other information accessible to him.[3] Although we did not derive or assume values for these

[3] Other variables such as dummies for quarters and regions in the mean equation, a time by weight interaction in the variance equation, or a variable signaling economic conditions after signing the contract and before delivery could have been included. However, we keep the model simple since we assume the buyer cannot access all this information. Weather for instance needs to be predicted in order to include seasonality in the model. When seasonal dummies were included, parameters changed substantially across periods.

parameters in the theory section, it is possible to obtain them empirically with the data available.

The following equations are used to estimate weight mean and variance:

$$y - y_0 = \alpha_0 + \alpha_1 y_0 + \alpha_2 y_0^2 + \alpha_3 w\_tol + \alpha_4 slide + \upsilon \qquad (10)$$

where the υ's are independent and normally distributed with mean zero and variance

$$\begin{aligned} \sigma_v^2 = \exp(&\beta_0 + \beta_1 y_0 + \beta_2 y_0^2 + \beta_3 slide + \beta_4 w\_tol + \beta_5 head^{-1} + \\ &+ \beta_6 steers + \beta_7 time + \beta_8 time^2 + \beta_9 MidWest + \beta_{10} West + \\ &+ \beta_{11} South + \beta_{12} Upper + \beta_{13} WCoast + \beta_{14} LSW) \end{aligned} \qquad (11)$$

The variable y is actual weight, y_0 is base weight, *Steers* is a dummy variable for steers, *Time* denotes time to delivery, *slide* is the price slide, and *w_tol* is the weight tolerance. *MidWest, West, South, Upper*, *WCoast*, and *LSW*, are dummy variables representing the regions where the cattle are located.[4] The inverse of the number of head in the lot, $head^{-1}$, is included to capture the reduced variability from averaging over a large number of animals. Equations (10) and (11) are estimated in SAS using the Mixed procedure with the local=exp() option in the Repeated statement (i.e. maximum likelihood).

These equations are used to define the buyer's expectations about mean and variance of weight difference. With them and the base weight given by the seller, the buyer estimates cattle weight distribution at delivery and proposes a contract price.

---

[4] There are 7 regions: *MidWest:* Nebraska, Kansas, Colorado, Missouri, Illinois, and Iowa. *West*: Montana, Wyoming, Idaho, Utah, and Nevada. *South*: Mississippi, Florida, Louisiana, Alabama, Arkansas, North Carolina, Georgia, Tennessee, and Kentucky. *Upper*: South Dakota, North Dakota, Minnesota, and Wisconsin. *WCoast*: California, Arizona, Oregon, and Washington. *LSW*: Texas, Oklahoma, and New Mexico. *East*: States east of Illinois and north of Kentucky.

The parameter estimates of equations (10) and (11) are reported in table 2. The assumption that weight variability increases with time to delivery is also tested using these same equations. The parameter estimates of time to delivery in the cattle weight variance equation (table 2) indicate that time to delivery has a positive effect on the variance of base weight. The parameter for the price slide is not consistently significant in the mean equation but it is negative and significant in the variance equation for both periods, indicating that sellers do use slightly larger slides when more certain about weights. As expected, our variables are better able to explain the variability in the bias than the bias itself. In fact, Buse's R^2 is only 0.024 and 0.04 for the mean equation in the first and second period respectively. The parameter estimate for weight tolerance in the first period suggests that reducing the weight tolerance decreases the bias. The bias varies greatly with weight. With low base weights the bias tends to decrease, while high base weights tend to increase the bias. The variance is also heavily influenced by weight with sellers being much less accurate at estimating weight of light-weight cattle. The inverse of number of head per lot in the variance equation has a positive coefficient as expected indicating more error variance with small lot sizes.

3.3. Contract price

With weight mean and variance estimates obtained from equations (10) and (11), the buyer should be able to construct the expected discount per cwt., which, according to equation (6), is given by: $Ep = \gamma \int_{y_0+\delta}^{y_{\max}} (y - y_0 - \delta) y f_b(y) dy / E_b(y)$.

Assuming normality, these two parameters completely specify a distribution of weights at delivery time so that the expected discount can be estimated. The contract price is explained as follows:

$$\begin{aligned} p_0 = {} & \alpha_0 + \alpha_1 y_0 + \alpha_2 y_0^2 + \alpha_3 Head + \alpha_4 Head^2 + \alpha_5 Steers + \alpha_6 Time + \\ & + \alpha_7 Time^2 + \alpha_8 Futures + \alpha_9 MidWest + \alpha_{10} West + \alpha_{11} South + \\ & + \alpha_{12} Upper + \alpha_{13} WCoast + \alpha_{14} LSW + \alpha_{15} Ep + \sum_{i=16}^{34} \alpha_i Oc_i + u + \varepsilon \end{aligned} \qquad (12)$$

where ε has mean zero and variance σ_ε^2, whereas u is an error component associated with the day of the sale having zero mean and variance σ_u^2.

The variable p_0 is the contract price, *Ep* is the estimated expected discount per cwt.[5], *Futures* is the current price of the futures contract that will be the nearby futures at the time of delivery. *Oc* is other market and lot characteristics such as breed, flesh, and frame, *Head* is the number of head, and all other variables are defined as before. This random-effects model is estimated using the Mixed procedure in SAS. The estimated mean equation is used to plot the contract price against base weight and time to delivery.

The difference between the average expected discount and the average of actual discounts is small although statistically significant (table 1). Thus, the model used for the distribution of weights is imperfect. This could be due to some minor misspecification such as incorrect functional form or non-normality.

Buse's R^2 for the price equation is 0.83 and 0.80 for the first and second periods

[5] The use of the variable *Ep* creates a generated regressor problem. As Hoffman shows, parameter estimates are still consistent with a generated regressor, but estimates of standard errors are biased. Monte Carlo studies by Hoffman show that this bias is small (<10%) except when using a lagged dependent variable. No correction is made here since the estimated coefficient is several times its standard error.

respectively. Regression estimates are shown in table 3. The base weight was included in equation (12) in quadratic form. As seen in table 3, the parameter estimate of the base weight is negative while that of the square of the base weight is positive. Figure 1 shows the effect of the base weight on the contract price for periods 1 and 2. As expected, the contract price decreases as base weight is increased.

The effect of time to delivery on the contract price is plotted in figures 2 and 3. Figures 2 and 3 show, for periods 1 and 2 respectively, the contract price as a function of time to delivery resulting from estimating equations 12 and 11, i.e., when the estimated discount is entered as an explanatory variable. These relationships are compared, in the same figure, with the relationship between contract price and time to delivery when the estimated discount variable is not accounted for, but otherwise all other variables are kept in the model. For the first period, including the expected discount widens the range in which time to delivery has a negative effect on the contract price. Price increases with small values of time to delivery and decreases otherwise. In the actual data set, however, most of the values of time to delivery are within the range where the contract price slightly increases. An explanation of this could be that buyers in this market pay a premium to reduce their input risk or it could be that sellers wanting to sell cattle immediately must pay a liquidity cost in that most buyers are demanding cattle for future delivery. However, in the second period (figure 3), including the expected penalty does make the price negatively related to delivery time in the range of 0 to 40 days. So once the discount generated by the price slide is accounted for, the positive effect of time to delivery on price is reversed for a good part of the relevant range. The hypothesis that this positive relationship is due to the implicit option created by the price slide cannot be rejected for period 2.

The parameter for expected discount was significant and positive for both data sets (4.45 and 3.31). According to equation (6), however, a value close to one was expected. The coefficient estimate was also sensitive to changes in the specification of the weight-difference mean equation that was used to obtain the expected penalty. However, while the size of the coefficient was fragile, it always remained positive and significant, suggesting that the option-like value of the slide is indeed recognized by buyers. The parameter being greater than one means sellers have an incentive to use large price slides and low weight tolerances.

3.4. Comparing the actual price slide with the model's predictions

To see if in reality the price slide is set at the value that makes the seller want to provide an unbiased estimate of cattle's weight (according to the analytical model), an estimate of the weight discount is needed. This is taken to be the change in the contract price due to a unit increase in weight. In other words, the weight discount is estimated as the derivative of the price equation with respect to weight[6]. For periods 1 and 2 weight discounts were estimated as 4.36 and 4.87 (cents/cwt.)/lb. respectively. On average, the slide is around 1.29 times the market weight discount for the first period and 1.41 times the weight discount for the second period.[7] This suggests that there has been a tendency to increase the slide above the market weight discount over time. But when calculating

[6] We are using $\partial P_0 / \partial y_0$ as an estimate of $\partial v / \partial y$.

the slide that according to the model should give unbiased estimates of weight (equation 9), we obtain, for period 1, a 10.48 (cents/cwt.)/lb. optimal slide for the mean weight, which is about 2.4 times the corresponding weight discount. For period 2, a similar situation is shown: The optimal slide for the average base weight, according to equation (9) is 11.07 (cents/cwt.)/lb., about 2.3 times the corresponding weight discount. Thus in both cases, equation (9) indicates that either a bigger price slide, or (if the slide is to be kept at about the market weight discount) a smaller weight tolerance (since the slide is increasing in the weight slide) is needed to guarantee unbiased estimates of weight.

Since the model predicts that the price slides used at SLA are not big enough to avoid unbiased estimates of cattle weight, we would expect to see this difference in the data. In fact, as seen in table 1, actual weights are slightly larger than base weights.

4. Conclusions

Feeder cattle sold through video auctions and by private treaty are often for future delivery. Because delivery weights are not known when cattle are contracted, sellers must estimate them. Since sellers and buyers have asymmetric information about cattle weights, contracts need to be structured to provide sellers with an incentive to accurately represent their estimates of average delivery weights.

---

[7] We estimate the weight discount for each observation, obtain the slide-weight discount ratio, and average over all observations.

The usual approach to dealing with weight uncertainty is to adjust the contract price by a price slide. The analytical model provides the solution to how cattle should be valued in the presence of a price slide.

Comparative statics results show that the price slides used are not sufficient to impose unbiased predictions of cattle weights. Furthermore, empirical results confirm that the price slides used are too small to impose unbiasedness and that sellers underestimate weights.

Sellers often ask extension economists for advice on how to pick slides. The theoretical results suggest using price slides double the market's discount for weight if large weight tolerances are used, or to use small weight tolerances if a price slide close to the market's weight discount is to be used. Empirical results show sellers receive more than the expected discount in terms of higher prices. Thus, results suggest that larger price slides or lower weight tolerances should be encouraged.

Table 1. Summary Statistics of Selected Variables, Superior Livestock Video Cattle Auction Data, 1987-1989, 1993-1994.

		1987-1989 period				1993-1994 period			
Variable	**Units**	**Mean**	**Min.**	**Max.**	**Std. Dev.**	**Mean**	**Min.**	**Max.**	**Std. Dev.**
Base weight	lbs.	624.05	270.0	890.0	133.8	608.15	225.0	895.0	126.6
Actual weight	lbs.	629.70	302.8	960.4	134.2	613.52	92.5	933.3	128.4
Contract price	$/cwt.	82.65	57.0	130.0	10.4	87.42	55.0	134.5	10.8
Weight difference	lbs.	5.65	-360.1	282.4	42.1	5.37	-238.9	193.0	36.1
Price slide	(cents/cwt)/lb	5.39	3.0	10.0	2.9	7.04	1.0	15.0	2.7
Weight tolerance	lbs.	15.12	0.0	35.0	7.3	10.00	10.0	10.0	0.0
Head		131.11	9.0	2000.0	116.4	117.23	25.0	1000.0	84.7
Time to delivery	days	38.31	0.0	290.0	35.4	28.36	2.0	145.0	25.7
Discount	($/cwt.)	0.65	0.0	27.2	1.6	0.79	0.0	11.0	1.4
Estimated discount	($/cwt.)	0.69	0.2	3.5	0.5	0.90	0.2	6.7	0.5
Discount- Est. Discount		0.05	-25.7	2.2	1.5	0.12	-9.9	4.3	1.3

Note: The number of observations for the 1987-1989 period is 3688 and for the 1993-1994 period, 2364.

Table 2. Parameter Estimates of the Cattle Weight Difference (Actual Weight minus Base Weight) Mean and Variance Equations. Feeder Cattle Auction Data.

	1987-1989 period		1993-1994 period	
Variable	**Parameter Estimate**	**Standard Error**	**Parameter Estimate**	**Standard Error**
Mean equation				
Intercept	53.45*	13.52	160.94*	16.47
Slide	-0.35	0.34	-2.71*	0.50
Weight tolerance	0.45*	0.10		
Base weight	-12.61*	4.16	-38.09*	4.94
Base weight squared	0.63*	0.34	2.48*	0.41
Variance equation				
Exp(Intercept)	1071.63*	634.76	7754.48	6287.99
Slide	-0.09*	0.01	-0.11*	0.02
Weight tolerance	-5.8E-3	3.8E-3		
Base weight	-5.3E-2	0.18	-0.16	0.22
Base weight squared	-5.6E-4	0.01	5.3E-3	0.02
Steers	6.5E-2	0.05	-0.13*	0.07
Time	9E-3*	1.7E-3	1.8E-2*	3.2E-3
Time squared	-3E-5*	1.1E-5	-1.2E-4*	3.3E-5
$Head^{-1}$	0.16*	0.03	0.18*	0.08
MidWest	-6.7E-2	5.7E-2	-0.56*	0.25
West	7.5E-2	7.2E-2	-0.74*	0.26
South	-6.4E-2	8.9E-2	-0.69*	0.26
Upper	0.23	0.32	-0.58*	0.30
WCoast	0.20	0.15	-0.32	0.28
LSW			-0.74*	0.25

Notes: Asterisks denote significance at the 10 percent level.
Base weight in the regression is in cwt. and base weight squared in $cwt.^2$
Head is in hundreds and time in days.
The dependent variable is measured in lbs.
Weight tolerance is fixed at 10 lbs. for the 1993-1994 period.

Table 3. Parameter Estimates of the Contract Price Equation, Superior Livestock Video Cattle Auction Data.

	1987-1989 period		1993-1994 period	
Variable	**Parameter Estimate**	**Standard Error**	**Parameter Estimate**	**Standard Error**
Intercept	72.47*	4.59	65.31*	9.70
Futures price	0.80*	0.05	0.74*	0.09
Steers	-6.82*	0.12	-5.34*	0.19
Head	0.48*	0.09	0.74*	0.26
$Head^2$	-2.9E-2*	0.01	-6.7E-2	4.3E-2
Base weight	-12.24*	0.54	-10.33*	1.55
Base weight squared	0.63*	0.04	0.45*	0.11
Expected Discount	4.45*	0.35	3.31*	0.66
Time	5.6E-3	4.4E-3	-4.4E-3	1.2E-2
Time squared	-5.7E-5*	2.6E-5	6.2E-5	1.1E-4
English-Exotic-Cross	-0.95*	0.40	-0.20	0.66
English-Cross	-0.93*	0.41	1.36*	0.71
Exotic-Cross	-1.27*	0.45	0.92	0.75
Angus	-0.36	0.73	5.32*	1.10
Dairy	-8.85*	0.58	-14.71*	0.93
Heavy	-2.68*	1.14		
Medium Heavy	-2.73*	0.47	2.54	1.83
Medium Flesh	-2.38*	0.44	2.38	1.79
Light-Medium Flesh	-2.02*	0.47	2.63	1.79
Large Frame	6.65*	1.10	4.08*	0.87
Medium-Large Frame	5.89*	1.09	2.58*	0.62
Medium Frame	4.35*	1.17	2.00*	0.61
Horn	-1.64*	0.48	-1.11*	0.35
MidWest	0.93*	0.15	5.52*	0.75
West	1.56*	0.19	5.92*	0.80
South	-1.44*	0.22	-0.10	0.77
Upper	1.49*	0.83	5.99*	0.88
WCoast	-0.89*	0.37	1.68*	0.79
LSW			2.69*	0.77
Truck	-0.88*	0.25		
Mixed	1.62*	0.20	1.94*	0.24
Miles	-0.11*	0.02		
Sale (random effect)	4.55*	1.27	2.20*	1.01

Note: Asterisks denote significance at the 10 percent level. Base weight is in cwt., Base weight squared in $cwt.^2$, Head and Miles in hundred units.

References

Bailey, D., M.C. Peterson, and B.W. Brorsen. "A Comparison of Video Cattle Auction and Regional Market Prices." *American Journal of Agricultural Economics*, 73(May, 1991):465-475).

Bailey, D., B.W. Brorsen, and C. Fawson. "Buyer Concentration at Feeder Cattle Auctions." *Review of Agricultural Economics*, 15(January 1993):103-119.

Bailey, D., and L.N. Holmgren. "Price Slides for Feeder Cattle." Working paper, Utah State University, 1996.

Brorsen, B.W., J. Coombs, and K. Anderson. "The Cost of Forward Contracting Wheat." *Agribusiness*, 11(1995):349-354.

Elam, E. "Cash Forward Contracting Versus Hedging of Fed Cattle, and the Impact of Cash Contracting on Cash Prices." *Journal of Agricultural and Resource Economics*, 17(1992):205-217.

Fundenberg, D. and J. Tirole. *Game Theory* Cambridge, Massachusetts: The MIT Press, 1992.

Gibbons, R. *Game Theory for Applied Economists* Princeton, New Jersey: Princeton University Press, 1992.

Greene, W. *Econometric Analysis* 3rd Edition, New Jersey. Prentice-Hall, 1997.

Harris, M., and A. Raviv. "Optimal Incentive Contracts with Imperfect Information." *Journal of Economic Theory*, 20(April 1979):231-259.

Hoffman, D.L. "Two-Step Generalized Least Squares Estimators in Multi-Equation Generated Regressor Models." *Review of Economics and Statistics*, (1987):336-364.

Phlips, L. *The Economics of Imperfect Information.* Cambridge, Cambridge University Press, 1988.

Prevatt, J. W. "Marketing Feeder Cattle with a Price Slide." Circular ANR-800, Alabama Cooperative Extension Service, Department of Agricultural Economics, Auburn University, AL, 1992.

Superior Livestock Auction. Unpublished Data, Greeley, CO, 1987-1989, 1993-1994.

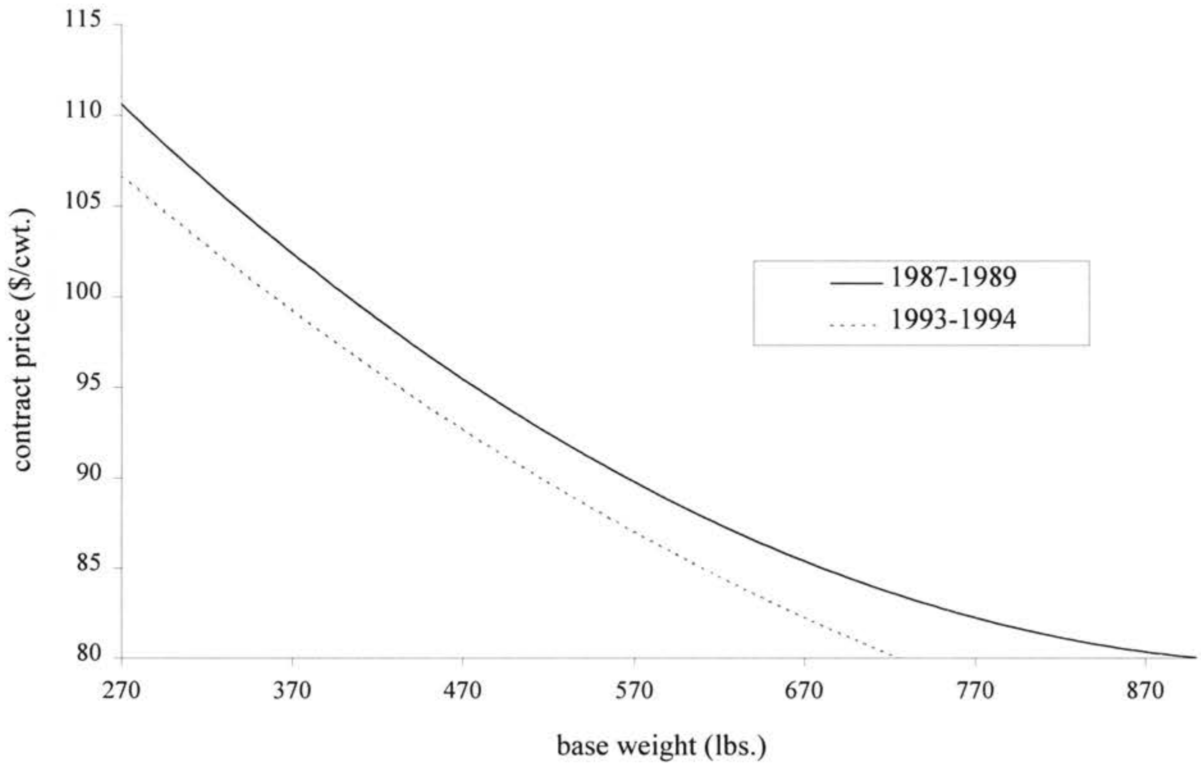

Figure 1. The effect of cattle weight on the contract price. Superior Livestock auction data.

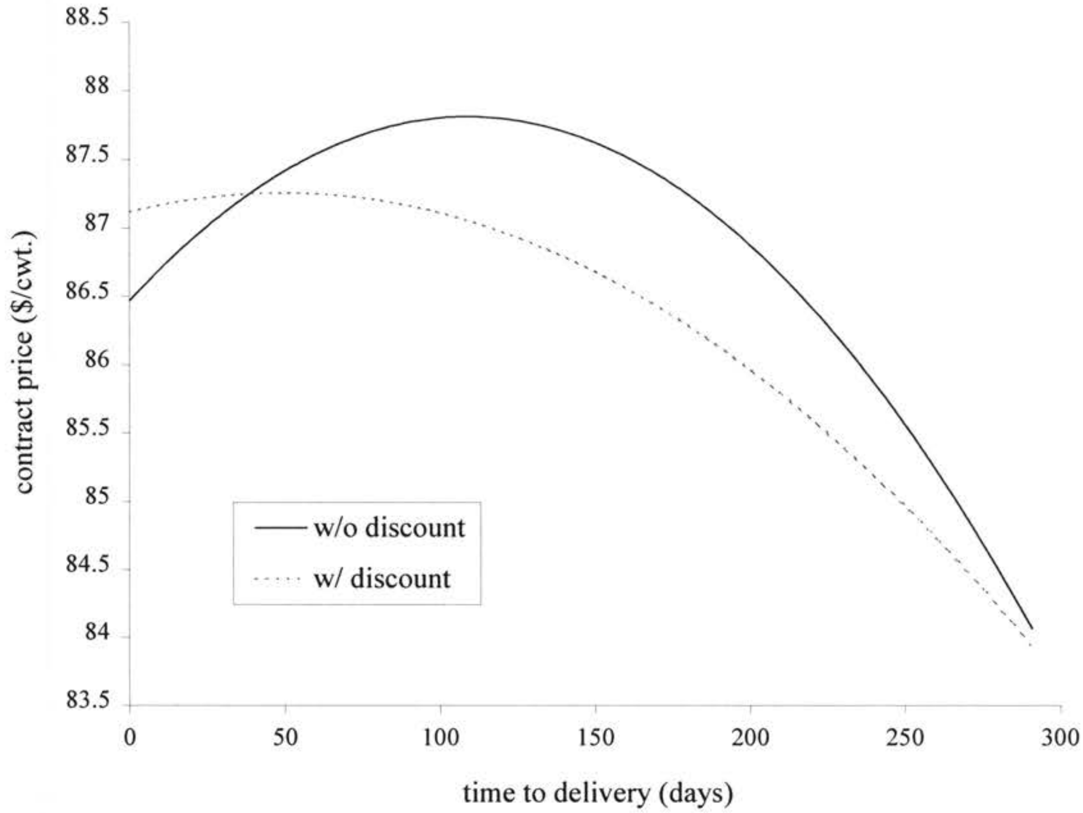

Figure 2. The effect of delivery time on the contract price, accounting and not accounting for the discount due to the slide. Superior Livestock auction data, period 1987-1989.

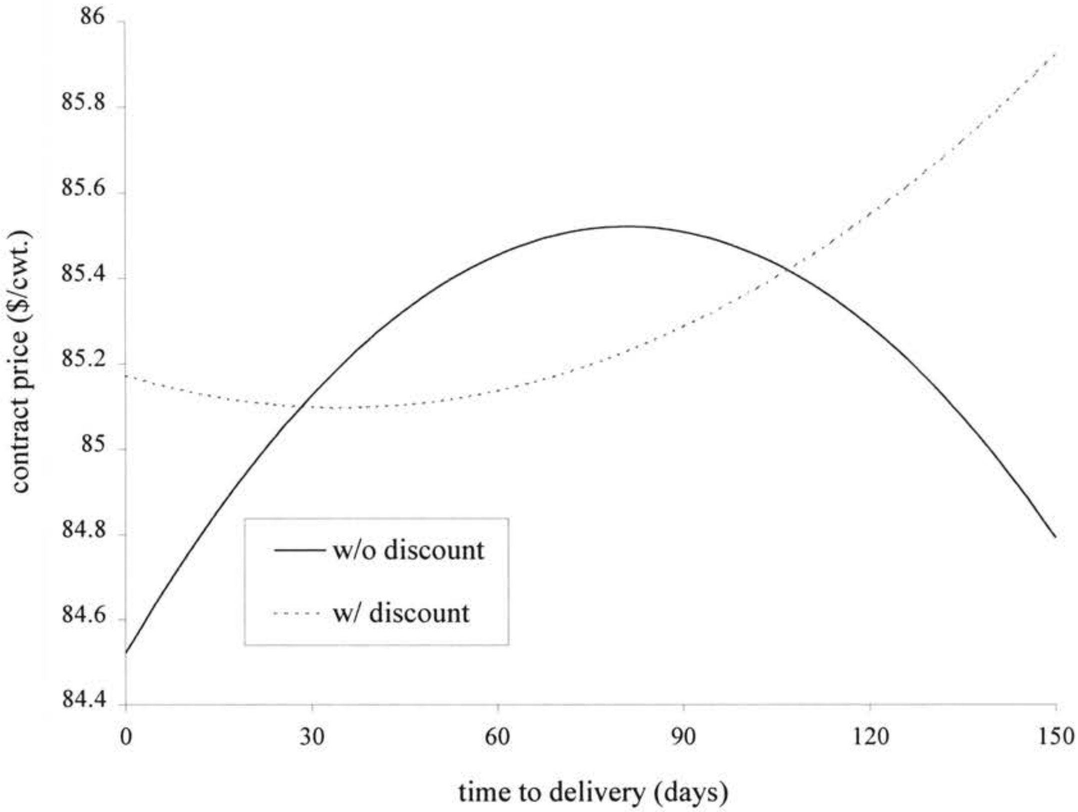

Figure 3. The effect of delivery time on the contract price accounting and not accounting for the discount due to the slide. Superior Livestock auction data, period 1993-1994.

Appendix

Derivation of equation 9.

The seller's problem is:

$$\max_{y_0} \quad E(r_s) = p_0 * (y_0, \gamma^{**})\mu - \frac{\gamma^{**}}{\sigma\sqrt{2\pi}} \int_{y_0+\delta}^{y_{\max}} (y - y_0 - \delta) y \exp(-\frac{(y-\mu)^2}{2\sigma^2}) dy, \qquad \text{(A.1)}$$

where the slide is fixed at γ^{**}.

The first order condition is:

$$\partial E(r_s) / \partial y_0 = \mu \; \partial p_0^* / \partial y_0 + \gamma^{**} \int_{y_0+\delta}^{y_{\max}} \tfrac{1}{\sigma\sqrt{2\pi}} y \exp(-\tfrac{(y-\mu)^2}{2\sigma^2}) \, dy = 0 . \qquad \text{(A.2)}$$

It is left to see what $\partial p_0^* / \partial y_0$ is. If the slide is given such that it is in the seller's best interest to accurately estimate cattle weight, then the buyer can trust the seller's estimate y_0, and take it as the mean of weight distribution. Thus, $E_b(y) = y_0$. Also, we assume both buyer and seller take the variance of weights to be σ^2.

With these assumptions we can obtain the derivative of the buyer's best response function with respect to the base weight:

$$\partial p_0^* / \partial y_0 = \frac{\partial}{\partial y_0}\left(\tfrac{1}{\sigma y_0 \sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \nu(y) y \exp(-\tfrac{(y-y_0)^2}{2\sigma^2}) dy + \gamma \int_{y_0+\delta}^{\infty} (y - y_0 - \delta) y \exp(-\tfrac{(y-y_0)^2}{2\sigma^2}) dy \right]\right)$$

$$= \tfrac{1}{\sigma y_0 \sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \nu(y) y \tfrac{(y-y_0)}{\sigma^2} \exp(-\tfrac{(y-y_0)^2}{2\sigma^2}) dy - \gamma \int_{y_0+\delta}^{\infty} y \exp(-\tfrac{(y-y_0)^2}{2\sigma^2}) dy + \right.$$

$$+ \gamma \int_{y_0+\delta}^{\infty} (y - y_0 - \delta) y \tfrac{(y-y_0)}{\sigma^2} \exp(-\tfrac{(y-y_0)^2}{2\sigma^2}) dy + \gamma (y_0 + \delta) \exp(-\tfrac{\delta^2}{2\sigma^2}) -$$

$$-\frac{1}{y_0}\int_{-\infty}^{\infty}\nu(y)y\exp(-\frac{(y-y_0)^2}{2\sigma^2})dy+\frac{\gamma}{y_0}\int_{y_0+\delta}^{\infty}(y-y_0-\delta)y\exp(-\frac{(y-y_0)^2}{2\sigma^2})dy\Bigg]$$

(A.3)

Replace (A.3) in (A.2). This gives us the base weight as a function of the slide. Therefore, if we want the value of the slide that makes the base weight equal to the real average cattle weight, μ, we need only to replace y_0 with μ, and solve for the slide, γ^{**}:

$$\partial E(r_s)/\partial y_0\Big|_{y=\mu}=\frac{1}{\sigma\sqrt{2\pi}}\Bigg[\frac{\gamma}{\mu}\int_{\mu+\delta}^{y_{\max}}(y-\mu-\delta)y\exp(-\frac{(y-\mu)^2}{2\sigma^2})\,dy- \qquad \text{(A.4)}$$

$$-\frac{1}{\mu}\int_{-\infty}^{y_{\max}}\nu(y)y\exp(-\frac{(y-\mu)^2}{2\sigma^2})\,dy+\int_{-\infty}^{\infty}\nu(y)y\frac{(y-\mu)}{\sigma^2}\exp(\frac{-(y-\mu)^2}{2\sigma^2})dy+$$

$$+\gamma^{**}\int_{\mu+\delta}^{\infty}(y-\mu-\delta)y\frac{(y-\mu)}{\sigma^2}\exp(\frac{-(y-\mu)^2}{2\sigma^2})dy+\gamma^{**}(\mu+\delta)\exp(-\frac{\delta^2}{2\sigma^2})\Bigg]=0$$

Thus, solving for the slide yields equation 9.

Paper III

Aggregate versus Disaggregate Data in the Measurement of School Quality.

Aggregate versus Disaggregate Data in the Measurement of School Quality

Abstract

Estimators of school quality based on aggregate or school level data are developed here. These estimators are compared with the commonly used OLS estimators as well as with the shrinkage estimators based on the more informative two-level data. The theoretical results are supported by a Monte Carlo experiment. Results show that for samples containing small schools (sample average may be about 100 students per school but sample includes several schools with about 30 students), the proposed estimator performs better than OLS and only slightly worse than the shrinkage estimator. While our aggregate data-estimates are not as precise as those with disaggregate data, the difference is smaller than previous research suggests. Thus, the proposed estimator should be used when school officials are unable to gather disaggregate data. The proposed estimator can also be used to measure efficiency of industrial or agricultural firms where the use of aggregate data can be a wise choice.

Aggregate versus Disaggregate Data in the Measurement of School Quality

1. Introduction

Over the last three decades, resources devoted to education have continuously increased while student performance has barely changed (Odden and Clune). In response to this fact, several states now reward and provide incentives for public schools that perform better than others, based on their own measures of school quality (Ladd). Test scores are used not only by policymakers in reward programs but are also presented in state report cards issued to each school. Already more than 35 states have comprehensive report cards reporting on a variety of issues including test scores and a comparison of school variables with district and state averages. But often the information presented is misleading or difficult to interpret. Accurate information on school performance is needed if report cards and reform programs are to succeed in improving the public school system.

Hierarchical linear modeling (HLM), a type of multilevel modeling, has been recognized by most researchers as the appropriate technique to use when ranking schools by effectiveness. As Webster argues, HLM recognizes the nested structure of students within classrooms and classrooms within schools, producing a different variance at each level for factors measured at that level. Multilevel data, also called disaggregate data is needed to implement HLM. For example, two-level data could consist of variables for students within schools. The value-added framework within the HLM methodology has become popular among researchers (Hanushek, Rivkin, and Taylor; Goldstein; Woodhouse and Goldstein). Value-added regressions are able to isolate school's effect on test scores during a given time period, by using regressors such as previous test scores, and student and school characteristics. But as of 1996, among the 46 out of 50 states that have accountability systems with some type of assessment, only 2 had used value-added statistical methodology in implementing such systems (Webster). Multilevel analysis has been said to involve complicated statistical analyses that school officials are unable to understand (Ladd).

A common approach is to use aggregate data. As opposed to having data for each student within each school, aggregate data refers to having only averages of these data over all students, within a school. School administrators may be able to obtain records of each student's individual test score but may not be able to match them with their parents' income, for example. Therefore, average test scores in a school are matched to the average income in the respective school district.

To obtain a measure of school quality with aggregate data, it is common to regress school mean outcome measures on the means of several demographic and school variables The residuals from this regression are totally attributed to the school effect, and thus, are used to rank schools. Although the use of aggregate data has been widely criticized in the literature (Webster; Woodhouse and Goldstein), many states use aggregate data, in part due to their inability to match data sets by student or computational problems such as the lack of software able to handle so many observations.

The purpose of this work is to propose a new and more efficient estimator of quality based on aggregate data, and then compare it with the commonly used OLS estimator as well as with the value-added-disaggregate estimator. Evidently, estimators based on disaggregate data will perform better than any estimator based on aggregate data. The questions that arise are: by how much will their performances differ? Should schools be using OLS, when they can use a more efficient aggregate estimate at no extra cost?

One of Goldstein's main oppositions to aggregate data models is that they say nothing about the effects upon individual students. Also, aggregate data does not allow studying differential effectiveness, which distinguishes between schools that are effective for low achieving students and schools that are effective for high achieving students. The inability to handle differential effectiveness is a clear disadvantage of aggregate as compared to disaggregate data. However, when aggregate data are all that schools have, is it still possible to detect the over and under performing schools? When using OLS on aggregate data, it has been observed that small schools are disproportionately rewarded (Clotfelter and Ladd). The estimator proposed here eliminates that bias.

Woodhouse and Goldstein argue that residuals from aggregate level regression analysis are highly unstable and therefore, unreliable measures of school efficiency. Woodhouse and Goldstein analyze an aggregate model used in a previous study and show how small changes in the independent variables as well as the inclusion of non-linear terms will change the rank ordering of regression residuals. However, their data set is small and they do not examine whether disaggregate data would have also lead to fragile conclusions.

As of today, most of the research has focused on criticizing the commonly used aggregate data model, which uses OLS residuals to estimate school quality. Goldstein, for example, illustrates the instability of aggregate data models with an example in which he compares estimates coming from an aggregate model versus estimates from several multilevel models showing they are different. The aggregate model, however, does not provide an estimate of the between-student variance, which suggests that the author does not use MLE residuals to estimate school effects. Maximum likelihood estimation is possible since the form of heteroskedasticity for the aggregate model is known (Dickens).

While it is expected that aggregation will attenuate the bias due to measurement error, few researchers have compared aggregate data models versus multilevel models while considering measurement error. Hanushek, Rivkin, and Taylor analyze the impact of aggregation on specified models aimed at measuring school resource effects on student learning, and find that aggregation produces an ambiguous bias on the estimated regression parameters. Thus they suggest an empirical examination of the effects of aggregation in the presence of measurement error.

Although it has become conventional wisdom that aggregate data should not be used to measure school quality, the literature on which this argument is based on, is insufficient to support the claim. Research comparing aggregate with disaggregate models, (Goldstein, Woodhouse) have used ordinary least squares rather than maximum likelihood estimators so the validity of their criticism is unclear. Standardized efficient estimators of school quality based on aggregate data, as well as their confidence intervals will be developed here and compared to multilevel estimators with and without measurement error. Since many states either continue to use aggregate data or use other less accurate measures to rank and reward schools, the relevance of this issue cannot be denied.

2. Theory

Estimators for the effect of schools on student achievement based on disaggregate data have been developed and reviewed extensively in the education literature, and will be presented only briefly here. However, since aggregate data have been disregarded due to the loss of information that aggregation implies, little effort has been devoted to develop appropriate estimators for aggregate data.

This section consists of three parts. The first part will show how aggregation of a 2-level error components model, with heterogeneous number of first-level units within second-level units, leads to a model with heteroskedastic error terms. Therefore, for estimators of the parameters of the model to be efficient, ML or GLS estimation is required. The aggregate data estimator is presented as well as its standardized version.

The second part derives confidence intervals for the aggregate data estimator and presents the confidence intervals commonly used for disaggregate data. The third part introduces measurement error in the model and derives the bias when estimating the parameters of the explanatory variables in both the disaggregate and aggregate models.

2.1. Aggregation of a Simple 2-Level Error Components Model

Consider the following model:

$$Y_{ij} = (\boldsymbol{X}\beta)_{ij} + u_j + e_{ij}, \quad i = 1, \dots, n_j \quad j = 1, \dots J, \tag{1}$$

where Y_{ij} is the test score of the i^{th} student in the j^{th} school, $(\mathbf{X}\beta)_{ij}$ is the fixed part of the model, likely to be a linear combination of student and school characteristics, such as previous test score (for a value added measure), parents' education, and average parents' income for each school, u_j is the random effect for school, that we are trying to estimate, and e_{ij} is the unexplained portion of the test score, with distributions given by

$u_j \sim iid\ N(0,\sigma_u^2),\ e_{ij} \sim iid\ N(0,\sigma_e^2),\quad \mathrm{cov}(u_j,e_{ij}) = 0$.

In matrix notation the model is:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \mathrm{Z}\ \boldsymbol{u} + \boldsymbol{e}, \tag{1.a}$$

where

$$\mathrm{Z} = \begin{bmatrix} \boldsymbol{1}_{n_1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{1}_{n_J} \end{bmatrix},$$

$$\mathrm{Z}\ \boldsymbol{u} + \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{V}),$$

$$\boldsymbol{V} = \begin{bmatrix} \sigma_e^2 \boldsymbol{I}_{n_1} + \sigma_u^2 \boldsymbol{J}_{n_1} & & 0 \\ & \ddots & \\ 0 & & \sigma_e^2 \boldsymbol{I}_{n_J} + \sigma_u^2 \boldsymbol{J}_{n_J} \end{bmatrix}.$$

The random effect u_j represents the departure from the overall mean effect of schools on students' scores. While the intercept contains the overall mean effect of schools, u_j measures by how much school j deviates from this mean.

The shrinkage estimator of u_j is (Goldstein):

$$\hat{u}_j = (\sigma_u^2 / (\sigma_u^2 + \sigma_e^2 / n_j))(\textstyle\sum_{i=1}^{n_j} \hat{y}_{ij}) / n_j \tag{2}$$

$$\hat{y}_{ij} = Y_{ij} - (\boldsymbol{X}\hat{\beta})_{ij},$$

where the $\hat{y}_{ij}$'s are called raw residuals and $\hat{\beta}$ is the MLE of β. So the school effect for school *j* is estimated by the raw residuals, averaged over all students, and 'shrunken' by a factor that is a function of the variance components and the number of students in the school. The larger the number of students in a school, the closer this factor is to one. But if school size is small, there will be less information to estimate the school effect. Thus, the shrinkage factor becomes smaller, making the estimate of the school effect deviate less from the overall mean.

Now let us see how the model changes with aggregation. Adding over all students within each school,

$$\sum_{i=1}^{n_j} Y_{ij} = \sum_{i=1}^{n_j} (X\beta)_{ij} + n_j u_j + \sum_{i=1}^{n_j} e_{ij}$$

and dividing by the number of students in each school, leads to the following model:

$$Y_{.j} = (X\beta)_{.j} + u_j + e_{.j}, \quad j = 1, \dots, J \tag{3}$$

$$u_j \sim iid\ N(0, \sigma_u^2), e_{.j} \sim N(0, \sigma_e^2 / n_j), \quad \text{cov}(u_j, e_{.j}) = 0,$$

where the dot is the common notation to denote that the variable has been averaged over the corresponding index; students in this case. The error term for the aggregated model will be $v_j \sim (0, \sigma_u^2 + \sigma_e^2 / n_j)$.

Again, in matrix notation the model is:

$$\boldsymbol{Y}_a = \boldsymbol{X}_a \beta + \boldsymbol{u} + \boldsymbol{e}_a, \tag{3.a}$$

$$\boldsymbol{X_a} = \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1'_{n_1}} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \frac{1}{n_J}\boldsymbol{1'_{n_J}} \end{bmatrix} \boldsymbol{X}, \quad \boldsymbol{Y_a} = \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1'_{n_1}} \\ \vdots \\ \frac{1}{n_J}\boldsymbol{1'_{n_J}} \end{bmatrix} \boldsymbol{Y}, \quad \boldsymbol{e_a} = \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1'_{n_1}} \\ \vdots \\ \frac{1}{n_J}\boldsymbol{1'_{n_J}} \end{bmatrix} \boldsymbol{e}$$

$$\boldsymbol{u} + \boldsymbol{e}_a \sim N(\boldsymbol{0}, \boldsymbol{V}_a),$$

$$\boldsymbol{V_a} = \begin{bmatrix} \sigma_u^2 + \sigma_e^2 / n_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_u^2 + \sigma_e^2 / n_J \end{bmatrix}$$

We are interested in estimating the random effects u_j's. For this, we estimate the MLE residuals of the error term v_j. Following a procedure similar to the one used with disaggregate data, we define our estimator as the conditional mean of u_j given v_j, i.e., $\tilde{u}_j = \hat{E}(u_j / v_j)$, This value can be shown to be (see appendix):

$$\tilde{u}_j = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2 / n_j)} (Y_{.j} - (\boldsymbol{X}\hat{\beta})_{.j}), \tag{4}$$

where $\hat{\beta}$ is the MLE of β for the aggregate model. Notice that this estimator has the same shrinkage factor as the disaggregate estimator.

However, the school effects in (4) are heteroskedastic, while the true school effects are not. Thus, to correct for heteroskedasticity, we divide the estimator by its standard deviation obtaining the standardized estimator of school effect:

$$\breve{u}_j = \frac{1}{\sqrt{\sigma_u^2 + \sigma_e^2 / n_j}} (Y_{.j} - (\boldsymbol{X}\hat{\beta})_{.j}) \tag{5}$$

Thus, the set of $\breve{u}_j$'s may also be used to rank schools.

2.2. Confidence Intervals for the Estimates of School Quality

A confidence interval for school effects is: $\hat{u}_j \pm t_{1-\alpha/2}\sigma_{u|\hat{u}}$. Thus, it is necessary to obtain the conditional variance of the random effect given its estimator; that is, $Cov(\boldsymbol{u} \mid \hat{\boldsymbol{u}})$.

For both, disaggregate and aggregate estimators, the covariance matrix is derived similarly. First it is necessary to obtain the joint distribution of the vector of school effects $\boldsymbol{u}$ and its estimator. For this, notice that in both cases, the estimator is a linear combination of the vector of dependent variables, test scores in our case. Thus the joint distribution can be derived from the joint of $\boldsymbol{u}$ and **Y**. Then, using a theorem from Moser (theorem. 2.2.1, page 29), the conditional covariance matrix of school effects is obtained. A derivation of this covariance matrix is given in the appendix.

The conditional covariance matrix based on the disaggregate estimator is:

$$Cov(\boldsymbol{u} \mid \hat{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 Z' \boldsymbol{V}^{-1} \left(\boldsymbol{V} - \boldsymbol{X} \left(\boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{X} \right)^{-1} \boldsymbol{X}' \right) \boldsymbol{V}^{-1} Z \,, \tag{6}$$

The conditional covariance matrix based on the aggregate estimator is:

$$Cov(\boldsymbol{u} \mid \tilde{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 \boldsymbol{V}_a^{-1} \left(\boldsymbol{V}_a - \boldsymbol{X}_a (\boldsymbol{X}_a' \boldsymbol{V}_a^{-1} \boldsymbol{X}_a) \boldsymbol{X}_a' \right) \boldsymbol{V}_a^{-1} . \tag{7}$$

2.3. Bias in Estimation Introduced by Measurement Error

Let us consider a two-level model with measurement error. Following Goldstein's notation. The model is:

$$y_{ij} = (\mathbf{x}\boldsymbol{\beta})_{ij} + u_j + e_{ij}, \qquad i = 1, \ldots, n_j \quad j = 1, \ldots, J \tag{8}$$

$$Y_{ij} = y_{ij} + q_{ij}$$

$$X_{hij} = x_{hij} + m_{hij}, \quad h = 1, \ldots, H$$

$$\mathrm{cov}(q_{ij}, q_{i'j}) = \mathrm{cov}(m_{hij}, m_{hi'j}) = 0$$

$$E(q_{ij}) = E(m_{hij}) = 0$$

$$\mathrm{cov}(m_{h_1 ij}, m_{h_2 ij}) = \sigma_{(h_1, h_2)m}$$

where y_{ij} is the real test score for the i^{th} student in the j^{th} school, q_{ij} is the measurement error for y_{ij}, $q_{ij} \sim N(0, \sigma_q^2)$, Y_{ij} is the observed test score, x_{hij} is the true measure of the

h^{th} student or school characteristic corresponding to the i^{th} student in the j^{th} school, m_{hij} is the measurement error for x_{hij}, u_j is the random component for school j, e_{ij} is the residual, and $\sigma_{(h_1,h_2)m}$ is the covariance of measurement errors from two explanatory variables, h_1 and h_2, for the same student. The covariance of measurement errors from any two variables is assumed to be equal for all students regardless of the school they attend.

Following Goldstein (1995), it can be seen that without measurement error, β could be estimated by the FGLS estimator $\hat{\beta} = (\boldsymbol{x}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{x})^{-1}(\boldsymbol{x}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y})$. But measurement error as defined by model (8) implies that $E(\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}) - E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$; so an unbiased estimator for β in the presence of measurement error is proposed by Goldstein to be:

$$\hat{\beta} = [\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X} - E(\boldsymbol{m}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{m})]^{-1}(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{Y}). \tag{9}$$

When measurement error is not taken into account, the matrix $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$ is omitted. Using Goldstein's derivation of $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$ and realizing that the inverse of V is also a block diagonal with elements $\dfrac{(n_j - 1)\sigma_u^2 + \sigma_e^2}{\sigma_e^2(n_j\sigma_u^2 + \sigma_e^2)}$ in the diagonal, each element (h_1, h_2) of the $H \times H$ matrix $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$ can be expressed as

$$\sum_{j=1}^{J} n_j \left\{ (n_j - 1)\frac{\sigma_u^2}{\sigma_e^2} + 1 \right\} \frac{\sigma_{(h_1,h_2)m}}{n_j\sigma_u^2 + \sigma_e^2}. \tag{10}$$

Now let us see how does this omitted matrix, $E(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{m})$, compares with the one to be obtained when aggregating the model. Aggregating the true disaggregate model, we obtain:

$$y_{.j} = (\mathbf{x}\beta)_{.j} + u_j + e_{.j},\ j = 1, \ldots, J \tag{11}$$

$$Y_{.j} = y_{.j} + q_{.j}$$

$$X_{h.j} = x_{h.j} + m_{h.j}$$

$$\text{cov}(q_{.j}, q_{.j'}) = 0$$

$$E(q_{.j}) = E(m_{h.j}) = 0$$

$$\text{cov}(m_{h_1 ij}, m_{h_2 ij}) = \sigma_{(h_1,h_2)m} / n_j$$

where notation is as in model (8).

Notice how the covariance of measurement error between any two fixed explanatory variables is reduced in the aggregate model. Now the covariance matrix of the true model is a diagonal matrix with elements defined in the first part of this section; and which will be denoted by $\boldsymbol{V}_a$. Following a procedure analogous to Goldstein's derivation for the disaggregate model, one can obtain the following unbiased estimator of β for the aggregate model:

$$\hat{\beta}_a = [\boldsymbol{X}'_a \hat{\boldsymbol{V}}_a^{-1} \boldsymbol{X}_a - E(\boldsymbol{m}'_a \hat{\boldsymbol{V}}_a^{-1} \boldsymbol{m}_a)]^{-1} (\boldsymbol{X}'_a \hat{\boldsymbol{V}}_a^{-1} \boldsymbol{Y}_a), \qquad (12)$$

where the subscript *a* denotes aggregate data. As can be seen, the bias now will depend on $E(\boldsymbol{m}'_a \boldsymbol{V}_a^{-1} \boldsymbol{m}_a)$, an $H \times H$ matrix whose (h_1, h_2) element is

$$\sum_{j=1}^{J} \frac{\sigma_{(h_1,h_2)m}}{n_j \sigma_u^2 + \sigma_e^2}. \qquad (13)$$

As can be seen by comparing values in (10) and (13), the bias in β due to measurement error is attenuated in the aggregate model. Bias in the estimation of β without accounting for measurement error, is likely to affect the estimators of school effects, as suggested in (2) and (4). This result is worth considering since adjustments for measurement error are seldom made and, as Woodhouse et. al. argue, different assumptions about variances and covariances of measurement error may lead to totally different conclusions (when ranking schools, for example). Therefore, when not correcting for measurement error, gains from aggregation may somewhat offset the negative consequences of aggregation. Then, at least asymptotically, aggregate estimates of school effects may be less inaccurate than what researchers have claimed.

However, to examine the properties of our aggregate and disaggregate estimators of school effects in small samples, a Monte Carlo study will be necessary. Also, from the study we will be able to compare the estimators' asymptotic and small sample behavior.

3. Data and Procedures

A Monte Carlo study was used to compare aggregate and disaggregate estimates of school effects with their true values. These values were also compared to OLS estimates with aggregate data since this is what is most often done. The model on which the data generating process was based, was taken from Goldstein's 1997 paper, table 3, page 387, since it was simple, and provided estimates of the random components for school and student, based on real data.

This model regresses test scores of each student against a previous test score, a dummy variable for gender, and a dummy for type of school (boys', girls', or mixed school). Test scores were transformed from ranks to standard normal deviates. The random part consists of the school effect and the student effect.

According to Goldstein, multilevel analysis provides the following estimated model:

$$\hat{T}score_{ij} = \text{-}0.09 + 0.52Pscore_{ij} + 0.14Girl_{ij} + 0.10GirlsSch_j + 0.09BoysSch_j,$$

$$i = 1, \ldots, n_j \quad j = 1, \ldots J. \tag{14}$$

The estimated variance of school effects, also called between-school variance, is $\hat{\sigma}_u^2 = 0.07$, and the variance of student effects, also called within-school variance, is $\hat{\sigma}_e^2 = 0.56$. These values and the estimates of the fixed part of the model were used to generate the disaggregate data using SAS. At each replication a number of n_j observations were generated for each school, where n_j was a random realization of a lognormal distribution with mean equal to 100 and variance equal to 50000. Lagged test scores were generated as the sum of two normally distributed components. The first was a school component, common for each student within a school. The second was an individual component, generated for each different student. Dummy variables were generated from binomial distributions. The random components of the model for school and student were generated using a normal with zero mean and variance $\hat{\sigma}_u^2 = 0.07$ and $\hat{\sigma}_e^2 = 0.56$ respectively, and the actual test score was obtained as in equation (2). Then measurement error was introduced to the previous and actual test scores. Measurement error was assumed to be a normal random variable with a zero mean and a standard deviation of 0.2. All dummy variables are assumed measured without error.

Once a disaggregate data set is generated, estimates for school effects and variance components are obtained using multilevel analysis as provided by the Mixed procedure in SAS. Then, the disaggregate data set is aggregated by schools. Residuals as well as the two components of the variance of the error term are estimated using NLMIXED in SAS. At this point, we will have a set of 100 true school effects (since the number of schools in the sample is 100), and two sets of estimated school effects using aggregate and disaggregate data. Each of these sets generates a ranking of the schools in the sample. The greater the school effect, the better the school's performance, and therefore, the higher its position will be in the ranking. We will also have standardized rankings for each estimate and the OLS estimate of school effects to see how this set compares to the alternative estimators and to the true ranking. Finally, we compute the estimated variance components under both approaches and compare them with the true values.

A comparison of the school effect estimators is done in several different ways. Spearman's correlation coefficient is calculated for all estimators in order to measure the degree of correlation of each ranking with the true schools ranking. Another measure used for comparison is the root mean squared error of the estimates, and finally we compare the top-ten set of schools obtained with each estimator, with the true top-ten set. The whole process described above constitutes a single iteration of the Monte Carlo study. As many as 1000 iterations were conducted.

As many iterations as needed can be performed for each set of parameter values of interest. In particular, outcomes with and without measurement error are compared in order to see if the aggregate estimator is in fact more robust to errors in measurement than the disaggregate estimator. The parameters used to randomly generate the number of students in each school are also changed, to corroborate the theory's suggestion that as schools in the sample grow larger, the difference in the estimators' performance will narrow[1].

---

[1] This is because the shrinkage factor tends to one and also because the larger the sample, the closer averages are to their true means.

4. Results

Table 1 shows the first set of results for 1000 samples, each of 100 schools whose size is distributed lognormal with mean 100 and variance 50000. As expected, the disaggregate estimator performs best on almost all measures. The aggregate estimator's performance, however, is surprisingly good, and clearly above the OLS estimator's performance. OLS in fact tends to reward small schools. The average school size for the top ten schools as estimated by OLS is about 76, while the true average for this group is about 99. However, table 1 also shows that both the aggregate and disaggregate estimators tend to reward large schools. This can be explained as follows: OLS estimators are based on residuals whose variance is $\sigma_u^2 + \sigma_e^2 / n$. So, small schools will have a larger variance and will be more likely to be either at the bottom or top of the rankings.

The aggregate and disaggregate estimators have a shrinkage factor that compensates for these large residuals by reducing the residuals of small schools. Recall the shrinkage factor is $\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 / n}$. This factor is always less than one, but decreases with school size, bringing down the absolute value of small school residuals. Results in table 1 suggest that the shrinkage factor may over-compensate for the residuals effect, and thus, leave only large schools in the extremes. Estimators with a smaller shrinkage factor (the factor is $\frac{1}{\sqrt{\sigma_u^2 + \sigma_e^2 / n}}$) such as the standardized aggregate (equation 5) and standardized disaggregate estimators seem to alleviate this problem. Table 1 shows how the average size for the top ten schools according to the standardized estimators only differs by two or three students from the true top-ten group size average. These standardized estimators also seem to give a somewhat better match than their non-standardized versions when determining how many of the real top ten schools are selected by the estimators.

When measuring the root mean squared error (RMSE) of the estimators with the true ranking we find again that the disaggregate estimator performs only slightly better than the aggregate estimator. For the standardized estimators, the RMSE's were calculated using the standardized true rankings, and thus, cannot be compared to the non-standardized versions. Since we are measuring the performance of the estimators by their ability to match the true ranking and not the true values of the school effects, the RMSE might not be as good of a measure as all the others presented in the table.

The between- and within-school variance estimates are presented in Table 1. Although the aggregate point estimates are very close to the true variances, by looking at the standard deviations of these estimates, it is clear that aggregation will always reduce the ability to estimate the within schools variance as compared to the disaggregate estimator.

Table 2 introduces measurement error as 20% of the highest possible test score. We had hypothesized that measurement error would have less effect on the aggregate estimators. This is true but almost unperceivable, considering that a 20% measurement error is high. Thus, measurement error is relatively unimportant in this case.

Finally, table 3 shows the results for ranking estimates when schools have on average 350 students. As school size increases, the variation in averaged residuals due to students (σ_e^2 / n) becomes insignificant. This implies that aggregation becomes less of a concern for estimating school effects (thus, the aggregate and disaggregate estimators should perform more alike now), and heteroskedasticity is almost insignificant (thus OLS is not as bad of a choice as before). In fact, table 3 shows differences among ranking measures have narrowed for all estimators, and that the problem with small or large schools being consistently rewarded, has almost disappeared. However, aggregate data will no longer be able to estimate the variance components of the model with any accuracy.

5. Conclusions

Researchers argue that value-added multilevel models provide the most accurate measures of school quality. But most states continue to use aggregate data (usually not in a value added framework) to rank and reward schools. Research criticizing aggregate models, by comparing them with disaggregate models, have used ordinary least squares rather than maximum likelihood estimators so part of their criticism is uncertain. States need to know the correct way to handle aggregate data and how much accuracy is lost by using aggregate data. Efficient estimators of school quality based on aggregate data and confidence intervals are derived here and compared to multilevel and OLS estimators with and without measurement error. A Monte Carlo study is used in order to perform this comparison that includes measuring the correlation of aggregate versus disaggregate estimates with the true values of school effects.

Results show that when many small schools are present in the data, the proposed aggregate data estimator performs better than OLS on aggregate data, and only slightly worse than the disaggregate data estimator. However, as school size increases, the three estimates perform more alike.

Even though the aggregate data estimator is only slightly worse than the disaggregate data estimator for ranking schools based on efficiency, we still want to encourage the collection of disaggregate data because of their many uses in understanding school quality and student learning.

Also, OLS estimators do tend to reward small schools over bigger ones, as the empirical literature has shown, while the shrinkage disaggregate estimator unexpectedly rewards large schools. A standardized version of this estimate is presented that eliminates this problem.

Thus, when school officials are able to collect multilevel data, this study suggests they consider standardizing the estimates of school quality before ranking schools. However, when disaggregate data are not available, and small schools are present in the sample the standardized aggregate estimator proposed here should be used over the OLS approach.

The methods proposed and evaluated here provide a one-dimensional measure that can be used to understand school quality. However, an efficiency measure based on standardized test scores is not the only measure that should be considered when evaluating schools. This study provides new information about the strengths and weaknesses of alternative methods and data. Our application is to schools, but these results are applicable to measuring efficiency in any industry where aggregate data may be the only data available.

Table 1. Comparison of estimates of school quality using aggregate vs. disaggregate data with no measurement error.

Measure	Type of estimator	Mean	Std.Dev.
Spearman	Disaggregate	0.8527	0.0341
	Std. disaggregate	0.8444	0.0375
	Aggregate	0.8431	0.0367
	Std. aggregate	0.8373	0.0399
	OLS	0.8167	0.0451
RMSE	Disaggregate	0.1332	0.0136
	Std. disaggregate	0.5378	0.0573
	Aggregate	0.1416	0.0164
	Std. aggregate	0.0591	0.0592
	OLS	0.1873	0.0284
Top Ten	Disaggregate	6.50	1.203
	Std. disaggregate	6.53	1.178
	Aggregate	6.33	1.284
	Std. aggregate	6.42	1.207
	OLS	6.01	1.258
School Size Avg.	Real Group	98.96	69.27
In Top Ten Group	Disaggregate	126.13	85.78
	Std. disaggregate	101.60	73.26
	Aggregate	126.93	82.45
	Std. aggregate	102.14	73.58
	OLS	76.34	60.19
Variance Estimates	Dis. Within Sch.	0.560	0.008
	Dis. Between Sch.	0.070	0.013
	Agg.Within Sch.	0.572	0.342
	Agg Between Sch.	0.067	0.016

Note: Results are for 1000 simulations, each including 100 schools. The number of students per school is a lognormal random variable with mean 100 and variance 50000. Mean is the average over all simulations, RMSE is root mean squared error, Top Ten is the average number of schools ranked in the top ten with the estimator, that belong to the true top ten set. Estimators compared are the disaggregate estimator, its standardized version, the aggregate estimator, its standardized version, and the OLS estimator of school effects. Variance estimates are also presented for the disaggregate and aggregate methods.

Table 2. Comparison of estimates of school quality using aggregate vs. disaggregate data with measurement error.

Measure	Type of estimator	Mean	Std.Dev.
Spearman	Disaggregate	0.8445	0.0346
	Std. disaggregate	0.8362	0.0381
	Aggregate	0.8391	0.0363
	Std. aggregate	0.8330	0.0394
	OLS	0.8119	0.0455
RMSE	Disaggregate	0.1364	0.0135
	Std. disaggregate	0.5513	0.0575
	Aggregate	0.1433	0.0165
	Std. aggregate	0.0561	0.0590
	OLS	0.1925	0.0302
Top Ten	Disaggregate	6.43	1.236
	Std. disaggregate	6.42	1.192
	Aggregate	6.30	1.260
	Std. aggregate	6.35	1.205
	OLS	5.94	1.244
School Size Avg.	Real Group	103.40	86.84
In Top Ten Group	Disaggregate	131.58	96.26
	Std. disaggregate	104.83	89.76
	Aggregate	132.50	96.79
	Std. aggregate	107.39	90.93
	OLS	78.79	75.18
Variance Estimates	Dis. Within Sch.	0.610	0.009
	Dis. Between Sch.	0.071	0.013
	Agg.Within Sch.	0.611	0.361
	Agg Between Sch.	0.067	0.016

Note: Results are for 1000 simulations, each including 100 schools. The number of students per school is a lognormal random variable with mean 100 and variance 50000. Measurement error is 20% in actual and previous scores. Mean is the average over all simulations, RMSE is root mean squared error, Top Ten is the average number of schools ranked top ten with the estimator, that belong to the true top ten set.

Table 3. Comparison of estimates of school quality using aggregate vs. disaggregate data for large schools.

Measure	Type of estimator	Mean	Std.Dev.
Spearman	Disaggregate	0.9620	0.0140
	Std. disaggregate	0.9620	0.0140
	Aggregate	0.9588	0.0178
	Std. aggregate	0.9606	0.0168
	OLS	0.9558	0.0185
RMSE	Disaggregate	0.0693	0.0114
	Std. disaggregate	0.2718	0.0440
	Aggregate	0.0854	0.0330
	Std. aggregate	0.2794	0.0521
	OLS	0.0746	0.0140
Top Ten	Disaggregate	8.17	0.967
	Std. disaggregate	8.16	0.974
	Aggregate	7.94	1.139
	Std. aggregate	8.09	1.029
	OLS	8.05	1.043
School Size Avg.	Real Group	348.73	73.10
In Top Ten Group	Disaggregate	351.07	71.59
	Std. disaggregate	347.13	70.57
	Aggregate	375.02	71.02
	Std. aggregate	359.47	67.65
	OLS	345.76	70.64
Variance Estimates	Dis. Within Sch.	0.610	0.005
	Dis. Between Sch.	0.071	0.011
	Agg.Within Sch.	2.344	3.447
	Agg Between Sch.	0.060	0.016

Note: Results are for 100 simulations, each including 100 schools. The number of students per school is a lognormal random variable with mean 350 and variance 50000. Measurement error is 20% in actual and previous scores. Mean is the average over all simulations, RMSE is root mean squared error, Top Ten is the average number of schools ranked top ten with the estimator, that belong to the true top ten set.

5. References

Clotfelter, C. T., and Ladd, H. F. "Recognizing and Rewarding Success in Public Schools" in *Holding Schools Accountable. Performance-Based Reform in Education* by Helen F. Ladd, Washington, D.C., The Brookings Institute, 1996.

Dickens, W. "Error Components in Grouped Data: Is it Ever Worth Weighting?" *Review of Economics and Statistics*, (1990): 328-333.

Goldstein, H. *Multilevel Statistical Models.* London: Edward Arnold, 1995.

Goldstein H. "Methods in School Effectiveness Research." *School Effectiveness and School Improvement* 8(1997): 369-395.

Hanushek, E. A., S. G. Rivkin, and L. L. Taylor. "Aggregation and the Estimated Effects of School Resources." *The Review of Economics and Statistics* 2(Spring 1996):611-627.

Ladd, H. F. "Catalysts for Learning. Recognition and Reward Programs in the Public Schools." *Brookings Review* 3(Summer 1996):14-17.

Moser, B. K. *Linear Models: A Mean Model Approach*. California, Academic Press, 1996.

Odden, A., and W. Clune. "Improving Educational Productivity and School Finance." *Educational Researcher* 9(December 1995):6-10.

Webster, W. J., R. L. Mendro, T. H. Orsak, and D. Weerasinghe. "The Applicability of Selected Regression and Hierarchical Linear Models to the Estimation of School and Teacher Effects." Paper presented at the annual meeting of the National Council of Measurement in Education, New York, NY, 8-12 April, 1996.

Woodhouse, G., and H. Goldstein. "Educational Performance Indicators and LEA League Tables." *Oxford Review of Education* 14(1998):301-320.

Woodhouse, G., M. Yang, H. Goldstein, and J. Rasbash. "Adjusting for Measurement Error in Multilevel Analysis." *Journal of the Royal Statistical Society* A, 159(1996):201-212.

Appendix

Derivation of the aggregate estimators of school effects

Recall equation (3.a) which shows the aggregate model:

$$\boldsymbol{Y}_a = \boldsymbol{X}_a \beta + \boldsymbol{u} + \boldsymbol{e}_a$$

However, the aggregate model has no way of differentiating among its random terms, thus we rewrite the model as:

$$\boldsymbol{Y}_a = \boldsymbol{X}_a \beta + \boldsymbol{w}.$$

We are to obtain the conditional mean of $\boldsymbol{u}$ given the total residual $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{e}_a$ based on the distributions of $\boldsymbol{u}$ and $\boldsymbol{e}$.

Since $\boldsymbol{u}$ and $\boldsymbol{e}$ are independent normal random vectors, its distribution is given by:

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \sim N(0,\ \boldsymbol{V}_{u,e})$, where $\boldsymbol{V}_{u,e} = \begin{bmatrix} \sigma_u^2 \boldsymbol{I}_J & 0 \\ 0 & \sigma_e^2 \boldsymbol{I}_N \end{bmatrix}$, N being the total number of students.

But $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix}$ is a linear combination of $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix}$, this is:

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix} = \boldsymbol{A}_1 \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_J & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \frac{1}{n_1}\boldsymbol{1}'_{n_1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \frac{1}{n_J}\boldsymbol{1}'_{n_J} \end{bmatrix} \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix}$, where $\boldsymbol{1}_{n_j}$ is an n_j vector of 1's. Thus, its distribution will be as follows:

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix} \sim N\left(\boldsymbol{0}, \boldsymbol{A}_1 \boldsymbol{V}_{u,e} \boldsymbol{A}_1'\right).$$

From this random vector, we construct $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix}$ pre-multiplying $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_a \end{pmatrix}$ by $\boldsymbol{A}_2 = \begin{pmatrix} \boldsymbol{I}_J & \boldsymbol{0} \\ \boldsymbol{I}_J & \boldsymbol{I}_J \end{pmatrix}$.

Then, its distribution will be:

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix} \sim N\left(\boldsymbol{0}, \boldsymbol{A_2}\ \boldsymbol{A_1}\ \boldsymbol{V_{u,e}}\boldsymbol{A_1}'\ \boldsymbol{A_2}'\right).$$

Having the joint distribution of $\boldsymbol{u}$ and $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{e}_a$, our estimator is easily derived (Moser, theorem 2.2.1) as:

$$E(\boldsymbol{u} \mid \boldsymbol{w}) = Cov(\boldsymbol{u}, \boldsymbol{w})\ Cov(\boldsymbol{w})^{-1}(\boldsymbol{w}) = \begin{bmatrix} \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 / n_1} w_1 \\ \vdots \\ \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 / n_J} w_J \end{bmatrix}$$

Derivation of the conditional covariance matrix $Cov(\boldsymbol{u} / \hat{\boldsymbol{u}})$

Disaggregate data: Recall equation (1.a):

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \mathrm{Z}\,\boldsymbol{u} + \boldsymbol{e}$$

$$\mathrm{Z} = \begin{bmatrix} \boldsymbol{1}_{n_1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{1}_{n_J} \end{bmatrix}$$

$$\mathrm{Z}\,\boldsymbol{u} + \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{V})$$

The shrinkage estimator of school effects (equation 2) in matrix notation is:

$$\hat{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}) \text{ or } \hat{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1})\boldsymbol{Y} \qquad (*)$$

This shows clearly that the shrinkage estimator is a linear combination of the independent variable vector.

Thus, we can derive the joint distribution of $(\boldsymbol{u}, \hat{\boldsymbol{u}})'$ by knowing the distribution of $(\boldsymbol{u}, \boldsymbol{Y})'$.

The distribution of $(\boldsymbol{u}, \boldsymbol{Y})'$ is: $\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{Y} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}\beta \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{Z}' \\ \sigma_u^2 \boldsymbol{Z} & \boldsymbol{V} \end{bmatrix}\right).$

In general, $\boldsymbol{u}$ and any linear combination of $\boldsymbol{Y}$ of the form $\hat{\boldsymbol{u}} = \boldsymbol{AY}$, will be jointly distributed as follows:

$$\begin{bmatrix} \boldsymbol{u} \\ \hat{\boldsymbol{u}} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}\beta \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{Z}'\boldsymbol{A}' \\ \sigma_u^2 \boldsymbol{AZ} & \boldsymbol{AVA}' \end{bmatrix} \right)$$

Then, by Moser's theorem 2.2.1, the conditional covariance is:

$$Cov(\boldsymbol{u} \mid \hat{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 \boldsymbol{Z}'\boldsymbol{A}'(\boldsymbol{AVA}')^{-1}\boldsymbol{AZ}.$$

Equation (6) is obtained by replacing $\boldsymbol{A}$ with $\sigma_u^2 \boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1})$, from (*), in the expression above.

Aggregate data: Again, we will use the same argument. First, re-express the aggregate estimators of school quality in matrix notation:

$$\tilde{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{V}_a^{-1}(\boldsymbol{Y}_a - \boldsymbol{X}_a\hat{\beta}), \text{ or } \tilde{\boldsymbol{u}} = \sigma_u^2 \boldsymbol{V}_a^{-1}(\boldsymbol{I} - \boldsymbol{X}_a(\boldsymbol{X}_a'\boldsymbol{V}_a^{-1}\boldsymbol{X}_a)^{-1}\boldsymbol{X}_a'\boldsymbol{V}_a^{-1})\boldsymbol{Y}_a \qquad (**)$$

The distribution of $(\boldsymbol{u}, \boldsymbol{Y}_a)'$ is:

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{Y}_a \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}_a\beta \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{I} \\ \sigma_u^2 \boldsymbol{I} & \boldsymbol{V} \end{bmatrix} \right)$$

So, the distribution of $\boldsymbol{u}$ and $\boldsymbol{AY_a}$, a linear combination of $\boldsymbol{Y_a}$ is:

$$\begin{bmatrix} \boldsymbol{u} \\ \tilde{\boldsymbol{u}} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}_a\beta \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \sigma_u^2 \boldsymbol{A}' \\ \sigma_u^2 \boldsymbol{A} & \boldsymbol{AV}_a\boldsymbol{A}' \end{bmatrix} \right),$$

and the conditional covariance matrix is:

$$Cov(\boldsymbol{u} \mid \tilde{\boldsymbol{u}}) = \sigma_u^2 \boldsymbol{I} - \sigma_u^4 \boldsymbol{A}'(\boldsymbol{AVA}')^{-1}\boldsymbol{A}.$$

When $\boldsymbol{A} = \sigma_u^2 \boldsymbol{V}_a^{-1}(\boldsymbol{I} - \boldsymbol{X}_a(\boldsymbol{X}_a'\boldsymbol{V}_a^{-1}\boldsymbol{X}_a)^{-1}\boldsymbol{X}_a'\boldsymbol{V}_a^{-1})$, we obtain equation (7).

VITA

Francisca G-C. Richter

Candidate for the Degree of

Doctor of Philosophy

Thesis: EXPECTED FEES FOR MANAGED FUTURES, OPTIMAL FEEDER CATTLE PRICE SLIDES, AND AGGREGATE vs. DISAGGREGATE DATA IN MEASURING SCHOOL EFFECTIVENESS

Major Field: Agricultural Economics

Bibliographical:

Personal Data: Born in Lima, Peru, on April 13, 1970, the daughter of Ramon and Josefa Garcia-Cobian.

Education: Graduated from Miguel Grau High School, Lima, Peru, in December 1987; received the Bachelor of Science degree in Statistics from the Catholic University of Peru, in December of 1994, and a Master of Science degree in Statistics from Oklahoma State University, Stillwater, Oklahoma in May of 1997. Completed the requirements for the Doctor of Philosophy degree with a mayor in Agricultural Economics at Oklahoma State University in December, 2000.

Experience: Worked as teaching assistant for the departments of Statistics and Economics, Catholic University of Peru, January 1993 to January 1994; analyst of quantitative studies at the Lima Consulting Market Research Group, January 1994 to May 1995; teaching assistant, Statistics Department, Oklahoma State University, August 1996 to May 1997; research assistant Agricultural Economics Department, Oklahoma State University, August 1997 to December 2000.