Performance Incentives in Government Subcontracting: Evidence from the Job Training Partnership Act (JTPA)

Michael Ian Cragg*
Department of Economics
Columbia University

July 3, 1995

Abstract

It is curious that more government program do not use performance based subcontracts for human resource oriented programs. Theoretical explanations for their limited use are that: agents' risk aversion to idiosyncratic variation in compensation beyond their control constrains the effectiveness of performance incentives; and, moral hazard can restricit the efficacy of performance incentives if the performance measures do not perfectly reflect program goals. This paper examines the validity of these explanations by studying the performance management system used in the major federal job training program, the Job Training Partnership Act (JTPA). Existing JTPA performance measures lead to problems of moral hazard. The paper provides empirical evidence for the notion that problems of moral hazard preclude the wide-spread use of performance incentives in government programs.

*I have benefited greatly from discussion with Steve Cameron, John Cragg, Andrew Dick, Alexander Dyck, Wei-Yin Hu, Glenn Hubbard, Matt Kahn, Tom MaCurdy, John Pencavel, John Shoven and Frank Wolak. The generosity of Carol Romero at the National Commission for Employment Policy is gratefully acknowledged without whose data this study would not have been possible. Financial support is thankfully recognized from the Social Sciences and Humanities Research Council of Canada and the Lynde and Harry Bradley Foundation.

1 Introduction

While performance based contracts are routinely used in the private sector, it is curious that few government programs use performance incentives in their subcontracting relationships. In an increasing number of government programs,[1] legislators require that operators of welfare, training and education programs be compensated according to a set of performance standards. In an era when the cries of those rallying for government accountability have become more ardent, it seems particularly important to answer the question of why there are so few government programs using contracting relationships based on performance contingent compensation. Although many authors explain the theoretical problems with performance contracting, few studies examine the problems empirically in the context of welfare, education and job training programs. I use the case of subcontracting for government job training to empirically examine the importance of moral hazard as an explanation for the limited use of performance based pay in government programs.

Informational limitations theoretically may minimize the use of full performance-based contracts. Although government objectives may be clearly stated, practical performance measures which perfectly reflect objectives may either not exist or are bungled in their inception. This can lead to two problems that might limit the spread of performance based contracts. First, if subcontractors are risk averse, pay that varies both due to variation in their effort and also for idiosyncratic reasons may lead to a suboptimal provision of effort when pay is performance based. Second, measured performance may be affected by both desired and undesired efforts. If undesirable efforts are unobservable then a moral hazard may arise if the cost of undesirable activities are less than desirable activities. Theoretical work by Holmstrom and Milgrom (1991), Baker (1992) and Cragg (1993) examine these issues more rigorously.

While the theoretical explanations for the government's limited use of performance based contracts may be parsimoniously stated, few empirical studies document their relevance. The problem for the researcher is that government subcontracting typically requires the contractee (agent) to provide a *service* to another group of individuals. Job search, legal services, and technical aid are a few examples where this arises. This is different from performance contracts with

[1] The Omnibus Budget Reconciliation Act (1988) and the Family Support Act (1988) required that by 1993 performance management systems be incorporated into the Food Stamp and AFDC programs.

the production of goods. The empirical problem in measuring the degree of moral hazard with service provision is that both the agent and those they are serving may be sensitive to incentive parameters in the contract. The resulting empirical problem is illustrated by a simple example. Suppose an educational authority wants to elicit higher teacher effort and therefore bases teachers' salaries upon student performance. Finding that those enrolled have higher ability is not evidence of moral hazard whereby teachers substituted better students for less intelligent students applying to the program. It may be that better students applied. To solve this identification problem, the researcher needs a variable which exogenously alters the teacher's incentives but does not alter the students' propensity to apply to the program. State variation in the implementation of the Job Training Partnership Act (JTPA) provides such a research environment and thereby allows tests of agency theory as an explanation for the limited use of government incentive contracts.

The JTPA, which replaced its precursor the Comprehensive Employment and Training Act (CETA) in 1982, is the major federal job training program for the poor. Its main innovation over CETA was the introduction of a performance-contingent incentive system. The critical design element in the JTPA system is the selection of standards which reflect policy goals explicitly stated in terms of changes in employment and earnings and a reduction in welfare dependency. However, implementation problems precluded adopting a set of value-added performance measures and instead the adopted standards only measure levels of unemployment and earnings following training. This design problem provides an opportunity for program operators to cheat by enrolling those who would have high post-training earnings and employment even without training. However, this cream-skimming may not arise if program operators react to the intentions of the legislation and maximize value-added through better teaching methods. Of course, these activities are not mutually exclusive: high effort combined with cream-skimming may lead to providing training to those for whom the value-added is largest. In short, the performance standards used in practise may either raise or decrease "the basic return on investment (as) measured by the increased employment and earnings of participants and the reduction in welfare dependency. (Section 1. of the Job Training Partnership Act)." Thus, the JTPA provides an ideal case study to understand the importance of moral hazard in limiting government use of performance contracts.

This paper investigates training provider responses to the JTPA performance incentives by studying whether state variation in incentive policy has led to differences in earnings growth and cream-skimming. I find evidence that higher

incentives encourage "cream-skimming." This may reflect service providers targeting services to individuals for whom the value-added is largest or it may be cheating permitted by standards which measure only post-training outcomes. This ambiguity means that the researcher must use measures of labor market outcomes commensurate with program goals to analyze whether stronger incentives lead to higher value-added. As such a value-added measure, I use the increase in earnings growth. I find that in states with higher incentives, value-added is higher. Thus, the JTPA performance standards elicit both higher value-added and cheating. Although the positive incentive effect dominates the negative selection effect, the fact that moral hazard exists might explain why the JTPA is one of the few programs that have adopted performance contracts.

This paper adds to the growing literature on implementation problems of providing education and other public goods through decentralized institutional arrangements. In particular, it is the first study that empirically investigates the moral hazard problem associated with using performance based pay in educational systems. Thus this work adds empirical content to the burgeoning literature on the problems associated with using incentive contracts to resolve principal-agent problems. It complements the empirical work on contracting incentives in franchising by Krueger (1990) and Lafontaine (1992), in defense procurement by Anton-Yao (1987,1989,1990), managerial incentives by Gibbons-Murphy (1991) and in air-traffic control by Staten-Umbeck (1982). This paper is also a natural complement to the work of Anderson-Burkhauser-Raymond (1993) and Dickinson-West (1988) which investigates the impact of JTPA performance policy on clients, services and costs. While both papers provide interesting results, neither uses data which allow them to investigate the notions of cream-skimming, moral hazard and efficient targeting of JTPA services using measures of performance consistent with JTPA policy goals. The JTPA case suggests that the limited use of performance contracts in government programs providing human resource services is partly due to moral hazard problems that arise from imprecise performance measures.

2 Conceptual Framework for Measuring Program Operators' Responses to Performance Incentives

2.1 Why the JTPA Allows a Test of Moral Hazard?

The JTPA provides a separate role for each member of the federal-state-local hierarchy of government. At the federal level, the Department of Labor chooses performance standards to be used by States for monitoring and rewarding local SDAs.[2] Although the DOL was mandated to choose standards that stimulate gains in wages, employment and reductions in welfare receipt, standards based upon absolute levels following training were adopted in the period analyzed in this paper (1983-1987).[3] This creates the moral hazard problem: although the program goals clearly emphasize the concept of value-added, SDAs and their subcontractors may violate these intentions by enrolling individuals who will achieve high post-training outcomes even if the training adds nothing, thereby maximizing their measured performance.

The reason absolute level-standards rather than value-added standards were adopted appears associated with the problem of how to initially set performance thresholds. A number of earlier CETA demonstration projects had experimented with post-training level standards and hence there existed experience which guided setting national performance standards based upon absolute levels but not value-added standards.[4] This historical fluke provides a research opportunity. While both the researcher and program operators are clear about the goals of the JTPA (they are so explicitly stated in the legislation and rhetoric of the DOL),[5] program operators are only legally bound to respond to the performance measures. The moral hazard problem arises because training providers receive the same credit for placing individuals with significantly different attributes: results for a person with limited skills would be treated identically to those for an accomplished student.

[2]The country has been partitioned into 620 Service Delivery Areas (SDAs).

[3]Four adult performance measures have been used through 1987: the employment rate following training, the employment rate of welfare recipients, the average wage rate of those employed following training, and the cost per trainee who entered employment. There are also three youth (less than 21) standards: the employment rate following training, the "positive-termination rate," and the cost per positive termination where a "positive termination" is defined as a youth entering employment, full-time school or achieving schooling goals.

[4]See the account by Levitan and Gallo (1988).

[5]See the introductory section of the Job Training Partnership Act.

The DOL recognized the problem of enrolling different types of people and therefore also introduced a set of adjustments that mechanically correct the thresholds for *observable* differences in the treatment population. However, this "adjustment model" cannot account for *unobservable* differences in the applicants. Given rewards are based on the wages and employment of trainees following training, program operators have the incentive to enroll smarter applicants with more work experience. This aspect of the JTPA is the fundamental feature which allows me to study the moral hazard problem. This is one case where the econometrician observes more than the federal and state JTPA officials. While the DOL "adjustment model" corrects for many differences in the applicant pool, it does not correct for applicant ability and prior work experiences which are measures available in my data. That the state and federal JTPA officials observe less than the econometrician forms the basis for my tests for moral hazard.

As discussed in the introduction, to link directly moral hazard to features of the subcontracting relationship, a variable is necessary which measures differences in the incentives faced by program operators but not differences in the application propensities of the individuals they train. The JTPA gives state governors discretion over how to use the DOL standards to design incentives. They establish policies regarding the distribution of rewards to and sanctions on SDAs. SDAs make decisions regarding who is enrolled, the types of services that are offered, the types of contracts used and the selection of service providers. Although an SDA may be thought of as a not-for-profit organization, one imagines that they are sensitive to performance incentives for a number of reasons. Because the SDA is managed by a council of public servants and private industry members, bureaucratic self-preservation, reputation and empire building motivate the public officials to respond to the incentives. Thus, cross-state differences in the incentives faced by JTPA program operators form a natural experiment for studying the impact of performance incentives on service provider behavior.[6]

The assumption that the JTPA allows the potential to cream-skim (moral hazard) is based upon the presumption that the program is underfunded and services must be rationed. JTPA services are restricted predominantly to low-income individuals.

[6]The importance of incentive contracts is seen in a survey of SDAs that found that 30 percent desired rewards to maintain budgets while 10 percent cited public relations value as a significant reason to do well vis-a-vis other SDAs. (Dickinson and West (1988), p.117.). Bureaucratic self-preservation and reputation are important because if an SDA fails to achieve a minimum threshold for two-consecutive years, the State is mandated to replace all the public officials associated with the SDA.

After an income test, participant selection strategies range from first-come/first-served policies, to taking everyone willing to participate, to giving applicants priority based upon test scores, employment experience, or educational attainment. JTPA intake may be conducted by the SDA, public schools, the Employment Service, community organizations, for-profit agencies or government agencies. The contractors providing training can be community colleges, high schools, non-profit providers, for-profit groups or employers. Given that both the SDAs and their subcontractors face the same incentives, I assume they may be thought as conceptually linked.[7] Thus, I do not differentiate who generates the moral hazard.

2.2 Measuring Performance Incentives

To measure the effect of JTPA performance incentives on the behavior of SDAs and their subcontractors, I need to develop measures of the cross-state variation in incentive intensity. This variation arises because state governors have discretion about the design of the incentive system within the parameters established by the DOL. Variation in incentive intensity can be characterized by the fraction of SDA budgets exposed to performance evaluation, the idiosyncratic variation in performance standards, and the accuracy of the performance assessment.[8]

I use a data set collected in 1986 for the National Commission on Employment Policy to develop state level proxies for incentive variation in: the size of awards; the risk associated with winning awards; and the risk associated with being punished.[9] While the JTPA incentive programs can vary along many dimen-

[7] As evidence of this assumption note that 78% of SDAs use performance-based contracts when hiring sub-contractors to prvide training. In these contracts, more than three quarters of all contracts make at least 25 percent of the payment contingent on performance (National Commission for Employment Policy Survey, question F1).

[8] See Cragg (1993) for models of the contracting relationship between states and service providers.

[9] The NCEP survey was conducted as part of a standard program audit. Questionnaires were sent to all 51 state JTPA administrative heads. The state questionnaires asked question related indirectly to the three ideal measures:

POLICIES EXPOSING SDAS TO RISK: Questions addressed the use of the DOL adjustment model, use of policies in addition to the DOL model, the number of incentives that must be passed to qualify for an award, the number of standards that may be failed before being sanctioned, the threshold levels determining failing and passing, and the weight given to each standard in the calculation of awards.

INCENTIVE INTENSITY: Questions addressed the proportion of the state budget used for incentives, percentage of the budget received for marginally exceeding standards, whether there were caps on

sions, the data show that in fact a simple set of variables will capture most of the state program variations. Differences in the size and distribution of awards arise through the percentage of the state budget dedicated to performance awards and the allocation formulas dictating their distribution. Most states use a formula which relates award size to the amount by which standards are exceeded. While I would like to be able to define a variable which measures the proportion of per-trainee reimbursement which is contingent on performance, the NCEP data does not allow its definition. However, the fraction of the state budget dedicated to awards should be related to this ideal measure. The variable defined to capture this is called AWARDSIZE. Because JTPA funds are rationed, this variable will be directly correlated with the ideal measure under the assumption that across states SDAs have the same average expenditure per enrollee.[10] This variable is viewed as a proxy for the degree of incentive intensity and the proportion of SDAs' budget which is contingent on performance.

The likelihood of an SDA receiving an award in a state depends on the levels of standards and the number of standards that must be passed. This likelihood falls as the standards become stricter and as the number of qualifying standards rises. There is little variation in the levels of these standards but there are considerable differences in the number of standards that must be passed. This variation is captured by the policy variable CARROT which indicates the number of standards

---

the incentive awards, use of bonuses for extraordinary performance.

POLICIES TO REDUCE CREAMING: Questions addressed whether there were additional standards used beyond the 4 adult and 3 youth DOL standards, other hard-to-serve policies, and policies for adjustments beyond the DOL model.

[10]This is illustrated in the following accounting identities

$$
\begin{aligned}
\text{\% state budget used for awards} &= \frac{\sum_{sdas} \text{Awards}}{\sum_{sdas} \text{Awards} + \sum_{sdas} \text{Noncontingent Payments}} \\
&= \frac{\sum_{sdas} \text{Awards}}{\text{Number of State Trainees}} * \frac{\text{Number of State Trainees}}{\sum_{sdas} \text{Awards} + \sum_{sdas} \text{Noncontingent Payments}} \\
&= \frac{\text{State Average per Trainee Award}}{\text{State Average Cost per Trainee}} \\
&= K * \text{State Average Cost per Trainee}
\end{aligned}
$$

where in the last step I assume that the state average cost per trainee is a constant K across states. This would be unlikely if the JTPA were over-funded so that some states were unable to recruit enough trainees to use the funds efficiently.

that must be passed to qualify for an award.[11] This variable is related to the amount of risk to which a SDA is exposed. SDA risk is sensitive to local economic conditions and attributes of the applicant pool[12] because low trainee wages and placement rates might be observed either due to poor services or because of a temporary regional depression. In addition, economic fluctuation and regional factors may affect the ability levels of the applicant pool.[13] The Department of Labor's adjustment model corrects for observed variation in fourteen factors specific to the SDA. Almost all states use the DOL adjustment model thereby exposing SDAs to less risk. However, using multiple standards to determine award eligibility is likely to increase the amount of idiosyncratic risk to which SDAs are exposed because most of the performance standards are probably positively correlated.[14] Like AWARDSIZE, a higher value of CARROT corresponds to greater incentives to perform well.

Analogously, the probability of being punished depends on the number of standards that must be passed. This probability rises as the standards become more exacting and falls as the allowable number of failed standard rises. This variation is captured by the variable STICK, which measures the number of standards that may be failed before sanctioning is required. Again, a higher value of STICK corresponds to higher incentives.

Finally, some states believe the DOL model is inadequate for correcting for both temporal and local geographic economic factors which affect SDA performance. For SDA's that have marginally failed, these states have formalized application processes for SDAs to make special appeals for unexpected difficulties associated with their clients served, services offered or economic conditions. The variable ADJ indicates whether a state used formalized adjustment procedures beyond the DOL model. SDAs in states with adjustment policies should reduce the level of

[11]The NCEP data do not allow me to identify which standards must be passed.

[12]For instance, outcomes might vary not only because trainers expend different levels of effort but also because of idiosyncratic shocks to local economies. A region dominated by sunset industries are susceptible to unexpected plant closures which affect local wages and unemployment rates.

[13]Locations in urban slums where educational achievement is low may provide an intake population that is difficult to train: abnormally high unemployment in the area might lead to more skilled applicants.

[14]If an SDA is required to pass only 1 standard from a set of 7, it will focus on the one with the lowest variance beyond its control, and the one that is easiest to pass (call is P_A with variance σ_A). The risk associated with this is σ_A. If a second standard is required, the risk moves from σ_A to $\sigma_A + \sigma_B = 2\sigma_{AB}$. If the covariance between the two standards σ_{AB} is positive, the risk is unambiguously increased.

cream-skimming and offer riskier but higher return services because they are exposed to less idiosyncratic risk and thus can afford to take chances.

3 Section 3 - Empirical Models to Capture Local Responses to JTPA Performance Policy

Variation in the state level incentive structure provides a "natural experiment" to understand whether higher incentives motivate the provision of more effective services, and/or encourage enrollment of more employable individuals. If the designed standards successfully mirror program goals, then an increase in incentives as reflected by higher values of AWARDSIZE, CARROT and STICK should induce program operators to provide better training services: increases in employment and earnings should be greater in higher incentive states. However, higher incentives may also encourage cream-skimming because the standards adopted measure the levels of post-training employment and earnings. Thus, higher values of AWARDSIZE, CARROT and STICK may also lead to the enrollment of individuals with higher levels of human capital from the applicant pool but for whom the value added is lower, contrary to program goals. Finally, states where ADJ takes on a value of 1, should have less cream-skimming because the problems associated with just failing to attain standard thresholds are minimized. In addition, the use of adjustment policies for extraordinary circumstances may encourage program innovation because the consequences of SDA failure might be mitigated by appealing to the state for allowances.

3.1 Is there Cream-Skimming?

The impact of performance policy on the enrolled population is measured by the probability that an individual eligible for training is observed to be enrolled. Thus, data on the eligible and the enrolled population are necessary. I use data from the NLSY[15] to determine whether individuals eligible during a one-year period[16] were enrolled in a training program. I focus on the 1983-1987 JTPA introductory

[15]See Appendix for details on how the sample is constructed.

[16]Federal training programs during both the CETA and JTPA periods mandated that 90 percent of training recipients be disadvantaged which is defined as a point in time as being a member of a family whose income for the last six months is below a cut-off or whose income includes government welfare payments.

period because this period predates the first major revision of the JTPA incentive system in 1987-88. I also use 1979-1982 for comparing the CETA to the JTPA. The observation of an individual entering a training program is dependent on three conditions being satisfied. First, the individual must be from the eligible population. Second, the net expected benefits from participating in the program must be positive for the individual to apply. Third, the applicant must be chosen by the program administrator. The NLSY generates a sample of 467 JTPA/CETA enrollees from a potential pool which range from 737 to 3,945 annually. To these data I merge the JTPA incentive measures.

An individual applies for training if the net benefits are positive. Net benefits to the individual may be thought of as the difference between the discounted stream of earnings increases for the training program and the costs of participation. Costs of participation include forgone wages, transportation and child care. The level of forgone wages depends on the attained level of human capital, which is measured by educational attainment and past labor force experience. The forgone wage level also varies by age, race and sex. Costs of participation are likely to be lower if the individual lives at home and is married and higher if he has children. Because many states require that some welfare recipients participate in a training program, an indicator of welfare recipiency is also a control variable.

Cream-skimming refers to administrators enrolling individuals who are likely to do well even without training. To measure this, I need variables which reflect local administrators' and program operators' expectations as to whether an individual will have high post-training earnings. These variables must be observable to the econometrician but not to the state and federal JTPA administrators. The NLSY yields two types of variables with these attributes: previous work EXPERIENCE and the Armed Forces Qualification Test (AFQT). For each point in time, the NLSY has a very accurate measure of previously accumulated work experience because start and end dates for 5 jobs each year are collected. All of the members of the NLSY were administered the Armed Forces Qualification Test in the summer of 1980. The AFQT is a battery of 10 subtests which are used by the armed services to determine whether individuals are suited to enlisted work. The 10 subtests cover: general science, arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations, coding speed, auto and shop information, mathematics knowledge, mechanical comprehension, and electronics information This test is a measure of ability but it varies by the age at which the test was taken. Therefore, rankings within an age cohort are used as the ability measure.

Three events dictate the observation that an eligible individual is enrolled or

not enrolled: (1) an eligible individual does not apply; (2) an eligible individual applies but is not accepted for enrollment; and, (3) an eligible individual applies for enrollment and is accepted into the program. However, the data used in this study allow the identification of only state (3) and the joint event, either (1) or (2). The joint decision is important. If we examine the sample means, the first observation to be made is that the level of human capital as measured by the AFQT score or the EXPERIENCE variable is lower in the enrolled evidence. Is this evidence of cream-skimming? This inference is incorrect because it could simply reflect differences in application propensities across the EXPERIENCE and AFQT spectrum.

A natural framework for studying this problem is simply a reduced form model of the probability an eligible individual is enrolled. Define a latent index $I_i = X_i\beta + u_i$ which is composed of an observed portion $X_i\beta$ and an unobserved part u_i. When I_i is positive an individual is enrolled in a training program, and the probability that an individual is enrolled is $P(I_i > 0)$. In this framework, a positive value β for indicates that for a higher value of X_i an individual is more likely to be enrolled. If u_i has a normal distribution with the variance normalized to one for identification, then the model is a simple probit and maximum likelihood estimates for this model are consistent and traditional Wald and likelihood ratio tests may be used.[17]

The previous section discussed the factors that influence program administrators' choice of applicants. The discussion is summarized in the following set of predictions if the cream skimming hypothesis is correct.

> **Prediction 1**: If cream-skimming occurs, then in high incentive states, individuals with more work experience and higher ability will be enrolled. States with higher values for AWARDSIZE, CARROT and STICK provide greater incentives for service providers to meet and exceed performance standards. In states where ADJ is equal to 1 there should be less cream-skimming. Therefore, an interaction between the incentive variables and EXPERIENCE and/or AFQT should be positive.

> **Prediction 2**: JTPA funding formulas ensure that states with high unemployment receive greater funding.[18] Thus, the STATE UNEM-

[17] Amemiya (1985), Chapter 9.

[18] Two-thirds of JTPA funds are allocated to states based on the distribution of unemployment and the remaining one-third according to the low income population.

PLOYMENT RATE should be positive.

Prediction 3: States with higher unemployment are likely to exhibit more cream-skimming because layoffs both swell the applicant pool and raise its quality therefore allowing administrators to be more selective. Therefore, an interaction between the STATE UNEMPLOYMENT RATE and EXPERIENCE and/or AFQT should be positive.

Prediction 4: Temporal variation in incentives arises through the regime shift from CETA to JTPA in 1983. Enrollment probabilities should be higher during the period in which CETA was in place because it was more generously funded therefore allowing higher enrollment.

Prediction 5: Cream-skimming should be lower during CETA because there was no performance-dependent incentive system in place. Therefore an interaction between a time dummy CETA and EXPERIENCE and/or AFQT should be negative.

The model is estimated for two samples of eligible people drawn from the NLSY. In the first sample, I use only the JTPA eligible population and first explore whether states with higher unemployment rates tend to allow administrators to choose higher ability individuals. Second, I examine whether there is a tendency to enroll individuals with higher levels of human capital in states with more intensive incentive policy, as indicated by higher values of CARROT, STICK and AWARDSIZE. In the second sample, I include both the CETA and JTPA eligible populations and explore the differences in enrollment probabilities across the two time periods. I examine whether there was less cream-skimming during CETA and whether the unemployment effect is smaller. In addition, I investigate whether the JTPA targeting policies for enrolling more minorities, welfare recipients and high school dropouts altered the intake probabilities.

The first three columns of Table 3.1 provide probit enrollment models based upon the JTPA sample while the other columns used the combined CETA/JTPA sample. The first column presents the basic specification which includes controls for whether the individual was a welfare recipient, a high school dropout, two age dummies indicating whether the individual was less than 20 and whether the

individual was between 20 and 25, and sex and race dummies. The comparison group is a white high school graduate who is over 25 and does not collect welfare. Recalling that a positive coefficient indicates that an increase in the covariate raises the enrollment probability, we see that the probability of being enrolled is higher for welfare recipients and minorities. In addition, as expected it is more likely that younger individuals are enrolled. Finally, males and high school dropouts are less likely to be enrolled though this effect is insignificant.

The second column introduces the first three covariate of interest: the state employment rate, work experience[19] and an interaction between the unemployment rate and work experience. Jointly and individually the three variables are significant at the 5% level. We see that higher unemployment rates reduce the probability that an individual is enrolled in a JTPA program. Thus, while higher state unemployment increases the JTPA funds in a state, they are inadequate to overcome the increased eligible pool as more families slip below the poverty line. Surprisingly, we see that a higher level of work experience reduces the likelihood that an individual is enrolled in a training program. Thus it seems that JTPA policy reduces cream-skimming. However, recall that the observation that an individual is enrolled in a program is based upon two decisions: first an eligible individual applies for training, and second, he is enrolled by the administrator. To conclude that JTPA policy reduces cream-skimming by simply examining the coefficient estimate on work experience is incorrect. The relevant coefficient is the interaction between work experience and the state unemployment rate: it is both positive and significant. Thus, in states with higher unemployment rates, there is a greater tendency for more experienced applicants to be enrolled. This effect may be attributed to cream-skimming because administrators are able to be more selective because unemployment increases the size and 'quality' of the applicant pool (as in a Roy model). Note that while estimates are only presented where EXPERIENCE is the only measure of labor market ability observed by the econometrician and sda officials and not the state and federal officials, extensive testing was also performed on the AFQT scores. Although the coefficients were uniformly supportive of the cream-skimming hypothesis, except for test scores on word knowledge, paragraph comprehension , auto and shop information and mechanical comprehension, most tended to be insignificant at the 10% level.

I exploit the second source of geographic variation in incentives to test for

[19]EXPERIENCE is defined as the fraction of time spent employed since the individual has left school.

cream-skimming by interacting the three policy variables CARROT, STICK, and AWARDSIZE with work experience. Positive coefficients on these interactions are an indication of cream-skimming as is a negative coefficient on the interaction with ADJ. It is also important to also include the policy variables alone because states with higher incentives may also have higher enrollment rates. After extensive testing, only a specification which includes both ADJ and CARROT seems warranted because AWARDSIZE and STICK have little explanatory power. Column 3 presents the specification that includes both CARROT and ADJ. The coefficient estimate on the interaction between ADJ and EXPERIENCE is negative and significant at the 5% level indicating that in states with an adjustment policy there tend to be less experienced individuals selected into the JTPA programs. The interaction between CARROT and work experience is positive which indicates that in states with more intensive incentives there is a tendency to select more able individuals into the program. Both of these results are consistent with the cream-skimming hypothesis.

The final three columns in Table 3.1 present evidence for cream-skimming based upon the temporal variation arising in the shift from the CETA to the JTPA training regime. These estimates are based upon a pooled population from both the CETA and JTPA periods. The first column presents a specification that merges the CETA and JTPA effects. Using the same base controls as the previous section, the covariates include whether the individual was: a welfare recipient, a high school dropout, less than 20 years old, between 20 and 25 years old, black, hispanic, or male. If we average across the two periods, we see that minorities, high school dropouts, welfare recipients and younger individuals are more likely to be enrolled. As before, we see that states with higher unemployment rates have lower enrollment rates and in states with high unemployment more experienced individuals tend to be enrolled. The second column presents interactions between a dummy variable indicating that the observation is from the JTPA period and basic characteristics of the population. Positive coefficients on the JTPA interaction variables indicate that relative to the CETA period, these individuals had a higher selection probability during the JTPA period. The positive value of the interaction JTPA*WELFARE indicates that during the JTPA period, welfare recipients were more likely to be enrolled relative to nonrecipients. We also see that younger, whites, females or high school graduates were more likely to enroll during the JTPA period than during the CETA period. Testing suggested that interactions with Welfare, Dropout, and Age < 20 were significant while those with Black, Hispanic, and Male were not. The final column provides estimates for the interactions JTPA*EXPERIENCE,

JTPA*UNEMPLOYMENT, and JTPA*UNEMPLOYMENT*EXPERIENCE. A likelihood ratio test statistic of 2.22 suggests that from a statistical perspective across the CETA and JTPA policy shift, the average amount of cream-skimming relative to CETA did not change. While the average level of selection bias between the two periods may have been similar, the previous analysis indicates that across states the degree of cream-skimming increased.

To gauge the magnitude of the cream-skimming effects, Table 3.2 provides simulated probabilities based on the parameter estimates from columns 2, 3 and 6 of Table 3.1. Estimated probabilities are presented for two base populations, low- versus high-experience individuals in two environments, low- versus high unemployment.[20] The concept relevant to the cream-skimming hypothesis is the REP: the ratio of the enrollment probabilities for low- versus high-experience individuals in each environment. The REP is a summary statistic for cream-skimming because as the ratio decreases, there is more selection bias in favor of high-experience individuals. If the ratio is 1, there is no selection bias between the two groups. The first row shows that a low-experience,black, male high school dropout in his early twenties (the base case for the policy simulations), is almost three times as likely to be enrolled as his high-experience counterpart in low unemployment areas. In a high unemployment area this REP falls to 0.9 indicating more cream-skimming in high incentive states.

The middle of the table illustrates the impact of incentive policy on the likelihood of enrollment. The basic trends are partially masked by the fact that enrollment probabilities are higher in states with adjustment policies and in states with higher incentives. The impact of incentive policy on cream-skimming is seen by comparing the REPs of low- versus high-experience individuals across incentive regimes. Thus, to understand the impact of adjustment policies in low-unemployment states, compare the two pairings for adjustment policies in low- versus high-incentive states. In low-incentive states, we see that the adjustment policies have little effect: the relative selection probabilities for a low- versus high-experience person are 2.9 in a state with adjustment policies versus 2.8 for a state without. However, in high incentive states these ratios change from 3.3 to 2.3. Thus, in a high-incentive state with an adjustment policy, a low-experience individual is 3.3 times more likely to be enrolled in a program whereas in a state without an adjustment policy, the relative likelihood falls to 2.3.

[20]Low and high experience correspond to 0.1 and 0.6 respectively while low and high unemployment correspond to unemployment rates of 4% and 12%.

The impact of incentive policies on selection probabilities is understood by selecting either the rows where an adjustment policy is in place or those where it is not and then comparing the relative enrollment probabilities of low- versus high-experience person. For states without adjustment policies, we see that: in low-unemployment states, the relative probabilities fall from 2.8 in low-incentive states to 2.3 in high incentive states; in high-unemployment states, the relative selection probabilities are the same across the incentive regimes. For states with adjustment policies, we see that: in low-unemployment states, the relative probabilities rise from 2.9 in low-incentive states to 3.3 in high-incentive states: in high-unemployment states, the relative selection probabilities are close to being the same across the incentive regimes. Thus, in high-incentive states without an adjustment policy, cream-skimming effects are large, but when an adjustment policy is introduced it overcomes the selection bias introduced by higher incentives.

The bottom of the table makes it readily apparent that selection probabilities during the CETA period are much higher. During the CETA period, low-experience individuals were always more likely to be enrolled in a program than their higher experience counterparts. In the JTPA period, this was true only in low-unemployment states: in high-unemployment states, more experienced individuals were more likely to be enrolled than their less experienced counterparts. While this may have been the result of higher application rates for more experienced individuals during the JTPA period, it is consistent with the hypothesis that the JTPA incentive system gives an incentive to cream-skim more experienced applicants and that this incentive is higher in states with more unemployment.

This section has found evidence that cream-skimming is an important phenomenon. Given the small samples and the likely measurement error in the incentive variables which biases them to zero, it is safe to conclude that the measured cream-skimming effect is a lower bound. I emphasize that I have been careful in choosing the cream-skimming label rather than immediately declaring this as evidence of moral hazard. The reason is that cream-skimming may be consistent with the stated policy goals. Although individuals with more experience and higher AFQT scores are likely to do well even without training, they may also be the ones for whom the value-added is also highest. Thus, the next section explores the possibility that cream-skimming is targeting services to those for whom the value-added and measured performance scores are also highest.

3.2 Is there a Positive Incentive Effect?

In the previous section, I found that more intensive JTPA incentives lead to the enrollment of individuals with more work experience and higher AFQT scores. While I interpreted this as evidence that service providers were cream-skimming better qualified applicants instead of providing better services, it is equally plausible that higher incentives induce service providers to take care in providing services to those for whom value-added is highest. If value-added is greatest for the more experienced, then cream-skimming is a commendable activity. Because the goals of the JTPA program are to maximize value-added in the sense that providers are to be rewarded for increased earnings and unemployment and reduced welfare dependency, simply using the JTPA standards as a measure of performance is clearly an inadequate measure of value-added. To understand whether JTPA incentives increase performance and not just cream-skimming, I use a measure of performance consistent with the policy goals: the increase in annual earnings from before to after the training. Although the training literature[21] has focused on the measuring the counter-factual "what would earnings have been if the individual did not get training," I do not use this measure. Clearly, the increase in annual earnings from before to after training is a measure which better reflects the program goals than the actual performance measures utilized. As will be seen, econometrically specifying this is not trivial and the added complication of controlling for self-selection in estimating the classic counter-factual would likely make the problem intractable. Given that the goal of this section is to understand whether the cream-skimming effects are "good" or "bad", solving for self-selection is unnecessary.

An analysis of training on earnings conditional on participating in training assumes that pre-training earnings (Y_{i1}) and post-training earnings (Y_{i2}) are distributed

$$Y_{it} = X_{it}\beta + \epsilon_{it}$$

where $X_{it}\beta$ and ϵ_{it} ($t = 1$ or 2) are respectively the observed and unobserved portions of earnings. Under the assumption that ϵ_{it} are independently distributed with zero mean and variance , the impact of training on participant earnings is the difference

$$\Delta_i = E[Y_{i2} - Y_{i1}] = X_{i2}\beta_2 - X_{i1}\beta_1$$

a positive difference in the parameters on EXPERIENCE and/or AFQT is evidence that effective program targeting is being mislabelled as cream-skimming. JTPA

[21] See the excellent reviews by Barnow (1987) and Moffit (1987) and the work by Heckman and Robb (1985) and Ashenfelter (1978).

incentives affect program performance if the difference $\Delta(q(D))$ is a function of service quality q which in turn is a function of the incentive regime D such that $\frac{\partial q}{\partial D} > 0$ and $\frac{\partial \Delta}{\partial q} > 0$. Empirically, this can be captured by including the performance measures as explanatory variables in X_{it} and testing whether the pre- and post-training coefficient estimates are significantly different from zero and each other. This leads to the following:

> **Prediction 6**: If post-training coefficients on EXPERIENCE and AFQT in the earnings regression are positive and greater than their corresponding pre-training values, then cream-skimming is good and should not be labelled as moral hazard.

> **Prediction 7**: If the post-training coefficients on the incentive variables CARROT, STICK and AWARDSIZE in an earnings regression are positive and greater than their corresponding pre-training values, then "value-added" is higher in high incentive states. This is evidence of a positive incentive effect.

> **Prediction 8**: If the pre-training coefficients on the incentive variables CARROT, STICK and AWARDSIZE are positive and greater than their post-training values, then higher incentives lead to cream-skimming not captured by the other variables.

> **Prediction 9**: If the post-training coefficients and the JTPA dummy are positive and greater than the pre-training value, then value-added is higher during the JTPA period. This is evidence that overall, the JTPA had a positive incentive effect relative to CETA.

These statement are only valid if the vector X_{it} includes controls for factors that both raise earnings and are correlated with policy variables. X_{it} must also include variables that influence administrative cream-skimming. Therefore, the vector X_{it} also contains controls for age, education, race, sex, and previous market experience. In addition, to avoid incorrectly attributing state and time effects to the policy variables, I include state unemployment rates and average earnings as

covariates. This reasonably assumes that the program has a negligible effect on the local state unemployment rates and average earnings.

The previous model of average earnings is inadequate because a significant fraction of individuals both before and after training have zero earnings. In addition, the idiosyncratic portion of individual earnings is likely to be correlated over time. A simple solution to these problems is to assume a joint distribution $f(\epsilon_{i1}, \epsilon_{i2})$ for ϵ_{it} and derive the likelihood function based on the assumption that $Y_{it} = 0$ if $X_{it}\beta_t + \epsilon_{it} < 0$. While this bivariate tobit assumption makes for a statistically coherent model that provides a solution to the problem of a mass point at zero and the cross-time correlation in earnings, it is unappealing from an economic perspective. The problem is that the same parameters predict both the level of earnings and the likelihood that earnings are equal to zero. Therefore, I estimate the following 2-period version of the traditional generalized tobit model

$$\ln Y_{i1} = \left\{ \begin{array}{c} X_{i1}\beta_1 + \epsilon_{i1} \text{ if } X_{i1}\Gamma_1 + \epsilon_{i3} > 0 \\ 0 \text{ if } X_{i1}\Gamma_1 + \epsilon_{i3} < 0 \end{array} \right\}$$
$$\ln Y_{i2} = \left\{ \begin{array}{c} X_{i2}\beta_2 + \epsilon_{i2} \text{ if } X_{i2}\Gamma_2 + \epsilon_{i4} > 0 \\ 0 \text{ if } X_{i2}\Gamma_2 + \epsilon_{i4} < 0 \end{array} \right\}$$

This model allows the parameters predicting the level of earnings conditional on having worked in the year to differ from those predicting the likelihood of zero earnings. Cross-time correlation can be introduced by assuming that $\epsilon_1, \epsilon_2, \epsilon_3$, and ϵ_4 are jointly distributed $f(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4)$with mean zero. Under this assumption, the maximum likelihood estimates of the parameters are consistent, and traditional likelihood ratio and Wald tests are valid.[22] Under a multivariate normal assumption, the impact of training on earnings is the difference

$$\Delta_i = E[Y_{i2} - Y_{i1}] = \Pr(I_{i2} > 0)E(Y_{i2}|I_{i2} > 0) - \Pr(I_{i1} > 0)E(Y_{i1}|I_{i1} > 0)$$

where I define $I_{ij} = X_{ij}\Gamma_j + \epsilon_{ij}$.[23]

Estimates using this model with controls for sex, race, welfare recipiency, being a high school dropout, and two age dummies indicating whether an individual was

[22] Amemiya (1985), Chapter 4.

[23] The bivariate tobit was tested against the bivariate generalized tobit using a chi-square goodness-of-fit test developed in Cragg (1993). I calculated the test statistic by partitioning the sample space into 7 cells based on annual earnings brackets of ($0, $0-$2500, $2500-$5000, $5000-$7500, $7500-$10000, $10000-$15000, >$15000). The tobit model soundly rejects the null that the parametric specification is correct while the generalized tobit did not.

less than 20, or between 20 and 25 are presented in Table 3.3. The comparison group is a white high school graduate who is over 25 and does not collect welfare. In addition, to control for cyclical year effects and to recognize that the policy variables might also proxy for non-training state effects, I include both the state average unemployment rate and the state average earnings as control variables. Finally, to control for cream-skimming and to see to whom the greatest value added may be attributed, I include work force experience and/or AFQT scores as the final basic control variable.[24] In addition to the control variables, policy variables are included in both X_1 and X_2. State incentive policies are permitted only to shift the means of the pre- and post-training earnings distributions. I allow the policy variables to shift both the pre-enrollment and post-training means for the following reason: while the policy variables are intended to proxy for trainers' efforts and the quality of their training, they are also likely to be correlated with unobserved factors which inflate or deflate earnings which are not controlled for by the state, demographic, and human capital variables. Unlike in the previous probit equations, I exclude interactions between the policy variables because of the paucity of data.

Both the significance and magnitude of the parameter estimates on the policy and experience variables are important attributes in judging whether the incentive policy has a beneficial impact on JTPA performance. In the first column, the estimated model excludes the policy variables while in the second column they are included. From top to bottom, the parameter estimates are divided into the pre- and post-training estimates. A positive coefficient corresponds to an increase in the level of expected earnings. Across the pre-enrollment and the post-enrollment periods, the parameters are qualitatively consistent and jibe with economic intuition: older white male high school graduates who have more work experience[25] have higher annual earnings before and after training.[26] Individuals who live in states with high levels of unemployment or low-average earnings have lower annual earnings both before and after training. In addition, the covariance between past and future earnings is positive in the two tobit models.

With regards to the cream-skimming hypothesis, the interesting coefficients

[24]EXPERIENCE is again defined as the fraction of time spent employed since the individual has left school.

[25]Estimates which only include EXPERIENCE and not AFQT scores are again presented because although the effects are similar, the coefficientes on AFQT scores were typically insignificant.

[26]The findings presented in this section are consistent with those found with either a simple OLS or a bivariate tobit specification.

are those for work experience. For both the employment probability and the level of earnings conditional on working, we see that the difference between the pre-training and post-training experience parameters are less than zero.[27] For both the levels and the probability of positive earnings, this difference is significantly different from zero. While this indicates that there is a negative selection effect, at the same time there may be a positive incentive effect in that holding all else equal, higher incentives may lead to greater value-added. Therefore, the other parameters of interest are the four policy variables AWARDSIZE, CARROT, STICK, and ADJ in the second set of estimates presented in Table 3.3. The theory predicts that the value added measure might be either higher or lower in high-incentive states because of the moral hazard problem. On the other hand, if higher incentives lead to both a selection effect and better services, then the parameter difference should be positive when we control for the selection effect (by including work experience). For each pre- and post-training pair, the difference is positive but not significantly different to zero. Thus, higher incentives lead to both a negative selection effect and a positive selection effect: higher incentives encourage enrollment of individuals for whom the value-added is lower; and, higher incentives generate higher value-added holding all else equal. While the negative selection effect is statistically significant, the positive incentive effect is not.

Because the magnitudes of the parameter estimates are difficult to interpret, simulation results for a variety of scenarios are presented in Table 3.4. In particular, I focus on the expected earnings differentials for low- and high-experience individuals.[28] In the first two columns, the expected value is calculated from the specification that omits the policy variables and includes the JTPA dummy. To be comparable to the previous section, the other four columns present the earnings differential in states with and without an adjustment policy and in low- versus high-incentive states as measured by CARROT.[29]

---

[27] I am using the term *difference in the parameters* loosely. Strictly speaking one needs to refer to

$$\frac{\partial[E(Y_{i2}) - E(Y_{i1})]}{\partial X_i} = \begin{array}{l} \phi(X_{2i}\Gamma_2)\exp(X_{2i}\beta_2 + \frac{1}{2}\sigma_2^2) - \phi(X_{1i}\Gamma_1)\exp(X_{1i}\beta_1 + \frac{1}{2}\sigma_1^2) + \\ \Phi(X_{2i}\Gamma_2)\exp(X_{2i}\beta_2 + \frac{1}{2}\sigma_2^2) - \Phi(X_{1i}\Gamma_1)\exp(X_{1i}\beta_1 + \frac{1}{2}\sigma_1^2) \end{array}$$

where X refers to the policy variable and ϕ and Φ are the density and cummulative densities for a standard normal random variable.

[28] Again I calculate the simulated probabilities and expected values for a white male high school dropout between 20 and 25 years of age living in a state with an unemployment rate of 7.3% and state average earnings of $21,000.

[29] Both qualitatively and quantitatively, the simulation results are similar for the variables STICK

It should be noted that the earnings increase due to training is considerably higher than that found in the traditional training literature. The reason is that the traditional literature estimates the counter-factual "what would earnings have been without training." These estimates are meant to generalize to the whole population.[30] I estimate the simple difference in pre- and post-training earnings conditional on being trained thus ignoring the selection effect.

The first pattern that emerges is that the value-added measure is always higher for low-experience individuals than high-experience individuals. During the CETA period the increase from high- to low-experience individuals is $1,500 whereas during the JTPA period it is $2,100.

The second pattern is that in all cases (low- versus high-experience and with or without an adjustment policy), the predicted earnings difference in high-incentive states is higher than that for low-incentive states. In all cases, the difference between the value-added is between $1,100 and $2,700 which is 20% to 100% higher than the low-incentive measure. While one must be cautious regarding their statistical significance, the point estimates suggest that higher incentives lead to a rise in the value-added from training programs despite incentives being based only upon post-training outcome measures. The third pattern is that states with an adjustment policy always have a higher value-added than those without. The increase in the value-added ranges from $800 to $3,000. This increase arises mainly from a reduction from pre-training earnings measures and not from differences in post-training outcomes suggesting that adjustment policies significantly reduce the incentives to cream-skim and thus tend to enroll individuals with lower earnings.[31]

---

and AWARDSIZE.

[30] See the excellent reviews by Barnow (1987) and Moffit (1987) and the work by Heckman and Robb (1985) and Ashenfelter (1978).

[31] I assume that the differences between the CETA and JTPA trained individuals could be modelled as a simple intercept shift. An alternative model would allow the parameters on the basic socio/economic variables to differ across the CETA and JTPA trained individuals. I estimated a fully interactive model without the policy variables and formed a likelihood ratio test of

$$H_0 : \text{CETA and JTPA can be pooled except for the intercept}$$

The test statistic was 11.75 with 16 degrees of freedom. I also focused on the variables which were individually significant in this regression to see whether they can be pooled. A similar likelihood ratio test was formed for the specification that allowed the four variables state unemployment rate, state average earnings, male and experience to differ across the two periods. The likelihood ratio test statistic was 9.8. Thus, we can be confident that the estimates can be pooled across the two training regimes.

3.3 Does the Positive Incentive Effect Dominate the Negative Selection Effect?

The implemented JTPA performance standards are not perfectly correlated with the Act's training objectives. While I find that higher standards lead to performance increases as measured by the increase in annual earnings, they are also associated with a significant moral hazard problem. Higher incentives encourage the enrollment of individuals with more work experience for whom measured performance is higher but for whom value-added performance is lower. Thus, in implementing higher incentives, there is a trade-off between increasing value-added and shifting the enrolled population to groups with lower value-added. The question of whether the positive incentive effect dominates the negative selection effect depends on the population distribution of high- and low-experience individuals. If there are relatively more high-experience individuals to whom the training population is shifted through cream-skimming responses by administrators, then the selection effect is likely to dominate the positive incentive effect as relatively more individuals are enrolled for whom the value-added is lower. However, if the eligible population is predominately low-experience then increasing incentives is likely to increase overall program performance because most of the enrolled population will still be individuals for whom value-added is higher. I resolve this trade-off by simultaneously simulating both the selection effect and value-added across the population. This relies on data which characterizes the joint distribution of work experience and afqt scores (E) and individual attributes (X) within the JTPA eligible population. The total value-added for a particular incentive regime is the following sum over all individuals of the product of the probability that they participate in a JTPA training program and the total value-added from participating

$$\int_{-\infty}^{\infty}\int_{0}^{1} \Pr(\text{enrolled} \mid X,E)E(Y_2 - Y_1 \mid E,X)\, f(E,X)dEdX$$

The previous two sections developed estimates of $\Pr(\text{enrolled}|X,E)$ and $E(Y_2 - Y_1|X,E)$. The NLSY provides a measure of the distribution $f(E,X)$ so that it is possible to calculate the net trade-off between the positive incentive effect and the negative selection effect of using higher incentives.

Policy simulations for four incentive regimes are presented in Table 3.5. Two patterns are apparent: first, the high-incentive with an adjustment policy regime provides the greatest value-added, and, the low-incentive without an ADJustment policy regime serves the greatest fraction of the eligible population while the

low-incentive without an ADJustment policy regime serves the fewest number. Thus, using the JTPA eligible population from the NLSY, I find that whereas higher incentives could have reduced JTPA performance, they in fact both increase total value-added, value-added per individual, and overall access to the program. Although it was apparent that employing higher incentives led to a trade-off between performance increases as measured by value-added and cream-skimming that shifts the enrolled population to a group for whom value-added is lower, it was unclear which of the two effects dominated. The evidence from this simulation suggests that although the moral hazard problems associated with imperfect performance measurements could have overwhelmed the increases in training quality and effort associated with higher incentives, in reality, they do not.

4 Is this a Natural Experiment?

Depending on temporal and geographic variation in JTPA incentive policy as a means of identifying cream-skimming and positive incentive effects may have several shortcomings. With regards to the enrollment probabilities, what I interpret as selection differences arising from rational variation in program administrator's behavior, can also arise from other spurious factors. First, the selection propensities may actually represent application differences across various regimes. For instance, more experienced individuals might be more likely to apply in high-unemployment areas and an administrator randomly choosing enrollees would still choose a more experienced group for training. Alternatively across incentive regimes, higher application rates for more experienced individuals might arise because higher incentives might generate a reputation effect whereby it becomes a common perception within high-incentive communities that high-incentive individuals are more likely to be enrolled. This would dissuade less experienced individuals from applying. In addition, higher incentives might lead to more SDA advertising which might be more persuasive for high-experience individuals. Any of these effects would shift the composition of the applicant pool in favor of more experienced workers. Both of these arguments suggest that what I report as a cream-skimming effect instead reflects spurious correlation between the policy variables and the state average levels of work experience.

As a simple check that what I report as a cream-skimming effect does not simply reflect spurious correlation between the policy variables and the state average levels of work experience, I calculated the correlation between the policy

variables and the state average level of work experience. These correlations are close to zero in most cases: CARROT -0.074 (0.60), STICK -0.034 (0.81), AWARDSIZE 0.005 (0.97), and ADJ -0.188 (0.19).[32] As a second check of how well the policy variables capture solely JTPA policy differences, I estimated the JTPA specifications using data from the CETA period. If the policy variables are truly proxies for JTPA state incentive differences and not simply state differences, then the policy variables should be insignificant during the CETA period. The interactions between AWARDSIZE, CARROT, STICK, and EXPERIENCE or AFQT have very little explanatory power while the interaction of experience and ADJ can significantly explain a small fraction of the selection differences during the CETA period. States which use adjustment policies during the JTPA period also displayed similar selection patterns during the CETA period.

The method adopted relies on exogenous variation in the geographic and temporal incentive variables. Because the geographic and temporal variation may be spuriously correlated with other factors affecting enrollment and employment outcomes controlling for all other factors correlated with the policy variables is critical. I rely on the state average earnings and state unemployment rates to control for all time varying influences which systematically alter the selection and incentive effects across the JTPA and CETA periods. Alternatively, the cross-incentive regime deviations are identified by essentially geographic differences which might simply arise from regional wage effects.

I performed a number of checks to understand whether the incentive variable spuriously reflect regional wage and employment effects not captured by the state average earnings and unemployment rates. First, maps of the four primary variables AWARDSIZE, CARROT, STICK, and ADJ show that they are not geographically correlated thereby alleviating the fear that the policy variables are simply regional dummies. In addition, the only pairs of variables significantly correlated at the 10 percent level of significance are [CARROT, STICK] and [AWARDSIZE,STICK] which suggests that linear combinations of the policy variables are not capturing strictly regional wage and employment effects. Another simple test of the assumption that the average state unemployment rate and the average state earnings adequately control for state effects is to estimate the earnings regressions replacing the policy variables with state dummies for the five states with the largest state population. I found their coefficient estimates to be both small and insignificant. More convincing evidence of the exogeneity is derived by performing popula-

[32]The numbers in parentheses are p-values on the hypothesis H_0: $\rho = 0$.

tion earnings regressions based upon data from the entire NLSY sample using models similar to those estimated where I include both the policy variables and demographic/economic variables as explanatory variables. While the explanatory variables (EXPERIENCE, DROPOUT, MALE, STATE AV. EARNINGS, STATE UNEMPLOYMENT, MINORITY, and age dummies) had t-statistics which ranged from 30 to 350, the policy variables were jointly and usually individually insignificant. Occasionally, the t-statistics for ADJ and AWARDSIZE occasionally crept up close to 2.5 but the coefficient estimates still only explained $200 of the differences in the JTPA incentive regimes and not simply state effects.

5 Conclusions

It is a puzzle that more government contracts do not use performance incentives to create competitive pressures similar to private markets. The theoretical explanations have pointed to agency problems associated with imprecise performance criteria and/or agent risk aversion to idiosyncratic risk. As one of few government programs that use performance contingent payments to create competitive pressures, the Job Training Partnership Act provides a valuable case study of the strengths and weaknesses of performance contracting in the public sector. A particular institutional feature of the JTPA provides a test for why other programs have not adopted such performance systems. This study exploits the fact that the performance measures adopted are only partially correlated with performance objectives. Although servants of the public may react to the intentions of the JTPA and maximize value added, they may alternatively meet the performance objectives by enrolling individuals who will perform well even without training. This is moral hazard because it is an unmonitored activity which clearly violates the intentions of the JTPA. This design oversight provides a natural experiment for studying whether moral hazard is a potential explanation for the limited use of performance management systems by the government.

I find that more intensive incentives lead to moral hazard: training providers tend to enroll individuals who are more likely to have higher earnings and employment even without training, thereby potentially inflating the measured success of the program and hampering achievement of program goals. This does not reflect the targeting of services for whom the value is greatest, rather it is a form of "cream-skimming" because individuals with more work experience tend to have lower increases in annual earnings from participation in JTPA training. However,

this negative selection effect is balanced by a positive incentive effect: holding all else equal, I find that more intensive incentives induce greater performance using a measure consistent with policy goals. This raises the question of whether the JTPA's positive incentive effects leading to earnings increases dominate the negative motivation to target services to whom the value-added is lower. The answer to this question depends on the population distribution of individual characteristics correlated with these two effects. If the JTPA eligible population has relatively few people to whom the value-added is lower even though their enrollment probability is higher, then the positive incentive effect will dominate. Simulations examining the relative magnitude of two effects finds that the positive incentive effect dominates the moral hazard problem.

These findings have important policy implications. The JTPA policy reforms introduced in 1988 which reduced the number of performance standards and focused on collecting longer-run outcomes measures are unlikely to reduce cream-skimming or produce performance increases. Instead, my findings suggests that standards based upon changes in employment and earnings relative to what they were before training are likely to spur greater performance increases. Generally, important lessons should be drawn for the design of future systems like educational voucher systems and the performance management systems mandated for the training components for the Food Stamp Program and the AFDC.[33] One needs to be wary of the moral hazard and adverse selection problems associated with using performance contingent pay.

References

[1] Amemiya, Takeshi (1985). *Advanced Econometrics.* Cambridge: Harvard University Press.

[2] Anderson, Kathryn H., Richard V. Burkhauser, and Jennie E. Raymond (1993). "The Effect of Creaming on Placement Rates Under the Job Training Partnership Act," *Industrial and Labor Relations Review,* Vol. 46, No. 3, pp. 613-624.

[33]The Omnibus Budget Reconciliation Act (1988) and the Family Support Act (1988) require that by 1993 performance management systems be incorporated into the Food Stamp and AFDC programs.

[3] Anton, J.J. and D.A. Yao (1987). "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement," *RAND Journal of Economics*, V. 18, pp. 57-75.

[4] Anton, J.J. and D.A. Yao (1990). "Measuring the Effectiveness and Competition in Defense Procurement: A Survey of the Empirical Literature," *Journal of Policy Analysis and Management*, v.9, pp.60-79.

[5] Ashenfelter, Orley (1978). "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, V.60, pp. 47-57.

[6] Baker, George P. (1992). "Incentive Contracts and Performance Measurement." *Journal of Political Economy*, Vol. 100, No. 3, pp. 598-614.

[7] Barnow, Burt S. (1987). "The Impact of CETA Programs on Earnings: A Review of the Literature." *Journal of Human Resources*, Vol. 22, No. 2, pp. 155-193.

[8] Barnow, Burt S. (1992). "The Effect of Performance Standards on State and Local Programs." in *Evaluating Welfare and Training Programs*. eds. Charles F. Manski and Irwin Garfinkel. Cambridge: Harvard University Press.

[9] Cragg, Michael Ian (1993). "The Impact of Performance Incentives on Providing Job Training to the Poor: The Job Training Partnership Act (JTPA)," Stanford University Ph.D. Dissertation.

[10] Dickinson, Katherine P., and Richard W. West (1988). *Evaluation of the Effects of JTPA Performance Standards on Clients, Services and Costs.* Washington: U.S. Government Printing Office.

[11] Gibbons, Robert and Kevin J. Murphy (1992). "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy*, Vol. 100, No. 3, pp. 468-505.

[12] Goldberg, V.P. (1976). "Regulation and Administered Contracts," *Bell Journal of Economics*, v.7, pp. 41-58.

[13] Heckman, James J. and Richard Robb Jr. (1985). "Alternative Methods for Evaluating the Impact of Interventions," in *Analysis of Labor Market Data*, eds. James J. Heckman and Burton Singer, Cambridge: Cambridge University Press.

[14] Holmstrom, Bengt and Paul Milgrom (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization*, Vol.7, pp.24-52.

[15] Krueger, Alan (1990). "Ownership, Agency and Wages: An Examination of Franchising in the Fast Food Industry." *National Bureau of Economic Research Working Paper*, 3334.

[16] Levitan, Sar A. and Frank Gallo (1988). *A Second Chance Training for Jobs.* Kalamazoo: W.E. Upjohn Institute for Employment Research.

[17] Lafontaine, Francine (1992). "Agency Theory and Franchising: Some Empirical Results," *RAND Journal of Economics,* v.23, pp. 263-283.

[18] Moffitt, Robert (1987). "Symposium on the Econometric Evaluation of Manpower Training Programs." *Journal of Human Resources,* Vol. 22, No. 2, pp. 149-155.

[19] Moffitt, Robert (1992). "Incentive Effects of the U.S. Welfare System: A Review." *Journal of Economic Literature,* Vol. 30, No. 1, pp. 1-61.

[20] National Commission for the Employment Policy (1987). *The Job Training Partnership Act.* Washington: U.S. Government Printing Office.

[21] Staten, Michael E. and John Umbeck (1982). "Information Costs and Incentives to Shirk: Disability Compensation of Air Traffic Controllers." *American Economic Review*, Vol. 72, pp. 1023-1037.

6 Appendix - Data Description

The NLSY provides data that matches longitudinal earnings and welfare data with the timing of training, thereby allowing one to determine the population eligible for training all within a representative sample. The NLSY is an eleven year longitudinal panel that began annual interviewing of 12,686 youths between the ages of 14 and 21 in January, 1978. The panel used for estimation retained all observations until the first missed interview. Thus, by the 11th interview, just over a quarter of the sample were no longer observed. As is typically done in studies using the

NLSY, the military sample was dropped altogether. The eligible sample is further restricted to only those individuals for whom a complete labor market history from the last date of schooling is available. Including individuals with incomplete work history data would introduce initial conditions problems for both labor market experience measures and the potential presence of past training. Over the 11 years of the resulting panel ending in 1988, the annual sample ranges from 737 to 3,945 observations for whom there are 467 CETA or JTPA training episodes.

At least 90 percent of CETA and JTPA participants are restricted to be from disadvantaged families. A disadvantaged family is defined as one which either is collecting welfare or whose income excluding welfare payments in the last six months fall below the larger of the BLS's lower living standard or the Department of Agriculture's Poverty Index. An individual is included in the annual eligible population if the family collected welfare during the year or if in one quarter of the year, the family's previous six months' income fell below the JTPA and CETA qualifying threshold.[34] The eligible sample characteristics for the 12,026 annual observations on 3,252 individuals are presented in Table A1.

Pre- and post-training annual earnings were collected for those individuals who participated in a training program and for whom there was a full year of earnings information either before or after the training episode. Data for the 382 individuals who met this criteria are summarized in Table A2. The last drop from pre-training to post-training average earnings for those who worked reflects significant sample composition differences due to missing data: the post-training population is younger and has accumulated less human capital. If we isolate the 203 individuals for whom there is a full year of earnings data for both the period

[34]Income and family size are defined as:

If married, income in the last six months is the sum of individual income from earnings in the last six months, half of the spouse's annual income, half of the annual farm or self-employment income, half of annual financial aid, half of income recieved from others excluding welfare income. Family size is 2 + the number of children.

If single and not living with parents, income in the last six months is the sum of individual earnings from income from the last six months, half of the annual farm or self-employment income, half of annual financial aid, half of income recieved from others excluding welfare income. Family size is 1 + the number of children.

If single, living with parents and less than 18, income in the last six months is the sum of individual income from earnings in the last six months, half of the annual farm or self-employment income, half of annual financial aid, half of income recieved from others excluding welfare income, and half of the family's annual income. Family size is 1 + the number of children, the number of siblings, and the number of parents present.

prior to training and that following training, we see that there is a $1,637 earnings increase and we no longer see the aberrant age structure found in the full sample.